

Insights into the activation of the spliceosomal helicase Prp43

Dissertation

for the award of the degree

"Doctor rerum naturalium"

of the Georg-August-Universität Göttingen

submitted by

Henning Christian

from Husum

Göttingen, 2013

Thesis committee:

Prof. Dr. Ralf Ficner (Reviewer)
Department for Molecular Structural Biology
Institute for Microbiology and Genetics
Georg-August-Universität Göttingen

Prof. Dr. Reinhard Lührmann (Reviewer)
Department of Cellular Biochemistry
Max-Planck-Institut for Biophysical Chemistry, Göttingen

Prof. Dr. Holger Stark
3D-Cryo Electron Microscopy Group
Max-Planck-Institut for Biophysical Chemistry, Göttingen

Date of the oral examination: April 9th, 2013

I hereby declare that the available PhD thesis "Insights into the activation of the spliceosomal helicase Prp43" has been written independently with no other aids or sources than quoted. This thesis has not been submitted elsewhere for any academic award or qualification.

Henning Christian, March 2013

Contents

Abstract	5
1 Introduction	7
1.1 Classes of introns	7
1.2 The splicing reaction	9
1.3 The spliceosome	10
1.3.1 The splicing cycle	10
1.3.2 Components of the spliceosome	14
1.3.3 The catalytic center of the spliceosome	18
1.4 RNA helicases	18
1.4.1 Classification of helicases	19
1.4.2 DEAD-box, DEAH-box and Ski2-like helicases	21
1.5 The spliceosomal DExD/H-box helicases	25
1.5.1 Promoters of the splicing cycle	26
1.5.2 Proofreading during the splicing cycle	27
1.6 The DEAH-box protein Prp43	28
1.6.1 The role of Prp43 during the splicing cycle	28
1.6.2 The role of Prp43 in ribosomal biogenesis	29
1.6.3 The crystal structure of Prp43	29
1.6.4 The G-patch motif	31
1.7 Aims of this work	32
2 Material and methods	35
2.1 Material	35
2.1.1 Cell media	35
2.1.2 Chromatography systems, columns and accessories	36
2.1.3 Computer programs and data banks	36
2.1.4 Crystallisation screens	36
2.1.5 Equipment	37

Contents

2.1.6	Fine chemicals	38
2.1.7	Genomic DNA	38
2.1.8	Organisms	38
2.1.9	Plasmids	38
2.1.10	RNA oligonucleotides	40
2.1.11	Size standards for DNA and Proteins	40
2.1.12	Supplemental material	40
2.2	Molecular biological methods	41
2.2.1	Agarose gel electrophoresis	41
2.2.2	Cell harvest and lysis	41
2.2.3	Chromatographic methods	42
2.2.4	Concentrating of protein solutions	44
2.2.5	Deletion of introns	44
2.2.6	Determination of the DNA concentration	44
2.2.7	Determination of the protein concentration	45
2.2.8	DNA sequencing	45
2.2.9	Generation of chemical competent <i>E. coli</i> cells	45
2.2.10	Generation of glycerol cell stocks	46
2.2.11	Molecular cloning	46
2.2.12	Polymerase chain reaction	47
2.2.13	Preparation of plasmid DNA	48
2.2.14	Protein identification	48
2.2.15	Proteolytic digestion	48
2.2.16	Recombinant expression of proteins in <i>E. coli</i>	49
2.2.17	Restriction enzyme digest	49
2.2.18	SDS polyacrylamide gel electrophoresis	49
2.2.19	Transformation of chemical competent <i>E. coli</i> cells	51
2.3	Biophysical methods	51
2.3.1	Characterisation of the ATPase activity	51
2.3.2	Characterisation of the helicase activity	52
2.3.3	Circular dichroism spectroscopy	54
2.4	Crystallographic methods	54
2.4.1	High throughput screening	54
2.4.2	Optimisation of crystals	55
2.4.3	Cryocrystallography	55
2.4.4	Data collection	55

3	Results	57
3.1	Preparation of the NTR proteins from <i>S. cerevisiae</i>	57
3.1.1	Purification of yPrp43	57
3.1.2	Purification of yNtr2	58
3.1.3	Copurification of yNtr1 and yNtr2	59
3.1.4	Purification of yNtr1(1-120)	61
3.1.5	Purification of yNtr1(51-110)	62
3.2	Preparation of yPrp43/yNtr1 complexes	63
3.2.1	Preparation of the yPrp43/yNtr1(1-120) complex	63
3.2.2	Preparation of the yPrp43/yNtr1(51-110) complex	65
3.2.3	Crystallisation attempts of yPrp43/yNtr1 complexes	66
3.3	Functional characterisation of the interaction between yPrp43 and yNtr1	69
3.3.1	Determination of ATPase and helicase activities	69
3.3.2	Determination of dissociation constants	71
3.3.3	Fusion of yPrp43 and yNtr1	73
3.3.4	Identification of the minimal required yNtr1-fragment	74
3.4	Structural characterisation of the interaction between yPrp43 and yNtr1	76
3.4.1	Chemical cross-linking of the yPrp43/yNtr1 complexes	76
3.4.2	Hydroxyl radical footprinting of the yPrp43/yNtr1(1-120) complex	79
3.4.3	The role of the N-terminal extension of yPrp43	80
3.5	Interaction of yPrp43 with its substrates	81
3.5.1	UV-induced RNA cross-linking of yPrp43	81
3.5.2	Nucleotide binding properties of yPrp43	83
3.6	Analysis of the secondary structure of the G-patch motif	84
3.6.1	Analysis of the secondary structure of yNtr1(51-110)	84
3.6.2	Bioinformatical structural predictions for yNtr1(51-110)	85
3.7	Surface charge calculation of human helicase models	87
3.8	Prp43 of <i>C. thermophilum</i>	89
3.8.1	Identification and cloning of the NTR complex orthologs	89
3.8.2	Functional characterisation of ctPrp43	91
3.8.3	Crystallisation and structure determination of ctPrp43	94
3.8.4	Structural comparison of ctPrp43 and yPrp43	97

4 Discussion	101
4.1 Functional studies on the disassembly of the spliceosome	101
4.2 Properties and regulation of helicases	102
4.2.1 The spliceosomal DEAH-box helicases	102
4.2.2 Mechanisms for the activation of helicases	103
4.3 Crystallographic studies on Prp43	106
4.3.1 Crystallisation trials of the yPrp43/yNtr1 complex	106
4.3.2 Functional and structural properties of ctPrp43	107
4.4 The interaction of yPrp43 with RNA	108
4.4.1 The effect of RNA on the ATPase activity	108
4.4.2 Structural insights into RNA binding	110
4.5 The interaction of yPrp43 with yNtr1	111
4.5.1 Functional effects of yNtr1-binding to yPrp43	111
4.5.2 The minimally required yNtr1-fragment	112
4.5.3 Insights into the binding site of yNtr1 on yPrp43	113
4.5.4 Structural analysis of the G-patch motif	114
4.6 A model for the activation of yPrp43	116
Appendix	119
A.1 Protocols of expression and purification	119
A.2 Supplemental tables	122
A.3 Supplemental figures	128
Bibliography	140
Acknowledgements	171

Abstract

The excision of non-coding sequences that are present in the pre-mRNA is a hallmark of eukaryotic cells. This process is termed as splicing and is catalysed by the spliceosome, a single-turnover enzyme which consists of five Uridine-rich small nuclear snRNAs (UsnRNAs) and more than 100 associated proteins. For each intron to be excised, the spliceosome is assembled, catalyses the two-step splicing reaction and is disassembled afterwards.

The dynamics of the spliceosome is controlled by at least eight conserved DExD/H-box ATPases which promote the transitions between the different steps of the splicing cycle and are also implicated in the quality control. Hence, the activity of these molecular motors has to be highly and precisely regulated.

Subsequent to the release of the mature mRNA, the DEAH-box ATPase Prp43 catalyses the disassembly of the intron-lariat spliceosome. During this process, Prp43 is activated by the heterodimer Ntr1/Ntr2, which form together the ternary NTR complex. The binding of the N-terminal part of Ntr1 which contains a G-patch motif stimulates the helicase activity of Prp43. Recently, the crystal structure analysis of Prp43 from yeast in complex with ADP has revealed that besides the helicase core, Prp43 contains an N-terminal extension and a winged helix, a ratchet and an OB-fold domain. As the structure represents Prp43 in its post-catalytic state, it provides no information about the conformational changes that lead to the activation upon binding of Ntr1.

Here it is shown that the fragment Ntr1(51-110) is both required and sufficient for the activation of Prp43 and thereby stimulates its helicase as well as its ATPase activity. Furthermore, evidence is provided that the interaction with Ntr1 is mediated by the OB-fold domain of Prp43 and that the ratchet domain of Prp43 is involved in RNA binding. The results presented here suggest that the G-patch motif of Ntr1 is intrinsically unstructured and directly implicated in RNA binding. Additionally, Prp43 from the thermophile eukaryote *Chaetomium thermophilum* is biochemically and structurally characterised. In summary, these results provide new insights into the activation of Prp43.

1 Introduction

In eukaryotic cells, the genetic information is stored nearly exclusively in form of chromosomal DNA which is located in the nucleus. Upon induction of gene expression, it is transcribed to a precursor messenger RNA (pre-mRNA). Before the mRNA can be exported into the cytoplasm and translated into a polypeptide chain by means of the ribosome, it has to be processed (Moore and Proudfoot, 2009; Hocine *et al.*, 2010).

During processing, the pre-mRNA obtains a 5'-m⁷G cap which contributes to its stability as it protects the mRNA against 5'-3' exonucleases and serves as recognition motif before translation (Hsu and Stevens, 1993; Tarun and Sachs, 1996). At its 3'-end, a poly(A) tail is added to the pre-mRNA which is important for stability and translation efficiency (Drummond *et al.*, 1985). A third processing event is necessary for pre-mRNAs that contain non-coding intervening regions (introns): This process is termed splicing and consists of the excision of the intron and the subsequent ligation of the coding sequences (exons). The process of splicing is a hallmark of eukaryotic cells (Will and Lührmann, 2011) and is carried out by the spliceosome, a multi-megadalton, highly dynamic ribonucleoprotein (RNP) complex (Wahl *et al.*, 2009; Will and Lührmann, 2011).

There are two different modes of splicing: constitutive splicing, where all exons are present in each processed mRNA, and alternative splicing, where the linkage of a distinct number of spliced exons gives rise to different species of mRNA, generated from the same pre-mRNA. This phenomenon is encountered especially in higher eukaryotes and increases the genomic complexity, as many different proteins can result from a single gene (Ast, 2004; Chen and Manley, 2009). The following sections will focus on constitutive splicing.

1.1 Classes of introns

In total, three different classes of introns have been described in eukaryotes so far. These are, besides the spliceosomal introns, the group I/II self-splicing introns and

1 Introduction

the tRNA introns.

Group I introns are the first self-splicing ribozymes that have been discovered (Kruger *et al.*, 1982) and represent one of the first examples for catalysis performed by RNA molecules (Guerrier-Takada *et al.*, 1983). Group II introns have no phylogenetic relationship to group I introns and are primarily found in bacteria, but also in plants, protists and fungi (Peebles *et al.*, 1986). They form a highly complex RNA structure consisting of six structural domains. The crystal structures of a group II intron from *Oceanobacillus iheyensis* in different catalytic steps (Toor *et al.*, 2008; Marcia and Pyle, 2012) have strengthened the consideration that they share a common ancestor with the spliceosomal introns (Dayie and Padgett, 2008; Toor *et al.*, 2009).

The second class of introns is present in the tRNA of bacteria, archaea and eukaryotes. Of the 272 tRNA genes identified in *Saccharomyces cerevisiae*, 59 are interrupted by introns (Trotta *et al.*, 1997). These introns span between 14 and 60 nucleotides and interrupt the anticodon loop directly next to the anticodon in 3'-direction (Ogden *et al.*, 1984). The mechanism of tRNA splicing consists of three steps in which an endonuclease, a ligase and a phosphotransferase are involved (Abelson *et al.*, 1998).

Within the three different classes of introns, spliceosomal introns represent the most abundant class. In yeast, only 3 % of the genes, but 27 % of the total transcripts contain short introns (Ares *et al.*, 1999; Lopez and Séraphin, 1999). In mammals, only 8 % of the genes are intron-free (Fedorova and Fedorov, 2005). Mammalian introns cover several hundred to several thousands of nucleotides (Deutsch and Long, 1999), whereas the length of the exons is, quite constantly, about 120 nucleotides (Ast, 2004).

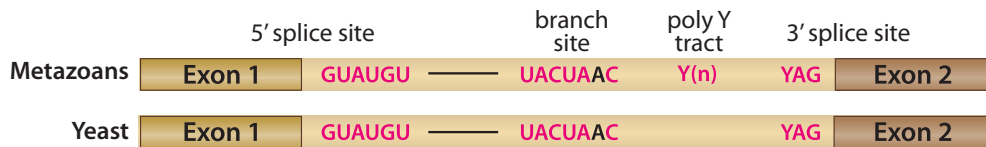


Figure 1.1: Conserved motifs which define the position and the borders of exonic and intronic sequences in pre-mRNA of metazoans and yeast. These motifs are present at the 5' splice site, the branch site, the polypyrimidine tract (Y(n)) and the 3' splice site. The branch point adenosine is highlighted in black. This figure was adapted from Will and Lührmann (2011).

Spliceosomal introns contain several characteristic sequence motifs, namely a 5' splice site (ss), a 3'ss and a branch site (bs), that is located 18 to 40 nucleotides

upstream of the 3'ss (see figure 1.1). Introns of metazoans contain furthermore a conserved polypyrimidine tract between the branch site and the 3'ss, which is supposed to play a role during spliceosomal assembly (Sharma *et al.*, 2008). On the pre-mRNA, additional *cis*-acting elements are present in form of exonic or intronic splicing enhancers (ESEs/ISEs) and exonic or intronic splicing silencers (ESSs/ISSs) which interact with *trans*-acting splicing factors and thereby support or inhibit the splice site recognition or the assembly of the spliceosome (Smith and Valcárcel, 2000; Wang and Burge, 2008).

1.2 The splicing reaction

The excision of an intron occurs in two consecutive S_N2 -type transesterification reactions for which no energy is required (Moore and Sharp, 1993). The conserved 5'ss, 3'ss and the branch site play an important role in this process (see figure 1.2).

During the first reaction step, the 2'-hydroxyl group of the branchpoint adenosine attacks the phosphate group which is located at the 5'ss in a nucleophilic manner (Grabowski *et al.*, 1984; Padgett *et al.*, 1984). This reaction produces a free 5'-exon and an intron-lariat-3'-exon (or lariat intermediate), which is formed by the 2'-5'-phosphodiester bond of the branch point adenosine and the guanine at the 5'-cleavage site. Subsequently, in the second reaction step, the 3'-hydroxyl group of the exon performs a nucleophilic attack on the phosphate ester bond at the 3'ss; the two exons are connected by a new phosphodiester bond, and the excised intron forms a lariat (Ruskin *et al.*, 1984).

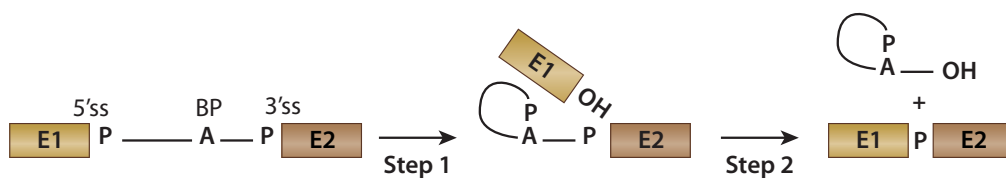


Figure 1.2: Diagram showing the two reaction steps of pre-mRNA splicing. The exons are shown as boxes (E1 and E2), the intron as a line. Additionally, the 5'- and the 3'-splice site (5'/3'ss), the phosphate groups (P) and the branchpoint adenosine (BP) are indicated. This figure was adapted from Will and Lührmann (2011).

Based on the finding that the reaction intermediates and products of pre-mRNA splicing resemble those of the group II self-splicing introns, it was suggested that the mechanism of pre-mRNA splicing could be also RNA-based (Jacquier, 1990).

1 Introduction

Interestingly, both splicing steps can be reversed to a high extent under certain conditions, which shows another similarity to group II self-splicing introns (Tseng and Cheng, 2008).

In the cell, the excision of spliceosomal introns does not occur on its own but is performed by a macromolecular machinery, the spliceosome.

1.3 The spliceosome

The spliceosome is formed by five uridine-rich small nuclear RNAs (UsnRNAs) and numerous associated proteins (Wahl *et al.*, 2009). It is a single turn-over machinery and assembles stepwise on the pre-mRNA with numerous changes in its composition between the different steps. Its assembly results in a catalytically active spliceosome which is disassembled after the catalysis of the two reaction steps (see figure 1.3). There are two different types of spliceosomes: the major and the minor spliceosome.

The U1, U2, U4/U6 and U5 snRNPs are part of the major (or U2-dependent) spliceosome which is responsible for the excision of most introns in pre-mRNAs (Will and Lührmann, 2001; Patel and Bellini, 2008). In some metazoan and plant genes an additional class of rare introns is excised by the minor (or U12-dependent) spliceosome which is formed by the U11, U12, U4atac/U6atac and U5 snRNPs (Tarn and Steitz, 1996; Turunen *et al.*, 2012). In addition to specific UsnRNAs, the minor spliceosome contains a set of proteins absent in the major spliceosome (Will *et al.*, 2004). The following explanations will solely refer to the major or U2-dependent spliceosome.

Research on the spliceosome and the mechanistics of splicing is motivated by gaining insights into the regulation of gene expression, and also because it displays a therapeutic target (Ward and Cooper, 2010; Tazi *et al.*, 2005).

1.3.1 The splicing cycle

The different steps of the splicing cycle can be divided into the spliceosomal assembly, the catalysis of the two reaction steps and the disassembly of the intron-lariat spliceosome and will be described in the following sections.

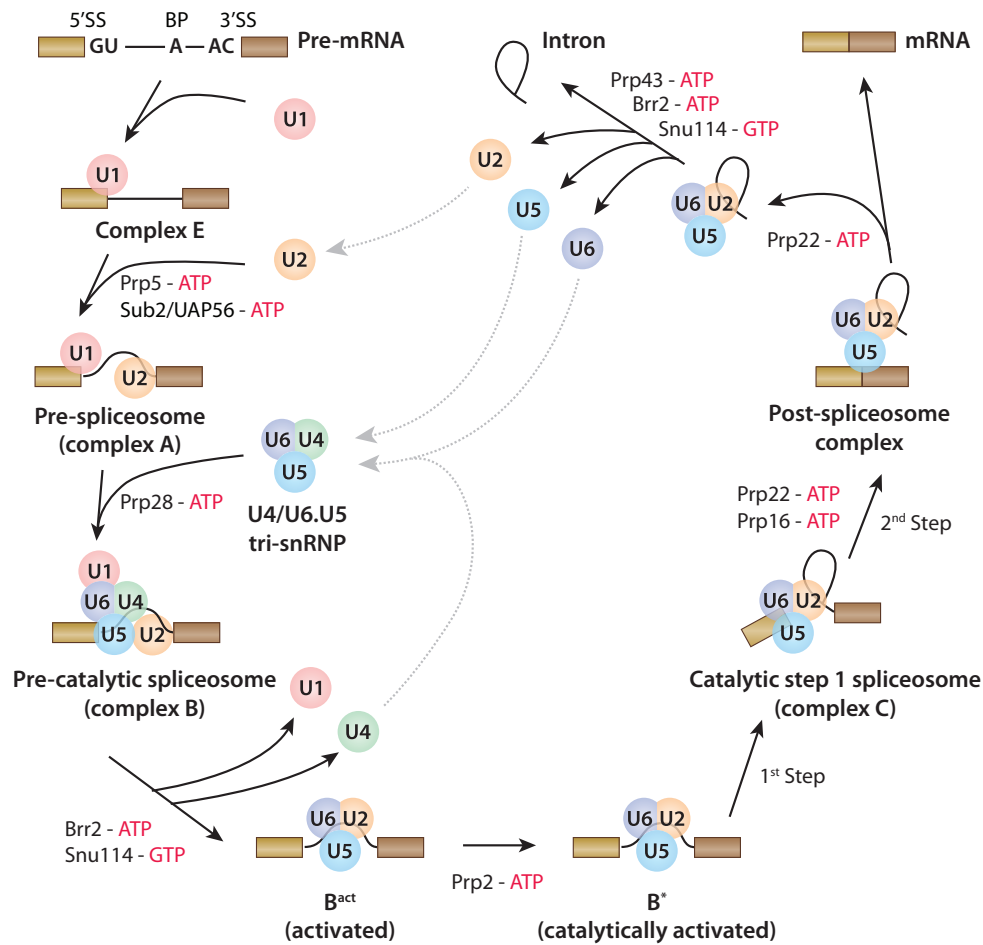


Figure 1.3: Schematic representation of the assembly, catalysis and disassembly steps of the U2-dependent spliceosome. The pre-mRNA substrate, the snRNPs (circles) and the eight conserved DExD/H-box proteins (along the arrows) are shown. For the different spliceosomal complexes, the names according to the metazoan nomenclature are indicated. This figure was adapted from Will and Lührmann (2011).

1.3.1.1 Assembly and activation of the spliceosome

For all introns that are not larger than about 200 to 250 nucleotides, the spliceosome assembles across the intron. The starting event is the formation of the complex E, which is characterised by the ATP-independent interaction of the U1 snRNP with the 5'ss (Fox-Walsh *et al.*, 2005). This interaction is stabilised in higher eukaryotes by the proteins of the U1 snRNP and by SR proteins. Furthermore, additional sequence motifs of the pre-mRNA are recognised during the formation of complex E: the splicing factor SF1/mBBP binds to the branch site, the U2 auxiliary factor (U2AF) interacts with the polypyrimidine tract (ppt)

1 Introduction

and the 35 kDa subunit of U2AF associates with the 3'ss (Berglund *et al.*, 1997; Zamore and Green, 1991).

If the size of an intron exceeds about 200 to 250 nucleotides, alternative ways of spliceosomal assembly exist. The assembly occurs across an exon by formation of an exon-defined complex, through a process called exon definition that is found in mammals to a higher extent than in other metazoans (Berget, 1995; Fox-Walsh *et al.*, 2005). Exon definition is initiated by the binding of the U1 snRNP to the 5'ss downstream of an exon and the subsequent binding of the U2AF protein to the ppt. The U2 snRNP interacts with the BP adenosine, and SR proteins are recruited via ESEs (Hoffman and Grabowski, 1992; Das *et al.*, 2000).

There are several suggested pathways for the conversion of an exon-defined to an intron-defined complex, but this phenomenon is so far poorly understood. Interactions between the U1 and U2 snRNP with the 5'ss may lead to the formation of the A complex (Das *et al.*, 2000; Smith and Valcárcel, 2000) which could be directly transformed into a cross-intron B complex (Schneider *et al.*, 2010). Interestingly, there are indications that the choice of exons during alternative splicing is coupled to the transition from an exon-defined to an intron-defined complex (House and Lynch, 2006; Sharma *et al.*, 2008).

Once the complex E is formed, the U2 snRNP joins the spliceosome by forming base pairs with the 3'ss and gives rise to the complex A, in which the BP adenosine is exposed. This interaction is supported by the U2 snRNP proteins SF3a/SF3b which bind near the branch site (Gozani *et al.*, 1996) and by U2AF65 that interacts directly with the branch site (Valcárcel *et al.*, 1996). Moreover, two DEAD-box ATPases facilitate the binding of the U2 snRNP: Sub2 induces the removal of the U2AF subunit U2AF65 (Kistler and Guthrie, 2001), whereas Prp5 remodels the U2 snRNP itself (Xu and Query, 2007).

Subsequently, the binding of the preassembled U4/U6.U5 tri-snRNP leads to the formation of the catalytically inactive B complex. During the activation of the spliceosome and the formation of the activated complex B^{act} which is not yet catalytically activated, the U4/U6 interaction is disrupted, and the 5'-end of the U6 snRNP forms base pairs with the 5'ss, thereby replacing the U1 and U4 snRNPs (Konforti *et al.*, 1993). The release of the U1 and U4 snRNPs is catalysed by the DEAD-box ATPase Prp28 and the Ski2-like ATPase Brr2 (Raghunathan and Guthrie, 1998). The U6 snRNP interacts with the U2 snRNP which positions the 5'ss near the branch site, and hence prepares the first step of the splicing reaction (Madhani and Guthrie, 1992; Sun and Manley, 1995). The Prp19-related

complex (NTC for NineTeen complex) joins the spliceosome and contributes to the activation by inducing interactions between the pre-mRNA and the U5 snRNP (Chan *et al.*, 2003; Chan and Cheng, 2005). Additionally, the binding of the NTC results in the release of the LSm proteins from the U6 snRNP, thereby establishing interactions between the LSm site on the U6 snRNP and the intron region located 30 bp downstream from the 5'ss (Chan *et al.*, 2003). This complex contains a catalytic core and is able to catalyse both steps of the splicing reaction.

1.3.1.2 Catalysis of the splicing reaction

The transition from the B^{act} to the catalytically activated B* complex is promoted by the DEAH-box protein Prp2. It was demonstrated that at high salt concentrations, the ATP-dependent helicase Prp2 induces the release of the U2 snRNP components SF3a and SF3b from the spliceosome (Lardelli *et al.*, 2010), but it has also been shown that the U2 proteins might be destabilised, leaving the branch site accessible to attack the 5'ss (Warkocki *et al.*, 2009). A recent study has confirmed that the U2 proteins stay associated with U2 snRNA at physiological salt conditions (Ohrt *et al.*, 2012).

Subsequently, Prp2 dissociates from the spliceosome (Teigelkamp *et al.*, 1994). Furthermore, the first step of the splicing reaction is dependent on the presence of at least three additional proteins, Yju2 which interacts with the NTC (Liu *et al.*, 2007), Spp2 (Silverman *et al.*, 2004) and Cwc25 (Chiu *et al.*, 2009).

Prior to the second reaction step, the DEAH-box ATPase Prp16 interacts with the pre-mRNA on the intron sequence near the 3'ss (Umen and Guthrie, 1995) and induces the release of Yju2 and Cwc25 (Tseng *et al.*, 2011). Whereas the 3'ss seems to be dispensible for the first transesterification reaction (Rymond and Rosbash, 1985), it has to be identified before the second reaction by the splicing factors Slu7, Prp18 and Prp22 (Horowitz and Abelson, 1993; McPheeters *et al.*, 2000). The U5 snRNP is a key player for the alignment of the 5'- and the 3'-exon for their ligation (Newman and Norman, 1992). The interactions of the U5 snRNP have been suggested to be stabilised by the U5 component Prp8 (Teigelkamp *et al.*, 1995).

In contrast to the release of the mRNA (see section 1.3.1.3), the action of the DEAH-box ATPase Prp22 during the second transesterification reaction is ATP-independent (Company *et al.*, 1991; Schwer and Gross, 1998).

1.3.1.3 Disassembly of the spliceosome

The release of the mature mRNA from the spliceosome is supported by the DEAH-box helicase Prp22 in an ATP-dependent manner (Wagner *et al.*, 1998). Prp22 binds to the 3'ss during the second transesterification reaction (McPheeters *et al.*, 2000) and is thought to disrupt the U5-exon base pairing by translocating in 3'-5' direction on the mRNA after ligation of the exons (Schwer, 2008).

The disassembly of the intron-lariat spliceosome, including the lariat RNA as well as the U2, U5 and U6, is dependent of the action of an additional DEAH-box ATPase, Prp43 (Arenas and Abelson, 1997; Martin *et al.*, 2002). Prp43 interacts with Ntr1 that itself forms a heterodimer with Ntr2; the complex of these three proteins has been termed NTC-related (NTR) complex (Tsai *et al.*, 2005; Boon *et al.*, 2006). Besides Prp43, also the ATP-dependent Ski2-like helicase Brr2 and the GTPase Snu114 have been proposed to be required in this process (Small *et al.*, 2006).

Once liberated, the lariat intron is cleaved by the debranching enzyme Dbr1 and degraded in order to prevent a reverse splicing mechanism (Chapman and Boeke, 1991), while the UsnRNPs are recycled in order to reinitiate a new splicing cycle (Raghunathan and Guthrie, 1998).

1.3.2 Components of the spliceosome

1.3.2.1 UsnRNPs biogenesis and assembly

All UsnRNAs except the U6 snRNA are transcribed by the RNA-polymerase II in the nucleus where they acquire posttranscriptionally a m⁷GpppG (m⁷G) cap (Hamm and Mattaj, 1990). After being exported to the cytoplasm, seven Sm proteins (B/B', D3, D2, D1, E, F and G) assemble to a ring-like structure on the highly conserved Sm-site of the UsnRNAs (Raker *et al.*, 1996). Subsequently, the m⁷G cap of the snRNA is dimethylated to a 2,2,7-trimethylat-guanosine (m₃G) cap by a cap dimethyltransferase (Mattaj, 1986; Mouaikel *et al.*, 2002). This complex is imported back into the nucleus, where the binding of the additional snRNP-specific proteins occurs (Will and Lührmann, 2001). Additionally, several base modifications such as 2'-O-methylation and pseudouridylation occur in a posttranscriptional manner in the cajal bodies (Jády *et al.*, 2003).

The maturation of the U6 snRNP is different compared to the other UsnRNPs. The U6 snRNA is transcribed by RNA polymerase III (Paule and White, 2000)

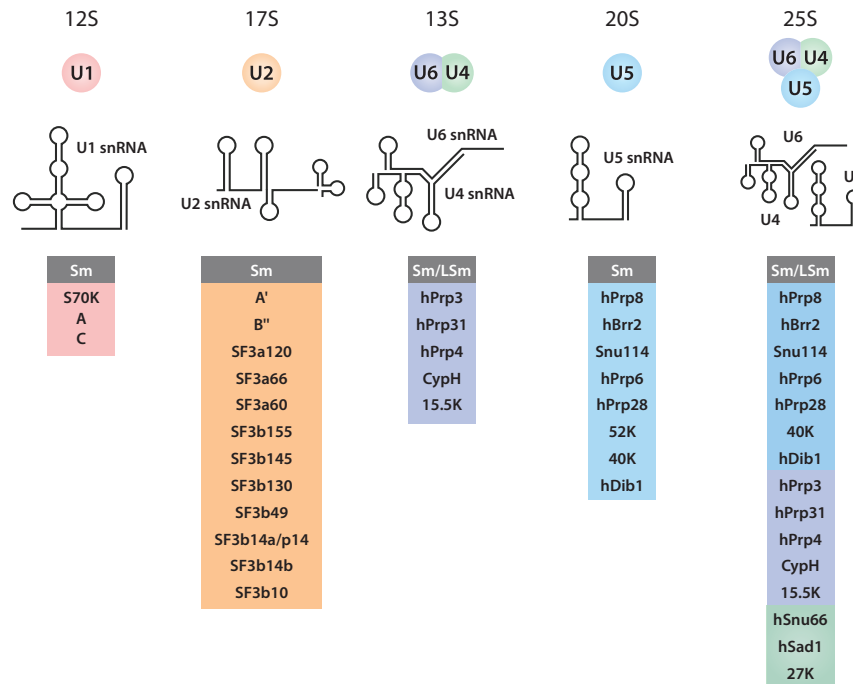


Figure 1.4: Protein composition and snRNAs of the human UsnRNPs. Besides the names of the proteins, the most probable secondary structure of the UsnRNAs is provided which however is remodelled during the splicing cycle. The Sm/LSm proteins are indicated as a group of common proteins on top, followed by the names of the proteins that specifically associate with the individual snRNP. The size of each UsnRNP is shown on top in Svedberg units. This figure was adapted from Will and Lüthmann (2011).

and obtains a unique γ -monomethyl cap (Singh and Reddy, 1989). In contrast to the other UsnRNAs, the U6 snRNA lacks an Sm-site; instead Sm-like proteins (LSm 2-8) bind to the 3'-end of the U6 snRNA (Achsel *et al.*, 1999) and are thought to be responsible for the nuclear retention of the U6 snRNA (Spiller *et al.*, 2007).

Within each UsnRNP, several proteins associate with the UsnRNAs and the Sm proteins (see figure 1.4).

1.3.2.2 Protein factors of the spliceosome

Besides the proteins that are part of the UsnRNPs, many additional proteins associate with the spliceosome, showing that the spliceosome is a protein-rich RNP complex (Jurica and Moore, 2003; Wahl *et al.*, 2009; Chen and Cheng, 2012).

The distinct stages of the splicing cycle possess different protein compositions,

1 Introduction

and in metazoans, the number of implicated different proteins averages between 110 and 170 (Fabrizio *et al.*, 2009). In yeast, the number of spliceosomal proteins is about 90 and therefore significantly lower compared to other species. For about 85 % of these proteins, homologs in the human spliceosome have been identified, indicating that the principal design of the spliceosome is conserved in yeast (Fabrizio *et al.*, 2009). Most of the spliceosomal proteins absent in yeast are suggested to be involved in the regulation of alternative splicing in the human spliceosome, a process which does not exist in yeast (Fabrizio *et al.*, 2009).

In a study performed by Fabrizio *et al.* (2009), the protein composition of affinity-purified and *in vitro* assembled complexes B, B^{act} and C was determined by mass spectrometry, providing a descriptive example for the dynamics of the spliceosome as well as the reduced number of proteins present in yeast spliceosome (shown in figure 1.5). During the transition from the B to the B^{act} complex, about 35 proteins and two UsnRNPs are released from the spliceosome, whereas 12 proteins are recruited. The changes leading to the formation of the complex C are significantly smaller: two proteins were found to disassemble from and nine to join the spliceosome. Interestingly, homologous proteins that bind or dissociate during the transition of the B to the C complex have been identified in yeast and metazoans, indicating that the main features of the spliceosome is evolutionary conserved (Deckert *et al.*, 2006; Bessonov *et al.*, 2008; Fabrizio *et al.*, 2009; Herold *et al.*, 2009).

The dynamics of the splicing machinery is regulated and controlled by several mechanisms. Different post-translational modifications of spliceosomal proteins like phosphorylation, acetylation, methylation, hydroxylation and ubiquitination have been described (Will and Lührmann, 2011; Korneta and Bujnicki, 2012). Additionally, eight DExD/H-box ATPases which are conserved in yeast and humans have been found to promote the splicing cycle by supporting or disrupting RNA/RNA, RNA/protein and protein/protein interactions in an ATP-dependent manner (Staley and Guthrie, 1998), some of them are also involved in the rejection of erroneous substrates (for further details see section 1.5 on page 25). In humans, there are at least eight peptidyl-prolyl *cis/trans* isomerases (PPIases), which are associated with the spliceosome; however their concrete function remains to be elucidated (Wahl *et al.*, 2009).

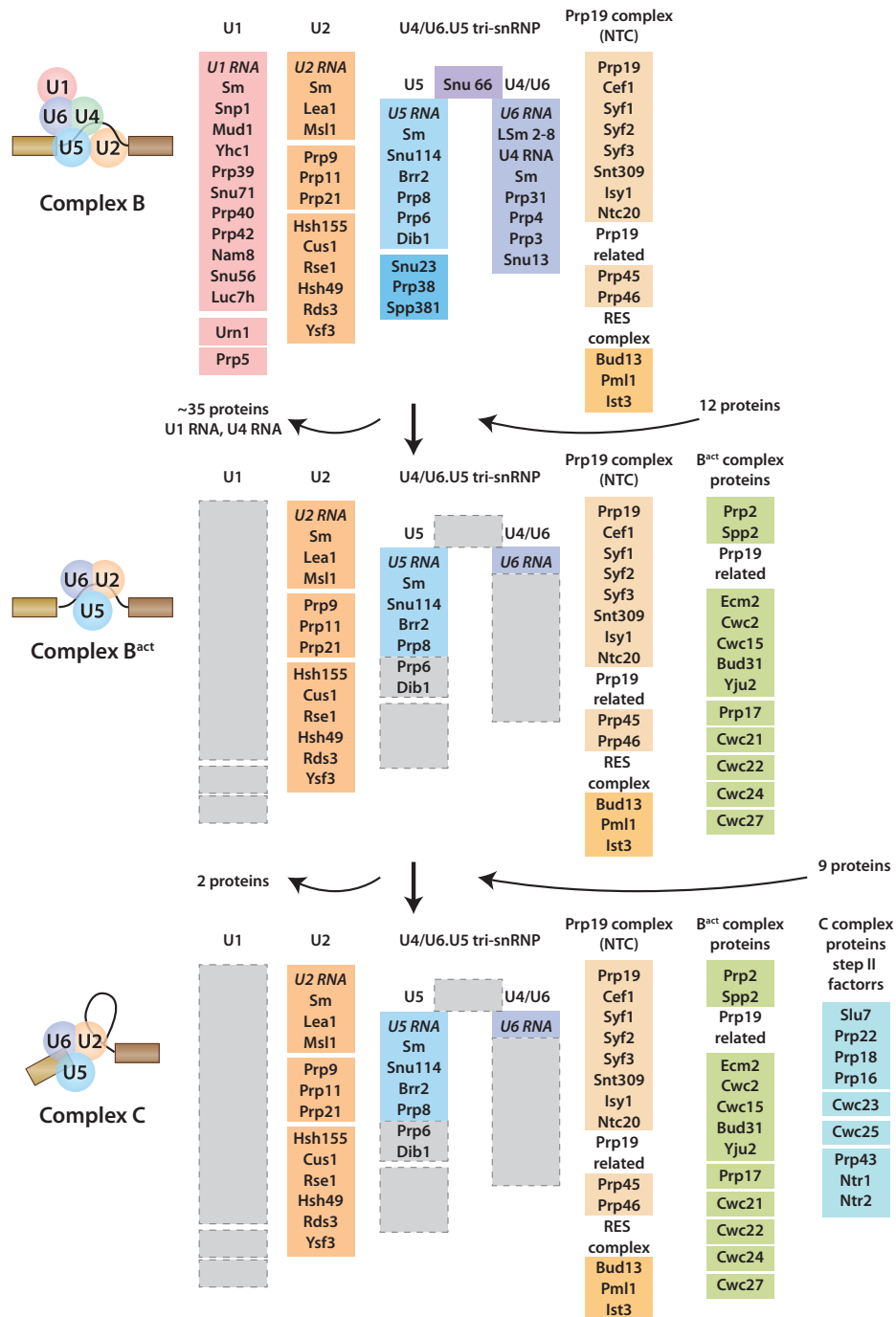


Figure 1.5: Changes in the composition of the yeast spliceosome during complex B, complex B^{act} and complex C. Proteins are shown according to their association with a spliceosomal complex or to their function, respectively. The identification of these proteins was performed via mass spectrometry. This figure was adapted from Fabrizio *et al.* (2009).

1.3.3 The catalytic center of the spliceosome

Based on the finding that the splicing reaction consists of two consecutive transesterification reactions, it was suggested that the spliceosome contains two distinct active sites (Moore and Sharp, 1993). Currently, the evidence is increasing that there is only one catalytic center in the spliceosome which changes the conformation during the two reactions, with an equilibrium between the two states (Query and Konarska, 2004; Konarska and Query, 2005). This model is also supported by the reversibility of the splicing reaction (Tseng and Cheng, 2008) and by data indicating that the transition between the two states could occur in several phases, attained by multiple remodelling events (Smith *et al.*, 2008; Liu *et al.*, 2007).

There are several indications that the active site of the spliceosome is stabilised by proteins, but that the catalysis is mainly RNA-based, performed by the U2 and U6 snRNP (Valadkhan, 2005; Wachtel and Manley, 2009) and two metal ions (Sontheimer *et al.*, 1997). An important catalytic role is assumed for an intramolecular stem loop located in the central region of the U6 snRNA (Wolff and Bindereif, 1993). The interaction sites of the U2, U5 and U6 snRNA with the pre-mRNA resemble those of the group II introns (Keating *et al.*, 2010). On the other hand, crystallographic studies on Prp8 have revealed that it contains domains resembling a bacterial group II intron reverse transcriptase, a type II restriction endonuclease and an RNase H-like domain (Pena *et al.*, 2008; Ritchie *et al.*, 2008; Yang *et al.*, 2008; Galej *et al.*, 2013). These domains are organised in a comparable manner known for group II self-splicing introns, which indicate that they share the same evolutionary origin and that tasks of the catalysis have been transferred to proteins in the spliceosome (Toor *et al.*, 2008; Chan *et al.*, 2012; Galej *et al.*, 2013).

1.4 RNA helicases

RNA helicases are present in eukaryotes, in prokaryotes and in viruses and not only essential components during splicing, but also implicated in various processes like translation initiation, ribosomal biogenesis, RNA transport and replication of viruses (Cordin *et al.*, 2006; Bleichert and Baserga, 2007). Initially, the term RNA helicase was used for enzymes that couple NTP hydrolysis to the unwinding of RNA duplexes, but there is evidence that members of this family are also able to remodel protein-RNA complexes without RNA unwinding, to anneal RNA

strands, to act as RNA clamps and to stabilise folding intermediates (Bleichert and Baserga, 2007; Pyle, 2008).

1.4.1 Classification of helicases

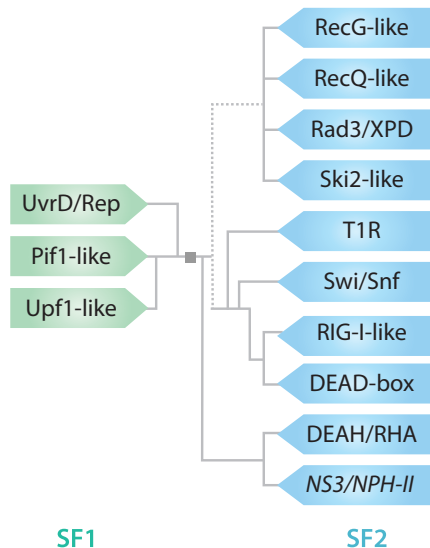


Figure 1.6: Cladogram showing the phylogenetic relations between members of the SF1 (left) and of the SF2 helicases (right), which comprise in total twelve families and one group (shown in italic). The analysis was based on sequence features and the domain organisation. The starting point is represented by a square, the length of the branches is not scaled. The families are named after prominent members (Ski2, Upf1 etc.) or conserved sequences in the Walker B motif (motif II, DEAD and DEAH). The dotted line indicates an uncertain region of the cladogram. This figure was adapted from Fairman-Williams *et al.* (2010).

The most recent classification of both DNA and RNA helicases was based on sequence motifs as well as structural and mechanistic features and has defined six superfamilies (SFs) (Fairman-Williams *et al.*, 2010). Hexameric and toroidal helicases comprise SF3 to SF6 whereas members of the SF1 and SF2 do not exploit such kind of quarternary structures. The phylogenetic alignment of all SF1 and SF2 members from *H. sapiens*, *S. cerevisiae*, *E. coli* and several viral helicases has revealed the presence of three families¹ in the SF1 and of nine families as well as one group² in the SF2 (see figure 1.6).

The central feature of all SF1 and SF2 helicases is a helicase core that is formed by two domains resembling the bacterial protein recombinase A (RecA), and hence are called RecA-like domains (Subramanya *et al.*, 1996). The interface of these two domains is responsible for ATP and RNA binding and contains several conserved motifs (Bork and Koonin, 1993; Rocak and Linder, 2004). Solely the DEAD-box family contains exclusively RNA helicases, whereas among the other families of the SF2, RNA as well as DNA helicases are present. Several viral helicases of the

¹a family includes three or more members

²a group contains less than three members

1 Introduction

NS3/NPH-II group and the Upf1-like family have been shown to unwind both DNA and RNA substrates (Fairman-Williams *et al.*, 2010).

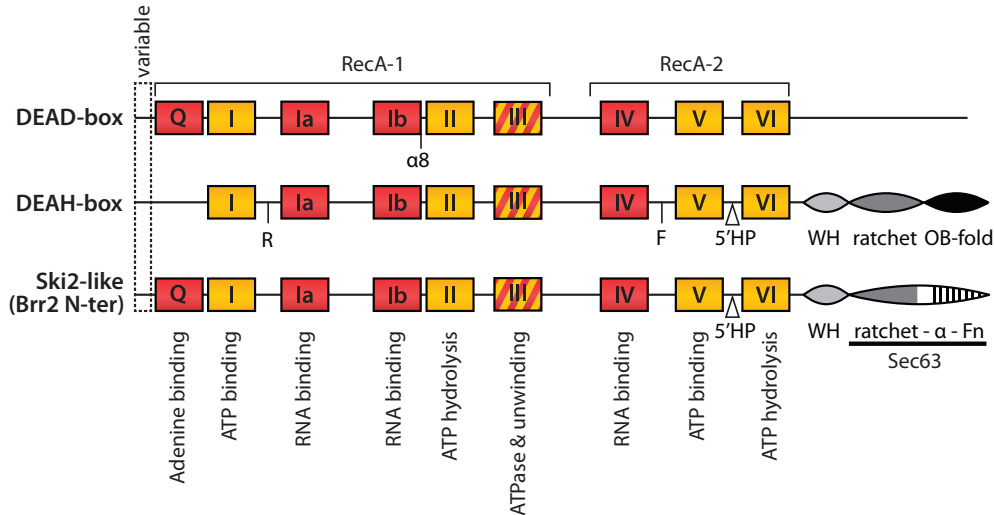


Figure 1.7: Domain overview of the SF2 helicase families DEAD-box, DEAH-box and Ski2-like. For the helicase core (RecA-1 and RecA-2) the conserved motifs for the interaction with ATP (red), the RNA (yellow) or both (yellow/red) as well as specific features are shown (α -helix-8 ($\alpha 8$), conserved arginine (R) and phenylalanine (F) residues and the 5'-hairpin (5'HP)). While for the DEAD-box proteins, both N- and C-terminal parts are variable, the members of the DEAH-box and Ski2-like families contain conserved domains in their C-terminal regions. This figure was adapted from Cordin *et al.* (2012).

As all spliceosomal helicases are either member of the DEAD-box, the DEAH-box or the Ski2-like family, it is focused in the following sections on the features and characteristics of these families. The helicase core of the spliceosomal helicases contains up to nine different conserved motifs (see figure 1.7). Motifs I (or Walker A), II (or Walker B), V and VI are presumably responsible for ATP binding and hydrolysis (Walker *et al.*, 1982; Pause *et al.*, 1993), whereas motifs Ia, Ib and IV are supposed to be implicated in RNA binding (Cordin *et al.*, 2006). The motif III has a dual implication in both ATP hydrolysis and RNA unwinding (Pause and Sonenberg, 1992).

A motif containing a conserved glutamate residue, the Q-motif, has been identified to be responsible for the interaction with the adenine of ATP. This motif is present exclusively in DEAD-box and Ski2-like helicases and leads to their specificity for ATP (Tanner *et al.*, 2003). In contrast, members of the DEAH-box family lack this motif and hence are able to hydrolyse any NTP. Additional conserved residues or motifs seem to be important in the process of unwinding performed by the helicases, depending on their family.

1.4.2 DEAD-box, DEAH-box and Ski2-like helicases

Concerning the mechanistic view of unwinding, a high degree of processivity during unwinding has been predicted for DNA helicases that are implicated in the replication process (Singleton *et al.*, 2007). There is much evidence that the unwinding mechanism of the DEAD-box proteins occurs in a non-processive manner (Hilbert *et al.*, 2009). The other members of the SF2, including DEAH/RHA, NS3/NPH-II and Ski2-like families, are thought to show a moderate to high processivity (Pyle, 2008). These differences are due to specific features of these families which will be described in this section.

1.4.2.1 Characteristics of DEAD-box helicases

As they have been proposed to be non-processive, DEAD-box proteins act in general on short duplexes and are thought to function as conformational switches (Pyle, 2008). There is evidence that DEAD-box proteins bind to their RNA substrate, induce a local destabilisation and thereby provoke a base pair disruption (Yang *et al.*, 2007).

3D structures of several DEAD-box proteins in absence of RNA and ATP show a very variable positioning of the two RecA-like domains relative to each other which are connected by a flexible linker (Caruthers *et al.*, 2000; Story *et al.*, 2001; Cheng *et al.*, 2005). This conformation has been defined as "open conformation" (Hilbert *et al.*, 2009). In the helicase YxiN from *B. subtilis*, no nucleotide binding occurs when the two RecA-like domains are expressed separately and the two monomers are subsequently mixed (Karow *et al.*, 2007). For the DEAD-box protein Hera, in contrast, evidence exists that the isolated N-terminal RecA-like domain is able to bind AMP, ADP and ATP on its own (Rudolph *et al.*, 2006). Furthermore, recent studies of the DEAD-box helicase Mss116 have revealed an ATP affinity of the N-terminal and an RNA affinity of the C-terminal RecA-like domain (Mallam *et al.*, 2012). These findings reveal that the members of the DEAD-box helicases have, besides their similarities, also differences e. g. concerning their interaction modes with substrates.

In presence of RNA and ATP, the two RecA-like domains of DEAD-box proteins are interacting very tightly. The existence of this "closed conformation" has been shown within crystal structures containing DEAD-box proteins in complex with ATP analogues and RNA (Sengoku *et al.*, 2006; Andersen *et al.*, 2006; von Moeller *et al.*, 2009) as well as by FRET experiments (Theissen *et al.*, 2008). One example

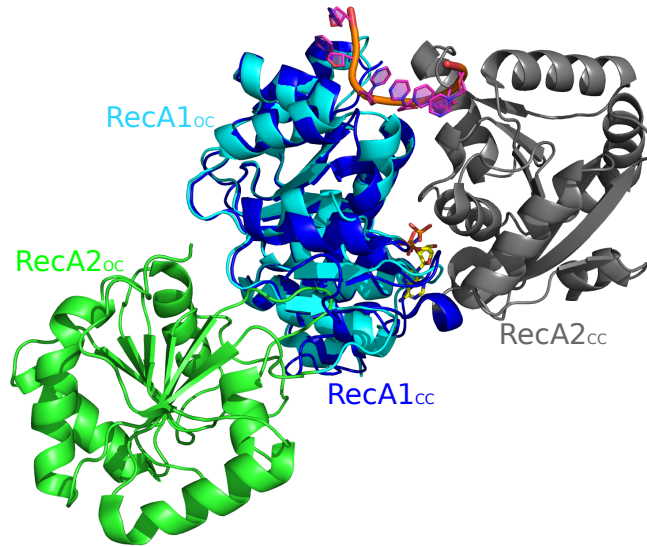


Figure 1.8: Superposition of the open and the closed conformation of the DEAD-box ATPase eIF4AIII. The two RecA-like domains of the open conformation are shown in cyan/green, whereas those of the closed conformation are shown in blue/grey. The closed conformation state is adopted in the presence of an ATP analogue and a polyU oligonucleotide and causes an enormous conformational change of the helicase core (Andersen *et al.*, 2006). The two N-terminal RecA-like domains of both conformations have been superposed using Coot. The structures have been visualised using PyMOL, the pdb-IDs are 2hxy for the open and 2hyi for the closed conformation.

for the differences between the open and the closed conformation is provided by the superposition of the DEAD-box protein eIF4AIII in the ligand-free form and in complex with ATP and ssRNA (see figure 1.8). The relative positioning of the two domains in the open conformation is probably due to crystal packing effects. In the closed conformation, the presence of the nucleotide and the nucleic acid induces the tight interaction of both RecA-like domains with each other as well as with the nucleotide and the RNA (Andersen *et al.*, 2006).

There is no unique model describing the sequential arrangement for the binding of ATP and RNA by DEAD-box proteins. Cooperativity during the binding events has been described as well as its absence for different helicases (Hilbert *et al.*, 2009), showing again the heterogeneity among the members of the DEAD-box family. On the contrary to this, an integrative mechanism of unwinding has been proposed. In all crystal structures representing a DEAD-box protein in presence of ATP and RNA, helix $\alpha 8$ of the N-terminal domain was identified to interact with the nucleic acid strand and induces a kink in the RNA. Within this kink, base pairs of the RNA duplex are separated. This event has been proposed to initiate the unwinding process (Sengoku *et al.*, 2006; Karow and Klostermeier,

2009). The conformation of this particular helix is obviously specific for DEAD-box proteins, as the corresponding helices in other members of the SF2 are unable to distort the substrate RNA (Hilbert *et al.*, 2009).

The currently most widely accepted model proposes that the binding of ATP and RNA induces the formation of the closed conformation that could lead to an activation step. The subsequent interaction between helix $\alpha 8$ and one RNA strand induces the local strand separation (Hilbert *et al.*, 2009). Subsequently, the hydrolysis of ATP leads presumably to a decrease in the affinity for RNA which could trigger its release (Lorsch and Herschlag, 1998; Nielsen *et al.*, 2009), but the exact mechanism of strand release remains elusive.

1.4.2.2 Characteristics of DExH-box helicases

Although all DExH-box helicases contain also a helicase core consisting of two RecA-like domains, they do not share similarities with the DEAD-box proteins regarding the mechanism of unwinding. In order to reveal this mechanism of these processive helicases, intensive studies have been performed during the last years especially on members of the DEAH-box/RHA and the Ski2-like families as well as the the NS3/NPH-II group. These three families are traditionally referred to as DExH-box helicases.

Two principal models describing the physical unwinding mechanism of processive helicases will be presented in the following: the Brownian motor model and the backbone stepping motor model.

The Brownian motor model

The central feature of the Brownian motor model is the assumption that the helicases perform cycles of states with strong and weak binding to RNA, which would allow a movement on the nucleic acid template. These cycles could be regulated by ATP hydrolysis. Duplex unwinding could hence be induced by the motion on the strand (Pyle, 2008). For the helicase NS3 of the hepatitis C virus (HCV) it has been shown, that RNA affinity is low in the ATP bound state and increases in the ADP bound state or when no bound nucleotide is present (Levin *et al.*, 2003). A comparable behavior has been shown for the spliceosomal DEAH-box helicases Prp22 and Prp43, although in this case with low RNA affinity in presence of bound ADP and high affinity when ATP is bound (Tanaka and Schwer, 2005; He *et al.*, 2010).

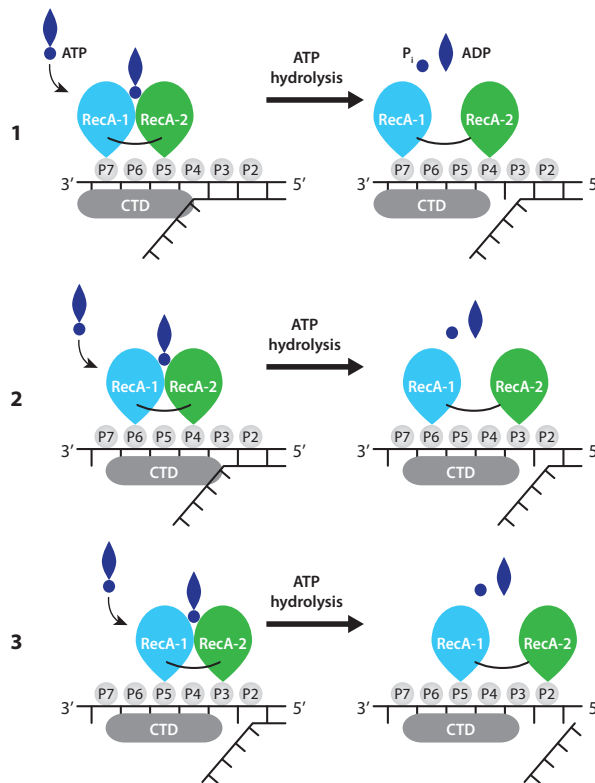
1 Introduction

This basic model is predicated solely on thermodynamic properties and does not refer to the flexibility and rigidity of the helicase itself (Pyle, 2008).

The backbone stepping motor model

Within the backbone stepping motor model it is proposed that the helicase moves directionally along the backbone of the single-stranded part of the RNA in a distinct step size, also driven by ATP hydrolysis. The two RecA-like domains get close to each other in the ATP-bound state and more remote in the ADP-bound state. The change of the affinities for RNA in the different nucleotide states is thought to result in an increased flexibility of the helicase and not in releasing the substrate RNA as proposed by the Brownian motor model. This behaviour would induce a movement along the phosphate backbone of the nucleic acid (Pyle, 2008).

Figure 1.9: The backbone stepping motor model proposed for the HCV NS3 helicase. Three cycles of the helicase stepping in 5'-direction along the RNA tracking strand upon ATP hydrolysis are shown (numbers 1-3), leading to strand separation. The two helicase core domains are colored in cyan (RecA-1) and in green (RecA-2), whereas the C-terminal domain (CTD) is colored in grey. The tracking strand is numbered with reference to the phosphate backbone, the contact between the two RecA-like domains and the nucleic acid is guided by two conserved threonine residues. This figure was adapted from Appleby *et al.* (2011).



Evidence for the backbone stepping motor model was revealed by several crystal structures and FRET experiments of the HCV NS3 helicase (Myong *et al.*, 2007; Appleby *et al.*, 2011) and the crystal structure of the Ski2-like helicase Hel308 in complex with dsDNA (Hopfner and Michaelis, 2007; Büttner *et al.*, 2007). A tryptophan residue located in an additional C-terminal domain of the HCV

NS3 helicase was identified that might be stacked between the dsRNA and be a physical reason for unwinding when the RecA-like domains move (Myong *et al.*, 2007). Using single molecule techniques, it has been shown that the HCV NS3 helicase performs unwinding in regular steps of three nucleotides, occurring in three sub-steps (Dumont *et al.*, 2006; Myong *et al.*, 2007). The according model proposed by Appleby *et al.* (2011) is shown in figure 1.9. The stepping itself could be guided by two highly conserved threonine residues that are located on each RecA-like domain, contacting the phosphate backbone of the RNA substrate (Hopfner and Michaelis, 2007). These threonine residues have been reported to be within a distance of three nucleotides in the ATP-free and of two nucleotides in the ATP-bound state (Appleby *et al.*, 2011). Thus, every small step could be driven by the hydrolysis of one molecule ATP (Pyle, 2008; Gu and Rice, 2010; Appleby *et al.*, 2011).

A β -hairpin protruding from the C-terminal RecA-like domain was observed in the crystal structures of the Ski2-like proteins Hel308 (Büttner *et al.*, 2007), Mtr4 (Weir *et al.*, 2010) and Ski2 (Halbach *et al.*, 2012) as well as in the DEAH-box/RNA helicase A-like (RHA) family member Prp43 (He *et al.*, 2010; Walbott *et al.*, 2010). The β -hairpin has been proposed to be a double strand breaker by interacting with the 5'-end of the RNA substrate, comparable to the conserved tryptophan in the HCV NS3 helicase, and is therefore called 5'-hairpin (5'HP). Furthermore the presence of 5'-HP provides an indication for similar mechanisms of the Ski2-like and DEAH-box/RHA helicases. On the other hand, the different lengths of the 5'HP in the distinct families might also be an indication for different functional roles which remain elusive.

1.5 The spliceosomal DExD/H-box helicases

As introduced earlier in section 1.3, eight conserved DExD/H-box ATPases in yeast and humans act as molecular motors of the spliceosome (Wahl *et al.*, 2009; Cordin *et al.*, 2012). Although several of them show RNA unwinding activity *in vitro*, there is not much information available about their precise target sequences within the spliceosome. Furthermore, there are indications that they might act as remodelling ATPases or RNPsases (Staley and Guthrie, 1998; Schwer, 2001).

1.5.1 Promotors of the splicing cycle

During the splicing cycle, the three DEAD-box proteins Sub2, Prp5 and Prp28 are involved in the assembly of the spliceosome, while the four DEAH-box proteins Prp2, Prp16, Prp22 and Prp43 play a role in catalysis and product release (Cordin *et al.*, 2012). The Ski2-like helicase Brr2 has been suggested to play a dual role in the assembly as well as during disassembly. The steps in which these proteins are involved are summarised in table 1.1.

Table 1.1: Overview of the functions of the eight spliceosomal DExD/H-box proteins from *S. cerevisiae*. For each protein, the family and the function is provided. Brr2 is shown twice as it is suggested to play a role in both the assembly and the disassembly of the spliceosome. All proteins indicated with an asterisk are furthermore involved in proofreading during the splicing process.

Protein	Family	Function
<i>Assembly</i>		
Sub2	DEAD-box	Recruitment of the U2 snRNP
Prp5*	DEAD-box	Remodeling of the U2 snRNP
Prp28	DEAD-box	Release of the U1 snRNP
Brr2	Ski2-like	Release of the U4 snRNP
<i>Catalysis</i>		
Prp2	DEAH-box	Release of SF3a and SF3b
Prp16*	DEAH-box	Release of Yju2 and Cwc25
Prp22*	DEAH-box	Identification of the 3'ss
<i>Disassembly</i>		
Prp22*	DEAH-box	Release of the mature mRNA
Prp43*	DEAH-box	Disassembly of the intron-lariat spliceosome
Brr2	Ski2-like	Disassembly of the intron-lariat spliceosome

Crystallographic as well as biochemical studies on the human Sub2-orthologue, UAP56, have revealed similarities to the helicase eIF4A (Zhao *et al.*, 2004; Shi *et al.*, 2004). UAP56 seems to have a second function in nuclear export (Shen, 2009). The ATPase activity of the second spliceosomal DEAD-box protein, Prp5, is specifically enhanced in the presence U2 snRNA *in vitro*, indicating that this stimulation might assure the specificity of its action *in vivo* (O'Day *et al.*, 1996). The phosphorylation of the third DEAD-box protein, Prp28, is required for the formation of the B complex in humans (Mathew *et al.*, 2008), but neither ATPase nor helicase activity of yPrp28 have been detected *in vitro* so far (Andreas Schmitt, personal communication).

The Ski2-like helicase Brr2 has a remarkable domain organisation as it consists

of two helicase cassettes, both followed by a Sec63 homology unit. Interestingly, exclusively the N-terminal helicase cassette was reported to be required for activity *in vivo* (Kim and Rossi, 1999). Crystallographic and functional studies of human Brr2 have revealed recently that the C-terminal cassette is able to stimulate the helicase activity of the N-terminal cassette (Santos *et al.*, 2012). Moreover, the helicase activity can be inhibited through the RNase H-like domain of Prp8 which interrupts the interaction between Brr2 and the U4 snRNA (Mozaffari-Jovin *et al.*, 2012).

For two of the four spliceosomal DEAH-box proteins, namely Prp2 and Prp43, interaction partners have been identified which both contain a G-patch motif (see section 1.6.4). Prp2 is recruited by Spp2 (Silverman *et al.*, 2004), whereas Prp43 is recruited and activated by Ntr1 (Tsai *et al.*, 2005). For the two other DEAH-box proteins, Prp16 and Prp22, no interaction with a G-patch protein has been reported so far.

1.5.2 Proofreading during the splicing cycle

Besides their functions in promoting the splicing cycle, the DExD/H-box ATPases Prp5, Prp16 and Prp22 have been reported to be implicated in the proofreading during pre-mRNA splicing (Egecioglu and Chanfreau, 2011; Semlow and Staley, 2012). These proteins act by promoting optimal and rejecting erroneous substrates, thereby competing with productive events and leading to branch pathways in case of suboptimal substrates (Burgess and Guthrie, 1993; Mayas *et al.*, 2010; Koodathingal *et al.*, 2010). While the proofreading action of Prp16 is reversible, Prp43 acts irreversibly due to its general disassembly function (Koodathingal *et al.*, 2010). Recently, it was reported that Prp43 as part of the NTR complex can bind to the spliceosome at distinct stages of the splicing cycle, supporting its disassembly function during proofreading (Chen *et al.*, 2013).

Interestingly, the 3'-end of telomerase RNA in *S. pombe* was suggested to be generated by the first catalytic step of the spliceosome (Box *et al.*, 2008). Although there is a consensus 3'ss, the second reaction step is not performed and the RNA is discarded by the spliceosome, indicating that local regulation of the proofreading process is also possible (Semlow and Staley, 2012).

1.6 The DEAH-box protein Prp43

In the following, further details will be provided concerning the function and structure of the terminator of the splicing cycle, the DEAH-box ATPase Prp43.

1.6.1 The role of Prp43 during the splicing cycle

Prp43 has been identified as an essential component of the spliceosome due to the finding that its deletion leads to the accumulation of excised intron lariats (Arenas and Abelson, 1997). Like all other seven spliceosomal DExD/H-box proteins, Prp43 is an RNA-dependent ATPase and possesses furthermore a marginal unwinding activity *in vitro* (Martin *et al.*, 2002; Tanaka and Schwer, 2006). During the disassembly of the intron-lariat spliceosome, the action of Prp43 is dependent on the interaction with the heterodimer Ntr1/Ntr2 (Tsai *et al.*, 2005, 2007). The NTR complex is relatively dynamic *in vivo* as about 30 % of the Ntr1/Ntr2 heterodimer and 70 % of Prp43 have been found to be unbound *in vivo* (Tsai *et al.*, 2007). Tsai and colleagues have proposed a model for the NTR-mediated disassembly of the intron-lariat spliceosome (see figure 1.10).

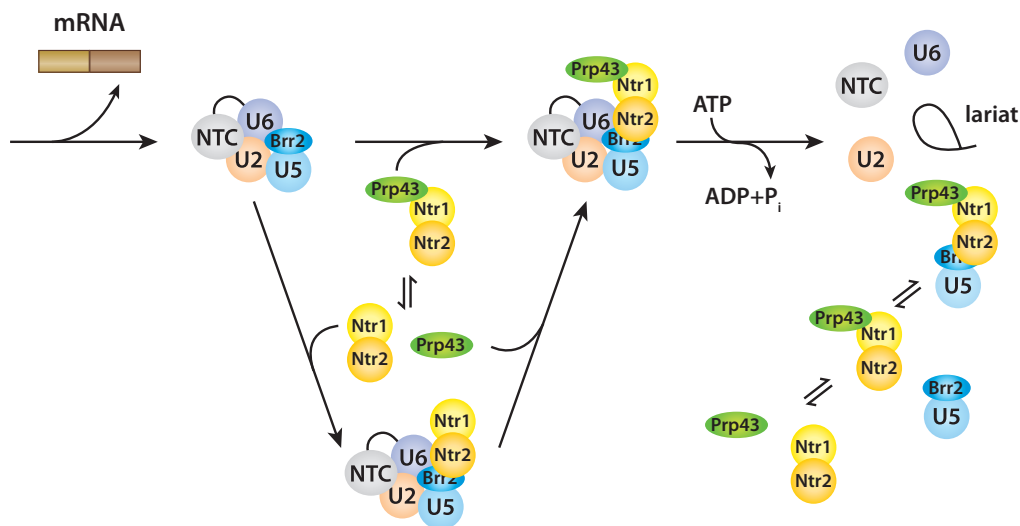


Figure 1.10: Schematic representation of the interactions of Prp43 with the Ntr1/Ntr2 dimer which leads to the disassembly of the intron-lariat spliceosome after the release of the mature mRNA. Equilibrium arrows indicate states of permanent association and dissociation of the components. Whereas the NTR proteins can bind to each other and to the postspliceosome ATP-independently, ATP hydrolysis is required for the release of the lariat intron, the Nine-Teen complex (NTC), the U2, U5 and U6 snRNP and the Ski2-like helicase Brr2. This figure was adapted from Tsai *et al.* (2007).

Yeast two hybrid interaction studies have revealed that the interaction between Ntr1 and Prp43 is mediated by the N-terminal part of Ntr1 whereas the middle part of Ntr1 is responsible for the interaction with Ntr2 (Tsai *et al.*, 2005). It has been shown by *in vitro* studies that the 120 N-terminal residues of Ntr1 are sufficient for the interaction and that they strongly stimulate the helicase activity of Prp43 – therefore Ntr1 has been proposed to act as an activator of Prp43 (Tanaka *et al.*, 2007). In the absence of Ntr1, Prp43 was reported to show a 5' to 3' unwinding activity (Tanaka and Schwer, 2006), whereas in the presence of Ntr1, it is capable to unwind substrates in both 5' to 3' as well as 3' to 5' direction (Tanaka *et al.*, 2007).

The N-terminal part of Ntr1 that interacts with Prp43 contains a G-patch motif which is located between residues 60 and 106 (Conserved Domain Database, CDD), 59 and 106 (Simple Modular Architecture Research Tool, SMART) or 61 and 105 (Protein families database, Pfam), respectively, depending on the database. A direct interaction of Prp43 and Ntr2 has not been reported yet.

1.6.2 The role of Prp43 in ribosomal biogenesis

Besides the dissociation of the intron-lariat spliceosome, Prp43 is also implicated in the biogenesis of the ribosome. Several binding sites of Prp43 to the 18S and 25S pre-rRNAs have been reported (Lebaron *et al.*, 2005; Bohnsack *et al.*, 2009). Interestingly, Prp43 interacts during this process with another G-patch protein, Pfa1 (Pandit *et al.*, 2009; Pertschy *et al.*, 2009).

Pfa1 is able to stimulate both the ATPase as well as the helicase activity of Prp43 *in vitro* and interacts with the helicase through its N-terminal part and with its C-terminal part which contains the G-patch motif (Lebaron *et al.*, 2009). Whereas Pfa1 is not required for the integration and retention of Prp43 in pre-ribosomal particles, it seems to be essential for the processing of the 20S pre-rRNA *in vivo* (Lebaron *et al.*, 2009).

1.6.3 The crystal structure of Prp43

The crystal structure of Prp43 from *S. cerevisiae* in complex with ADP has been published independently by two groups several months after this thesis was started – it was the first crystal structure analysis of a DEAH-box protein (Walbott *et al.*, 2010; He *et al.*, 2010). The structure reveals that the two RecA-like domains are followed by a winged helix (WH), a ratchet and a C-terminal OB-fold domain (see

1 Introduction

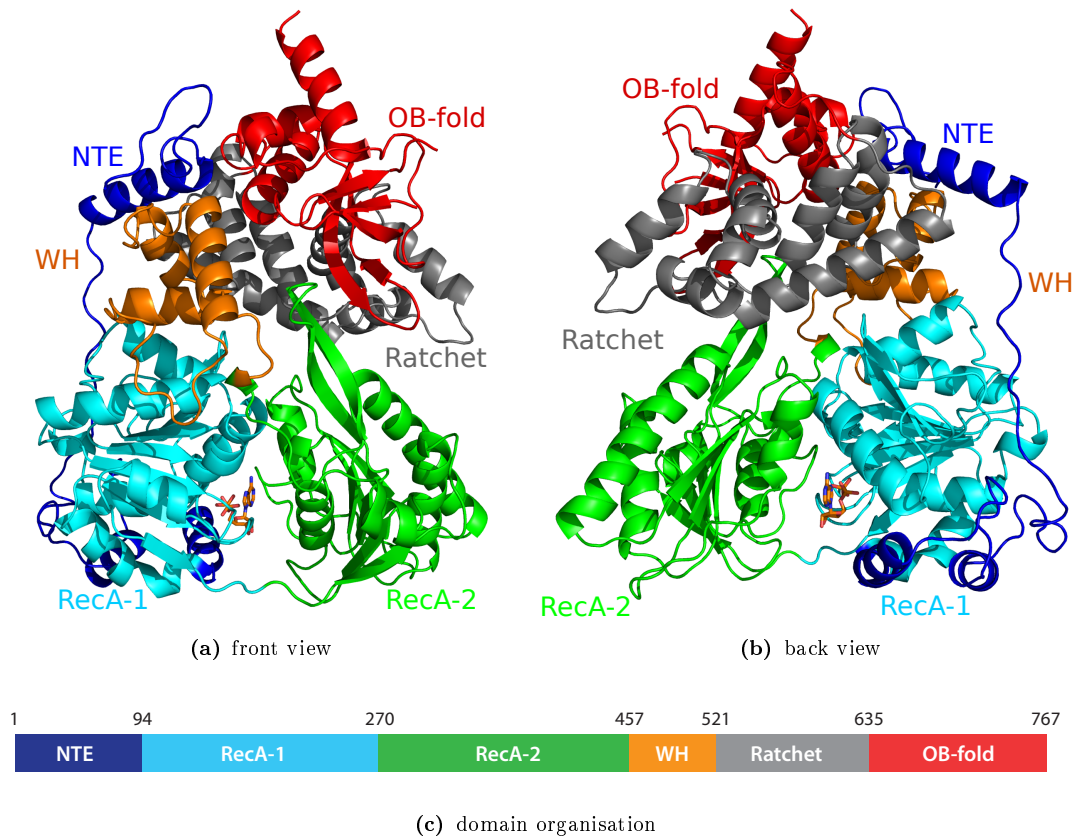


Figure 1.11: The crystal structure of yPrp43 (pdb-ID 2xau) in front (a) and back (b) view and the domain organisation of yPrp43 (c). Besides the two central RecA-like domains, a Winged helix (WH), a Ratchet and a C-terminal OB-fold domain is present. Additionally, Prp43 contains an N-terminal extension consisting of about 100 amino acids. The ADP molecule is bound between both RecA-like domains. The structure was visualised using PyMOL.

figure 1.11). While OB-fold domains have been shown to be able to interact with both, nucleic acid as well as proteins (Murzin, 1993), the role of the winged helix domain, which has been found in several helicases, remains elusive (Woodman and Bolt, 2011). Additionally, an N-terminal extension (NTE) is present which includes about 100 residues and links the first RecA-like domain to the WH domain. An important feature in the structure is furthermore a 5'HP that protrudes from the second RecA-like domain and interacts with the OB-fold domain via a salt bridge between residues R407 and D675.

In the crystal structure of Prp43, the ADP molecule is localised in the interface of the two RecA-like domains. Its adenine moiety is stacked between the two domains, a binding mode which is different to the DEAD-box protein Vasa (Sengoku *et al.*, 2006) and was therefore proposed to be specific for DEAH/RHA

and NPH-II helicases, as the base is not recognised specifically (Walbott *et al.*, 2010).

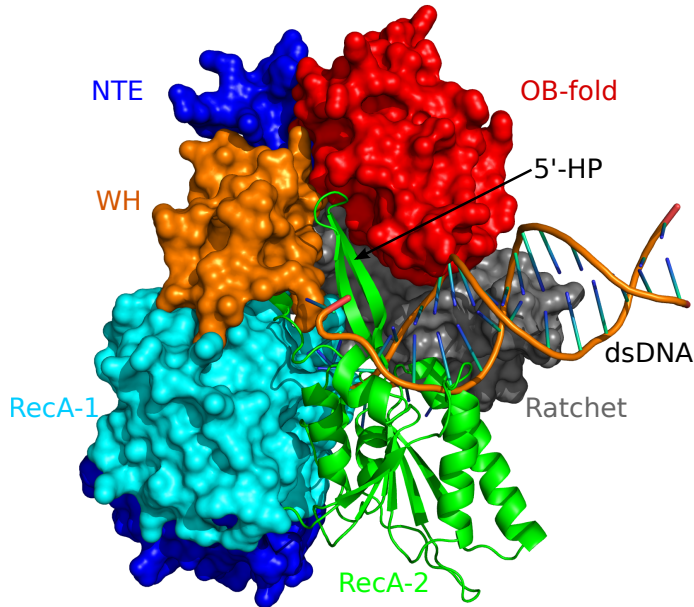


Figure 1.12: Superposition of the dsDNA present in the structure of the Ski2-like helicase Hel308 (pdb-ID 26pr) and the structure of Prp43 (2xau). Solely the C-terminal RecA-like domain is presented as cartoon, whereas the other domains are shown as surfaces. The color code for the domains is the same like in figure 1.11, the position of the 5'-hairpin (5'-HP) is indicated by an arrow. The structures were visualised using PyMOL.

Prp43 shares, besides the two RecA-like domains, the WH and the Ratchet domains with the Ski2-like helicase Hel308 whose the structure was solved in complex with a dsDNA substrate (Büttner *et al.*, 2007). The homology of these two proteins allows their superposition and hence the modelling of the dsDNA from the structure of Hel308 into the structure of Prp43 (see figure 1.12). This model provides an idea how a double-stranded RNA substrate might interact with Prp43. The 3'-overhang of the dsDNA is bound in a tunnel between the interface between the two RecA-like domains and the C-terminal domains of Prp43, especially along the extended Ratchet helix. The 5'HP is located between the two nucleic acid strands and therefore might play an important role during the strand separation.

1.6.4 The G-patch motif

The G-patch motif has been identified using a bioinformatical approach and encompasses about 50 residues including at least five highly conserved glycine residues in the consensus sequence $\text{hxx}_3\text{Gax}_2\text{GxGhGx}_4\text{G}^3$ (Aravind and Koonin, 1999). The G-patch motif is found exclusively in eukaryotic and some viral pro-

³h=hydrophobic, a=aromatic, x=any residue

1 Introduction

teins. First, it was suspected that the G-patch motif is an RNA-interacting module, but there is evidence that it interacts with RNA as well as with other proteins.

Besides Ntr1 and Pfa1, a third G-patch containing protein, Gno1, is suspected to interact with Prp43 during ribosome biogenesis (Guglielmi and Werner, 2002; Lebaron *et al.*, 2009). During the splicing cycle, the G-patch protein Spp2 has been identified to recruit the DEAH-box protein Prp2 to the spliceosome (Silverman *et al.*, 2004). In contrast to the interaction between Prp43 and Ntr1, no stimulatory effect of Spp2 of the ATPase or helicase activity of Prp2 has been reported yet.

There is evidence that two human G-patch proteins that interact with the human Prp43 are implicated in cancer disease: The GPATCH2 protein is over-expressed in breast cancer cells (Lin *et al.*, 2009), whereas the tumor suppressor and G-patch protein RBM5 has been reported to stimulate the helicase activity of human Prp43 *in vitro* (Niu *et al.*, 2012).

To date, no 3D structure of a protein containing a G-patch motif has been reported neither in an isolated state nor in complex with an interaction partner and hence the structural basis for the interaction of the G-patch motif with other proteins or nucleic acids is unknown so far.

1.7 Aims of this work

When this work was started, poor information was available regarding the regulation of the spliceosomal DExD/H box proteins by their interaction partners. The crystal structure of Prp43 reflects the structure of the helicase in its ADP-bound and hence in its post-catalytic state. The goal of this thesis was the investigation of the structural and functional basis for the changes that lead to the activation of the helicase activity of Prp43 upon binding of its activator Ntr1.

A major aim of this thesis was the crystallisation of Prp43 in its activated state, either in form of a ternary NTR complex or in a binary Prp43/Ntr1 complex. Furthermore, the mechanism of unwinding should be elucidated for Prp43 by the crystallisation of the complexes in presence of different nucleotides and RNAs. Therefore, the purification of the single components of the NTR complex as well as the purification of truncated variants of these proteins should be established.

As it could not be predicted if crystallisation conditions for these protein complexes will be identified or not, additional functional studies should be performed regarding the ATPase and unwinding activity of Prp43/Ntr1 complexes, using

1.7 Aims of this work

fluorescence-based, non-radioactive assays. Furthermore, as no structural and poor functional information was available for the G-patch motif, the investigation of the structure and the function of the G-patch motif using biochemical as well as biophysical methods was a major aim of this work.

2 Material and methods

In this chapter, information about the utilised material and the performed biochemical, biophysical and crystallographic methods is provided.

2.1 Material

2.1.1 Cell media

The cell media components were dissolved in desalted H₂O and autoclaved at 121 °C for 20 min prior to usage. If required, the solutions were supplemented with antibiotics to final concentrations of 100 µg/ml for ampicillin, 33 µg/ml for chloramphenicol and 30 µg/ml for kanamycin.

In this work, LB and 2YT media were used. Their composition is shown below.

LB medium		2YT medium	
1 % (w/v)	Tryptone	1 % (w/v)	Tryptone
0.5 % (w/v)	Yeast extract	1 % (w/v)	Yeast extract
0.5 % (w/v)	NaCl	1 % (w/v)	NaCl

To produce agar plates, the media was heated up in the microwave, cooled down, supplemented with the required antibiotics and poured into cell plates. After solidification, the plates were stored at 4 °C. The recipe for LB agar is shown below.

LB agar	
500 ml	LB medium
1.5 % (w/v)	Agar

2 Material and methods

2.1.2 Chromatography systems, columns and accessories

For the procedure of chromatographic methods, the machines Äkta Prime, Äkta Explorer and Äkta Purifier (GE Healthcare) were used. In order to load samples on the columns, superloops (150 ml, 50 ml and 10 ml) and loops (5 ml and 1 ml) were utilised. The used columns are mentioned in the following list.

GSH sepharose 5 ml	GE Healthcare
GSH sepharose 15 ml	GE Healthcare
HisTrap 5 ml Ni-NTA sepharose	GE Healthcare
HiPrep Desalting 5 ml	GE Healthcare
HiPrep Desalting 26/10	GE Healthcare
Source 30 Q	GE Healthcare
StrepTactin HP sepharose 5 ml	GE Healthcare
StrepTactin HP sepharose 22 ml	GE Healthcare
Superdex 75 (26/60)	GE Healthcare
Superdex 200 (26/60)	GE Healthcare
Superdex 200 (10/300)	GE Healthcare

2.1.3 Computer programs and data banks

The databases and bioinformatical tools used during this work are listed below.

APBS	Baker <i>et al.</i> (2001)
BLAST	Altschul <i>et al.</i> (1990)
ClustalW	Larkin <i>et al.</i> (2007)
CDD	Marchler-Bauer <i>et al.</i> (2011)
Coot	Emsley <i>et al.</i> (2010)
EMBOSS	Rice <i>et al.</i> (2000)
ESPrpt	Gouet <i>et al.</i> (1999)
MolProbity	Chen <i>et al.</i> (2010)
pdb2pqr	Dolinsky <i>et al.</i> (2007)
Pfam	Finn <i>et al.</i> (2010)
ProtSkin	Deprez <i>et al.</i> (2005)
PSIPRED	Buchan <i>et al.</i> (2010)
SMART	Letunic <i>et al.</i> (2012)

2.1.4 Crystallisation screens

The following crystallisation screens were routinely used during initial screening for potential crystallisation conditions (see section 2.4.1). Each screen contained 96 different buffer conditions.

Ammonium sulfate suite	Qiagen
JBScreen Classic (1, 2, 4, 5)	Jena BioScience
JBScreen Classic (6-8, 10)	Jena BioScience
JBScreen Nuc-Pro HTS	Jena BioScience
JCSG+ Suite	Newman <i>et al.</i> (2005)
MIDAS (1× and 0.5×)	Molecular Dimensions
Morpheus	Molecular Dimensions
Natrix HT	Hampton Research
PGA	Molecular Dimensions
ProPlex	Radaev <i>et al.</i> (2006)

2.1.5 Equipment

The machines and equipment used during this work are shown below.

Adjustable pipets	Eppendorf
Agarose gel electrophesis chamber	Bio-Rad
Binoculars	Carl Zeiss
Centrifuge Allegra 21R	Beckman Coulter
Centrifuge Avanti J-20 XPI	Beckman Coulter
Centrifuge Avanti J-30 I	Beckman Coulter
Fine balance	Sartorius
Fluorimeter Fluoromax III	Horiba Jobin Yvon, USA
HLC block thermostat BT1302	Techne
Incubator Mytron	Schütt
Innova 4230 incubator	New Brunswick
Magnetic stirrer IKAMAG REO	IKA
PCR mastercycler gradient	Eppendorf
PCR-Whatman Biometra T personal	Biometra
pH-Meter Beckman	Beckman Coulter
Photometer	Eppendorf
Pipette Controller Accu-Jet	Brand
Platform shaker Promax 1020	Heidolph
Rotor JA-20	Beckman Coulter
Rotor JA-30.50 Ti	Beckman Coulter
Rotor JA-8.1000	Beckman Coulter
Rotor S4180	Beckman Coulter
SDS-PAGE system Hoefer miniVE	Amersham Pharma
Table mixer RM5	Schütt
Table top centrifuge 5417R	Eppendorf
Table top centrifuge Micro centrifuge II	Sylvania
Thermomixer comfort	Eppendorf
Unitron incubator shakers	Infors

2 Material and methods

2.1.6 Fine chemicals

All fine chemicals and organic substances were purchased from the companies AppliChem, Bio-Rad, Fluka, IBA, Merck, MWG Biotech, Oxoid, Roth or Sigma-Aldrich and featured the purity grade *pro analysis*.

Generally, the provider with the lowest price was chosen.

2.1.7 Genomic DNA

The genomic DNA libraries of the organisms *Saccharomyces cerevisiae S288c* and *Chaetomium thermophilum DSM 1495* were part of the collection of the Department for Molecular Structural Biology (Prof. Ralf Ficner), Georg-August-Universität Göttingen.

2.1.8 Organisms

The used *E. coli* strains BL21(DE3), Rosetta 2(DE3), XL1-Blue, SoluBL(21) and ArcticExpress (DE3) were part of the strain collection of the Department for Molecular Structural Biology (Prof. Ralf Ficner), Georg-August-Universität Göttingen.

The *E. coli* TOP10 cells were purchased from IBA.

2.1.9 Plasmids

In table 2.1, the IBA StarGate entry vectors are shown. In table 2.2, all expression plasmids are listed that were used during this work.

Table 2.1: IBA StarGate entry vectors that have been designed and synthesised in this thesis for IBA StarGate cloning (see section 2.2.11.2). The coded protein and the involved residues for truncated proteins is provided within the name. The sequences of the oligonucleotides are listed in the supplemental table A.4.

Plasmid	Resistance	Oligonucleotides
pENTRY-IBA10_yPrp43	Kanamycin	S1, S2
pENTRY-IBA10_His6-yNtr1	Kanamycin	S4, S6
pENTRY-IBA10_yNtr1(1-120)	Kanamycin	S3, S8
pENTRY-IBA10_yNtr1(51-110)	Kanamycin	S5, S7
pENTRY-IBA10_yNtr2	Kanamycin	S9, S10

Table 2.2: Expression vectors that have been designed and synthesised in this thesis. The name of every plasmid contains the used vector, the protein name and the included residue numbers. Affinity tags included in the plasmid name have been introduced by the oligonucleotide during cloning; the position of all affinity tags is shown in parentheses, either N-terminal (N) or C-terminal (C). Besides the resistance and the oligonucleotides, the cloning method is provided with StarGate (SG), restriction enzyme (RE), restriction enzyme (RE) or In-fusion (IF) cloning. The sequences of the oligonucleotides used are shown in the supplemental tables A.5 and A.6.

Plasmid	Tag(s) (location)	Resistance	Cloning method	Oligonucleotides
pASG-IBA3_yPrp43	Strep (C)	Ampicillin	SG	-
pETM-13_yPrp43(58-767)-Strep	Strep (C)	Kanamycin	RE (NcoI, SalI)	R1, R2
pETM-13_yPrp43-yNtr1(1-120)-Strep	Strep (C)	Kanamycin	IF	I1, I2, I14, I22
pETM-13_yPrp43-yNtr1(51-110)-Strep	Strep (C)	Kanamycin	IF	I1, I4, I7, I13
pETM-13_yPrp43-yNtr1(51-102)-Strep	Strep (C)	Kanamycin	IF	I1, I4, I7, I11
pETM-13_yPrp43-yNtr1(51-99)-Strep	Strep (C)	Kanamycin	IF	I1, I4, I7, I12
pETM-13_yPrp43-yNtr1(59-110)-Strep	Strep (C)	Kanamycin	IF	I1, I5, I9, I13
pETM-13_yPrp43-yNtr1(62-110)-Strep	Strep (C)	Kanamycin	IF	I1, I6, I10, I13
pETM-13_yPrp43(1-745)-yNtr1(51-110)-Strep	Strep (C)	Kanamycin	IF	I1, I3, I8, I13
pASG-IBA3_His6-yNtr1	His (N), Strep (C)	Ampicillin	SG	-
pPSG-IBA5_yNtr1(1-120)	Strep (N)	Ampicillin	SG	-
pASG-IBA25_yNtr1(51-110)	GST (N)	Ampicillin	SG	-
pASG-IBA3_yNtr2	Strep (C)	Ampicillin	SG	-
pETM-13_ctPrp43-Strep	Strep (C)	Kanamycin	RE (NcoI, SalI)	R3, R5
pETM-13_ctPrp43(61-764)-Strep	Strep (C)	Kanamycin	RE (NcoI, SalI)	R4, R5
pET21a_ctNtr1-Strep	Strep (C)	Ampicillin	RE (NdeI, SalI)	R6, R7
pETM-13_ctNtr2-Strep	Strep (C)	Kanamycin	IF	I15, I16, I17, I18
pETM-13_ctPrp43-yNtr1(51-110)-Strep	Strep (C)	Kanamycin	IF	I19, I20, I21, I13

2.1.10 RNA oligonucleotides

In table 2.3, the sequences of the RNA oligonucleotides used for the ATPase assay (A₂₀-RNA), the protein-RNA-cross-linking (ssRNA-XL) and both strands of the substrates for the helicase unwinding assay (uw-l and uw-ul) is provided.

All RNA oligonucleotides except the ssRNA-XL (Roche Kulmbach GmbH) were purchased from IBA. The grade of purity was "PAGE grade" for the polyA RNAs and "HPLC grade" for the helicase assay substrate.

Table 2.3: Names and sequences of the RNA oligonucleotides used in this work. The oligonucleotide "uw-ul" was labelled with a Cyanine 5 (Cy5) and a BlackBerry Quencher 650 (BBQ).

Name	Sequence (5'-3')
A ₂₀	AAAAA AAAAA AAAAA AAAAA
ssRNA-XL	pCAGGUAGUGGUUGUCGGGCAG
uw-l	GCGCCUACGGAGCUGGUGGCGUAGGCGCAAAAAAAAAAAAAAAAAAAAA
uw-ul	(Cy5)-GCGCCUACGCCACCAGCUCCGUAGGCGC-(BBQ)

2.1.11 Size standards for DNA and Proteins

The size standards that were used during agarose gel electrophoresis (see section 2.2.1) and SDS-PAGE (see section 2.2.18) are listed below.

GeneRuler 1kb DNA Ladder	Fermentas
Prestained Protein Molecular Weight Marker	Fermentas
Protein Molecular Weight Marker	Fermentas

2.1.12 Supplemental material

The additionally used material is listed below.

Amicon Ultra centrifugal filter units	Millipore
Crystal clear tape	Henkel
Glass ware	Merck
Nylon loops for crystals	Hampton Research
Pipet tips	Sarstedt
Reaction tubes (0.5-50 ml)	Sarstedt
Sterile filters	Millipore

2.2 Molecular biological methods

2.2.1 Agarose gel electrophoresis

Agarose gel electrophoresis was used for the separation of DNA fragments resulting of a PCR (see section 2.2.12) or a restriction enzyme digest (see section 2.2.17). 40 ml of TAE buffer were supplemented with 0.4 mg agarose and boiled in the microwave. The agarose solution was cooled to about 50 °C and then poured into a gel mold with a comb. The samples were mixed with DNA loading dye (6×).

Once the gel had solidified, it was covered completely with TAE and the samples were transferred into separate lanes. The gel was run at 10-12 V/cm until a suitable separation was reached.

The recipes for the DNA loading dye and the TAE buffer are shown below.

DNA loading dye (6×)	TAE buffer (50×)
10 mM Tris/HCl pH 7.6	2 M Tris acetate pH 8.0
0.03 % Bromophenole blue	50 mM EDTA
0.03 % Xylene cyanol ff	
60 % Glycerol	
60 mM EDTA	

In order to visualise the DNA fragments on the agarose gel, the gel was incubated for 15 min in TAE buffer supplemented with GelRed (biotium, VWR) and documented at UV light (254 nm) on a GelDoc EQ System (Bio-Rad).

2.2.2 Cell harvest and lysis

Cells were harvested from expression cultures by centrifugation at 4800×g at 4 °C for 15 min. The cell pellet was resuspended in the cold lysis buffer; per 1 g of cell pellet, 5 ml buffer were used. Cell pellets were either directly lysed or stored at -20 °C.

Cell lysis was performed using a Microfluidizer 110S (Microfluidics) in which the cell suspension was pushed for five times through a chamber at high pressure (0.55 MPa) and passed through a cooling coil. The lysate was centrifuged at 30000×g at 4 °C for 45 min and the protein of interest then was purified by applying the supernatant to chromatography.

2.2.3 Chromatographic methods

All columns were stored in 20 % (v/v) ethanol (see section 2.1.2). Prior to usage, the columns were washed with 1.5 column volumes (CV) desalted H₂O and equilibrated in the appropriate buffer. After usage, they were, if required, regenerated, and then washed with 1.5 CV H₂O and 1.5 CV 20 % (v/v) ethanol.

2.2.3.1 Preparation of GST fusion proteins

The isolation of N-terminal GST fusion proteins from the cell lysate was performed through glutathion which was immobilised on a column matrix (glutathion sepharose 4B). The fusion protein binds specifically to this matrix, while other proteins do not and can be washed off the column.

The protein was eluted from the column using a buffer containing 30 mM reduced glutathion; the column was washed with 1 CV 6 M guanidine hydrochloride.

2.2.3.2 Preparation of His fusion proteins

The His₆ fusion proteins contained an additional sequence of six histidines and were purified by binding to Ni-NTA sepharose. The fusion proteins were eluted by a gradient of the lysis buffer supplemented with 500 mM imidazol.

2.2.3.3 Preparation of Strep fusion proteins

The Strep fusion proteins contained an additional sequence of eight amino acids (WSHPQFEK) and could be purified by binding to a mutated variant of streptavidine (StrepTactin). The fusion proteins bound specifically to immobilised StrepTactin and were eluted using a buffer containing 2 mM D-desthiobiotin. The column was regenerated through washing it with 1.5 CV of 0.5 M NaOH.

2.2.3.4 Anion exchange chromatography

In order to bind proteins to an anion exchange column, the protein has to be negatively charged. This is reached by using a buffer with a pH higher than the isoelectric point of the protein.

During this work, a Source 30 Q column was used. The protein sample was bound to the column in presence of a low-salt buffer and eluted afterwards with a gradient of salt concentration.

2.2.3.5 Desalting of protein samples

To exchange the buffer or to desalt a protein solution, HiPrep Desalting 26/10 and HiPrep Desalting 5 ml columns have been used during this work. The injected sample volume had to be below 12.5 ml or 1.5 ml, respectively. After injection, the protein was eluted before the sample buffer.

2.2.3.6 Size exclusion chromatography

By using a size exclusion column, proteins can be separated according to their size; large proteins elute earlier from the column than small proteins.

During this work, Superdex 75 and Superdex 200 columns in XK 26/60 containers were used for preparative runs with a column volume between 300 ml and 320 ml. The volume of the injected protein sample was 10 ml or below.

For analytical runs, a Superdex 200 column in a 10/300 container was used, with a column volume of 24 ml and a sample volume of 500 μ l.

The molecular weight and the oligomerisation state of the protein could be determined according to a calibration run with proteins of known mass.

2.2.3.7 Nucleotide loading state analysis

Anion exchange chromatography was used to analyse the nucleotide binding state of a protein.

50 nmol of the protein were desalted in buffer A (100 mM NaCl, 10 mM Tris pH 7.5, 2 mM EDTA) in order to remove the magnesium ions and thereby also bound nucleotides. In a second desalting step, the protein was transferred into buffer B (100 mM NaCl, 10 mM Tris pH 7.5, 2 mM MgCl₂) and afterwards incubated for two hours with a nucleotide which was present in a ten-fold molar excess. A consecutive desalting step in buffer B was performed in order to get rid of unbound nucleotides.

The protein was precipitated by incubation at 95 °C for 5 min and removed from the sample by centrifugation at 16000 \times g for 10 min. The supernatant was diluted 1:5 with H₂O and consecutively loaded on a Nucleosil 4000-7 HPLC anion exchange column (Macherey-Nagel) which had been equilibrated with buffer C (10 mM Tris pH 8.0). The nucleotides were eluted by a 20 ml gradient to buffer D (2 M NaCl, 10 mM Tris pH 8.0) and were identified according to their elution volumes.

2 Material and methods

Calibration of the column was performed by injection and elution of 5 nmol ADP, AMPPNP and AMPPCP.

In contrast to the other columns used in this work, this column was stored in 15 % (v/v) ethanol.

2.2.4 Concentrating of protein solutions

To concentrate a protein solution, the sample was transferred into a channel filtration chamber of a Vivaspin concentrator (Vivascience) which contained a vertical membrane. During centrifugation at $4000\times g$ and $4\text{ }^{\circ}\text{C}$, the proteins were retained depending on their size in the chamber, while smaller particles like water and ions could pass through the membrane.

The protein concentration of the sample was determined regularly, the concentrating was stopped as soon as the desired concentration was reached.

2.2.5 Deletion of introns

To clone genes containing introns, in a first step the whole gene sequence was cloned into a suitable vector. For the deletion of the intron via PCR, a forward oligonucleotide was designed that contained the first nucleotides of the second exon, and, as 5'-overhang, the last ten nucleotides of the first exon. A reverse oligonucleotide was designed accordingly, containing the end of the first exon and as overhang the first ten nucleotides of the second exon.

Using a PCR, the whole plasmid was amplified (see section 2.2.12) and afterwards digested with DpnI (Fermentas) for 1 h at $37\text{ }^{\circ}\text{C}$. Subsequently, $10\text{ }\mu\text{l}$ of the digested PCR sample were transformed into *E. coli* XL1-Blue cells (see section 2.2.19).

The sequences of the oligonucleotides used for the deletion of introns are listed in the supplemental table A.3.

2.2.6 Determination of the DNA concentration

The concentration of DNA samples was determined by the absorption at 260 nm using a photometer. For double-stranded DNA, 1 absorption unit at 260 nm (A_{260}) corresponds to $50\text{ }\mu\text{g/ml}$ and for single-stranded DNA to $40\text{ }\mu\text{g/ml}$.

The purity of DNA samples was determined by the ratio of the A_{260} and the A_{280} values which is between 1.8 and 2.0 for pure DNA; if the sample contained

also proteins, the value was below than 1.8.

2.2.7 Determination of the protein concentration

The Bradford assay uses the change of the absorbance maximum for an acidic solution of Coomassie Brilliant Blue G-250 from 465 nm to 595 nm if binding to protein occurs.

20 μ l of the protein solution were mixed with 980 μ l of 1:5 diluted Bradford reagens (Bio-Rad) and incubated for 5 min at room temperature. Subsequently, the OD₅₉₅ of the sample was measured against the blank value. The photometer had been calibrated that the OD₅₉₅ was identical to the protein concentration in mg/ml in the range of 0.1 and 0.9.

2.2.8 DNA sequencing

The sequencing of DNA was performed either in the Department of Developmental Biochemistry (Prof. Tomas Pieler) or in the Department of Genomic and Applied Microbiology (Prof. Rolf Daniel) of the University of Göttingen. The sequences of the oligonucleotides that were used for the amplification of the insert are listed in the supplemental table A.2.

2.2.9 Generation of chemical competent *E. coli* cells

5 ml of LB medium were inoculated with the appropriate *E. coli* strain and incubated overnight at 37 °C. This preculture was used for the inoculation of 500 ml LB medium in a 1:500 ratio and was incubated at 37 °C and 180 rpm until an OD₆₀₀ of 0.6 was reached.

Cells were harvested via centrifugation at 3000 \times g for 10 min at 4 °C. The cell pellet was resuspended in cold TFB1 buffer and incubated for 5 min on ice. The resuspended cells were centrifuged at 3000 \times g for 10 min at 4 °C for a second time. The cell pellet was resuspended carefully in 50 ml cold TFB2 buffer and aliquots of 50 μ l were flash cooled in liquid nitrogen and stored at -80 °C.

TFB1 buffer		TFB2 buffer	
30 mM	KAc pH 7.0	10 mM	Na-MOPS pH 7.2
50 mM	MnCl ₂	75 mM	CaCl ₂
10 mM	CaCl ₂	10 mM	RbCl
100 mM	RbCl	15 % (v/v)	Glycerol
15 % (v/v)	Glycerol		

2 Material and methods

TFB1 buffer was adjusted to pH 5.8 with acetic acid, TFB2 buffer was adjusted to pH 6.5 with NaOH.

2.2.10 Generation of glycerol cell stocks

For long term storage of cells, a 5 ml LB culture was inoculated with the appropriate *E. coli* strain and incubated at 37 °C overnight. 400 µl of this culture were mixed with 400 µl of 80 % (v/v) glycerol, flash cooled in liquid nitrogen and stored at -80 °C.

2.2.11 Molecular cloning

Three different methods for molecular cloning were used during this work.

2.2.11.1 Restriction site based cloning

The gene of interest was amplified performing a PCR with a forward and a reverse oligonucleotide containing suitable restriction sites and purified afterwards (see section 2.2.12). The insert as well as the vector were digested by the appropriate restriction enzymes and inactivated (see section 2.2.17). The digested vector was dephosphorylated via incubation with 0.01 unit per pmol DNA of the Calf intestinal alkaline phosphatase (CIAP, Promega) for 30 min at 37 °C. The phosphatase was afterwards inactivated for 20 min at 80 °C.

The ligation reaction was performed at 16 °C overnight and was composed of 200 ng of the digested and dephosphorylated vector, the digested insert in a fivefold molar excess, 1 µl T4 DNA ligase (Fermentas), 1 µl of the supplemented 10× T4 ligation buffer (Fermentas) and 1 µl of 10 mM ATP.

The entire ligation reaction was transformed into *E. coli* XL1-Blue cells (see section 2.2.19) and plated on a LB agar plate, supplemented with the appropriate antibiotic(s).

2.2.11.2 Cloning using IBA StarGate

The cloning technology IBA StarGate[®] (IBA) is based on recombinatorial cloning: a donor or entry vector is generated from which the gene can be transferred to different expression vectors.

The gene of interest was amplified by PCR using oligonucleotides containing the upstream or downstream combinatorial sites, respectively (see section 2.2.12).

After purification of the PCR product, it was added to a reaction tube containing the pENTRY-IBA10 vector and supplemented with specific solutions following the IBA StarGate instruction manual (version december 2008). After 1 h of incubation at 30 °C, the reaction mixture was transformed into *E. coli* TOP10 cells (see section 2.2.19) and plated on a LB agar plate containing kanamycin.

After verification of correct cloning by sequencing, the gene of interest was transferred from a donor vector to an acceptor vector by adding the donor vector to a tube containing the acceptor vector and supplementing it with specific solutions as described in the manual. This mixture was incubated for 1 h at 20 °C and transformed into *E. coli* TOP10 cells.

2.2.11.3 In-fusion cloning

The In-fusion[®] HD Cloning Kit (Clontech) allows cloning of several DNA fragments into one vector. The fusion of the DNA fragments is performed by means of the poxvirus DNA polymerase in the presence of 15 bp overlaps which are introduced to the fragments by the oligonucleotides (Zhu *et al.*, 2007).

The DNA fragments of interest were amplified by PCR and purified subsequently (see section 2.2.12). The reaction mixture was mixed following the User Manual (version October 2011), incubated at 50 °C for 15 min and finally transformed into *E. coli* XL1-Blue cells (see section 2.2.19).

2.2.12 Polymerase chain reaction

The polymerase chain reaction (PCR) allows the selective amplification of a specific DNA sequence. A standard PCR mixture included 1 µl Phusion polymerase (Finnzyme), 10 µl of the supplemented 5× HF buffer (Finnzyme), 1 µl dNTPs (10 mM each), 10 pmol of each forward and reverse oligonucleotide, 200 ng of DNA template (genomic DNA, a plasmid or a PCR purification product) and desalted H₂O, added to a final volume of 50 µl.

PCR cycles were performed as shown in the following scheme.

2 Material and methods

	Temperature (Cycle)	Time
1×	95 °C (Initial denaturation)	120 s
25-30×	95 °C (Denaturation)	30 s
	52-58 °C (Annealing of the oligonucleotides)	20 s
	72 °C (Elongation)	30 s/kb
1×	72 °C (Final elongation)	240 s

The result of the PCR was analysed by agarose gel electrophoresis (see section 2.2.1). The PCR product was purified using the peqGOLD MicroSpin Cycle-Pure Kit (PEQLAB).

2.2.13 Preparation of plasmid DNA

In order to amplify plasmid DNA, the plasmid was transformed into *E. coli* XL1-Blue cells (see section 2.2.19). A preculture of 5 ml LB medium was inoculated with the desired strain and incubated at 37 °C overnight.

The preparation of plasmid DNA was performed using the peqGOLD Plasmid Mini Kit (PEQLAB) following the user manual. The method was based on the alkaline lysis of the cells and the binding of the DNA plasmid to a silica membrane. After several washing steps, the DNA was eluted by 10 mM Tris/HCl pH 8.5.

2.2.14 Protein identification

Protein bands on an SDS-PA gel were identified using mass spectrometry in the Department of Biochemistry II (Dr. Bernhard Schmidt) of the University of Göttingen.

2.2.15 Proteolytic digestion

PreScission Protease (GE Healthcare) was applied for the removal of the GST affinity tag which had been recombinantly fused to the target protein. For the cleavage of the affinity sequence, 0.01 mg of the protease was added per 1 mg of target protein and incubated at 4 °C overnight. The cleavage efficiency was determined subsequently by means of SDS-PAGE (see section 2.2.18).

2.2.16 Recombinant expression of proteins in *E. coli*

In order to express recombinant proteins, the gene of interest was cloned into a suitable plasmid in a way that it was positioned behind a promoter. Two different promoter systems were used during this work: the *lac*-promoter, which allows to induce the expression by adding the artificial inductor IPTG (from 0.1 to 1 mM), and the *tet*-promoter, which allows to induce the expression by adding 200 µg anhydrotetracycline per liter expression culture.

100 ml of 2YT medium were inoculated with the *E. coli* cells containing the plasmid and incubated overnight at 37 °C. The expression cultures were inoculated in a 1:100 ratio with the preculture, supplemented with the appropriate antibiotics and incubated at 37 °C and 225 rpm. As soon as an OD₆₀₀ between 0.6 and 0.9 was reached, the temperature was decreased to 16 °C or 30 °C, and the expression of the target protein was induced upon the addition of the inductor.

The cell cultures were incubated at the appropriate temperature between 3 h and 16 h.

2.2.17 Restriction enzyme digest

In order to digest a DNA fragment or vector DNA, it was incubated for 1 h at 37 °C with one or two suitable restriction enzymes (Fermentas) in presence of the appropriate buffer (according to the manual), afterwards the restriction enzymes were inactivated by denaturation at 80 °C for 20 min. Per µg DNA, 1 unit of the restriction enzyme was used.

For test digestions, the total reaction volume was 20 µl, for insert digestions 50 µl.

2.2.18 SDS polyacrylamide gel electrophoresis

SDS polyacrylamide gel electrophoresis (SDS-PAGE) was used to separate proteins according to their size. Depending on the size of the proteins, acylamide contents of 7.5 %, 12.5 %, 15 % and 17.5 % were chosen.

The utilised gel chamber allowed the preparation of four gels at the same time. The chamber was filled to 80 % with the resolving gel which was covered with 2-propanol subsequently. After the solution had polymerised, the 2-propanol was removed, the stacking gel solution was added to the chamber and the combs were inserted. As soon as the polymerisation was complete, the gels were used or stored

2 Material and methods

at 4 °C.

The recipe for the gel solutions is shown in the following.

Resolving Gel (7.5-17.5 %)	
7.5/12.5/15/17.5 % (w/v)	Acrylamide
0.2/0.33/0.4/0.46 % (w/v)	Bisacrylamide
0.375 M	Tris/HCl pH 8.8
0.1 % (w/v)	SDS
0.1 % (v/v)	TEMED
0.05 % (w/v)	(NH ₄) ₂ S ₂ O ₈

Stacking Gel	
5 % (w/v)	Acrylamide
0.13 % (w/v)	Bisacrylamide
0.125 M	Tris/HCl pH 6.8
0.1 % (w/v)	SDS
0.1 % (v/v)	TEMED
0.05 % (w/v)	(NH ₄) ₂ S ₂ O ₈

To perform electrophoresis, the samples were mixed with Laemmli sample buffer, heated for 5 min at 95 °C and loaded into the gel. The gel was run for 75 min at 30 mA. For detection of the proteins on the SDS-PA gel, it was incubated with staining solution and afterwards destained using H₂O. For documentation, gels were scanned and saved as TIFF-files.

The recipes of the used buffers and solutions are provided below.

Running buffer		2×Laemmli sample buffer	
25 mM	Tris/HCl	62.5 mM	Tris/HCl pH 6.8
192 mM	Glycine	70 mM	SDS
0.1 % (w/v)	SDS	55 % (v/v)	Glycerol
		0.1 % (w/v)	Bromophenol blue
		5 % (v/v)	2-Mercaptoethanol

Staining solution	
10 % (v/v)	Ethanol
5 % (v/v)	Acetic acid
0.002 % (w/v)	Coomassie Brilliant Blue G/R 250

2.2.19 Transformation of chemical competent *E. coli* cells

In order to transform plasmid DNA into *E. coli* cells, 200 ng of plasmid or an entire ligation reaction, respectively, was added to 50 μl of chemical competent cells. After 30 min of incubation on ice, cells were heat-shocked at 42 $^{\circ}\text{C}$ for 1 min. Following to additional 5 min on ice, 900 μl of 2YT-medium were adjoint and the cell suspension was incubated for 45 min at 37 $^{\circ}\text{C}$.

Finally, the cell suspension was distributed on LB agar plates containing the appropriate antibiotic to select for transformed cells.

2.3 Biophysical methods

2.3.1 Characterisation of the ATPase activity

The ATP hydrolysis rate was determined using the EnzCheck[®] Phosphate Assay Kit (Invitrogen). The purine nucleoside phosphorylase (PNP) catalyses the re-action of inorganic phosphate with 2-amino-6-mercapto-7-methylpurine riboside (MESG), forming a methylpurine with an absorption maximum at 360 nm (shown in figure 2.1). The increase of absorption at 360 nm is thus directly proportional to the hydrolysis of ATP into ADP and inorganic phosphate by a protein with ATPase activity (Webb, 1992).

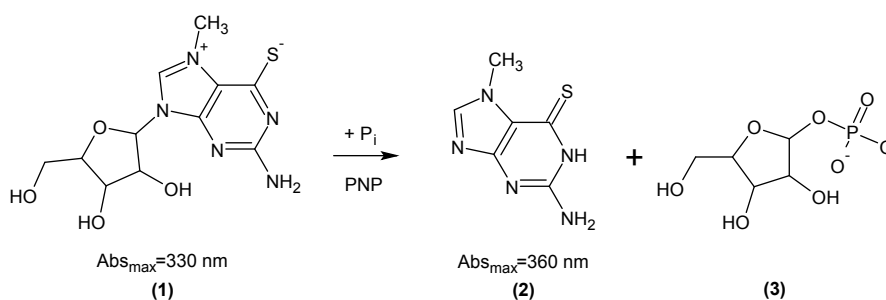


Figure 2.1: Principle of the EnzCheck[®] Phosphate Assay Kit. The release of inorganic phosphate (P_i) is coupled to the cleavage of 2-amino-6-mercapto-7-methylpurine riboside (1) into 2-amino-6-mercapto-7-methylpurine (2) and ribose-1-phosphate (3). The absorption maximum changes to 360 nm due to the formation of the purine derivative and allows the quantitation of the release of inorganic phosphate by a protein with ATPase activity. The coupled reaction is catalysed by the purine nucleoside phosphorylase (PNP).

2 Material and methods

The 100 μl reaction mixtures contained 10 mM Tris/HCl pH 7.5, 100 mM NaCl, 2 mM MgCl_2 and 1 μM protein as well as 1 μl PNP, 20 μl MESG and 0 to 5 μM RNA; they were incubated at 20 $^\circ\text{C}$ for 10 min prior to the experiment. Measurement was started upon addition of ATP in concentrations between 0 and 200 μM , the absorbance at 360 nm was measured every two seconds during ten minutes at 20 $^\circ\text{C}$ using a Ultrospec 2100pro Photometer (GE Healthcare). The initial slope of the change of absorption, thus the starting velocity v_0 ($\Delta\text{A}/\text{min}$), was determined for different ATP concentrations, each concentration was measured independently for three times. Negative controls without ATP and without protein were performed to assure that the origin of signal was the ATPase activity of the protein.

The amount of released phosphate was determined by interpolation of the measured values to a standard phosphate curve. Kinetic parameters were calculated by non-linear regression to a hyperbola using SigmaPlot (Systat Software Inc.). The hyperbola was expressed in a form identical to the Michaelis-Menten equation and therefore it provided the Michaelis constant K_m and the maximum velocity v_{max} with the respective standard deviations directly.

2.3.2 Characterisation of the helicase activity

The helicase assay was performed based on a principle published by Belon and Frick (2008). Two RNA strands were annealed of which one was labelled at the 5'-end with a Cyanine 5 (Cy5) and at the 3'-end with a BlackBerry Quencher 650 (BBQ). Upon excitation at 643 nm, the Cy5 emitted light at 667 nm. Upon strand separation, the labelled RNA oligonucleotide formed a hairpin structure in which the fluorophor got in close proximity to the BBQ which quenched the signal. Therefore, the decrease in the fluorescence signal at 667 nm was directly proportional to the rate of strand separation (see figure 2.2).

For the annealing of both RNA strands, a mixture including 4.5 μl of the non-labelled RNA strand (100 μM) and 4.5 μl of the labelled RNA strand (100 μM) and 1 μl 100 mM Tris pH 8.5 was incubated at the thermocycler at 95 $^\circ\text{C}$ for 2 minutes (for the sequences of the RNA oligonucleotides see section 2.1.10). Subsequently, the sample was cooled down from 80 $^\circ\text{C}$ to 20 $^\circ\text{C}$ by decreasing the temperature by 1 $^\circ\text{C}$ per 20 seconds and afterwards incubated on ice. The annealed RNA was diluted 1:45 to obtain a 1 μM -solution of dsRNA.

For all proteins tested in this work, the unwinding velocity at dsRNA con-

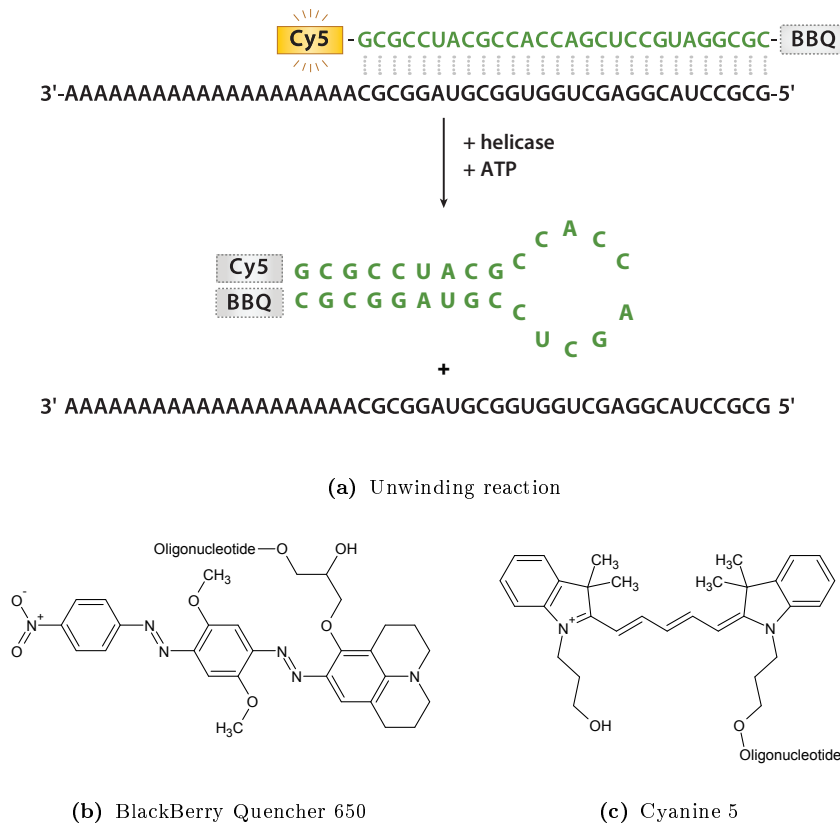


Figure 2.2: Principle of the helicase activity assay.

(a) The dsRNA educt consists of an non-labelled strand (black) and a second strand (green) that is labelled with a Cy5 and a BBQ. The strand separation leads to the formation of a hairpin in the labeled strand. Within this hairpin structure, the Cy5 gets in closed proximity to the quencher which quenches the signal.

(b), (c) The chemical structures of the BlackBerry Quencher 650 and the Cyanine 5.

centrations of 10, 50 and 200 nM was identical, indicating that the maximum velocity was reached. This assay was not suitable for dsRNA concentrations below 10 nM due to a too signal to noise ratio. Therefore, the maximum velocity was determined for all different proteins and protein complexes using a substrate concentration of 50 nM.

The 150 μ l reaction mixtures contained 20 nM protein, 50 nM dsRNA and 10 mM Tris pH 7.5, 100 mM sodium chloride and 2 mM $MgCl_2$. The samples were incubated at 20 $^{\circ}C$ in the fluorescence spectrometer for 15 min, the measurements were started by adding ATP to a final concentration of 1 mM. The excitation wavelength was set to 643 nm, the change in the fluorescence at 667 nm was measured every 2 s for 10 min. For the determination of the maximal reaction velocity (v_{max}), the initial slope v_0 ($\Delta F/\text{min}$) of the change in fluorescence was

2 Material and methods

determined three times independently for each sample.

The amount of unwound dsRNA was calculated with respect to the signal at 667 nm for the annealed and completely unwound dsRNA, respectively.

2.3.3 Circular dichroism spectroscopy

Circular dichroism spectroscopy is a method to detect the secondary structure content of a protein and is based on the differences in the absorption of left and right circularly polarised light by optically active molecules (Kelly *et al.*, 2005). Different secondary structure elements as α -helices, β -sheets and random coils show distinct absorption edges in the far UV range.

The protein sample was desalted in buffer CD (100 mM Sodium phosphate pH 7.5) and diluted to 0.1 mg/ml. 300 μ l of the sample were pipetted into a cuvet with a coat thickness of 1 mm and transferred into a Chirascan CD spectrometer (AppliedPhotophysics). The spectra were recorded in far UV range from 195-260 nm at 20 °C with a time-per-point value of 1 s and a path length of 1 nm. Every measurement was performed for five times and the mean values were taken. For the final absorption curve, the values of the negativ control which contained exclusively the sample buffer were subtracted from the values of the protein sample. The small organic molecule trifluorethanol was added in concentrations between 10 to 40 % (v/v) to the samples.

2.4 Crystallographic methods

2.4.1 High throughput screening

High throughput screening of crystallisation conditions was performed using a Phoenix IRE pipetting robot (Phenix). Protein samples were pipetted into MRC-96-3well plates and mixed with the reservoir solution in a ratio of 1:1 (0.25 μ l protein : 0.25 μ l reservoir) and 2:1 (0.25 μ l protein : 0.125 μ l reservoir); additionally, for each condition, the reservoir solution was mixed with the sample buffer in 1:1 ratio as a negative control. The screen solutions had been mixed by an Alchemist HT-Robot (Rigaku) or purchased (see section 2.1.4). The volume of the reservoir solution was 35 μ l.

The plates were sealed and incubated at 4 °C and 20 °C, respectively. By default, screening was performed using all eleven screens shown in section 2.1.4,

at both temperatures and for protein concentrations between 3 and 12 mg/ml. The wells were photographed regularly by a Minstrel System (Rigaku) and the photos were controlled by eye using the CrystalTrak software (Rigaku).

2.4.2 Optimisation of crystals

Upon identification of an initial crystallisation condition or a promising precipitation condition, optimisation was performed by altering the protein concentration, the drop size and the ratio of the reservoir solution and the protein solution.

Additionally, the variation of single components of the screen solution was performed using 24-well crystallisation plates (Hampton Research) in the sitting drop vapour diffusion method. The reservoir solution volume was 500 μ l, the drop was composed of 1 μ l of the protein solution and 1 μ l of the reservoir solution. Plates were sealed, incubated at 4 °C or 20 °C and regularly checked using a binocular microscope.

2.4.3 Cryocrystallography

Crystals were harvested with a nylon loop and transferred for 5 to 10 s into a cryo buffer containing the crystallisation buffer and 25 % (v/v) glycerol. Immediately afterwards, the crystals were flash cooled into liquid nitrogen.

2.4.4 Data collection

Nylon loops with fixed protein crystals were mounted on the goniometer head and exposed to X-ray radiation. The cooling of the crystal was maintained. Testing of the crystals was performed on a MicroMax 007 rotating anode (Rigaku). Data collection of the ctPrp43(61-764) crystal was performed at the beamline P14 at PETRA III, DESY, Hamburg.

Table 2.4: Parameters of the data collection of yPrp43(FL) and ctPrp43(61-764) crystals

	yPrp43(FL)	ctPrp43(61-764)
Beamline	BESSY BL 14.1	DESY P14
Wavelength (Å)	0.91841	1.23953
Detector	MarMosaic 225	Pilatus 6M
Distance of detector (mm)	263.740	443.705
Starting angle (°)	355.0	30.0
Oscillation (°)	0.250	0.100

3 Results

3.1 Preparation of the NTR proteins from *S. cerevisiae*

A goal of this thesis was to establish purification strategies for the three proteins of the NTR complex of *S. cerevisiae*, in the following referred to as yPrp43, yNtr1 and yNtr2. If possible, the expression of the full-length proteins should be performed using *E. coli* cells as expression system. For yNtr1, a purification strategy also for the smaller fragments encompassing the N-terminal residues 1 to 120, yNtr1(1-120), and the residues 51 to 110, yNtr1(51-110), should be developed.

In table 3.1, an overview of the properties of different proteins and truncations described in this section is shown. The detailed protocols for expression and purification of these proteins are provided in the supplemental section A.1.

Table 3.1: Proteins and truncated variants of the yeast NTR complex purified in this work. For each protein, the name, the size in kDa, the number of amino acids and the nature as well as the position of the affinity tag(s) is shown. For truncated proteins, the number of included residues is given in parentheses; all other proteins are full-length proteins and hence contain the whole sequence.

Protein	Size (kDa)	Length (AA)	Affinity tag(s)
yPrp43	87.6	767	Strep (C)
yNtr1	83.1	708	His (N), Strep (C)
yNtr1(1-120)	13.1	120	Strep (N)
yNtr1(51-110)	6.1	60	GST (N)
yNtr2	36.7	322	Strep (C)

3.1.1 Purification of yPrp43

The DNA fragment coding for yPrp43 was amplified from genomic yeast DNA and inserted into a vector fusing a C-terminal Strep affinity sequence to the full

3 Results

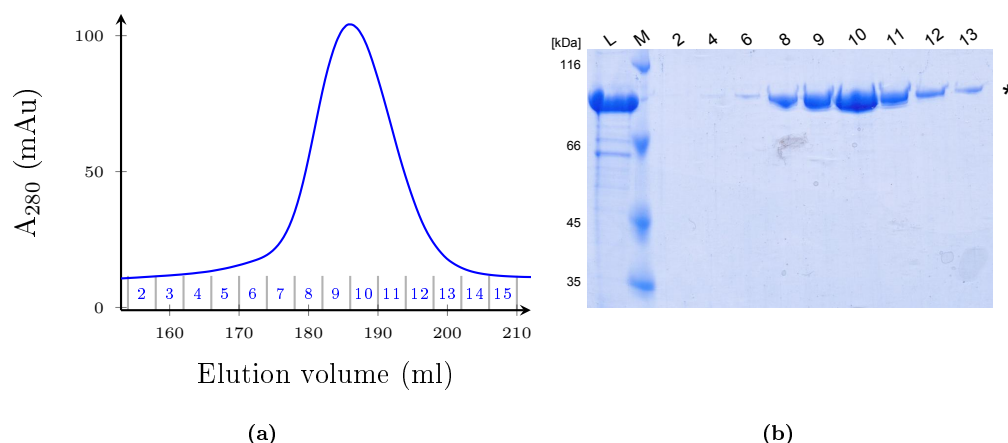


Figure 3.1: Preparative size exclusion chromatography of yPrp43-Strep, using a Superdex 200 26/60 column.

(a) In the chromatogram, the absorbance at 280 nm in a machine-specific unit (mAu, y-axis) is shown in blue referring to the elution volume (ml, x-axis). The numbers of the collected elution fractions are indicated in blue along the x-axis. The complete chromatogram is shown in the supplemental figure A.1.

(b) Coomassie-stained 12.5 % SDS-PA gel containing the loaded sample (L), the molecular weight marker (M) and elution fractions (numbers). The position of the bands corresponding to yPrp43-Strep is indicated by an asterisk.

length protein.

yPrp43-Strep was overexpressed in *E. coli* Rosetta 2(DE3) cells and isolated from the supernatant of the cell lysate using a StrepTactin column. yPrp43-Strep could be purified to homogeneity by performing a size exclusion chromatography using an S200 column (see figure 3.1). The peak containing the recombinantly produced helicase appeared at an elution volume corresponding to the molecular weight of the protein (88.7 kDa including the Strep-tag). The final yield was 15 mg of pure yPrp43-Strep per litre expression culture.

3.1.2 Purification of yNtr2

The DNA sequence coding for yNtr2 was amplified from genomic yeast DNA and cloned into a vector fusing a C-terminal Strep-tag to the full-length protein. yNtr2-Strep was overexpressed in *E. coli* Rosetta 2(DE3) cells and isolated from the soluble fraction of the lysed cells using a StrepTactin column. The eluted fractions of the StrepTactin column contained full-length yNtr2-Strep as well as bands according to proteins of smaller size, presumably C-terminal fragments of yNtr2 as they had bound to the StrepTactin column. For further purification, the sample was bound to an anion exchange column, from which pure full-length

3.1 Preparation of the NTR proteins from *S. cerevisiae*

yNtr2-Strep eluted at a different salt concentration than the impurities (see figures 3.2 and 3.3).

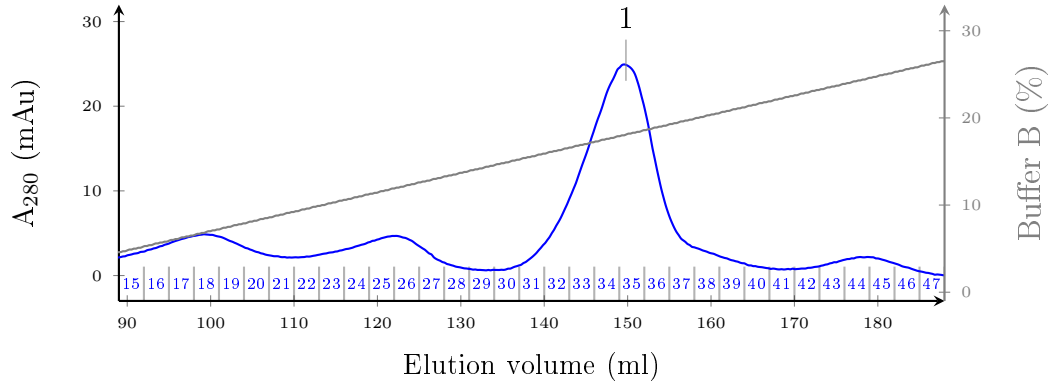


Figure 3.2: Chromatogram of the anion exchange chromatography of yNtr2-Strep, using a Source 30Q column. The absorption at 280 nm (mAu, left y-axis) is shown in blue with respect to the elution volume (ml, x-axis). The gradient of buffer B is shown on the right y-axis (in %) as a grey line, the numbers of the collected fractions are indicated in blue along the x-axis. The elution peak (1) corresponds to yNtr2. The complete chromatogram is shown in the supplemental figure A.2.

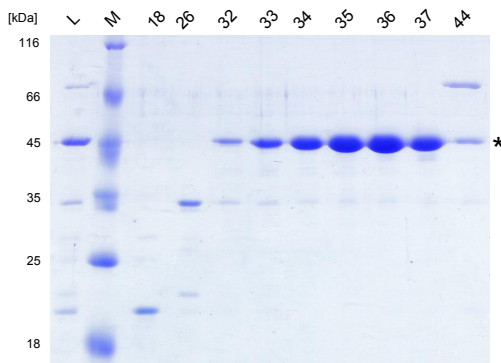


Figure 3.3: Coomassie-stained 12.5 % SDS-PAGE gel of the anion exchange chromatography of yNtr2-Strep, using a Source 30Q column. The loaded sample (L), the molecular weight marker (M) and elution fractions (numbers) are shown. The position of the bands corresponding to yNtr2-Strep is indicated by an asterisk.

yNtr2-Strep migrated slower during SDS-PAGE as expected from its molecular mass (38.7 kDa including the Strep affinity sequence). It could be verified by mass spectrometry that the band observed at about 45 kDa contained yNtr2 (data not shown). Size exclusion chromatography revealed that yNtr2 behaved as a monomer in solution (data not shown). The final yield was 2.5 mg yNtr2-Strep per litre expression culture.

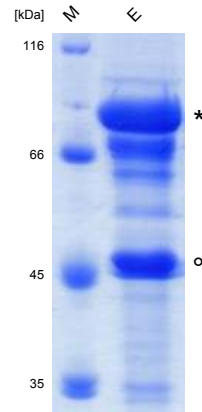
3.1.3 Copurification of yNtr1 and yNtr2

The DNA sequence coding for yNtr1 was amplified from yeast genomic DNA and cloned into multiple expression vectors, adding different affinity tags to the protein

3 Results

(Strep, His₆ and GST). After transformation of the plasmid into different *E. coli* expression cells (Rosetta 2(DE3), SoluBL(21), ArcticExpress (DE3)), no soluble yNtr1 could be expressed. In addition, the use of different buffer conditions during purification did not significantly increase the solubility of the protein (data not shown).

Figure 3.4: Coomassie-stained 12.5 % SDS-PA gel of the HisTrap purification of His-yNtr1-Strep and yNtr2-Strep. The lanes show the molecular weight marker (M) and the eluted sample (E) from the Ni-NTA column. The position of the band of yNtr1 is indicated by an asterisk, the position of yNtr2 by a circle.



Tanaka *et al.* (2007) had described previously that coexpression of yNtr1 and yNtr2 yields a soluble complex of both proteins. In this work, separately purified yNtr2-Strep was added before the lysis of *E. coli* Rosetta 2(DE3) cells containing His₆-yNtr1-Strep, resulting in a soluble yNtr1/yNtr2 complex. This complex was isolated from the soluble fraction of the cell lysate using a StrepTactin column and could be further purified by a subsequent HisTrap purification step to a purity of about 80 % (see in figure 3.4). Besides the bands for both full-length proteins and several impurities, a prominent band was present at about 70 kDa. Mass spectrometric analysis revealed that this band contained a fragment of yNtr1.

Although the yNtr1/yNtr2 complex behaved as an aggregate during analytical size exclusion chromatography (data not shown), it could be shown by Jean-Baptiste Fourmann (Department of Cellular Biochemistry, Prof. Reinhard Lührmann, Max-Planck-Institute for Biophysical Chemistry, Göttingen) that the preparation contained a functional yNtr1/yNtr2 complex: It was able to activate yPrp43 to disassemble the intron-lariat spliceosome in an *in vitro* yeast reconstitution and disassembly assay. These experiments revealed that the NTR complex in presence of ATP is capable to disassemble an isolated 35S intron-lariat spliceosome to a 20-25S U2 snRNP, an 18S U5 snRNP, the U6 snRNA and the intron-lariat RNA. Interestingly, the disassembly was also efficient when ATP was replaced by UTP. As Brr2 is not able to hydrolyse UTP, these results indicate that the ATPase activity of Brr2 is not required for the disassembly by using the

3.1 Preparation of the NTR proteins from *S. cerevisiae*

in vitro reconstitution system. This suggests that the RNA helicase activity of Brr2 is not involved in the disassembly (for further details see Fourmann *et al.* (2013)).

3.1.4 Purification of yNtr1(1-120)

As the purification of pure and monomeric yNtr1 failed, the truncated variant yNtr1(1-120) was expressed which was known to be sufficient for the activation of yPrp43 (Tanaka *et al.*, 2007). yNtr1(1-120) was not soluble when expressed in *E. coli* Rosetta 2(DE3) cells with a C-terminal Strep affinity sequence. Interestingly, soluble protein could be obtained by changing the position of the Strep affinity sequence to the N-terminus of yNtr1(1-120).

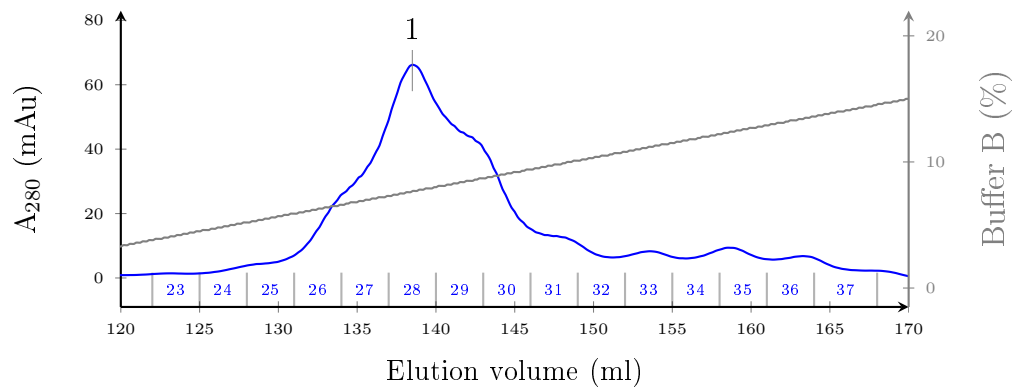


Figure 3.5: Chromatogram of the anion exchange chromatography of Strep-yNtr1(1-120), using a Source 30Q column. The absorption at 280 nm is shown in blue in a machine specific unit (mAu, left y-axis) referring to the elution volume in ml (x-axis). The gradient of buffer B is indicated on the right y-axis (in %) as a grey line; collected fractions are shown along the x-axis. The elution peak (1) corresponds to yNtr1(1-120). The complete chromatogram is shown in the supplemental figure A.3.

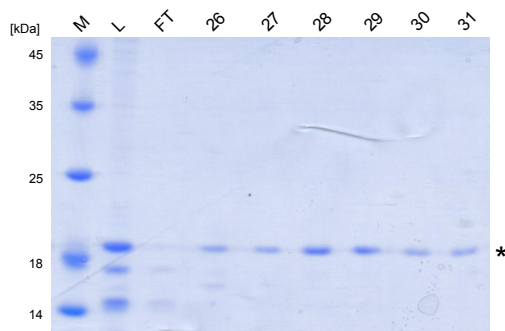


Figure 3.6: Coomassie-stained 15 % SDS-PAGE of the ion exchange chromatography of Strep-yNtr1(1-120), using a Source 30Q column. The lanes show the molecular weight marker (M), the loaded sample (L), the flow-through (F) and fractions collected during the exchange chromatography (numbers). The position of the bands of yNtr1(1-120) is indicated by an asterisk.

The elution fractions from the StrepTactin column contained, besides yNtr1(1-120), two shorter fragments which could be separated by means of an anion ex-

3 Results

change chromatography (shown in figures 3.5 and 3.6). These two fragments did not bind to the Source 30Q column. Remarkably, Strep-yNtr1(1-120) migrated slower during SDS-PAGE than expected from its molecular mass (14.1 kDa), as the band appeared slightly above the 18 kDa reference band of the molecular weight marker.

Analytical size exclusion chromatography revealed that yNtr1(1-120) behaved as a monomer in solution (data not shown). Per litre expression culture, 0.8 mg of pure Strep-yNtr1(1-120) were obtained.

3.1.5 Purification of yNtr1(51-110)

As it was already known that yNtr1(1-120) was sufficient for the activation of yPrp43 (Tanaka *et al.*, 2007), one goal of this thesis was to investigate whether the isolated G-patch motif was able to interact with yPrp43. Based on the different predictions for the location of the G-patch motif (see section 1.6.4), the fragment yNtr(51-110) was chosen for expression and purification.

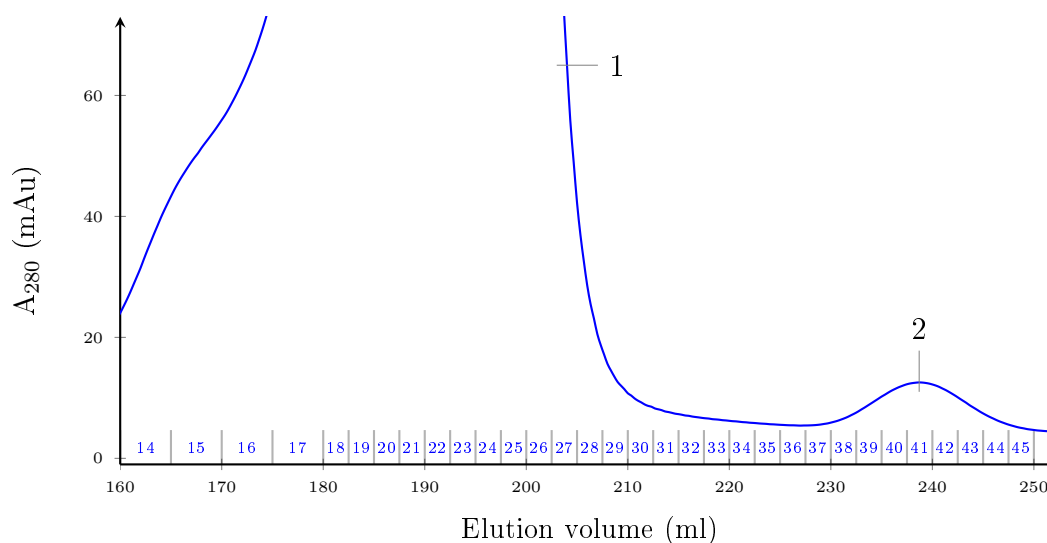


Figure 3.7: Chromatogram of the S75 size exclusion chromatography of yNtr1(51-110) after cleavage of the GST affinity sequence, using a Supderdex 75 26/60 column. The absorption at 280 nm in a machine specific unit (mAu, y-axis) is shown as a blue line with respect to the elution volume (ml, x-axis). The local maxima represent the free GST (1) and yNtr1(51-110) (2). Collected elution fractions are indicated along the x-axis. The complete chromatogram is shown in the supplemental figure A.4.

yNtr1(51-110) was not soluble when expressed in *E. coli* Rosetta 2(DE3) cells with a C- or an N-terminal Strep-tag, but was soluble when expressed as a GST fusion protein. After isolation from the cell lysate by a GSH sepharose purification

3.2 Preparation of yPrp43/yNtr1 complexes

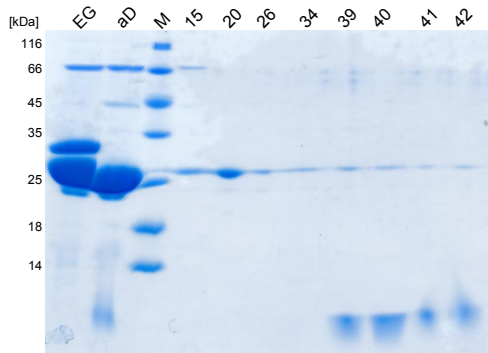


Figure 3.8: 17.5 % SDS-PA gel of the S75 size exclusion chromatography of yNtr1(51-110). On the lanes, the eluted sample from the GSH-sepharose chromatography (EG), the sample after proteolytic digest (aD), the molecular weight marker (M) and elution fractions of the size exclusion chromatography (numbers) are shown. The position of the bands of uncleaved GST-yNtr1(51-110) is labelled by a plus, of GST by a circle and of yNtr1(51-110) by an asterisk.

step, the GST-tag was removed proteolytically. yNtr1(51-110) was separated from free GST by an S75 size exclusion chromatography step (shown in figures 3.7 and 3.8).

yNtr1(51-110) eluted earlier from the size exclusion column than expected from its size, this behaviour can be explained by the fact that it presumably is unstructured (see section 3.6). Probably due to the low molecular mass of 6.1 kDa, the band corresponding to the yNtr1-variant appears not sharp, but smeary over a larger region on the SDS-PA gel. The yield per litre expression culture was 0.4 mg yNtr1(51-110).

3.2 Preparation of yPrp43/yNtr1 complexes

Once having the individual components of the NTR complex at hand, the next goal was the preparation of protein complexes of yPrp43 and yNtr1 (in the following yPrp43/yNtr1) for their functional and structural characterisation. As no full-length yNtr1 protein could be obtained, the work was focused on the complex formation between yPrp43 and the two truncated variants of yNtr1, namely yNtr1(1-120) and yNtr1(51-110).

3.2.1 Preparation of the yPrp43/yNtr1(1-120) complex

yPrp43 and yNtr1(1-120) were mixed in a molar ratio of 1:4 and incubated overnight. The binding of yNtr1(1-120) and thus the successful complex formation could be shown by means of analytical gel filtration chromatography (see figure 3.9). The elution volume of the complex was shifted by 0.3 ml compared to the elution volume of free yPrp43, indicating a higher molecular weight and, according to the calibration of the column, the formation of a complex in a 1:1

3 Results

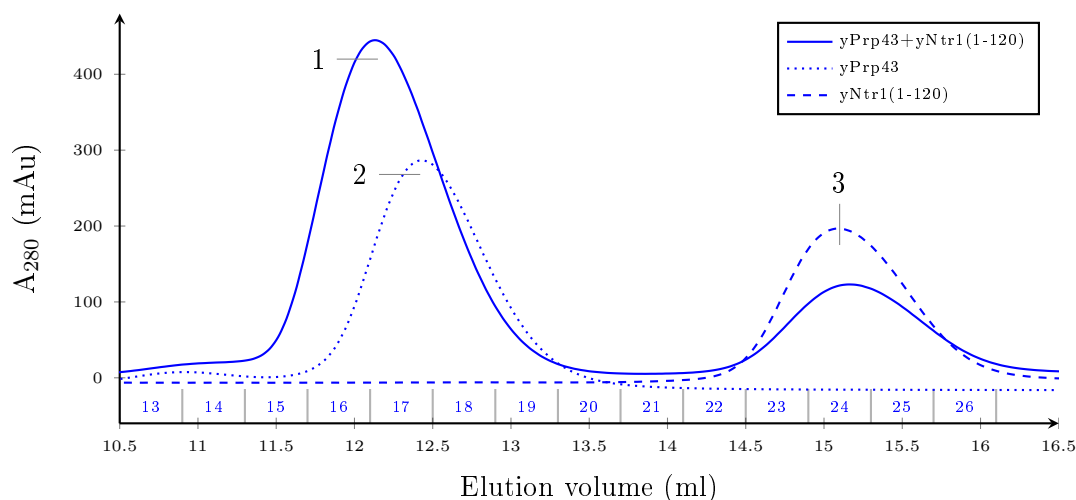
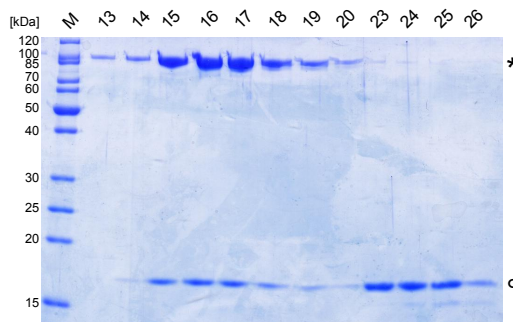


Figure 3.9: Chromatogram of the analytical size exclusion chromatography of the complex yPrp43/yNtr1(1-120), using a Superdex 200 (10/300) column. The absorption at 280 nm (machine specific unit, y-axis) is shown with respect to the elution volume (ml, x-axis) for three independent samples: yPrp43+yNtr1(1-120) (continuous line), free yPrp43 (dotted line) and free yNtr1(1-120) (dashed line). The local maxima correspond to the yPrp43/yNtr1(1-120) complex (1), yPrp43 (2) and yNtr1(1-120) (3). The fractions indicated along the x-axis have been collected during the run of yPrp43 and yNtr1(1-120). The complete chromatogram is shown in the supplemental figure A.5.

Figure 3.10: Coomassie-stained 15 % SDS-PAGE gel of the analytical S200 size exclusion chromatography of yPrp43 and yNtr1(1-120). The marker (M) and elution fractions (numbers) from the gel filtration are shown on the lanes. The position of yPrp43 is labelled by an asterisk, that of yNtr1(1-120) by a circle.



ratio. Additionally, the amount of the free yNtr1(1-120) decreased as indicated by the smaller local absorption maximum at 15.1 ml.

SDS-PAGE of the elution fractions from the size exclusion chromatography revealed that the first local UV maximum indeed contained both, yPrp43 and yNtr1(1-120), whereby the second peak corresponded to unbound yNtr1(1-120) (see figure 3.10). Large scale preparations of the complex have been performed using a preparative S200 size exclusion column.

3.2.2 Preparation of the yPrp43/yNtr1(51-110) complex

The complex of yPrp43 and yNtr1(51-110) was prepared following the same strategy as for the complex yPrp43/yNtr1(1-120) (see section 3.2.1). Both proteins were mixed in a molar ratio of 1:4 and incubated overnight. The complex was separated from the excess of yNtr1(51-110) by a size exclusion chromatography.

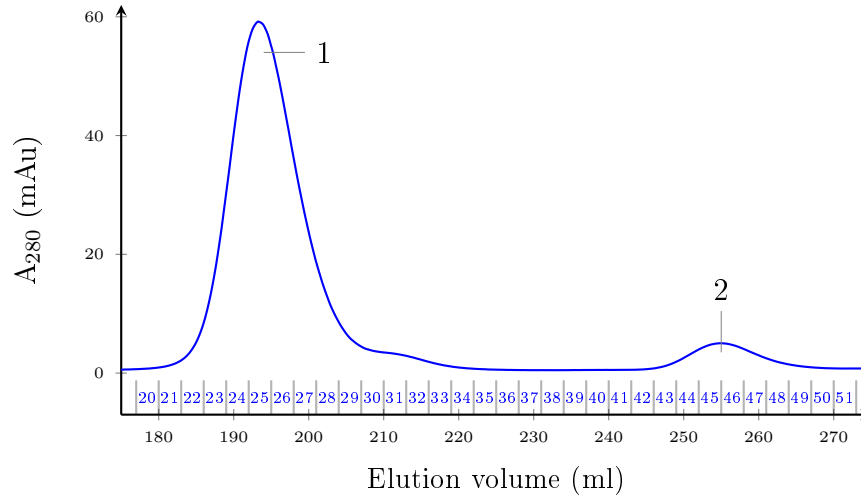


Figure 3.11: Chromatogram of the size exclusion chromatography of yPrp43 and yNtr1(51-110), using a Supderdex 200 26/60 column. The blue line reflects the UV absorption at 280 nm (machine specific unit, y-axis) with respect to the elution volume (ml, x-axis). The numbers of collected fractions are indicated along the x-axis. The first local maximum (1) corresponds to the yPrp43/yNtr1(51-110) complex and the second (2) to yNtr1(51-110), which was added in excess. The complete chromatogram is shown in the supplemental figure A.6.

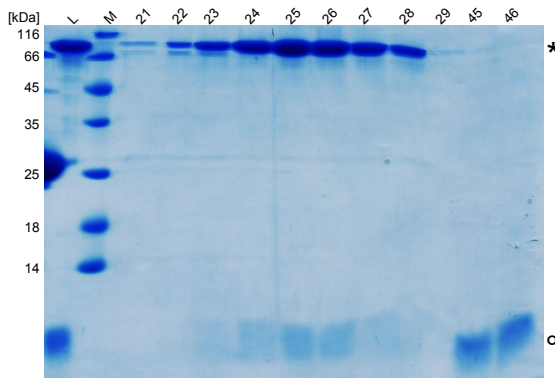


Figure 3.12: Coomassie-stained 17.5 % SDS-PAGE gel of the S200 size exclusion chromatography of the complex of yPrp43 and yNtr1(51-110). The lanes contain the loaded sample (L), the molecular weight marker (M) and samples collected during the chromatography (numbers). The position of yPrp43 is indicated by an asterisk, the position of yNtr1(51-110) by a circle.

The chromatogram of the preparative S200 size exclusion run of the complex yPrp43/yNtr1(51-110) shows two local maxima (see figure 3.11). Upon analysis of the collected elution fractions by SDS-PAGE it was revealed that the first

3 Results

peak contained both, yPrp43 and yNtr1(51-110), and that the second peak corresponded to the excess of yNtr1(51-110) (see figure 3.12).

In summary, the pure complexes of the helicase with both truncated yNtr1 variants could be prepared, revealing that not only yNtr1(1-120), but also yNtr1(51-110) is able to bind to yPrp43.

3.2.3 Crystallisation attempts of yPrp43/yNtr1 complexes

The prepared yPrp43/yNtr1 complexes were subsequently subjected to crystallisation trials. Screening for crystallisation conditions was performed for the complex of yPrp43 and yNtr1(1-120) in presence of a tenfold molar excess of ADP and at protein concentrations ranging from 3.5 to 10 mg/ml. At a protein concentration of 4.5 mg/ml, several conditions could be identified that led to protein crystals (shown in figure 3.13). However, these conditions were very similar to the crystallisation conditions of the already published structures of unliganded yPrp43 (see supplemental table A.1).

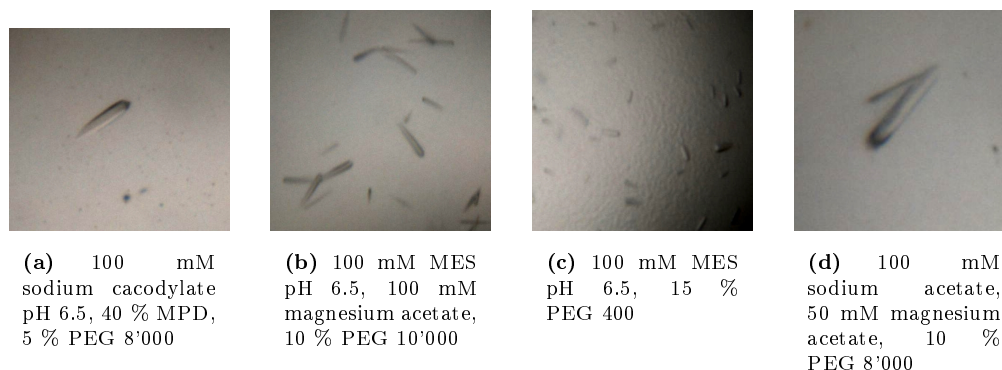


Figure 3.13: Protein crystals obtained during initial screening of the complex yPrp43/yNtr1(1-120) in presence of ADP at a concentration of 4.5 mg/ml. Crystals formed at 4 °C and in conditions with mostly large PEGs. After structure solution of (d), it turned out that the crystals contained yPrp43 exclusively.

The obtained crystals diffracted to a maximum resolution of 3.5 Å at the home source. During determination of the space groups and the cell parameters, it turned out that the values were virtually identical to the known crystal forms of yPrp43 alone. The presence of the 120 additional residues of yNtr1(1-120) in the crystal would have had a significant influence on the cell constants, thus the probability was very high that the crystals contained exclusively the helicase and not the complex of yPrp43 and yNtr1(1-120).

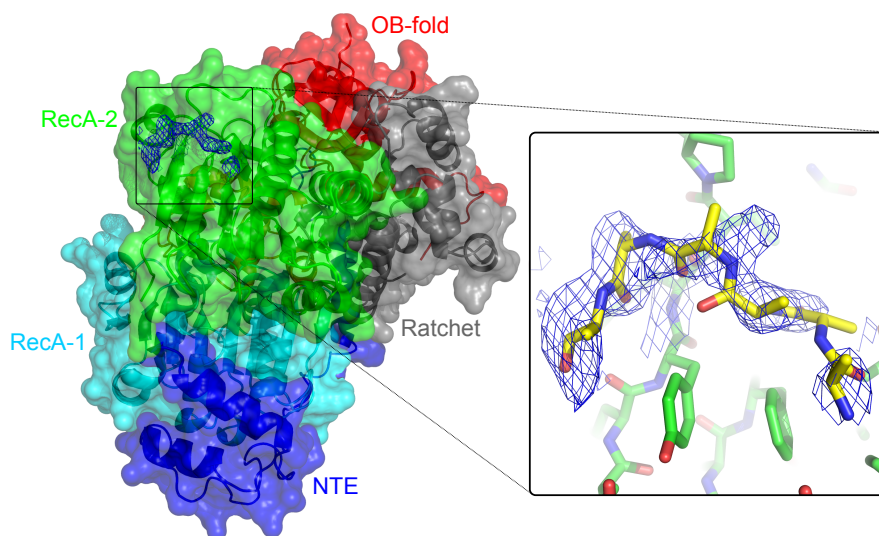


Figure 3.14: 3D structure model obtained from crystallisation attempts of the complex yPrp43/yNtr1(1-120) in the presence of ADP. The space group and unit cell parameters were identical to those that have been published for the crystallisation of unliganded yPrp43, and only yPrp43 could be traced into the electron density, with a structure identical to the one published by Walbott *et al.* (2010) and He *et al.* (2010). Some additional electron density could be identified that was localised on the surface the RecA-2 domain. The position of the $2F_oF_c$ electron density is shown, the zoom panel indicates that it was possible to model a peptide chain containing five residues into the electron density. As the nature of the residues could not be defined, they are represented exclusively by alanines.

To be sure about the content of the crystal, a dataset of a crystal obtained in the buffer condition (d) was collected at the synchrotron (for details of the data collection see table 2.4). As the space group was identical to that of the published structure of yPrp43 (pdb-ID 2xau), the structure was solved using this model and subjected to a rigid-body refinement, followed by simulated annealing with PHENIX (Adams *et al.*, 2002). The resulting model was completed by iterative cycles of model building in Coot and refinement in PHENIX.

The crystal diffracted to a maximum resolution of 2.29 Å and the structure could be refined to R-factors of 19.5 % and 24.6 %, respectively (see table 3.2). The structure determination confirmed only yPrp43 molecules had crystallised. The superposition of the solved crystal structure with the published structure of yPrp43 (pdb-ID 2xau) revealed that both are nearly identical, confirmed by a low RMSD for the C_α -atoms of 0.1803 Å.

Interestingly, additional electron density was present near residue P280 of the RecA-2 domain for both monomers in the unit cell. The shape of the electron density indicated that it did not represent a PEG molecule but a polypeptide, due to the fact that branches (for the residues) and density for carbonyl groups was

3 Results

Table 3.2: Data collection and structure refinement statistics for yPrp43. The values in parentheses refer to the highest resolution shell. The Ramachandran plot was calculated with MolProbity.

Data collection		Refinement	
Space group	P3 ₂ 21	Resolution range (Å)	30.0 - 2.29 (2.32-2.29)
<i>Unit cell parameters</i>		Number of reflections	92171
cell axes (Å)	a=b=118.04, c=254.13	Completeness (%)	100.0 (96.0)
cell angles (°)	$\alpha=\beta=90, \gamma=120$	Total atoms	12854
Resolution range (Å)	50.0-2.04 (2.16-2.04)	Protein residues	1504
Completeness (%)	92.2 (69.3)	Water molecules	810
Multiplicity	4.83 (3.33)	R/Rfree (%)	19.46/24.56 (26.57/31.30)
Unique reflections	120903 (14482)	Mean B-factor (Å ²)	38.26
I/ σ (I)	9.94 (0.85)	<i>RMSD from ideal</i>	
R _{merge} (%)	10.4 (53.6)	bond lengths (Å)	0.005
Wilson B-value (Å ²)	29.55	bond angles (°)	0.869
		<i>Ramachandran plot</i>	
		outliers (%)	0.2
		allowed (%)	3.09
		favored (%)	96.71

present. In total, a chain of five residues could be traced (see figure 3.14). The nature of these amino acids could not be clearly identified. The facts that only five residues could be traced, no structural changes occurred at any place of the yPrp43 structure and the G-patch motif is supposed to interact with the OB-fold domain of yPrp43 (see later in section 3.4.1), indicate that the presence of this electron density was only due to a non-functional affinity of several residues of yNtr1(1-120) to this surface part of yPrp43. As it is located at a large solvent channel, the interaction of those residues might have occurred after the full-length yPrp43 proteins already had crystallised.

Similar results were obtained during initial crystal screening with the complex yPrp43/yNtr1(51-110) in presence of ADP during which crystals formed in similar conditions and also contained only yPrp43 (data not shown). To prevent complex dissociation during crystallisation, yPrp43 was mixed with a ten-fold molar excess of yNtr1(1-120) and yNtr1(51-110), respectively. No crystals were obtained in any of the screens for these saturated complexes.

Additionally, crystal screening for yPrp43 in complex with either yNtr1(1-120) or yNtr1(51-110) in presence of the ATP analogues AMPPNP or AMPPCP and either poly(U)₈- or poly(U)₁₂-RNA did not lead to any crystals. The same was true for crystallisation trials with yPrp43 alone in presence of different ATP analogues and/or different substrate RNAs.

3.3 Functional characterisation of the interaction between yPrp43 and yNtr1

For the characterisation and comparison of the enzymatic activities of yPrp43 and yPrp43/yNtr1 complexes, real-time and fluorescence-based assays for determination of the ATPase and the helicase activities were used. Furthermore, the functional properties of the complexes should be used to identify the minimal fragment of yNtr1 that is required for the activation of yPrp43.

3.3.1 Determination of ATPase and helicase activities

The ATPase activity was determined using a commercially available photometric assay that quantitatively detects the phosphate which is released upon ATP hydrolysis (see section 2.3.1). The assay allowed the determination of the Michaelis constant K_m and the maximal ATP hydrolysis rate v_{max} .

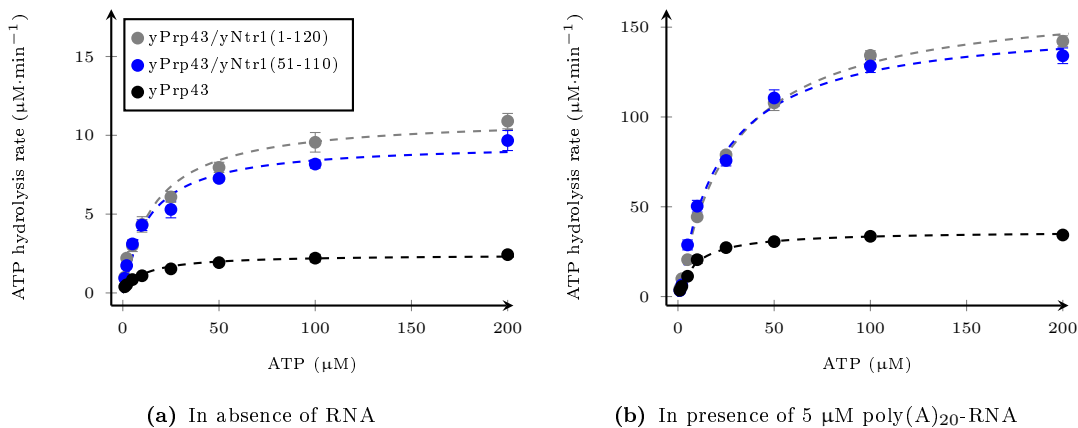


Figure 3.15: ATPase activity of yPrp43/yNtr1(1-120) (grey), yPrp43/yNtr1(51-110) (blue), yPrp43 (black) in absence of RNA (a) and in presence of 5 μM poly(A)₂₀-RNA (b). The ATP hydrolysis rate (in $\mu\text{M}\cdot\text{min}^{-1}$, y-axis) in dependence of increasing concentrations of ATP (x-axis, μM) is shown. The circles indicate the mean value, the error bars the root mean square deviation and the dashed lines the hyperbola fit.

Comparison of the ATPase activities of yPrp43 in the free state and in complex with yNtr1(51-110) or yNtr1(1-120) revealed that the fragments of yNtr1 stimulate the ATP hydrolysis rate of yPrp43 both, in absence and presence of poly(A)₂₀-RNA (see figure 3.15 and table 3.3). In both cases, the complex of yPrp43 and yNtr1(1-120) exhibited a slightly higher maximal ATP hydrolysis rate compared to the complex yPrp43/yNtr1(51-110).

3 Results

Table 3.3: Kinetic parameters describing the ATPase activity of yPrp43 alone as well as in complex with yNtr1(51-110) and yNtr1(1-120), in absence and in presence of poly(A)₂₀-RNA. For each complex, the Michaelis constant K_m and the maximal reaction velocity v_{max} is provided.

Protein	K_m (μM)	v_{max} ($\mu\text{M}\cdot\text{min}^{-1}$)
<i>In absence of RNA</i>		
yPrp43	11.07 \pm 2.07	2.43 \pm 0.12
yPrp43/yNtr1(51-110)	13.42 \pm 1.93	9.54 \pm 0.37
yPrp43/yNtr1(1-120)	16.37 \pm 1.81	11.18 \pm 0.34
<i>In presence of poly(A)₂₀-RNA</i>		
yPrp43	9.23 \pm 0.53	36.47 \pm 0.51
yPrp43/yNtr1(51-110)	22.48 \pm 1.50	153.40 \pm 3.08
yPrp43/yNtr1(1-120)	27.95 \pm 1.25	166.48 \pm 2.38

In absence of poly(A)₂₀-RNA, the presence of either yNtr1(51-110) or yNtr1(1-120) stimulated the maximal ATP hydrolysis rate 3.9-fold and 4.2-fold, respectively, compared to the intrinsic hydrolysis rate of yPrp43. In presence of RNA, the stimulation was similar: 4.2-fold in presence of yNtr1(51-110) and 4.5-fold in presence of yNtr1(1-120). Hence the extent of the stimulation of the hydrolysis rates were in a similar range in absence and presence of RNA for both yPrp43/yNtr1 complexes. Additionally, the Michaelis constants were increased for both complexes in absence and presence of RNA compared to free yPrp43. This does not automatically indicate a lower ATP affinity but can also be a consequence of the increased ATP hydrolysis rates.

The stimulation by poly(A)₂₀-RNA was in a similar range for the different samples. The ATP hydrolysis rate was stimulated 15.0-fold for yPrp43, 16.1-fold for yPrp43/yNtr1(51-110) and 14.9-fold for yPrp43/yNtr1(1-120) compared to the maximal hydrolysis rates measured in the absence of RNA.

Moreover, a fluorescence-based unwinding assay was adapted that was based on a dsRNA substrate containing one strand with a Cy5 fluorescent dye at the 5'-end and a BlackBerry Quencher (BBQ) at the 3'-end (see section 2.3.2). Together those strands formed 28 base pairs with a 3'-overhang. Upon strand separation, the signal of the Cy5 dye is quenched by the BBQ, thus the decrease in signal corresponded to the unwinding process. Due to the very low unwinding activity of yPrp43 in absence and presence of yNtr1 paired with the medium sensitivity of the assay, exclusively the maximal velocity of dsRNA unwinding v_{max} could be determined.

No helicase activity was detected for yPrp43 alone. Apparently, the intrinsic

3.3 Functional characterisation of the interaction between yPrp43 and yNtr1

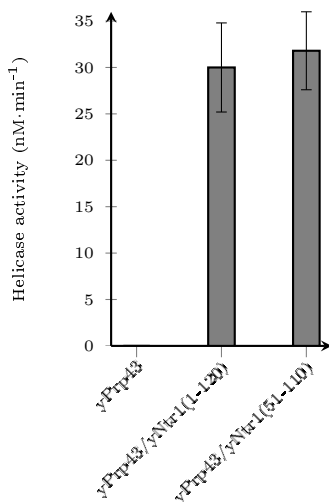


Figure 3.16: Histogramm showing the maximal unwinding velocity (v_{max}) for the complexes yPrp43/yNtr1(1-120) and yPrp43/yNtr1(51-110). No helicase activity was detected for free yPrp43.

Table 3.4: Kinetic parameters describing the helicase activity. For each protein/complex, the maximal reaction velocity (v_{max}) is provided. No unwinding could be detected for yPrp43.

Protein	v_{max} (nM·min ⁻¹)
yPrp43	n. d.
yPrp43/yNtr1(1-120)	30.0±3.6
yPrp43/yNtr1(51-110)	31.8±4.2

helicase activity reported for yPrp43 by Tanaka and Schwer (2006) was too low to be detected in this assay. In complex with yNtr1(1-120) as well as in complex with yNtr1(51-110), yPrp43 exhibited the same dsRNA unwinding velocity, 30.0 ± 3.6 nM·min⁻¹ and 31.8 ± 4.2 nM·min⁻¹, respectively (see table 3.4 and figure 3.16).

In conclusion, the results presented here indicate that yNtr1(51-110) is able to stimulate both, the ATPase and the helicase activity of yPrp43. The extent of stimulation is for both activities similar to the effect of yNtr1(1-120). Hence, yNtr1(51-110) seems to contain all residues required for the activation of yPrp43 and the additional residues present in yNtr1(1-120) do not seem to be involved in the activation of yPrp43.

3.3.2 Determination of dissociation constants

Due to the finding that the yNtr1-truncations containing the G-patch motif have a stimulatory effect on the ATPase activity of yPrp43, the dissociation constants (K_d) for the yPrp43/yNtr1 complexes under steady-state turnover conditions could be determined (see section 2.3.1). yPrp43 was incubated with increasing

3 Results

amounts of the respective yNtr1-fragment for two hours, allowing the complex to reach an equilibrium state. Subsequently, the ATPase hydrolysis rate for an equal amount of ATP was determined. As the contribution of free yPrp43 to the total ATPase activity was low and could hence be ignored, the concentrations of yNtr1 were plotted against the relative ATPase activity with 0 % corresponding to the intrinsic ATP hydrolysis rate of free yPrp43 and 100 % to the theoretical activity of a completely saturated yPrp43/yNtr1 complex.

The dissociation constant determined was identical to the concentration of yNtr1 at which 50 % of the total ATPase activity was reached and hence 50 % of yPrp43 was bound to yNtr1. The K_d could be directly deduced from the equation of the hyperbolic fit (see figure 3.17).

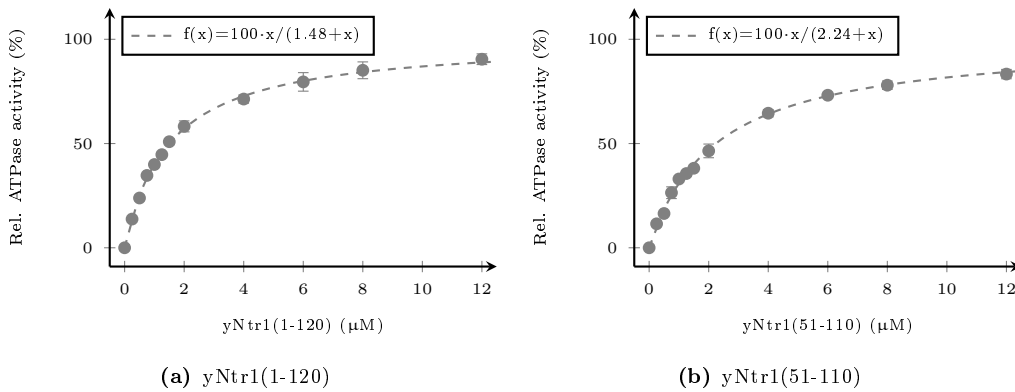


Figure 3.17: Determination of the dissociation constants of the complexes of yPrp43 and yNtr1(1-120) (a) and yNtr1(51-110) (b). The K_d was determined using the stimulatory effect of the yNtr1 truncations on the ATPase activity of yPrp43. The extent of stimulation (in %, y-axis) was plotted against increasing concentrations of the yNtr1-truncation (in μM , x-axis). The equation of the hyperbola fit is indicated in the graphs.

The dissociation constants of both complexes turned out to be in a micromolar range and the K_d of the yPrp43/yNtr1(1-120) complex ($1.49 \pm 0.05 \mu\text{M}$) was slightly higher than the K_d of the yPrp43/yNtr1(51-110) complex ($2.24 \pm 0.11 \mu\text{M}$). This difference indicates that there might be additional residues of yNtr1 beyond the stimulating fragment between the residues 51 and 110 that interact with yPrp43. These supplemental interactions could stabilise the binding to yPrp43 which results in a lower K_d , but do not contribute to the stimulation of the helicase or the ATPase activity of yPrp43 significantly (see section 3.3.1).

3.3.3 Fusion of yPrp43 and yNtr1

In order to prevent complex dissociation before crystallisation, a functional fusion protein was designed and purified that contained both yPrp43 and yNtr1. In a first approach, a construct was purified in which yNtr1(1-120) was directly fused to the C-terminus of yPrp43 with an C-terminal Strep-tag. This fusion protein, referred to as yPrp43-yNtr1(1-120), was soluble and could be prepared following the same protocol as for full-length yPrp43 (see section 3.1.1).

It turned out that the fusion protein exhibited ATPase as well as helicase activity (data not shown). Subsequently, the linker size was reduced which initially consisted of the first 50 residues of yNtr1. Several fusion proteins were purified and tested with respect to their ATPase and helicase activities (data not shown). It turned out that the direct fusion of yNtr1(51-110) to the C-terminus of yPrp43 resulted in a protein exhibiting significantly higher ATP hydrolysis and dsRNA unwinding rates than free yPrp43.

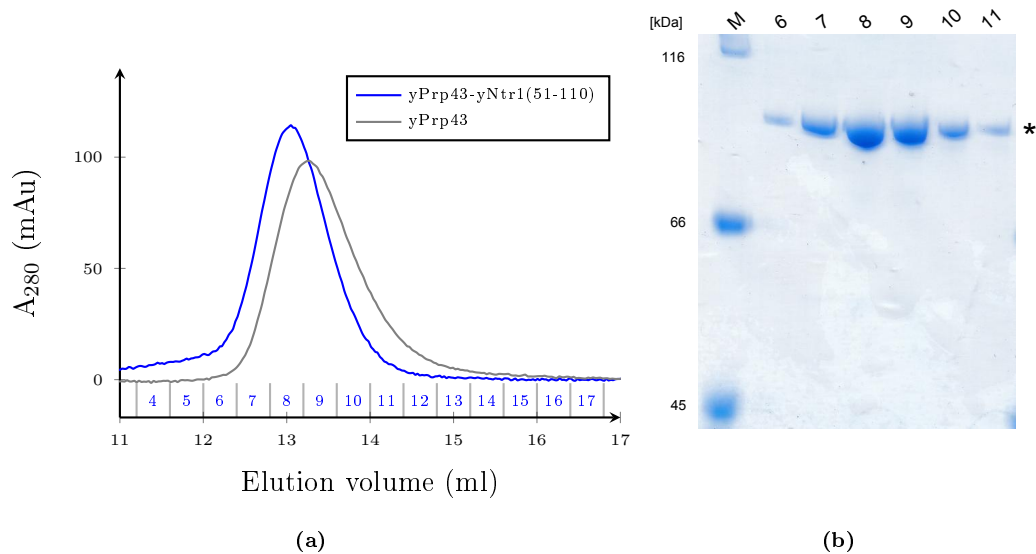


Figure 3.18: Result of the analytical S200 size exclusion chromatography of the fusion protein yPrp43-yNtr1(51-110) in comparison to yPrp43, using a Superdex 200 (10/300) column.

(a) The chromatogram shows the absorption at 280 nm in a machine specific unit (mAu, y-axis) with respect to the elution volume (ml, x-axis). The fusion protein yPrp43-yNtr1(51-110) is shown in blue, yPrp43 in grey. The collected fractions from the run of the fusion protein are indicated along the x-axis. The complete chromatogram is shown in the supplemental figure A.7.

(b) Coomassie-stained 7.5 % SDS-PA gel showing the molecular weight marker (M) and elution fractions (numbers) collected during the gel filtration of yPrp43-yNtr1(51-110). The position of the fusion protein is indicated by an asterisk.

By analytical size exclusion chromatography it could be shown that the fusion

3 Results

protein yPrp43-yNtr1(51-110) behaves as a monomer in solution (shown in figure 3.18), leading to the perception that every G-patch motif interacted with the yPrp43 molecule it was fused to. The elution volume of the fusion protein was slightly decreased compared to free yPrp43, corresponds to a molecular mass of 94.7 kDa for the fusion protein and of 88.6 kDa for free yPrp43. An intermolecular interaction of the G-patch motif with yPrp43 would have led to either a greater shift in case of dimerisation or to aggregation.

The maximal ATP hydrolysis rate of the fusion protein was $18.79 \pm 0.66 \mu\text{M}\cdot\text{s}^{-1}$ and therefore about twice as high compared to the complex yPrp43/yNtr1(51-110) (see figure 3.19 and table 3.5). Likewise, the helicase activity of the fusion protein is higher than the helicase activity of the complex: the maximal unwinding rate of yPrp43-yNtr1(51-110) is $127.2 \pm 8.4 \text{ nM}\cdot\text{min}^{-1}$ and thus about four times higher than the maximal unwinding velocity of yPrp43/yNtr1(51-110).

Taken together, these results show that the fusion protein yPrp43-yNtr1(51-110) has similar properties as the complex. The higher activities that were found for the fusion protein are most probably due to a decrease of the virtual K_d within the fusion protein.

Unfortunately, although extensive crystallisation attempts of the fusion protein yPrp43-yNtr1(51-110) have been carried out at different protein concentrations and in presence of different nucleotides and RNAs, they did not result in any crystals.

3.3.4 Identification of the minimal required yNtr1-fragment

In order to identify the minimal yNtr1-fragment that is required for the activation of yPrp43, it was tried to express yNtr1-truncations smaller than yNtr1(51-110), but no soluble protein was obtained using different tags and buffer conditions during cell lysis and purification. Synthetically produced yNtr1 peptides, namely yNtr1(51-70), yNtr1(71-90), yNtr1(91-110) and yNtr1(63-92), did not show any stimulatory effect on the ATPase activity of yPrp43. Furthermore, the presence of the first three peptides at the same time did not stimulate yPrp43, although they together span the required yNtr1-fragment 51-110. Obviously, these three isolated fragments have a too low intrinsic affinity for yPrp43 and are hence not able to bind to yPrp43 independently.

As all yNtr1-fragments smaller than 51-110 could not be purified successfully, fusion proteins of yPrp43 with different yNtr1 fragments were expressed and pu-

3.3 Functional characterisation of the interaction between yPrp43 and yNtr1

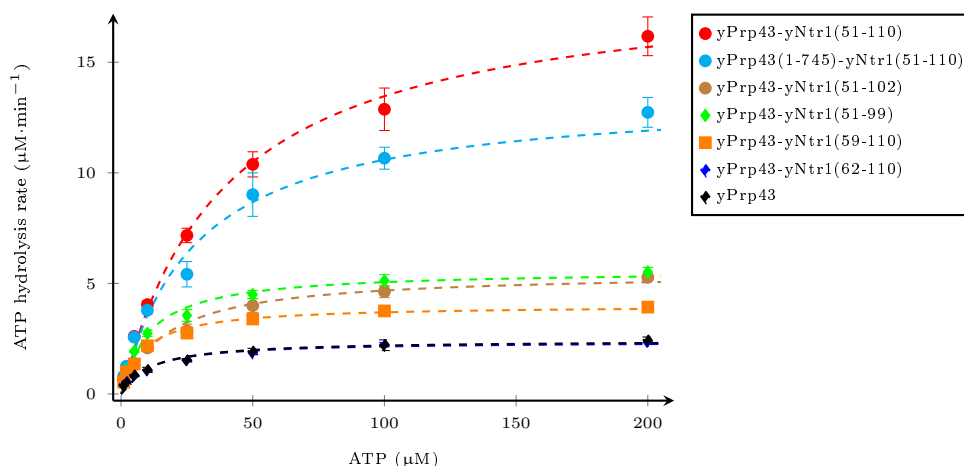


Figure 3.19: ATPase activity for yPrp43 and yPrp43-yNtr1 fusion proteins. The ATP hydrolysis rate (y-axis) in relation to increasing concentrations of ATP (x-axis, μM) is shown. The symbols indicate the mean value of three independent measurements, the error bars the root mean square deviation and the dashed line the hyperbola fit.

refined. The determination of the ATPase activity of these fusion proteins revealed that C- or N-terminal reduction of yNtr1(51-110) caused a dramatic decrease in the ATPase activity of the fusion protein (shown in figure 3.19 and table 3.5). To exclude that the reduced activity of the N-terminal truncated yNtr1-fragments was due to a too short linker, the yPrp43(1-745)-yNtr1(51-110) fusion protein was purified and characterised. In the ATPase activity assay, it exhibited a significantly higher activity compared to the N-terminal truncated yNtr1-fragments, indicating that the decrease in activity for the other fusion proteins was due to the lack of essential residues of yNtr1.

Table 3.5: Kinetic parameters describing the ATPase activity for yPrp43 and different yPrp43-yNtr1 fusion proteins. For each protein, the Michaelis constant K_m and the maximal reaction velocity v_{max} is provided.

Protein	K_m (μM)	v_{max} ($\mu\text{M}\cdot\text{min}^{-1}$)
yPrp43-yNtr1(51-110)	39.53 ± 3.92	18.79 ± 0.66
yPrp43(1-745)-yNtr1(51-110)	28.33 ± 4.61	13.59 ± 0.87
yPrp43-yNtr1(51-102)	17.50 ± 2.43	5.50 ± 0.22
yPrp43-yNtr1(51-99)	10.68 ± 1.38	5.60 ± 0.18
yPrp43-yNtr1(59-110)	8.77 ± 1.09	4.02 ± 0.12
yPrp43-yNtr1(62-110)	10.62 ± 2.18	2.38 ± 0.12
yPrp43	11.07 ± 2.07	2.43 ± 0.12

In conclusion, yNtr1(51-110) seems to be both necessary and sufficient for the activation of yPrp43. Longer fragments of yNtr1 lead to a slightly higher affinity,

whereas shorter truncations cannot stimulate yPrp43 to a comparable extent.

3.4 Structural characterisation of the interaction between yPrp43 and yNtr1

As all crystallisation trials for the yPrp43/yNtr1 complexes were not successful and no structural information about the interaction between yPrp43 and yNtr1 could be obtained, a cooperation was started with Romina Hofele (Laboratory of Bioanalytical Mass Spectrometry, Prof. Henning Urlaub, Max Planck Institute for Biophysical Chemistry, Göttingen). Based on the results of different mass-spectrometric methods and approaches, information regarding the binding site of yNtr1 and the movement of the domains upon the binding of yNtr1 to yPrp43 could be deduced.

3.4.1 Chemical cross-linking of the yPrp43/yNtr1 complexes

In order to identify the binding site of yNtr1 on yPrp43, chemical protein-protein cross-linking was performed by Romina Hofele. Two different lysine cross-linkers, BS3 and BS2G, spanning 11.4 Å and 7.7 Å, respectively, were used. Suitable ratios of protein and cross-linker were identified using SDS-PAGE and the cross-linked lysine residues were analysed via mass spectrometry (Sinz, 2006).

Protein-protein cross-linking data has to be interpreted cautiously for two main reasons: firstly, cross-links do not indicate direct interactions, but lysines which are within the distance determined by the spacer arm of the cross-linking reagents, in this case of 11.4 Å and 7.7 Å for BS3 and BS2G, respectively. Secondly, the absence of cross-links does not exclude the possibility of interactions, as this absence can be due to a lack of lysines in a suitable distance for being cross-linked. These restrictions have to be taken into account when this method is used for the identification protein-protein interaction sites.

For the complex yPrp43/yNtr1(1-120), in total 42 intramolecular and 16 inter-protein cross-links were identified. The two lysine residues K60 and K67 within the G-patch motif of yNtr1(1-120) were cross-linked with lysines located in the C-terminal part of yPrp43, residues 733-766 (see figures 3.20 and 3.21). As in the crystal structure no electron density for the residues 750-767 could be identified,

3.4 Structural characterisation of the interaction between yPrp43 and yNtr1

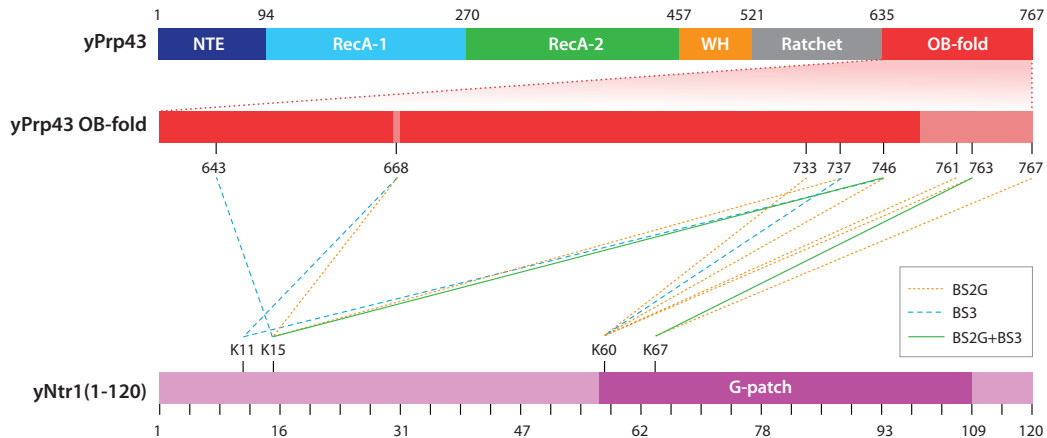


Figure 3.20: Result of the protein-protein cross-linking experiments for the yPrp43/yNtr1(1-120) complex. This scheme provides the position of the identified cross-links according to the primary structure of yPrp43 and yNtr1(1-120). The first bar shows the full-length domain organisation of yPrp43. As cross-links were solely found in the OB-fold domain of yPrp43, this region is enlarged (second bar) – regions in light red are absent in the crystal structure. yNtr1(1-120) is represented in purple (third bar). The cross-links are indicated as lines with respect using which cross-linker they have been found.

this part of yPrp43 is thought to be very flexible.

Two additional lysine residues of yNtr1(1-120) which are not part of the G-patch motif, namely K11 and K15, have been found to cross-link to K737 and K746, but also to K668 and K643 of the OB-fold domain. As only residues 51-110 of yNtr1 seem to play a functional role for the activation of yPrp43 (see section 3.3.4), the region in which these residues are located presumably only contributes to the binding to yPrp43 and not its activation. This finding is also consistent with the obtained dissociation constants which indicated a higher affinity of yNtr1(1-120) in comparison to yNtr1(51-110) for yPrp43 (see section 3.3.2).

The protein-protein cross-links identified within the complex yPrp43/yNtr1(51-110) confirm the results obtained for yPrp43/yNtr1(1-120): K60 and K67 cross-link with the C-terminal residues K733, K737, K747, K763 and K766 (data not shown). Additionally, a cross-link between yNtr1-K78 and yPrp43-K4 has been identified. K4 is located in the N-terminal tail of yPrp43 which encompasses 19 residues. Their location in the structure is presumably a crystallographic artefact; most probably they are very flexible and thus in close proximity to the C-terminal part of yPrp43. That is why the cross-link between yNtr1-K78 and yPrp43-K4 does not necessarily indicate a direct interaction between the G-patch motif and the N-terminal part of yPrp43.

3 Results

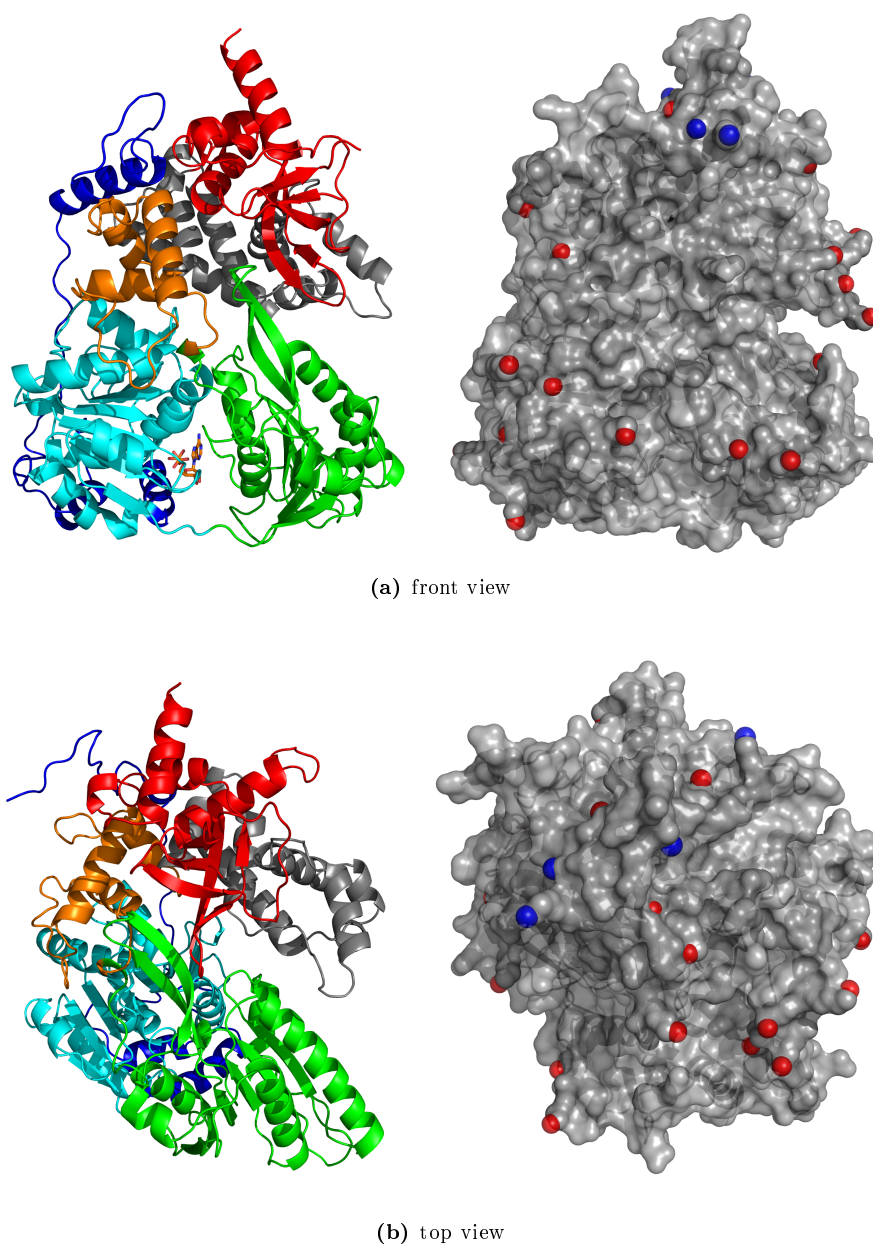


Figure 3.21: Location of the identified cross-linked lysine residues in the crystal structure of yPrp43 (pdb-ID 2xau) in front view (a) and top view (b). Each view is shown as cartoon (left) and in the surface representation (right). Within the surface representation, blue surface corresponds to cross-linked lysines and red surface to lysines that have not been cross-linked in the complex yPrp43/yNtr1(1-120). The four cross-linked lysine residues that are present in the crystal structure of yPrp43 (K643, K733, K737 and K746) are located in the OB-fold domain.

3.4.2 Hydroxyl radical footprinting of the *yPrp43/yNtr1*(1-120) complex

For the elucidation of conformational changes that are induced in *yPrp43* upon binding of *yNtr1*, hydroxyl radical footprinting experiments and data analysis were performed by Romina Hofele. In this method, the complex and the single proteins are exposed to hydroxyl radicals. The residues which have been modified by hydroxyl radicals can be identified and quantified by mass spectrometry. The comparison of the hydroxylation pattern between the single proteins and the complex provides information about residues that get more or less exposed upon complex formation (Wang and Chance, 2011). As there are residues like e.g. cysteine and tryptophan that are more prone to oxidation than e.g. asparagine and glycine, the obtained data from these experiment has to be interpreted in relative terms.

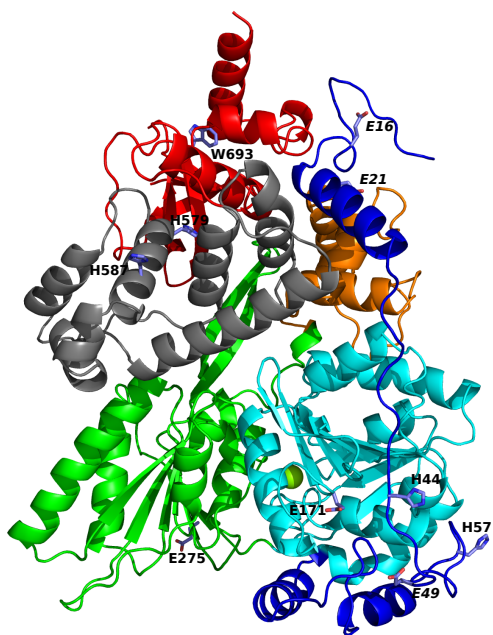


Figure 3.22: Cartoon representation of *yPrp43* (pdb-ID 2xau) in the front view. All residues that showed a different oxidation pattern between the complex *yPrp43/yNtr1*(1-120) and unliganded *yPrp43* are shown as sticks and are labelled. Residues in italics were detected to be less oxidised within the complex, whereas the other residues showed a higher oxidation level upon binding of *yNtr1*(1-120) to *yPrp43*.

The data obtained during the hydroxyl radical footprinting experiments of the complex *yPrp43/yNtr1*(1-120) revealed that residues in the RecA1 (E171), the RecA2 (E275), the ratchet (H579 and H587) and the OB-fold domain (W693) are oxidised to a higher extent in the complex, indicating significant movements in all domains of *yPrp43* (see figure 3.22). The evidence for conformational changes in the ratchet domain and in the C-terminal domain confirm the indications that the C-terminal domain of *yPrp43* interacts with the G-patch motif of *yNtr1* (see section 3.4.1).

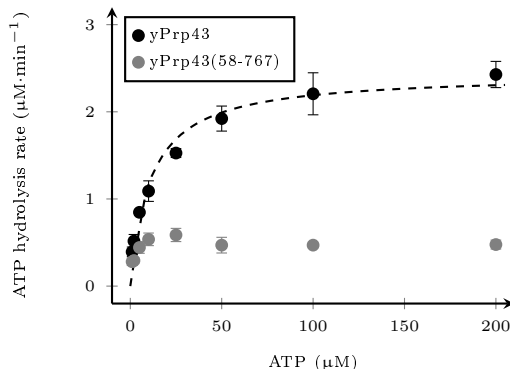
3 Results

Additionally, five residues showing different oxidation patterns in the complex are located in the N-terminal part of yPrp43: the residues E16 and E49 were less oxidised, the residues E49, H44 and H57 more oxidised in the complex. These results raised the question about the functional role of the N-terminal extension of yPrp43.

3.4.3 The role of the N-terminal extension of yPrp43

To investigate the function of the N-terminal extension of yPrp43, a deletion variant of the helicase was expressed and purified, namely yPrp43(58-767). This variant lacked the first 57 amino acids of yPrp43 and hence the N-terminal α -helix which interacts with the winged helix domain. The determination of ATPase activity of this variant revealed that it exhibited nearly no ATP hydrolysis activity (shown in figure 3.23), indicating that this helix is essential for the intrinsic ATPase activity of yPrp43. Interestingly, analytical gel filtration revealed that yNtr1(1-120) was still able to bind to yPrp43(58-767) (see figures 3.24 and 3.25) and to stimulate the ATP hydrolysis rate of yPrp43 (data not shown). Most importantly, the yPrp43(58-767)/yNtr1(51-110) complex exhibited a maximal unwinding rate of 29.4 ± 2.4 nM \cdot min $^{-1}$ and which is comparable to the yPrp43/yNtr1(51-110) complex (see section 3.3.1).

Figure 3.23: ATPase activity of yPrp43 (black) and yPrp43(58-767) (grey). The ATP hydrolysis rate (y-axis) is shown with respect to increasing concentrations of ATP (x-axis, μ M). The circles indicate the mean value, the error bars the root mean square deviation and the dashed line the hyperbolic fit. Due to the low ATP hydrolysis rate, no reasonable fitting was possible for the data points of yPrp43(58-767).



Therefore, the decreased oxidation level of residue E16 of yPrp43 seems to be a consequence of a protection which is not caused by direct binding of yNtr1, but by a general rearrangement of the C-terminal part of yPrp43. The deleted helix is in direct contact with the winged helix domain and it suggested to play no functional role when yPrp43 forms a complex with yNtr1 as its deletion does not influence the helicase activity in presence of yNtr1(51-110).

3.5 Interaction of yPrp43 with its substrates

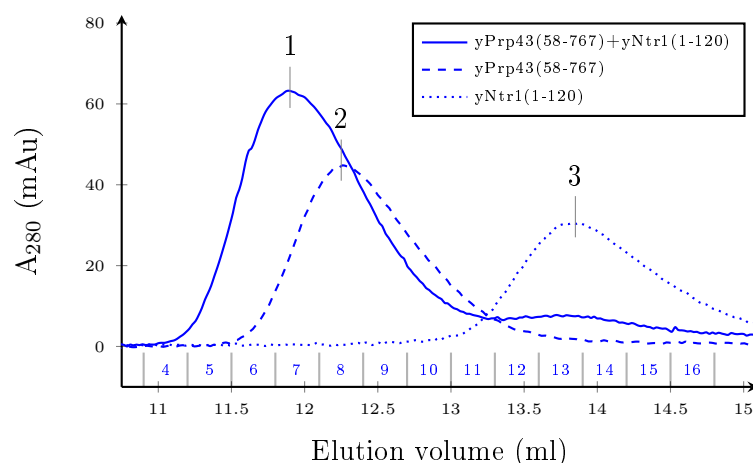


Figure 3.24: Chromatogram of the analytical size exclusion chromatography of the complex yPrp43(58-767)/yNtr1(1-120), using a Superdex 200 (10/300) column. The absorption at 280 nm (y-axis, machine-specific unit) during the three independent runs of yPrp43(58-767)+yNtr1(1-120), yPrp43(58-767) and yNtr1(1-120) is shown with respect to the elution volume (x-axis, ml). The fractions along the x-axis have been collected during the run of the complex. The local maxima are indicated by numbers: (1) corresponds to the complex, (2) to yPrp43(58-767) and (3) to yNtr1(1-120). The complete chromatogram is shown in the supplemental figure A.8.

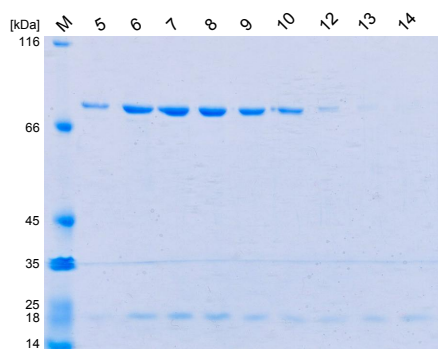


Figure 3.25: Coomassie-stained 15 % SDS-PAGE gel of the analytical gel filtration of the complex yPrp43(58-767)/yNtr1(1-120). On the lanes, the molecular weight marker (M) and elution samples from the chromatography (numbers) are shown. The position of yPrp43(58-767) is indicated by an asterisk, of yNtr1(1-120) by a circle.

3.5 Interaction of yPrp43 with its substrates

The characterisation of the interaction of yPrp43 with its substrates, namely nucleic acids and nucleotides, is essential for understanding its function. Based on protein-RNA cross-linking and nucleotide loading state analysis, the interactions on its own and the influence of yNtr1 on these interactions were investigated.

3.5.1 UV-induced RNA cross-linking of yPrp43

The identification of residues responsible for interaction with RNA was performed by Romina Hofele using UV-light induced protein-RNA cross-linking in combination with mass spectrometry. UV-light acts as a "zero-length cross-linker" as it

3 Results

generates a covalent bond between the interacting RNA base and the side chain of the protein.

The yPrp43/yNtr1(1-120) complex was incubated with a γ - ^{32}P -labelled ss-RNA oligonucleotide and exposed to UV radiation. The analysis of the samples by SDS-PAGE revealed that both, yPrp43 and yNtr1(1-120), could be cross-linked to RNA (see supplemental figure A.12). To obtain more detailed information about the molecular interaction between the complex and the nucleic acid, both yPrp43 and the yPrp43/yNtr1(1-120) complex were incubated with the RNA oligonucleotide and subsequently irradiated by UV-light. The samples were treated with RNases and trypsin and analysed by LC-ESI-MS/MS (Richter *et al.*, 2009).

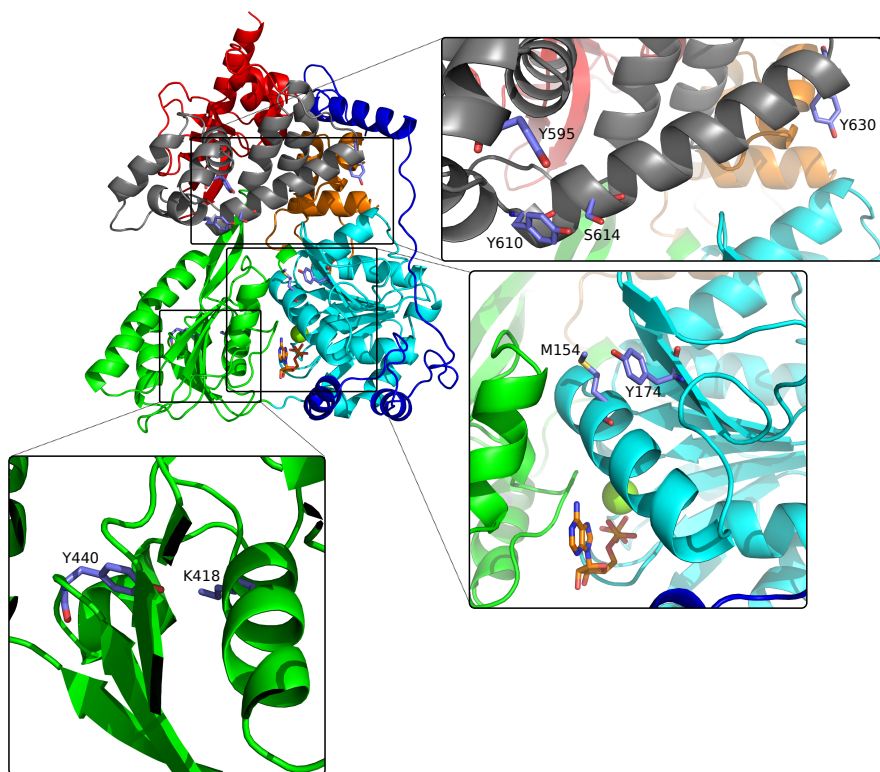


Figure 3.26: Cartoon representation of yPrp43 (pdb-ID 2xau) in back view with zoom panels showing the residues identified during protein-RNA-cross-linking experiments. The residues are located in the RecA-1 (cyan), the RecA-2 (green) and the ratchet domain (grey).

In yPrp43, eight cross-links between a side chain and the nucleic acid could be identified (see figure 3.26). The cross-linked residues were located in the RecA-like domains (M154, Y174, K418, Y440) and in the Ratchet domain (Y595, Y610, S614 and Y630). In addition, a cross-link of the residue K67 of yNtr1 to the ssRNA could be identified, indicating that the G-patch motif is directly involved in the RNA binding. No qualitative differences between the cross-linked residues

of the helicase alone and the complex was found.

A semi-quantitative analysis of the cross-linking data revealed that the presence of yNtr1 enhanced the cross-linking yield, especially for three residues located in the Ratchet domain (Y595, Y610 and S614). Although this result does not necessarily indicate a stimulation of RNA affinity induced by yNtr1, it suggests at least that the RNA is in a better position to be cross-linked in the complex.

3.5.2 Nucleotide binding properties of yPrp43

The nucleotide loading state of yPrp43 alone and in complex with yNtr1 was determined qualitatively in order to test if the presence of yNtr1 influences the nucleotide affinity of the helicase. Nucleotide-free protein was incubated with an excess of a certain nucleotide, desalted and denatured. As the nucleotide did not precipitate, it was separated from the protein and subsequently analysed by anion exchange column. The different nucleotides were distinguished by their different elution volumes with respect to a salt gradient.

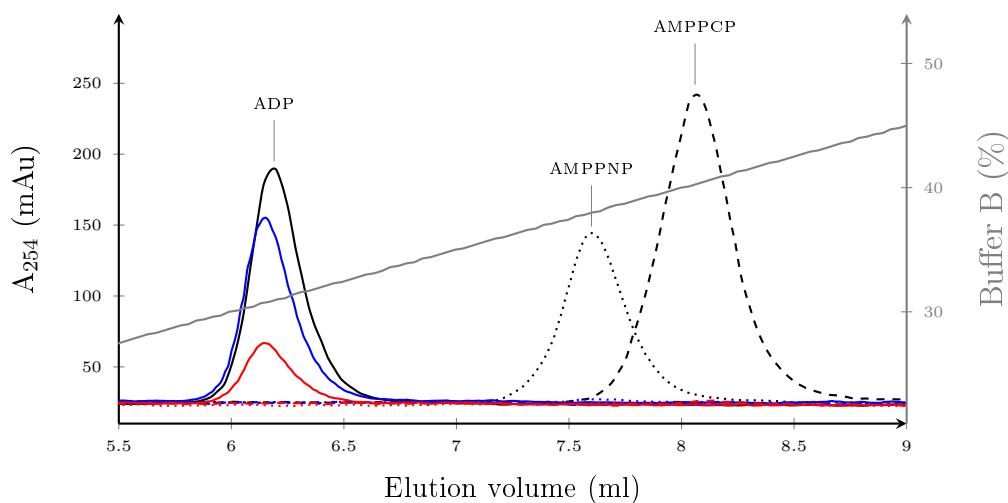


Figure 3.27: Elution profiles of nucleotide loading state determination using a Nucleosil 4000-7 HPLC column. The absorption at 254 nm (in a machine specific unit, left y-axis) is shown with respect to the elution volume (ml, x-axis). The grey line refers to the right y-axis and indicates the concentration of buffer B. The control runs of ADP (continuous line), AMPPCP (dashed line) and AMPPNP (dotted line) are colored in black. The blue lines correspond to free yPrp43 and the red lines to the fusion protein yPrp43-yNtr1(51-110), with continuous line for the ADP, dashed line for the AMPPCP and dotted line for the AMPPNP states, respectively.

The elution profiles from the anion exchange column were determined for yPrp43 and the fusion protein yPrp43-yNtr1(51-110) after the proteins had been

3 Results

incubated with ADP, AMPPNP and AMPPCP, respectively (see section 2.2.3.7). The comparison with the elution volumes of control runs containing only one nucleotide revealed that both yPrp43 and the fusion protein yPrp43-yNtr(51-110) were able to bind ADP, but did not bind the non-hydrolysable ATP analogues AMPPNP or AMPPCP (see figure 3.27). This result suggests that yPrp43 has a low affinity to ATP compared to ADP and that the G-patch motif of yNtr1 does not significantly influence the affinity of yPrp43 for ATP.

3.6 Analysis of the secondary structure of the G-patch motif

No structural information was available for the G-patch motif neither in form of a crystal nor an NMR structure (see section 1.6.4). During this work, all crystallisation attempts for both yNtr1(51-110) and yNtr1(1-120) in complex with yPrp43 as well as in their free form failed. Furthermore, the complex of both proteins could not be concentrated to concentrations suitable for NMR measurements which require at least a concentration of 200 μM . Therefore, circular dichroism spectroscopy and bioinformatical approaches were chosen to gain information about the structure of the G-patch motif.

3.6.1 Analysis of the secondary structure of yNtr1(51-110)

Far-UV circular dichroism (CD) spectroscopy was performed for the determination of content of secondary structure elements within yNtr1(51-110). The CD spectrum of this yNtr1-fragment indicated the absence of secondary structure elements (see figure 3.28, black line). The capacity of evolving secondary structure elements was tested by determination of far-UV CD spectra in presence of the small molecule trifluoroethanol (TFE) at different concentrations. TFE has been reported to support the formation of α -helices (Luo and Baldwin, 1997). For concentrations of 20 % (v/v) of TFE and beyond, the CD spectrum changed significantly and gave rise to two local minima at 208 and 222 nm, which was an indication for the presence of α -helices (see figure 3.28).

The CD spectra were hence indicating that yNtr1(51-110) is natively unstructured, but additionally that it has a capacity to form α -helices. Using different bioinformatical structure prediction tools, the presence of an α -helix was predicted

3.6 Analysis of the secondary structure of the G-patch motif

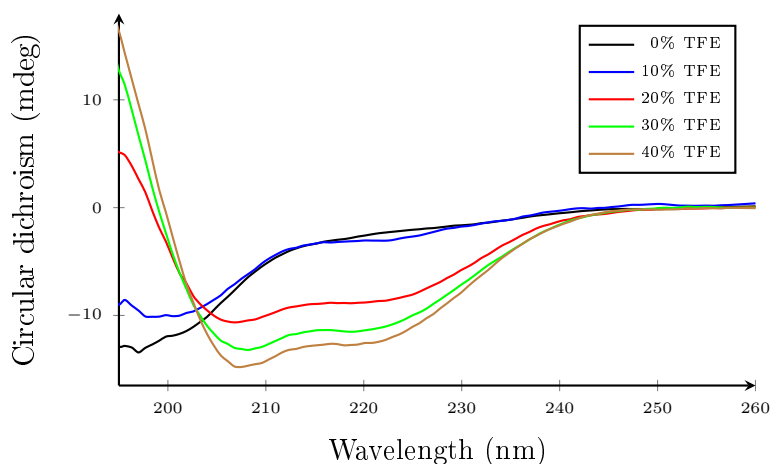


Figure 3.28: Far-UV CD spectra of yNtr1(51-110) with increasing concentrations of trifluoroethanol (TFE, 0-40% (v/v)). The circular dichroism (in mdeg, y-axis) is shown with respect to the wavelength (in nm, x-axis).

for residues 64-71 (3D-JIGSAW) and 63-71 (Phyre and SAM-TO2), respectively. Additionally, these programs predict the presence of β -strands in residues 91-95 (3D-JIGSAW), 91-96/102-106 (Phyre) and 86-89/91-95/105-107, respectively. The signal for β -strands in a CD spectrum can be detected by a decrease in signal between 200 and 210 nm and is similar to the signal indicating α -helices in this range, hence the presence of a β -strand in yNtr1(51-110) could not be proven in the measured CD spectra.

3.6.2 Bioinformatical structural predictions for yNtr1(51-110)

In a subsequent step, it was investigated whether the described potential secondary structure elements might evolve upon binding of yNtr1(51-110) to yPrp43. The IUPRED server has been reported to predict disordered regions in proteins by the calculation of unfavorable estimated energies. The ANCHOR server combines the disorder tendency predicted by IUPRED with the sensitivity to the structural environment and thus identifies regions within the protein that might form defined secondary structure elements when binding to a globular protein. The analysis of both servers is exclusively based on the primary structure of the protein.

While the predicted disorder probability was high for the N-terminal 150 residues (see figure 3.29), it was on a very low level for the residues 151 to 600 and in-

3 Results

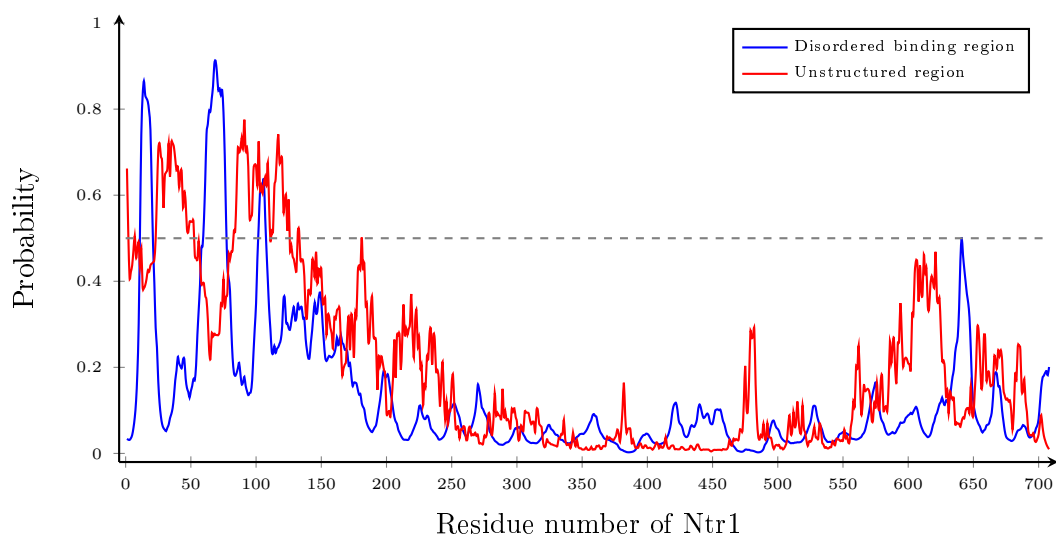


Figure 3.29: Prediction of disordered binding regions (blue) and unstructured region (red) for yNtr1. The residue numbers are indicated on the x-axis, the probability is shown on the y-axis. A probability higher than 0.5 (grey dashed line) indicates a disordered state (for IUPRED, red line) and a potential for forming structures in presence of a globular partner (for ANCHOR, blue line) for the respective residues. The values were obtained using the Anchor-server (Dosztányi *et al.*, 2009).

creased again for the last 100 C-terminal residues. Three regions of yNtr1 were predicted to be disordered binding regions: residues 11-20, 60-77 and 102-107; of these, the second and third are part of the minimal required activating fragment yNtr1(51-110). Residues 60-77 overlap with a region predicted to form α -helices – interestingly, this region was predicted to be unstructured at a probability of 0.3, which was lower compared to the residues 1-150, but still quite high compared to the region between 150-600.

Combining the experimental data collected by far-UV CD spectroscopy with the bioinformatical predictions, the following model for the G-patch motif of yNtr1 can be outlined: The residues forming the G-patch motif of yNtr1 are intrinsically unstructured, but they possess characteristics indicating that they have a tendency to form secondary structure elements, namely α -helices and β -strands. Regarding full-length yNtr1, the G-patch containing N-terminal 150 residues and the last C-terminal 100 residues are predicted to be unstructured, whereas the central part of the protein (amino acids 150-600) seems to form a structured region.

3.7 Surface charge calculation of human helicase models

The crystal structure of the C-terminal domains (CTDs) of human Prp22 (hPrp22) was solved by Denis Kudlinzki (Department of Molecular Structural Biology, Prof. Ralf Ficner, Universität Göttingen). The structure was highly similar to the C-terminal domains that are present in the structure of yPrp43 (He *et al.*, 2010; Walbott *et al.*, 2010). It is composed of a winged helix, a ratchet and an OB-fold domain. Furthermore, it was shown that the CTDs stimulate the ATPase activity of yPrp22 in absence and in presence of RNA and that the ATPase activity of yPrp22 is only stimulated by RNA if the C-terminal domains are present (for further details see Kudlinzki *et al.* (2012)). Due to the high sequence homology between the C-terminal domains of the spliceosomal DEAH-box helicases, the structures of hPrp2, hPrp16 and hPrp43 could be modelled.

As a part of this thesis, the obtained models were compared with the structures of the CTDs of hPrp22 and yPrp43. The overall folding of the CTDs was very similar. In order to detect differences between these domains, the surface charge distribution was calculated (see figure 3.30 on page 88). It turned out that the CTDs of the structure and the models contain conserved patches which are mainly positively charged, most probably they serve as interaction surface for the nucleic acid substrate.

Besides the conserved patches, there are also patches which are unique for the different CTD-models. Although it has to be taken into account that the surface charge calculation of modelled structures might be speculative, they suggest that these differences have evolved to assure the specific interaction with other proteins for each distinct spliceosomal DEAH-box helicase.

3 Results

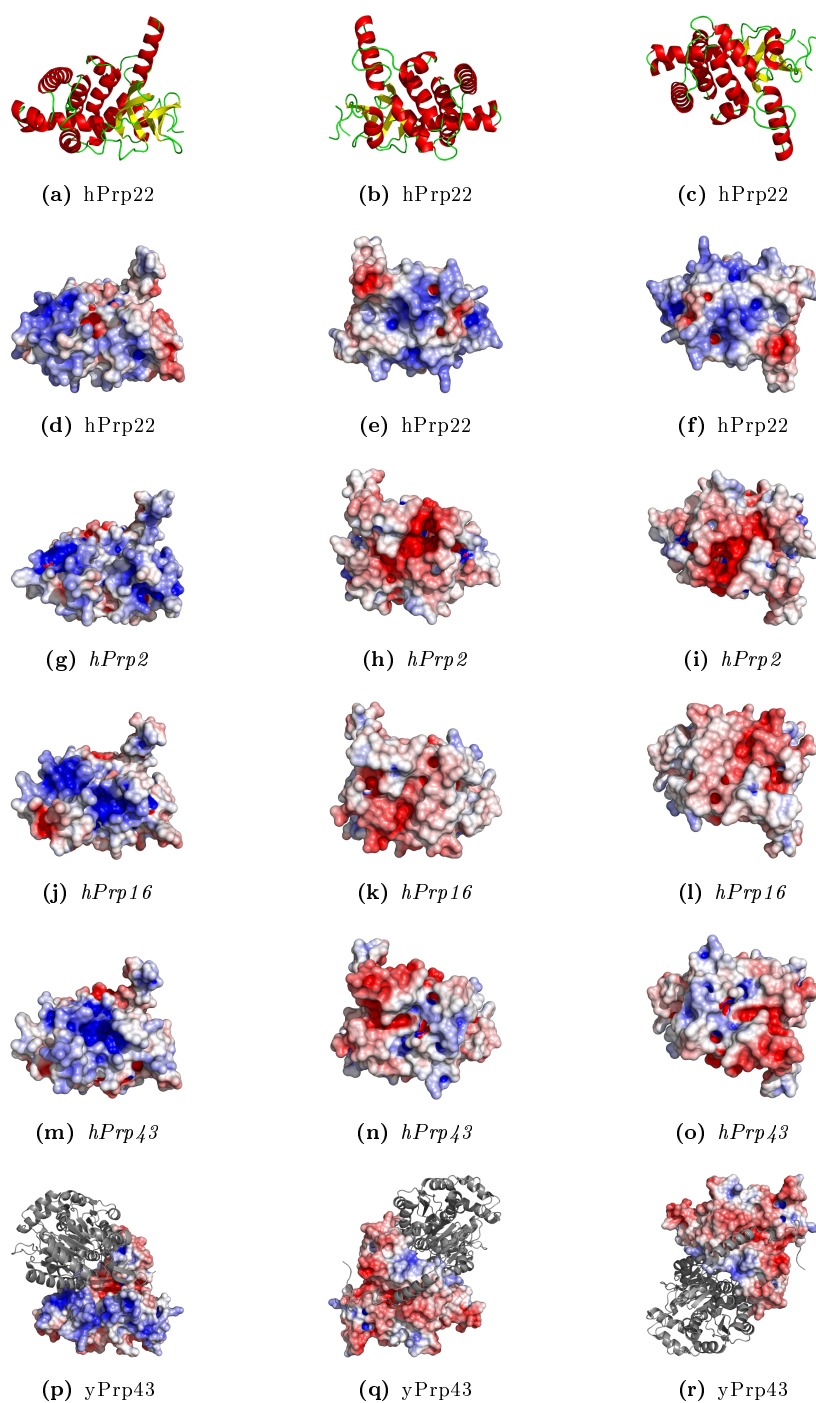


Figure 3.30: Comparison of the surface charge distribution of the CTDs of hPrp2, hPrp16, hPrp22, hPrp43 and yPrp43. Modelled structures are shown in italics. The first column shows the front view, the second column the back view and the third column the bottom view. The first row provides the structure of hPrp22 (pdb-ID 3i4u), the last row the surface charge of yPrp43 (pdb-ID 2xau) with the helicase core in grey ribbon representation. The surface charge distribution was calculated with pdb2pqr and visualised by the APBS plugin of PyMOL. The contour level is ± 5 kBT for all surface figures. Blue color indicates positive charge, red color negative charge.

3.8 Prp43 of *C. thermophilum*

While the practical work of this thesis was ongoing, reconstitution studies of the inner nuclear pore complex using proteins from the soil inhabiting fungus *Chaetomium thermophilum* were published (Amlacher *et al.*, 2011). In the same study, the complete genome of this thermophilic eukaryote was provided. Proteins of thermophilic organisms are thought to be in general more stable and hence are promising targets for crystallisation due to their biochemical and biophysical properties.

Ever since, several crystal structures of proteins from *C. thermophilum* have been published, e.g. the export receptor CRM1 (Monecke *et al.*, 2013) and the ribosome-associated complex RAC (Leidig *et al.*, 2013). These findings led to the motivation to identify and characterise the orthologs of the NTR complex in *C. thermophilum*.

3.8.1 Identification and cloning of the NTR complex orthologs

For the identification of a potential ortholog of Prp43 in *C. thermophilum* (ct-Prp43), a protein-protein BLAST search was performed aligning the sequence of yPrp43 against all proteins of *C. thermophilum var. thermophilum DSM 1495*. In this alignment, the best score was obtained for a protein annotated as "hypothetical protein CTHT_0005780". Sequence comparison revealed a sequence identity of 65.6 % and a similarity of 79.5 % between yPrp43 and the ctPrp43 candidate (see table 3.6 and supplemental figure A.13). Due to this high homology it was obvious that the identified protein represents indeed ctPrp43.

The gene coding for ctPrp43 contains one intronic sequence encompassing 64 nucleotides, therefore the whole gene was amplified from genomic DNA and inserted into an expression vector. Subsequently, the intron was deleted via mutagenesis PCR (see section 2.2.5).

For the identification of ctNtr1, in a first BLAST the G-patch motif of yNtr1 was aligned against all proteins of *C. thermophilum var. thermophilum DSM 1495*. In total, seven different genes coding for proteins containing a G-patch motif were identified in the genome of *C. thermophilum*. Among these, a protein annotated as "hypothetical protein CTHT_0020180" showed the highest homology to full-length yNtr1 with a sequence identity of 15.8 % and a similarity of 28.2 % (see

3 Results

Table 3.6: Sequence comparison of the orthologs of Prp43, Ntr1 and Ntr2 in *S. cerevisiae* and *C. thermophilum*. For each ortholog pair, the identity and the similarity of the sequence as well as the number of gaps are provided. The calculations have been performed using the EMBOSS needle server with the matrix EBLOSUM62 (gap penalty=10). The corresponding alignments are shown in the supplemental figures A.13, A.14 and A.15.

	Identity	Similarity	Gaps
Prp43	516/786 (65.5 %)	625/786 (79.5 %)	41/786 (5.2 %)
Ntr1	167/1063 (15.7 %)	300/1063 (28.2 %)	443/1063 (41.7 %)
Ntr2	81/559 (14.5 %)	144/559 (25.8 %)	286/559 (51.2 %)

table 3.6).

The G-patch motif which is located between residues 246 and 297 of ctNtr1 shows a high degree of conservation with a sequence identity of 37.9 % and a similarity of 43.9 % (see alignment in figure 3.31). Besides the G-patch motif, both proteins contain additional conserved residues and regions. Furthermore, the ctNtr1 candidate contains several insertions spanning eight to 31 amino acids in the middle and larger ones in the N-terminal region (see supplemental figure A.14).

As the gene of ctNtr1 was found to be intron-free, the entire sequence was amplified from genomic DNA and directly cloned into an expression vector.

```

yNtr1 (51-110)  . . N A P T I S K L T K T . . Y G I G A K L L S S M G Y V A G K G L G K D G S G I T T P I E T Q S R P M H N A G L G M F S N T N . .
ctNtr1 (242-306) S A F Q P T K D G K T K I N P N S F G A R M L A K M G Y K E G Q G L G K D G Q G R N V I I E V N L R P . Q G V G L G A V K E K T K Q

```

Figure 3.31: Alignment of the G-patch motifs of yNtr1 (residues 51 to 110) and the identified ctNtr1 candidate of *C. thermophilum* (residues 242-306). The alignment has been performed using ClustalW and was visualised with ESPRIPT. White letters on a red background indicate identity, red letters similarity. The numbers refer to the residues of yNtr1.

For the identification of a potential ortholog of yNtr2 in the genome of *C. thermophilum*, firstly an position-specific iterated BLAST (psi-BLAST) was performed in the group of fungi. In a second step, a psi-BLAST was conducted using the sequences of the best results of the first search, which hence represented the Ntr2 consensus sequence among the fungi. This approach resulted in the identification of a protein annotated as "hypothetical protein CTHT_0074550", which turned out to be a strong candidate for ctNtr2. The sequence identity between this protein and yNtr2 is 14.5 %, the similarity 25.8 % (see table 3.6). The ctNtr2 candidate contains about 60 % more residues than its potential yeast ortholog. A sequence alignment demonstrated that the ctNtr2 candidate has several large insertions as well as some regions with numerous conserved residues compared to yNtr2. This indicates that both proteins originate from the same precursor protein (see supplemental figure A.15).

The gene of the ctNtr2 candidate was found to contain one intronic sequence. The cloning of ctNtr2 was performed via the separate amplification of both exons and their ligation into an expression vector using the In-fusion cloning technique (see section 2.2.11.3).

Due to the time limits of this work, the candidates for ctNtr1 and ctNtr2 could not be characterised further, whereas the biochemical properties and the structure of ctPrp43 could be investigated.

3.8.2 Functional characterisation of ctPrp43

ctPrp43 was overexpressed in *E. coli* Rosetta 2(DE3) cells as a C-terminal Strep fusion protein and afterwards isolated from the supernatant of the cell lysate using a StrepTactin column. ctPrp43-Strep could be purified to homogeneity through a size exclusion chromatography (see figure 3.32). The peak fractions contained ctPrp43-Strep and appeared at an elution volume corresponding to the molecular weight of the protein, indicating a monomeric state (88.1 kDa, including the Strep-tag).

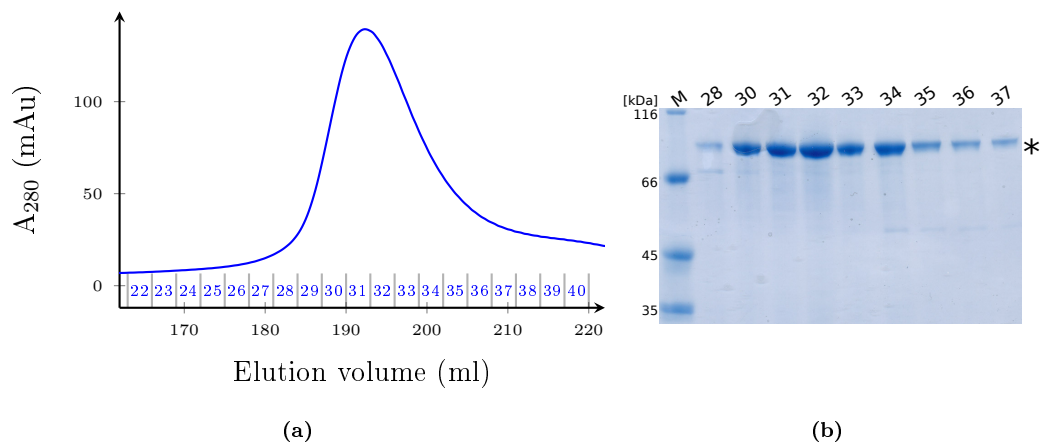


Figure 3.32: Preparative size exclusion chromatography of ctPrp43-Strep, using a Superdex 200 26/60 column.

(a) In the chromatogram, the absorbance at 280 nm in a machine-specific unit (mAu, y-axis) is shown in blue referring to the elution volume (ml, x-axis). The numbers of the collected elution fractions are indicated along the x-axis. The complete chromatogram is shown in the supplemental figure A.9.

(b) Coomassie-stained 12.5 % SDS-PAGE gel of the size exclusion chromatography of yPrp43-Strep. The molecular weight marker (M) and elution fractions (numbers) are shown. The position of the bands according to ctPrp43-Strep is indicated by an asterisk.

Per litre expression culture, 8 mg of pure ctPrp43-Strep were obtained. The detailed protocol of expression and purification for ctPrp43 is provided in the

3 Results

appendix in section A.1.2.

After the purification was established, the biochemical properties of ctPrp43 were investigated. The maximal ATP hydrolysis rate of ctPrp43 turned out to be $4.0 \pm 0.1 \mu\text{M} \cdot \text{min}^{-1}$ and hence 1.7-fold higher compared to Prp43 from yeast ($2.4 \pm 0.1 \mu\text{M} \cdot \text{min}^{-1}$). The Michaelis constant turned out to be $6.2 \pm 0.8 \mu\text{M}$ and therefore lower than for yPrp43 ($11.1 \pm 2.1 \mu\text{M}$) (see figure 3.33 and table 3.7). Additionally, ctPrp43 was tested for its helicase activity – like for yPrp43, no intrinsic unwinding activity for ctPrp43 could be detected (data not shown).

Figure 3.33: ATPase activity of ctPrp43 (blue) and yPrp43 (black). The ATP hydrolysis rate (in $\mu\text{M} \cdot \text{min}^{-1}$, y-axis) is shown in dependence of increasing concentrations of ATP (x-axis, μM). The circles indicate the mean value, the error bars the root mean square deviation and the dashed line the hyperbolic fit.

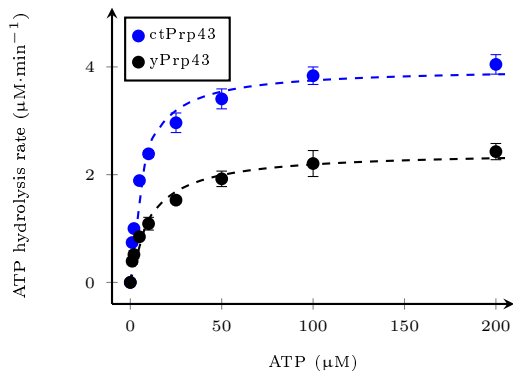


Table 3.7: Kinetic parameters describing the ATPase activity of ctPrp43 and yPrp43 in the absence of RNA. For both proteins, the Michaelis constant K_m and the maximal reaction velocity v_{max} is provided.

Protein	K_m (μM)	v_{max} ($\mu\text{M} \cdot \text{min}^{-1}$)
ctPrp43	6.21 ± 0.79	3.98 ± 0.12
yPrp43	11.07 ± 2.07	2.43 ± 0.12

In order to further investigate the similarities between ctPrp43 and yPrp43, proteins were designed which contained the different domains mixed from both species. A protein containing the NTE of ctPrp43 and the remaining domains from yPrp43 could be successfully purified as well as an additional polypeptide containing the helicase core domains of yPrp43 together with the NTE and the C-terminal domains from ctPrp43. Upon determination of the ATPase activity, it turned out that both designed proteins did not exhibit any ATPase activity (data not shown).

Due to the high homology between the G-patch motifs of yNtr1 and ctNtr1, ctPrp43 was tested for its ability to interact with the purified N-terminal fragments of Ntr1 from *S. cerevisiae*, namely yNtr1(1-120) and yNtr1(51-110). Analytical size exclusion chromatography revealed that ctPrp43 and yNtr1(1-120) were able

to interact and form a stable complex (see figures 3.34 and 3.35). The same result was obtained for the interaction of ctPrp43 and yNtr1(51-110) (data not shown).

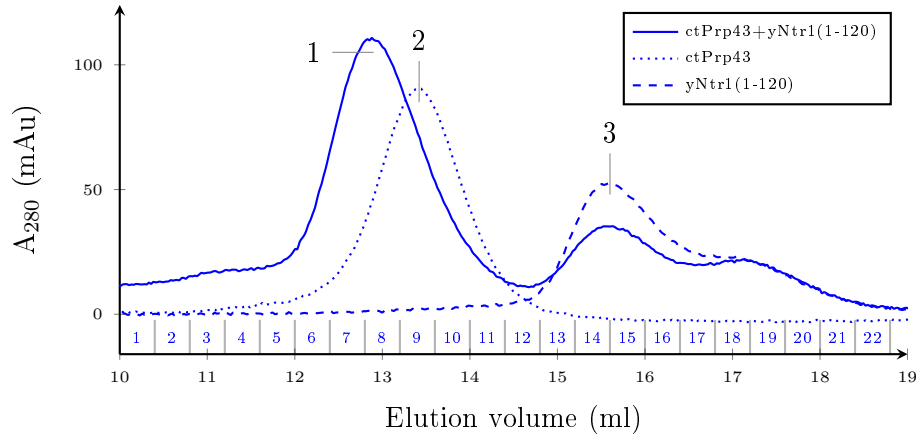


Figure 3.34: Chromatogram of the analytical size exclusion chromatography of the complex of ctPrp43 and yNtr1(1-120), using a Superdex 200 (10/300) column. The absorption at 280 nm (machine specific unit, y-axis) is shown with respect to the elution volume (ml, x-axis) for three independent samples: ctPrp43+yNtr1(1-120), free ctPrp43 and free yNtr1(1-120). The local maxima correspond to the ctPrp43/yNtr1(1-120) complex (1), unliganded ctPrp43 (2) and unliganded yNtr1(1-120) (3). The fractions indicated along the x-axis (numbers) have been collected during the run of ctPrp43 and yNtr1(1-120). The complete chromatogram is shown in the supplemental figure A.10.

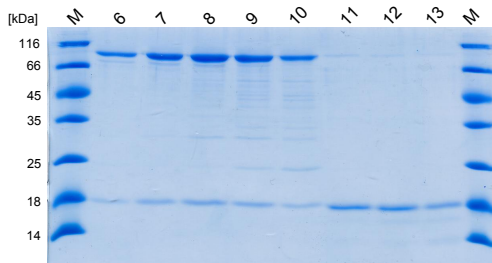


Figure 3.35: Coomassie-stained 15 % SDS-PA gel of the analytical size exclusion chromatography of ctPrp43 and yNtr1(1-120). The marker (M) and elution fractions (numbers) from the gel filtration are shown on the lanes. The position of ctPrp43 is labelled by an asterisk, that of yNtr1(1-120) by a circle.

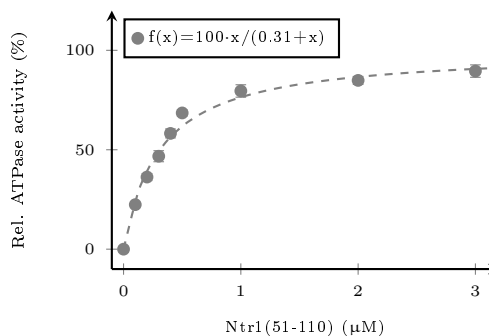
The complex ctPrp43/yNtr1(51-110) was tested for its helicase activity and the maximal unwinding rate was determined to be $80.4 \pm 6.6 \text{ nM} \cdot \text{min}^{-1}$, and hence 2.7-fold higher than for the yPrp43/yNtr1(51-110) complex (see section 3.3.1).

The binding of the G-patch motif of yNtr1 to ctPrp43 resulted in a stimulation of the maximal ATP hydrolysis rate. Therefore the dissociation constant under steady-state turnover conditions for the complex formation of ctPrp43 and yNtr1(51-110) could be determined, in the same way as for yPrp43/yNtr1 complexes (see section 3.3.2). The K_d was $0.31 \pm 0.03 \text{ } \mu\text{M}$ (see figure 3.36) and hence seven-fold lower than for the complex of yPrp43 and yNtr1(51-110), which was $2.24 \pm 0.11 \text{ } \mu\text{M}$. In conclusion, ctPrp43 shows a significantly higher affinity to this fragment of yNtr1 than its yeast ortholog.

3 Results

In contrast to yPrp43, a purification of the fusion protein ctPrp43-yNtr1(51-110) was not feasible because the protein aggregated. This behaviour indicates that predominantly the G-patch motifs were interacting intermolecular and not intramolecular as it was the case for the yPrp43-yNtr1(51-110) fusion protein (see section 3.3.3).

Figure 3.36: Determination of the dissociation constant for the complex of ctPrp43 and yNtr1(51-110). The K_d was determined using the stimulatory effect of yNtr1(51-110) on the ATPase activity of ctPrp43. The relative ATPase activity (in %, y-axis) is plotted against the concentration of yNtr1(51-110) (in μM , x-axis). The equation of the hyperbolic fit is indicated in the graph.



In conclusion, it was shown that ctPrp43 shows a higher ATP hydrolysis as well as unwinding rate and has a higher affinity for yNtr1(51-110) in comparison to yPrp43.

Furthermore, it was shown by Jean-Baptiste Fourmann (Department of Cellular Biochemistry, Prof. Reinhard Lührmann, Max-Planck-Institute for Biophysical Chemistry, Göttingen), that ctPrp43 was able to functionally replace yPrp43 in the yeast purified B^{act} spliceosomes and *in vitro* reconstitution and disassembly assay (data not shown).

3.8.3 Crystallisation and structure determination of ctPrp43

Once the complex formation of ctPrp43 with the yNtr1 truncations was established, extensive crystallisation trials of full-length ctPrp43 with different nucleotides and RNAs as well as in complex with yNtr1(51-110) and yNtr1(1-120) were performed. Unfortunately, these attempts did not lead to any crystals. As the N-terminal extension of yPrp43 was shown to be not involved in the binding of the G-patch motif of yNtr1 (see section 3.4.3), an N-terminal truncated variant of ctPrp43 was purified. Furthermore, the NTE is the part with the lowest homology between *S. cerevisiae* and *C. thermophilum*.

A variant of ctPrp43 lacking the first 60 residues, referred to as ctPrp43(61-764), was soluble after overexpression in *E. coli* Rosetta 2(DE3) cells with a C-terminal Strep-tag. ctPrp43(61-764)-Strep was purified successfully following

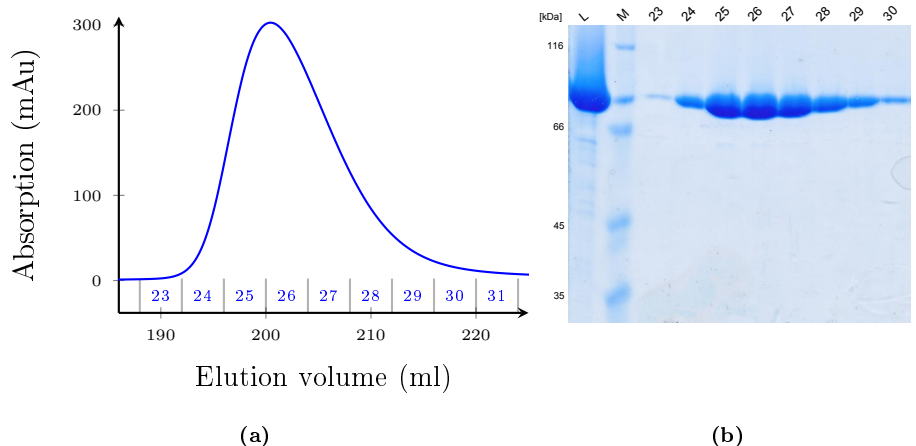


Figure 3.37: Size exclusion chromatography of ctPrp43(61-764), using a Superdex 200 26/60 column.

(a) In the chromatogram, the absorbance at 280 nm in a machine-specific unit (mAu, y-axis) is shown in blue referring to the elution volume (ml, x-axis). The collected elution fractions are indicated along the x-axis. The complete chromatogram is shown in the supplemental figure A.11.

(b) Coomassie-stained 12.5 % SDS-PA gel showing the loaded sample (L), the molecular weight marker (M) and elution fractions (numbers). The position of the bands according to ctPrp43(61-764) is indicated by an asterisk.

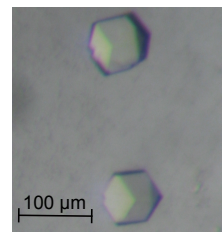
the same protocol that was used for full-length ctPrp43 (see section 3.8.2), consisting of a StrepTactin column run followed by a size exclusion chromatography (see figure 3.37). The yield after the final purification step was 14 mg of pure ctPrp43(61-764)-Strep per litre expression culture.

Size exclusion chromatography revealed that this truncated variant of ctPrp43 still was able to interact with the G-patch motif containing fragments of yNtr1 (data not shown). The maximal unwinding rate of the ctPrp43(61-764)/yNtr1(51-110) was determined to be $78.0 \pm 4.8 \text{ nM} \cdot \text{min}^{-1}$ and hence in the same range as for ctPrp43/yNtr1(51-110). Crystallisation attempts were performed for ctPrp43(61-764) in presence of different nucleotides and RNAs. Initial crystal screening for ctPrp43(61-764) at 5.0 mg/ml in presence of a ten-fold molar excess of ADP led to a crystallisation condition in the ProPlex screen which contained 0.15 M ammonium sulfate, 15 % (w/v) PEG4000 and 0.1 M MES pH 6.0. Upon fine screening based on this condition, the best diffracting crystals were obtained by increasing the concentrations of ammonium sulfate to 0.16 M and of PEG4000 to 18 % (w/v) (see figure 3.38).

ctPrp43(61-764) crystallised in the cubic space group I23 with cell axes of 118.9 Å. A complete dataset was collected at the beamline P14 at DESY, Ham-

3 Results

Figure 3.38: Crystals of ctPrp43(61-764). The crystals were obtained using the hanging drop technique and the precipitant 0.16 mM ammonium sulfate, 18 % (w/v) PEG4000 and 0.1 M MES pH 6.0.



burg (for further details of the data collection see section 2.4.4), the crystals diffracted to 3.0 Å. The crystallographic phase problem could be solved using the crystal structure of yeast Prp43 (pdb-ID 2xau) as a search model. Molecular replacement was conducted with PHASER, resulting in a final translation function Z score of 42 (McCoy *et al.*, 2007). Model building was conducted in Coot. The coordinates were refined to reasonable stereochemistry at a resolution of 2.99 Å using PHENIX. Alternating steps of refinement and structure adjustments were performed until the R-values converged. Simulated annealing omit maps were calculated in CNS (Brünger *et al.*, 1998).

The structure was refined to R/R_{free} -values of 20.5 % and 25.4 %, respectively (see data statistics in table 3.8). The whole amino acid sequence of ctPrp43(61-764) could be traced in the electron density, except the C-terminal residues from 758 to 764 and the C-terminal Strep-tag. Besides the ADP molecule, a magnesium ion was present in the nucleotide binding pocket.

Table 3.8: Data collection and structure refinement statistics for ctPrp43(61-764). The values in parentheses refer to the highest resolution shell. The Ramachandran plot was calculated with MolProbity.

Data collection		Refinement	
Space group	I23	Resolution range (Å)	47.23-3.00 (3.14-3.00)
<i>Unit cell parameters</i>		Number of reflections	22531
cell axes (Å)	a=b=c=188.9	Completeness (%)	100.0 (100.0)
cell angles (°)	$\alpha=\beta=\gamma=90$	Total atoms	5629
Resolution range (Å)	50.00-3.00 (3.10-3.00)	Protein residues	697
Completeness (%)	99.9 (100)	Water molecules	0
Multiplicity	7.9 (7.9)	R/R _{free} (%)	20.53/25.35 (29.57/32.39)
Unique reflections	22534 (2074)	Mean B-factor (Å ²)	38.26
I/ σ (I)	30.23 (3.62)	<i>RMSD from ideal</i>	
R _{merge}	5.3 (62.4)	bond lengths (Å)	0.003
		bond angles (°)	0.782
		<i>Ramachandran plot</i>	
		outliers (%)	0.72
		allowed (%)	5.90
		favored (%)	93.38

3.8.4 Structural comparison of ctPrp43 and yPrp43

The crystallised truncation of ctPrp43, lacking the N-terminal 60 residues, shows a domain organisation that is similar to its ortholog from yeast. The superposition of both structures revealed that the position of the secondary structure elements are nearly identical (shown in figure 3.39). The RMSD for the C $_{\alpha}$ -atoms of both structures is 1.14 Å, which indicates a very high degree of similarity within the organisation of the secondary structure elements. Key features as the nucleotide binding pocket, the 5'HP of the RecA-2 domain and the long ratchet helix exhibit no conformational differences in ctPrp43 compared to yPrp43.

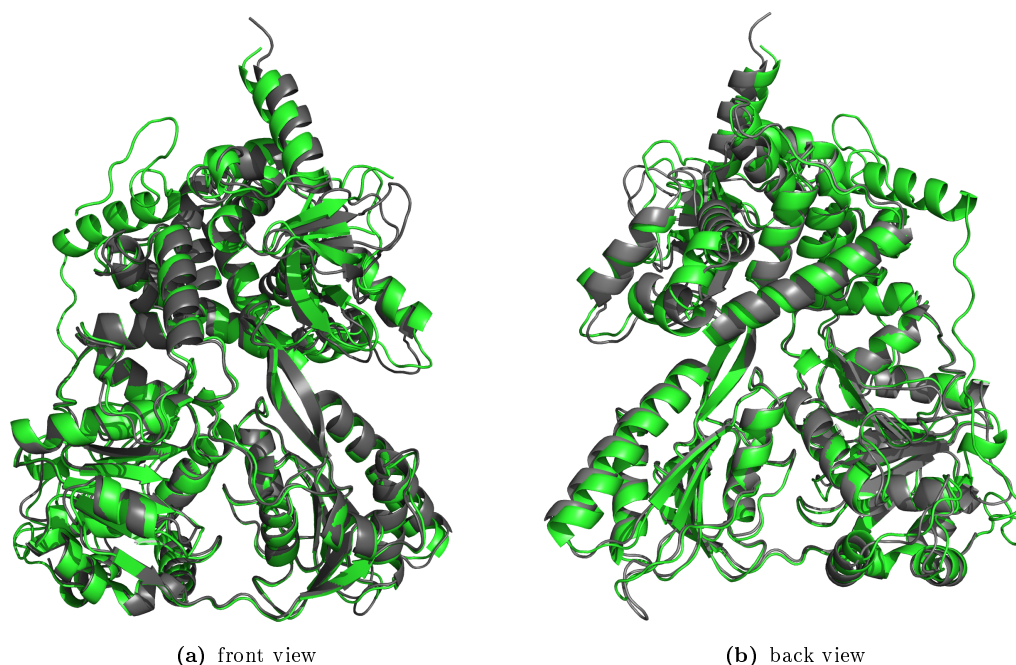


Figure 3.39: Comparison between the crystal structures of yPrp43 (pdb-ID 2xau) and ctPrp43(61-764). Both structures are shown in the cartoon representation in front view (a) and back view (b). yPrp43 is colored green and ctPrp43(61-764) is colored grey. The structures have been superposed using Coot and visualised with PyMOL.

Regarding the homology of the amino acid sequence of the domains of yPrp43 and ctPrp43, they reflect the high degree of similarity (see table 3.9 and alignment in the supplemental figure A.13). The two RecA-like and the winged helix domain show the highest degree of conservation with similarities for each domain about 90 %. The ratchet and the OB-fold domain which are thought to interact with the RNA and Ntr1, respectively, are more degenerated (similarities of 74 % and 68 %, respectively). The N-terminal extension is largely missing in the structure

3 Results

and is the least conserved part: It shows a similarity of only 49 %.

Table 3.9: Domain conservation between yPrp43 and ctPrp43. The calculations have been performed using the EMBOSS needle server with the matrix EBLOSUM62 (gap penalty=10)

	NTE	RecA-1	RecA-2	WH	Ratchet	OB-fold
Identity	29/109 (26.6 %)	142/176 (80.7 %)	149/190 (78.4 %)	55/66 (83.3 %)	64/116 (55.2 %)	71/136 (52.2 %)
Similarity	53/109 (48.6 %)	158/176 (89.8 %)	170/190 (89.5 %)	59/66 (89.4 %)	86/116 (74.1 %)	92/136 (67.6 %)
Gaps	29/109 (26.6 %)	1/176 (0.6 %)	5/190 (2.6 %)	4/66 (6.1 %)	4/116 (3.4 %)	12/136 (8.8 %)

In order to identify differences on the surface of both Prp43 orthologs, the ProSkin server was used for the visualisation of the conservation level of the surface residues (see supplemental figure A.17). No prominent region could be identified in which several non-conserved residues were present. Additionally, the surface charge distribution of both structures was calculated using pdb2pqr (see supplemental figure A.18). Regarding the surface charge, the different charged surface patches are quite similar. A closer look on the surface charge of the OB-fold domain, however, revealed differences for this region between yPrp43 and ctPrp43.

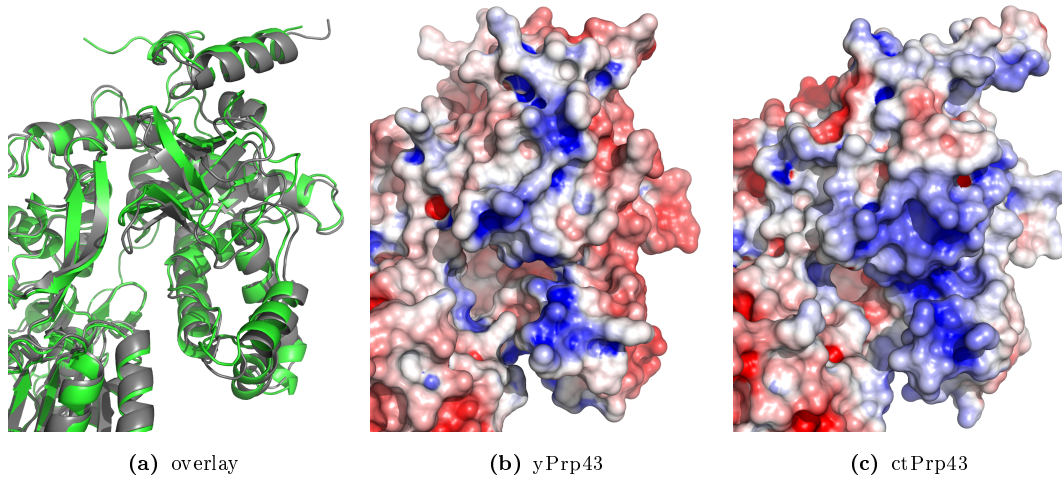


Figure 3.40: The OB-fold domains of yPrp43 (green) and ctPrp43 (grey) as an overlay (a) and the same domain in the surface charge representation for yPrp43 (b) and ctPrp43 (c). The contour levels were -6/+7 kBT for ctPrp43 and -8/+3 kBT for yPrp43, the different levels are due to the different overall charge of both orthologs. The surface charge has been calculated using pdb2pqr and has been displayed with the APBS plugin of PyMOL.

There are several different residues present in the OB-fold domain between yPrp43 and ctPrp43 (see alignment in the supplemental figure A.13). Although

the conformation of the main chain is quite similar, the surface charge distribution within this region reveals some differences (see figure 3.40). On the surface of the OB-fold in ctPrp43, a large positively charged patch is located which is not present in the same extent on the OB-fold domain of yPrp43. Therefore, this difference provides an explanation for the higher affinity for the G-patch motif of ctPrp43, as this region in yPrp43 (and hence presumably also for ctPrp43) is supposed to interact directly with the G-patch motif (see section 3.4.1).

Further structural and functional studies remain to be performed for ctPrp43 to reveal the reasons for the differences to its ortholog from yeast.

4 Discussion

The excision of non-coding sequences in pre-mRNAs is catalysed by the spliceosome, a single-turnover enzyme. The assembly, the catalysis and disassembly events during the splicing cycle are dependent of the action of at least eight DExD/H-box ATPases. After the release of the mature mRNA, the DEAH-box helicase Prp43 gets activated by the G-patch protein Ntr1 and catalyses the disassembly of the intron-lariat spliceosome in an ATP-dependent manner.

The goal of this thesis was the investigation of the molecular basis for the activation of Prp43 by Ntr1 by structural and biochemical characterisation of the interaction between both proteins.

4.1 Functional studies on the disassembly of the spliceosome

The involvement of Prp43 in the disassembly process of the intron-lariat spliceosome is known for more than a decade (Arenas and Abelson, 1997; Martin *et al.*, 2002). *In vitro* studies have proposed that the Ski2-like helicase Brr2 is also implicated in this process (Small *et al.*, 2006). In contrast to this suggestion, it turned out that the disassembly is complete when UTP is added instead of ATP in a purified splicing system (see section 3.1.2). Since Prp43 is able to hydrolyse UTP whereas Brr2 is not, these results indicate that the ATPase activity of Brr2 and hence probably the helicase activity is not needed for the disassembly (Fourmann *et al.*, 2013).

Furthermore, Prp43 was able in the purified splicing system to release the intron-lariat RNA and to separate the UsnRNPs into distinct components (see section 3.1.2). This finding raises the question if Prp43 has several target sequences in the intron-lariat spliceosome. These sequence could be identified in the future, e.g. by UV-induced RNA cross-linking.

Together, these results reflect the importance of the function of Prp43 itself

and of its regulation by Ntr1 during the splicing cycle.

4.2 Properties and regulation of helicases

This section discusses the functional and structural properties of the four spliceosomal DEAH-box ATPases and presents four well-studied examples for the activation of helicases.

4.2.1 The spliceosomal DEAH-box helicases

Whereas the four spliceosomal DEAH-box helicases yPrp2, yPrp16, yPrp22 and yPrp43 show an elevated sequence similarity in their helicase cores as well as in the three C-terminal domains, the N-terminal region preceding the first RecA-like domain is not conserved at all.

In yPrp43, the NTE contains 89 residues and is shorter than in yPrp2 (220 residues), yPrp16 (347 residues) and yPrp22 (480 residues). Martin *et al.* (2002) have shown that the deletion of the first 91 amino acids has no influence on the viability of yeast strains, indicating that it plays no essential functional role during the activation of yPrp43. Furthermore, the deletion of the α -helix which connects the N-terminal extension with the winged helix domain has no effect on the helicase activity of a yPrp43/yNtr1 complex (see section 3.4.3), confirming the results from Martin *et al.* (2002) on a functional level.

In vivo studies of yPrp2 and yPrp16 have indicated that the NTEs are functionally dispensable for both helicases (Edwalds-Gilbert *et al.*, 2004; Hotz and Schwer, 1998). The role of the N-terminal extension in yPrp22 is more ambiguous: the first 260 residues of the NTE are not required, whereas the residues 262 to 350 enhance the function of yPrp22 *in vivo* (Schneider and Schwer, 2001). Moreover, complete deletions of the NTE in yPrp22 result in higher ATPase and helicase activities *in vitro*, but also in a defect of its spliceosomal function (Schneider and Schwer, 2001). Obviously, a part of the N-terminal extension of yPrp22 is required for the recruitment to the spliceosome, but not for the enzymatic activities.

Despite the high degree of sequence conservation, there is a major difference between yPrp22 and yPrp43: Whereas yPrp22 shows intrinsic ATPase and helicase activities *in vitro* (Tanaka and Schwer, 2005) and presumably needs its interaction partners exclusively for the binding to the spliceosome, yPrp43 is dependent on the presence of yNtr1 for both, recruitment and activation (Tsai *et al.*, 2007). The

comparison of the C-terminal domains of the crystal structures of hPrp22 and yPrp43 to models of hPrp2, hPrp16 and hPrp43 indicates that their overall folding is similar (see section 3.7). Regarding the surface charge distribution of these C-terminal domains, both conserved and divergent patches can be found, indicating that this part of the proteins might be responsible for the different activities and for specific interactions with other spliceosomal components (Kudlinzki *et al.*, 2012).

Although no crystal structure of the C-terminal domains of yPrp22 is available, it can be speculated that differences in these domains might be a reason for the intrinsic activities of yPrp22. Further analysis of the functional differences between yPrp43 and yPrp22 requires deeper insights into the unwinding mechanism of both proteins, which would be provided e.g. by high resolution structures of either full-length yPrp22 or an activated yPrp43 in complex with yNtr1, in order to evaluate the structural changes that occur upon activation.

4.2.2 Mechanisms for the activation of helicases

In contrast to the limited structural data that is available to date for the spliceosomal DEAH-box helicases, several structural and functional studies on helicases in complex with interaction partners have been published during the last years. These studies provide insights into the different modes of regulation that have evolved.

The SF1 helicase Upf1 is involved in nonsense-mediated mRNA decay (NMD) and is stimulated concerning its ATPase and helicase activities by the interaction partner Upf2 (Chamieh *et al.*, 2008). Based on different crystal structures, it was demonstrated that the N-terminal cysteine-histidine-rich CH domain of human Upf1 (hUpf1) is displaced upon binding of hUpf2, leading to a decrease in the RNA affinity and thereby to an increased unwinding activity (Clerici *et al.*, 2009; Chakrabarti *et al.*, 2011). Interestingly, hUpf2 is intrinsically unstructured, but upon binding to hUpf1, it forms a β -hairpin and an α -helix which interact with hUpf1 (see figure 4.1a). Furthermore, it has been reported that the activity of hUpf1 is also regulated by phosphorylation (Okada-Katsuhata *et al.*, 2012) and by autoinhibition by its N- and C-terminal parts (Fiorini *et al.*, 2013), indicating that the control of the helicase activity of hUpf1 is assured by multiple mechanisms.

4 Discussion

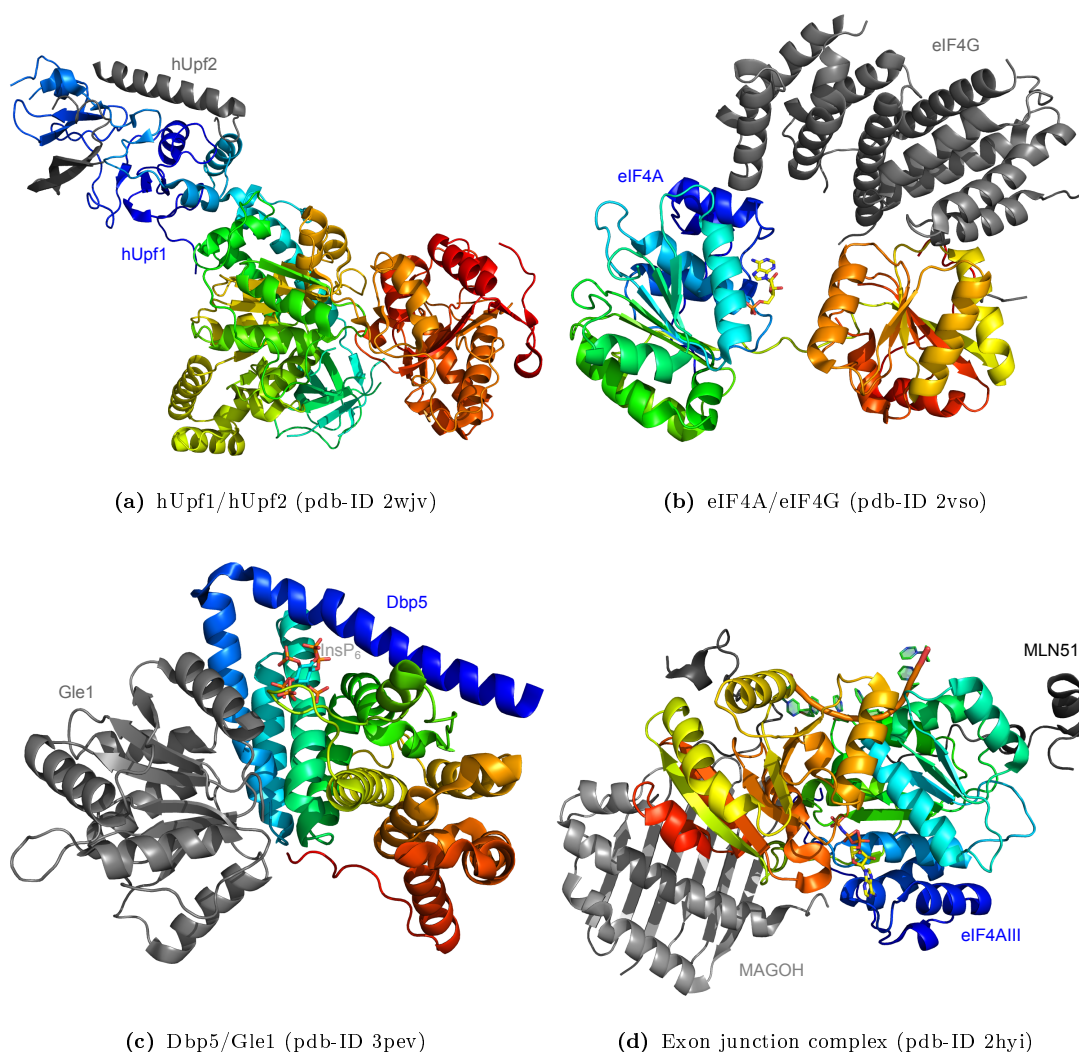


Figure 4.1: Crystal structures of helicases in complex with interacting proteins. The helicases molecules are shown in the rainbow representation (from the N-terminal region in blue to the C-terminal region in red), the interaction partners are colored in grey. The structures have been visualised using PyMOL.

(a) The SF1-helicase hUpf1 has an N-terminal CH-domain which is docked on the helicase core and interacts with hUpf2 (grey). The intrinsically unstructured hUpf2 forms an α -helix and a β -hairpin upon interaction with hUpf1.

(b) The conformation of the DEAD-box helicase eIF4A is modified upon binding of eIF4G (grey) in a way that both RecA-like domains get closer and align the different residues responsible for the nucleotide interaction.

(c) The helicase core domains of the DEAD-box protein Dbp5 are stabilised upon binding of Gle1 (grey). This interaction is dependent of the presence of inositol hexakisphosphate (InsP₆), which binds at the interaction site of both molecules.

(d) The DEAD-box protein eIF4AIII is the the central component of the exon junction complex (EJC). Within the EJC, eIF4AIII interacts with MLN51 (dark grey) and MAGOH (light grey) which stimulate the RNA binding and ATP hydrolysis of eIF4AIII.

4.2 Properties and regulation of helicases

The DEAD-box protein eIF4A is involved in translation initiation and represents a minimal helicase as it solely consists of two RecA-like domains. The ATPase and helicase activities of eIF4A are stimulated upon binding of eIF4G (Oberer *et al.*, 2005). Structural studies have shown that eIF4G interacts with both RecA-like domains of eIF4A and induces a proper alignment of the respective sequence motifs responsible for RNA and ATP binding (see figure 4.1b). This complex formation hence increases the affinity of eIF4A for RNA and ATP. It has been suggested that upon binding of ATP, only the C-terminal RecA-like domain remains in contact to eIF4G and therefore an enhanced flexibility of the N-terminal RecA-like domain contributes to increased ATPase and unwinding activities (Schütz *et al.*, 2008).

The DEAD-box protein Dbp5 plays a role during mRNA export. Dbp5 interacts with Gle1 which decreases its RNA affinity after strand separation and has been suggested to enhance the release of the RNA after strand separation (Weirich *et al.*, 2006). The structure analysis of Dbp5 in complex with Gle1 reveals that Gle1 is in contact with both RecA-like domains (see figure 4.1c). A special feature of this complex is the presence of the small molecule inositol hexakisphosphate (InsP₆), which is located at the interaction interface of both proteins and is required for stable complex formation (Montpetit *et al.*, 2011).

A fourth well-studied example for the activation of a helicase by interaction partners is provided by studies on the exon-junction complex (EJC). The EJC consists of four proteins, including the central DEAD-box protein eIF4AIII, which is formed solely by the two RecA-like domains. The heterodimer MAGOH-Y14 binds to eIF4AIII exclusively in the closed conformation and remodels motif I as well as the Q motif, leading to a stimulation of the ATPase activity (Andersen *et al.*, 2006; Bono *et al.*, 2006) (see figure 4.1d). The fourth component of the EJC, MLN51, has been shown to contribute to RNA binding and is suggested to stimulate eIF4AIII by supporting the formation of the closed conformation (Le Hir and Andersen, 2008).

In conclusion, these examples show many different modes of interaction that have been identified so far for helicases, e.g. folding upon interaction with a helicase, stabilisation of a productive conformation, implication of small molecules in the binding and the involvement of several interaction partners.

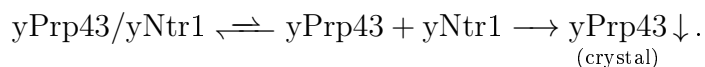
4.3 Crystallographic studies on Prp43

As the published structure of yPrp43 reveals only insights into its ADP-bound and hence into its post-catalytic state (He *et al.*, 2010; Walbott *et al.*, 2010), a central goal of this thesis was the crystallisation of yPrp43 in complex with yNtr1, representing yPrp43 in its activated state.

4.3.1 Crystallisation trials of the yPrp43/yNtr1 complex

All crystallisation trials for yPrp43 in complex with truncated yNtr1 fragments failed, mainly due to the fact that yPrp43 crystallised its apo-form (see section 3.2.3).

Obviously, yPrp43 crystallises that easily, that in the equilibrium state of a complex of yPrp43 and yNtr1 even a small amount of free yPrp43 was sufficient for crystallisation in a suitable buffer condition. This crystallisation event probably led to a shift in the equilibrium towards the dissociated complex as the crystalline yPrp43 was no longer part of the equilibrium reaction, as represented in the following reaction equation:



In order to shift the equilibrium towards the complex formation, crystallisation trials were performed with yPrp43 and a ten-fold molar excess of the respective yNtr1 fragment. However, no crystals of the complex were obtained using this approach (see section 3.2.3).

A possibility to prevent the dissociation of complexes during crystallisation is the fusion of the components. This strategy has already led to successful crystallisation, e.g. for the Arf/ArfGAP complex (Ismail *et al.*, 2010). Adopting this approach for yPrp43 and yNtr1, the fusion protein yPrp43-yNtr1(51-110) was found to be functional and to have higher ATPase and helicase activities than the complex of both proteins (see section 3.3.3). Crystallisation trials for this fusion protein were not successful, this might be due to structural changes that are induced on yPrp43 upon binding of yNtr1. These changes could lead to a conformation which is unfavorable for crystallisation, e.g. because of changes in the surface charge or the exposure of structurally flexible parts.

Further crystallisation trials should be performed in the future, e.g. by truncating the extended C-terminal helix of the OB-fold domain and/or the N-terminal extension of yPrp43 in fusion proteins as well as in yPrp43/yNtr1 complexes.

4.3.2 Functional and structural properties of ctPrp43

As proteins from thermophilic organisms might be more suited for crystallisation, intensive crystallisation attempts were performed with Prp43 from *Chaetomium thermophilum*. Whereas the crystallisation attempts for the ctPrp43/yNtr1(51-110) complex were not successful, the crystal structure of ctPrp43(61-764) in its ADP-bound and thus post-catalytic state could be determined (see section 3.8.3). The structure was very similar to the structure of yPrp43 in its ADP-bound state (see figure 3.39).

Functional studies have revealed that ctPrp43 has a seven-fold higher affinity for yNtr1(51-110) and a 1.7-fold higher ATPase activity than yPrp43 (see section 3.8.2), most probably as an adaptation for living at elevated temperatures. Upon comparison of the surface charge distribution of both orthologs, differences within the OB-fold domains could be identified (see section 3.8.3). As this region is supposed to mediate the interaction of yPrp43 with the G-patch motif of Ntr1 (see section 3.4.1), this difference in surface charge might enhance the direct interaction with the G-patch motif. However, the elevated ATPase activity of ctPrp43 could not be explained during the structural comparison, hence high resolution structures of ctPrp43 in complex with RNA and ATP are required to answer this question.

ctPrp43 and ctPrp43(61-764) were found to be able to replace yPrp43 in purified yeast B^{act} spliceosomes and *in vitro* reconstitution and disassembly assays (see section 3.8.2). This finding is in agreement with the high sequence conservation between both orthologs: solely the dispensable N-terminal domain is degenerated, whereas the helicase core and the C-terminal domains show a high level of conservation (see section 3.8.1). Furthermore, both Prp43 and Ntr1 seem to be functionally conserved among several organisms as the human ortholog of yNtr1, TFIP11, was reported to be able to complement the deletion of yNtr1 in a yeast strain (Herrmann *et al.*, 2007).

Several mutations have been described recently for thermophilic proteins in comparison to mesophilic proteins, that are supposed to enhance the thermostability, e.g. lysine to arginine and glycine to alanine mutations as well as an extended proline content in thermophilic organisms (van Noort *et al.*, 2013). Comparing the sequences of yPrp43 and ctPrp43 as well as the total amino acid content of both orthologs, no significant differences could be identified that would confirm these adaptations for ctPrp43 (see supplemental table A.7 and supplemental fig-

4 Discussion

ure A.13). This finding indicates that both Prp43 orthologs have a conformation which is stable also in case of elevated temperatures.

In contrast to Prp43, the potential Ntr1 and Ntr2 orthologs of *C. thermophilum* are more degenerated and hence provide promising targets for functional and structural studies of the activation of Prp43 (see section 3.8.1). As the preparation of a functional yNtr1/yNtr2 complex was successful, the complete NTR complex could not be prepared due to a high level of aggregation in the yNtr1/yNtr2 complex (see section 3.1.3). ctNtr1 and ctNtr2 might be more suited for purification and crystallisation experiments due to their potential adaptation to higher temperatures. Extensive crystallisation attempts should be performed in the future with the components of the NTR complex from *C. thermophilum*.

Although no high resolution structure of a Prp43/Ntr1 complex was obtained, several results from other methods revealed insights into the activation of Prp43. These results mainly provide information about the interaction of yPrp43 with RNA as well as with the G-patch motif of yNtr1 and will be discussed in the following sections.

4.4 The interaction of yPrp43 with RNA

In this section, the influence of RNA on the ATPase activity of yPrp43 will be discussed as well as the information obtained about the RNA binding by yPrp43.

4.4.1 The effect of RNA on the ATPase activity

The ATPase hydrolysis rate of yPrp43 is stimulated by RNA, both in presence and absence of yNtr1 *in vitro* (Schneider and Schwer, 2001; Tanaka and Schwer, 2006; Tanaka *et al.*, 2007). Comparable results were obtained for yPrp43 alone upon determination of the ATPase activity in absence and presence of RNA (see section 3.3.1). Furthermore it could be revealed that the extent of stimulation is in the same range in absence and in presence of the truncated yNtr1 variants. Hence, RNA seems to influence the ATP affinity and/or the mechanism of ATP hydrolysis of yPrp43.

The published crystal structure of yPrp43 represents the post-catalytic state of the helicase (He *et al.*, 2010; Walbott *et al.*, 2010). A structural comparison of yPrp43 and the HCV NS3 helicase provides information about the changes that might occur in yPrp43 upon RNA binding. Whereas the helicase core of both

4.4 The interaction of yPrp43 with RNA

proteins is similar in the ADP-bound state (see figure 4.2a), the superposition of the helicases cores of yPrp43/ADP and an NS3/ATP/DNA complex indicates that major structural changes occur in NS3 to allow efficient ATP binding (see figure 4.2b). The nucleotide binding pocket is remodelled especially within the RecA-2 domain, providing space for the additional phosphate group which cannot be accommodated in the ADP-bound conformation. Furthermore, the 5'-HP of the NS3 helicase seems to move significantly away from the nucleic binding cavity when the nucleic acid is bound (Appleby *et al.*, 2011; Gu and Rice, 2010).

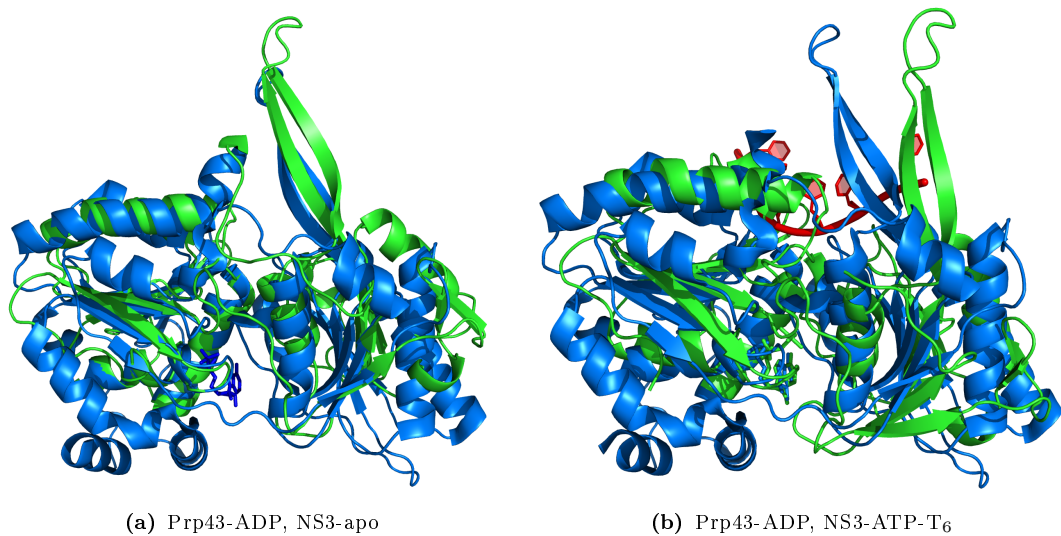


Figure 4.2: Superposition of the helicase core domains of yPrp43 (marine blue) and the HCV NS3 helicase (green). For yPrp43 (pdb-ID 2xau), the ADP-bound state is shown in both subfigures, whereas for the NS3 helicase, (a) provides the ADP bound state (pdb-ID 3o8b) and (b) the state with a bound ATP analog and poly(T)₆-DNA, colored in red (pdb-ID 3kql). The structures were superposed and visualised using PyMOL.

This comparison indicates that the RNA is able to induce structural changes in the RecA-2 domain, leading to a higher ATP affinity on the one hand and to a major movement of the 5'-HP on the other hand. For the spliceosomal DEAH-box protein yPrp22 it has been shown that RNA increases the affinity for ATP *in vitro* (Tanaka and Schwer, 2005), hence this could be a common feature of the DEAH-box ATPases. On the other hand, in the same study the authors have declared that the presence of ATP decreases the affinity for RNA, a property that would allow a movement of the RecA-like domains as proposed by the backbone stepping motor model (see section 1.4.2.2).

Nucleotide binding studies suggest that yPrp43 can bind ADP and not bind ATP, both in presence or absence of yNtr1 (see section 3.5.2). This indicates that

4 Discussion

solely the RNA might be responsible for the stimulation of the ATP affinity of yPrp43. Moreover, also a concerted binding of ATP and RNA is possible. Further studies need to be performed to investigate the influence of ssRNA and dsRNA binding on the nucleotide affinities of yPrp43 and yPrp43/yNtr1 complexes.

4.4.2 Structural insights into RNA binding

UV-induced protein-RNA cross-linking experiments have led to the identification of eight residues in yPrp43 which are supposed to interact directly with RNA (see section 3.5.1). A comparison to the other three spliceosomal DEAH-box helicases reveals that four of these residues are conserved within yPrp2, yPrp16, yPrp22 and yPrp43: the residues Y174 of the RecA-1 domain, K418 of the RecA-2 domain and two residues within the Ratchet domain (see figure 4.3). The other four cross-linked residues (M154, Y440, S614 and Y630) do not have conserved equivalents in the other spliceosomal DEAH-box helicases in yeast.

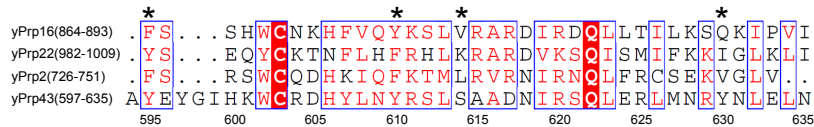


Figure 4.3: Alignment of the ratchet domains of the four spliceosomal DEAH-box ATPases yPrp16, yPrp22, yPrp2 and yPrp43. All residues labelled with an asterisk have been identified in yPrp43 to cross-link to RNA. The residue numbers refer to the sequence of yPrp43. White letters on red background indicate identical residues, whereas red letters in a blue box indicate similar residues. This ClustalW alignment has been visualised using ESPript.

Walbott *et al.* (2010) have demonstrated that the ATPase activity of a truncated yPrp43 lacking the OB-fold domain could not be stimulated by RNA, indicating that the OB-fold domain is interacting with RNA. Surprisingly, no RNA cross-link was detected for a residue within the OB-fold domain (see section 3.5.1). This could be due to the used oligonucleotide, a ssRNA comprising 21 nucleotides, including five uridines (see table 2.3 on page 40). In general, most cross-links are found with a uridine and, rarely, a guanine or a cytidin base (Romina Hofele, personal communication). This indicates that the nucleotide bases that interact with the OB-fold might have not been favorable to cross-linking. Alternatively, the OB-fold domain could be responsible for binding dsRNA and unable to interact with the ssRNA used in this experiment.

Besides the RNA-cross-links found for yPrp43, also a cross-link between the RNA and K67 of yNtr1 was identified, providing the first experimental evidence

for a residue within a G-patch motif that is directly interacting with a nucleic acid (see section 3.5.1). Interestingly, this residue is conserved within the G-patch motifs of yPfa1 and the human Ntr1 ortholog, TFIP11, supporting a potential functional role. Interactions of G-patch motifs with RNA have been reported for the protein TgDRE from the parasite *Toxoplasma gondii* by fluorescence anisotropy measurements with different short RNA oligonucleotides (Frénal *et al.*, 2006) as well as for the proteinase from the Mason-Pfizer monkey virus in electrophoretic mobility shift assays (Svec *et al.*, 2004). These findings propose a direct contribution of the G-patch motif to the RNA interaction.

4.5 The interaction of yPrp43 with yNtr1

In this section, the functional consequences of yNtr1 binding to yPrp43, the minimal yNtr1 binding domain and the structure of the G-patch motif will be discussed.

4.5.1 Functional effects of yNtr1-binding to yPrp43

Tanaka *et al.* (2007) have revealed by *in vitro* studies with the truncation yNtr1(1-120), that yNtr1 activates yPrp43 by the stimulation of its helicase activity. These findings were confirmed (see section 3.3.1), and it was demonstrated that an yNtr1-fragment comprising the residues 51-110 was required and sufficient to activate yPrp43 (see section 3.3.4). No helicase activity of unliganded yPrp43 was detected, probably due to the low sensitivity of the used assay together with the very low dsRNA unwinding rate of yPrp43.

The dsRNA substrate used contained 28 base pairs, due to the fact that the signal was detected upon formation of a hairpin within the labelled strand (see section 2.3.2). This hairpin was formed by 10 intramolecular base pairs in order to reach a melting temperature significantly higher than the temperature at which the assay was performed. Therefore for each substrate molecule, 28 base pairs had to be separated before the hairpin structure could form, which could lead to the consequence that the activity of weak helicases cannot be detected using this assay (see figure 2.2 on page 53).

The characterisation of the ATPase activity of yPrp43/yNtr1 complexes has shown clearly that the purified yNtr1 truncations, namely yNtr1(1-120) and yNtr1(51-110), were able to stimulate the ATPase activity of yPrp43 about four-

fold (see section 3.3.1). This result is in contrast to the findings of Tanaka *et al.* (2007), who purified an yPrp43/yNtr1 complex via glycerol gradient centrifugation and could not detect a difference of the ATPase activities compared to unliganded yPrp43, neither in absence nor in presence of RNA. Although Tanaka and colleagues used full-length yNtr1, it seems unlikely that additional domains of yNtr1 are able to inhibit the stimulatory effect of the G-patch motif on the ATPase activity completely.

Moreover, the G-patch protein yPfa1, the activator of yPrp43 in the process of ribosome biogenesis, was reported to stimulate besides the helicase also the ATPase activity of yPrp43 (Lebaron *et al.*, 2009). Therefore the stimulation of the helicase activity could be directly coupled to the stimulation of the ATPase activity and vice versa (Lebaron *et al.*, 2009). The mechanism for the activation of yPrp43 by yNtr1 is most probably similar and could hence be based on a parallel stimulation of both activities.

Using the stimulation effect of yNtr1 on the ATPase activity of yPrp43, for the first time binding constants for the interaction of a G-patch protein with an interaction partner were determined (see section 3.3.2). This method provides a simple way to determine the influence of mutations or deletions on the interaction between yPrp43 and yNtr1.

4.5.2 The minimally required yNtr1-fragment

Within the NTR complex, the interaction of yPrp43 and yNtr1 is mediated by the N-terminal part of yNtr1, as the G-patch containing yNtr1(1-122) was identified in yeast two-hybrid studies to be able to with yPrp43 (Tsai *et al.*, 2005). Additionally, yNtr1(1-120) was suggested to be sufficient for the stimulation of the helicase activity of yPrp43 *in vitro* (Tanaka *et al.*, 2007). Pandit *et al.* (2006) have reported furthermore that the deletion of the N-terminal 33 residues of yNtr1 seems to be dispensable for the function of yNtr1 *in vivo*.

By the characterisation of the ATPase and the helicase activity of yPrp43/yNtr1 complexes, the truncation yNtr1(51-110) was shown to stimulate the ATPase and the helicase activities of yPrp43 to the same extent as yNtr1(1-120) (see section 3.3.1). Furthermore, the ATPase activity of fusion proteins of yPrp43 and even smaller yNtr1-fragments was significantly lower than the ATPase activity of the fusion protein yPrp43-yNtr1(51-110), indicating that yNtr1(51-110) is both necessary and sufficient for the activation of yPrp43 (see section 3.3.4). This

fragment is half in size compared to the yNtr1 fragments that were known to be sufficient so far.

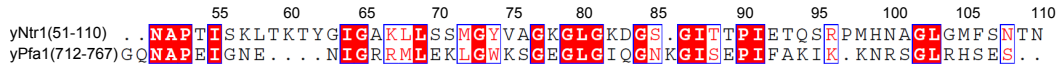


Figure 4.4: Alignment of the G-patch motifs in the N-terminal part of yNtr1 and in the C-terminal part of yPfa1. Besides the G-patch consensus sequence (Aravind and Koonin, 1999), which spans the yNtr1 residues 68 to 86, there are additional conservations in this region of yNtr1 and yPfa1. White letters on red background indicate identical residues, whereas red letters in a blue box indicate similar residues. The numbers refer to the residues of the G-patch motif of yNtr1. This ClustalW alignment has been visualised using ESPript.

This finding is supported by the sequence conservation within the G-patch motif of the two different activators of yPrp43, yNtr1 and yPfa1: Strikingly, conserved residues between the two proteins are present in the range of residue N51 to L103 of yNtr1, supporting that this residue range is needed for the activation of yPrp43 (see figure 4.4). As both proteins have the same functional effects on yPrp43, the residues which are responsible for the interaction are supposed to be highly conserved. This degree of sequence conservation confirms the experimental data and provides an explanation why the yNtr1 truncations 51-102 and 59-110 were not sufficient for the complete activation of yPrp43 (see section 3.3.4).

Admittedly, it cannot be excluded that the C-terminal residues 104 to 110 of yNtr1(51-110) might be dispensable for the activation as well, as there is no sequence conservation in this range between both G-patch proteins. Further activity studies should be performed to address this question.

4.5.3 Insights into the binding site of yNtr1 on yPrp43

In vivo studies in yeast strains have revealed that yPrp43(1-732) is still functional, whereas yPrp43(1-722) provokes a temperature sensitive and yPrp43(1-712) a lethal phenotype, indicating that the last 36 residues (732-767) of yPrp43 are most probably not involved in binding to yNtr1 (Martin *et al.*, 2002). Furthermore, a yPrp43-variant that lacked the complete OB-fold domain, yPrp43(1-657), was not able to interact with the G-patch containing C-terminal domain of yPfa1 *in vitro* (Walbott *et al.*, 2010). These findings are in agreement with the results obtained during the protein-protein cross-linking experiments of the yPrp43/yNtr1 complexes, as only residues within the OB-fold domain of yPrp43 have been found to cross-link with yNtr1(1-120) (see section 3.4.1).

Taken together, there is strong evidence that the C-terminal OB-fold domain

of yPrp43 mediates the interaction with the G-patch motif of yNtr1. Strikingly, Silverman *et al.* (2004) have reported that the interaction between the spliceosomal DEAH-box ATPase yPrp2 and the G-patch protein ySpp2 can be disrupted by point mutations in the predicted OB-fold domain of yPrp2. Obviously, the interaction between the OB-fold domain with the G-patch motifs is a conserved feature of both helicases, yPrp2 and yPrp43.

Although the cross-linkers were lysine-specific and relatively large (7.7 Å and 11.4 Å) which corresponds to distances between the C_α-atoms of cross-linked lysines of about 20 Å and 24 Å, respectively¹, only cross-links in a small surface area of yPrp43 were identified (see section 3.4.1). Most probably, no residues within the direct interaction surface could be cross-linked due to steric reasons. This could explain the absence of cross-links between lysines of the yNtr1 truncations and several lysine residues within the OB-fold domain of yPrp43, namely K650, K662, K663, K674 and K704. The identification of a cross-link between yPrp43-K4 and yNtr1(51-110) has most probably no functional relevance but was due to the flexibility of the N-terminal tail of yPrp43.

OB-fold domains are not only involved in nucleic acid binding (Murzin, 1993), but they can also serve as platforms for protein-protein interaction (Theobald *et al.*, 2003). Several studies have been performed to characterise these interactions, either by determination of crystal structures, like for the interaction between the OB-folds of the topoisomerase RMI1 and of its interaction partner RMI2 (Wang *et al.*, 2010), or by *in vivo* and *in vitro* studies like for the interaction between the SF1 helicase Pif1 and its activator Rim1 (Ramanagoudr-Bhojappa *et al.*, 2013). Therefore, it can be suggested that the OB-fold of yPrp43 is involved in both, interaction with the RNA substrate and with its activator, yNtr1.

4.5.4 Structural analysis of the G-patch motif

To date, no crystal or NMR structure of a G-patch motif is available. The determination of a high resolution structure of the G-patch motif failed, but information about its folding could be deduced (see section 3.6.1).

The absence of secondary structure elements has been suggested for the G-patch motif in the DNA repair protein TgDRE from the parasite *Toxoplasma gondii* using CD-spectroscopy (Frénal *et al.*, 2006). The CD-spectrum of yNtr1(51-110)

¹The length of a lysine side chain is estimated to be about 6 Å.

also indicates that there are no secondary structure elements in the G-patch motif of yNtr1, at least in the absence of binding partners (see section 3.6.1).

The presence of unstructured regions in proteins has been proposed to be a common feature of large complexes by facilitating their assembly (Hegyi *et al.*, 2007). Based on bioinformatical studies, Korneta and Bujnicki (2012) have predicted that nearly half of the residues of spliceosomal proteins are unstructured, supporting the suggestion of Hegyi and colleagues at least for the spliceosome.

Two major types of unstructured proteins have been described: On the one hand those that contain long and permanently disordered regions and that have been proposed to be a new structural type (Tompa *et al.*, 2009) – on the other hand those which have a "dual personality" and may perform transitions between an unstructured and a structured state (Zhang *et al.*, 2007). The results of the CD spectroscopy of yNtr1(51-110) in presence of TFE (see section 3.6.1) as well as the folding analysis from the IUPRED server (see section 3.6.2) indicate that the G-patch motif of Ntr1 could feature a dual personality by switching from an unstructured state in isolation to a structured state when bound to yPrp43. The folding of an unstructured protein upon binding to a helicase was already described for the hUpf1/hUpf2 complex (see section 4.2.2).

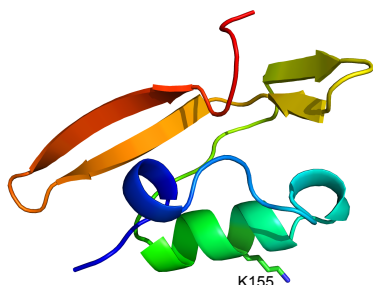


Figure 4.5: *De novo* model of the G-patch motif (residues 142 to 193) from the human ortholog of yNtr1, TFIP11 (Korneta *et al.*, 2012). In this model, the G-patch motif adopts a fold comparable to a dsRNA-binding domain. As the lysine K67 of yNtr1 is conserved in the G-patch motif of TFIP11 as K155, it is visualised as a stick here (see alignment in the supplemental figure A.16). The model has been visualised in the rainbow representation using PyMOL.

A *de novo* modeling of the G-patch motif from the human ortholog of yNtr1, TFIP11, indicates a folding comparable to a dsRNA-binding domain (Korneta *et al.*, 2012), which is formed by two α -helices and two pairs of antiparallel β -strands (see figure 4.5). As a cross-link between yNtr1 and an RNA was identified (see section 3.5.1), it is possible that at least a part of the G-patch motif is involved in RNA-interaction and hence adopts a fold suitable for the binding of nucleic acids.

A high resolution structure of a G-patch motif will allow further evaluation of this model and a more detailed characterisation of the folding state of the G-patch motif.

4.6 A model for the activation of yPrp43

The identification of four residues within the ratchet domain that directly interact with RNA supports the potential RNA-binding function of this domain (see section 3.5.1). Like for the HCV NS3 helicase and the Ski2-like protein Hel308 (Appleby *et al.*, 2011; Büttner *et al.*, 2007), it has been proposed for yPrp43 that the long ratchet helix is responsible for pulling one RNA strand into the nucleic acid binding cavity, whereas the 5'-HP is supposed to separate the two strands through a directed movement together with the OB-fold domain (Walbott *et al.*, 2010).

In order to transform yPrp43 from an inactive to an active state, structural changes are induced in yPrp43 upon binding of RNA and yNtr1. Taken together the indications for structural and functional influences of RNA and yNtr1 on yPrp43 which have been discussed above, the following speculative model is proposed here for the activation of yPrp43 (see figure 4.6).

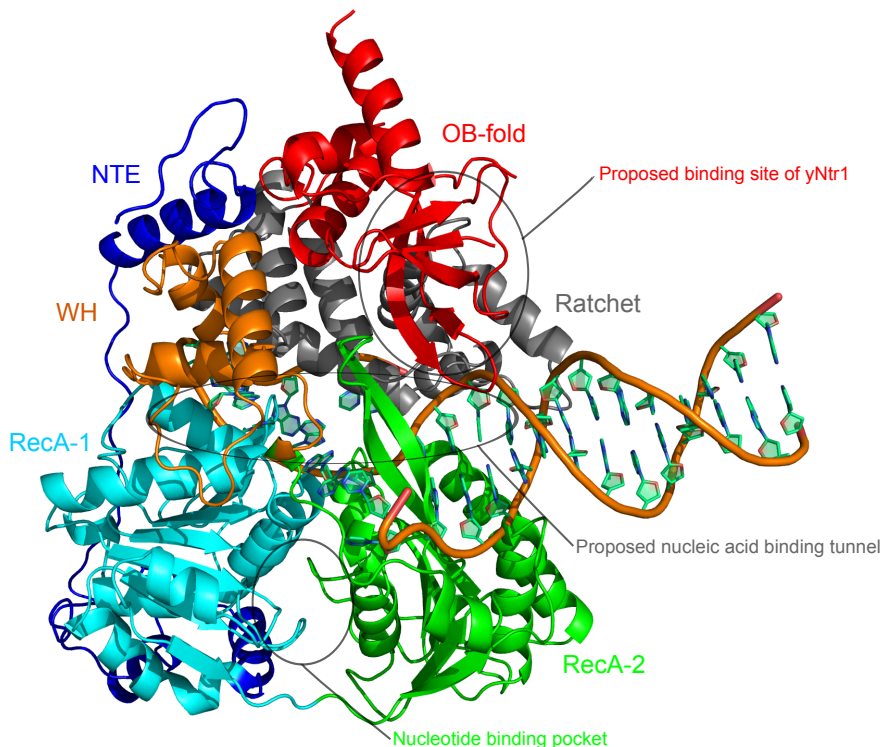


Figure 4.6: The crystal structure of yPrp43 (pdb-ID 2xau) with the modelled dsDNA substrate from the structure of the helicase Hel308 (pdb-ID 2p6r). Besides the domains, also the nucleotide binding pocket, the putative RNA binding tunnel and the potential interaction site of yNtr1 within the OB-fold domain is labelled.

4.6 A model for the activation of yPrp43

As free yPrp43 could be cross-linked to ssRNA, the first event could be the interaction of yPrp43 with an overhang of a dsRNA. The overhang might be bound in a similar manner like it was found for the Ski2-like helicase Hel308, in a tunnel between the helicase core and the C-terminal domains (Büttner *et al.*, 2007). This RNA binding could induce structural changes in the RecA-2 domain, which would on the one hand remodel the nucleotide binding pocket, thereby leading to a higher affinity for ATP, and on the other hand could induce a movement of the 5'HP towards the dsRNA. Subsequently, the G-patch motif of yNtr1 could bind to the OB-fold domain and thereby adjust this domain in that way that it directly interacts with the dsRNA and eventually with the 5'HP. Moreover, the binding of yNtr1 could induce conformational changes in several domains of yPrp43, thereby provoking a more productive conformation which, even in the absence of RNA, leads to a higher ATPase activity of yPrp43.

The overhang of the dsRNA could be stringently bound by the long helix of the ratchet domain. Upon hydrolysis of ATP, major structural changes could occur within the RecA-2 domain, which could also affect the position of the 5'HP and the affinity for the RNA in the nucleic acid binding tunnel. This could lead to a movement provoking the pulling of the dsRNA-overhang into the tunnel. Moreover, the OB-fold domain and the 5'HP could be also affected from domain movements upon ATP hydrolysis and conformational changes of these two structural features might induce a strand separation within the dsRNA and hence lead to the unwinding of the dsRNA substrate.

Further crystallographic and functional studies have to be performed to refine and modify this speculative model.

Appendix

A.1 Protocols of expression and purification

Large scale expressions were performed in ten 2l-flasks that each contained 500 ml of 2YT medium. All purifications procedures were performed at 20 °C. The purified protein samples were concentrated (see section 2.2.4) and subsequently stored in aliquots of 250 µl at -80 °C.

A.1.1 Expression and purification of yPrp43

The plasmid pASG-IBA3_yPrp43 was transformed in *E. coli* Rosetta 2(DE3) cells. The cells were grown at 37 °C in presence of ampicillin and chloramphenicol until an OD₆₀₀ of 0.9 was reached. The expression of yPrp43-Strep was induced by addition of 200 µg anhydrotetracycline per liter cell culture. The cells were incubated at 16 °C for 16 h.

Cells were harvested by centrifugation and resuspended in buffer S (400 mM NaCl, 50 mM Tris pH 7.5, 2 mM MgCl₂). Cell lysis was performed using a Microfluidizer, followed by a centrifugation step in order to separate the soluble from the insoluble fraction (see section 2.2.2).

yPrp43-Strep protein was isolated from the soluble fraction using a StrepTactin HP sepharose column and eluted with buffer S supplemented with 2 mM D-desthiobiotin (see section 2.2.3.3). The eluted fraction was further purified using an S200 size exclusion step in buffer G (100 mM NaCl, 10 mM Tris pH 7.5, 2 mM MgCl₂).

A.1.2 Expression and purification of yPrp43 variants, ctPrp43 and ctPrp43 variants

The plasmid coding for a yPrp43 variants, ctPrp43 or a ctPrp43 variant was transformed in *E. coli* Rosetta 2(DE3) cells. The cells were grown at 37 °C in

Appendix

presence of kanamycin and chloramphenicol until an OD₆₀₀ of 0.9 was reached. The expression of the protein was induced by addition of 1 mM IPTG. The cells were incubated at 16 °C for 16 h.

The purification steps were performed as described earlier for yPrp43 (see A.1.1).

A.1.3 Expression and purification of yNtr2

The plasmid pASG-IBA3_yNtr2 was transformed in *E. coli* Rosetta 2(DE3) cells. These cells were grown at 37 °C in presence of ampicillin and chloramphenicol until an OD₆₀₀ of 0.9 was reached. The expression of yNtr2-Strep was induced by addition of 200 µg anhydrotetracycline per liter cell culture, the cells were incubated at 16 °C for 16 h. Cells were harvested and lysed as described in section A.1.1, but prior to cell lysis, 15 mg pure yNtr2-Strep was added to the sample.

yNtr2-Strep was isolated from the soluble fraction using a StrepTactin HP sepharose column and eluted with Buffer S supplemented with 2 mM D-desthiobiotin. After a desalting step in buffer A (50 mM NaCl, 10 mM Tris pH 7.5, 2 mM MgCl₂), the sample was bound to an Source 30 Q column and eluted by a gradient over 200 ml to 100% buffer B (600 mM NaCl, 100 mM Tris pH 7.5, 2 mM MgCl₂).

A.1.4 Expression and copurification of yNtr1 and yNtr2

The plasmid pASG-IBA3_His₆-yNtr1 was transformed in *E. coli* Rosetta 2(DE3) cells. These cells were grown at 37 °C in presence of ampicillin and chloramphenicol until an OD₆₀₀ of 0.9 was reached. The expression of His₆-yNtr1-Strep was induced by addition of 200 µg anhydrotetracycline per liter cell culture, the cells were incubated at 16 °C for 16 h. Cells were harvested and lysed as described in section A.1.1.

The complex of His₆-yNtr1-Strep and yNtr2-Strep was isolated from the soluble fraction using a StrepTactin HP sepharose column and eluted with buffer S supplemented with 2.5 mM D-desthiobiotin. The elution fractions were applied to a HisTrap column and eluted using buffer S supplemented with 250 mM Imidazol (see section 2.2.3.2).

A.1.5 Expression and purification of yNtr1(1-120)

The expression plasmid pPSG-IBA5_yNtr1(1-120) was transformed into *E. coli* Rosetta 2(DE3) cells. The cells were grown at 37 °C in presence of ampicillin and chloramphenicol until an OD₆₀₀ of 0.9 was reached. The expression of Strep-yNtr1(1-120) was induced by addition of 1 mM IPTG. The cells were incubated at 30 °C for 3 h. Cells were harvested and lysed as described in section A.1.1.

The Strep-yNtr1(1-120) protein was isolated from the soluble fraction using a StrepTactin HP sepharose column and eluted with buffer S supplemented with 2 mM D-desthiobiotin. After a desalting step in buffer A (50 mM NaCl, 10 mM Tris pH 7.5, 2 mM MgCl₂), the sample was bound to an Source 30 Q column and eluted by a gradient over 200 ml to 100% buffer B (600 mM NaCl, 100 mM Tris pH 7.5, 2 mM MgCl₂).

A.1.6 Expression and purification of yNtr1(51-110)

The plasmid pASG-IBA25_yNtr1(51-110) was transformed in *E. coli* Rosetta 2(DE3) cells. These cells were grown at 37 °C in presence of ampicillin and chloramphenicol until an OD₆₀₀ of 0.9 was reached. The expression of GST-yNtr1(51-110) was induced by addition of 200 µg anhydrotetracycline per litre cell culture, the cells were incubated at 30 °C for 3 h. Cells were harvested and lysed as described in section A.1.1.

The GST-yNtr1(51-110) protein was isolated from the soluble fraction using a GSH sepharose column and eluted with buffer S supplemented with 30 mM reduced glutathion (see section 2.2.3.1). The GST-tag was cleaved off during incubation of the elution fractions with PreScission protease overnight at 4 °C (see section 2.2.15). yNtr1(51-110) was separated from the free GST by performing a size exclusion chromatography in buffer G (100 mM NaCl, 10 mM Tris pH 7.5, 2 mM MgCl₂), using an S75 gel filtration column.

A.2 Supplemental tables

Table A.1: Parameters of the published structures of yPrp43

	Walbott <i>et al.</i> (2010)	He <i>et al.</i> (2010)
Reference	Walbott <i>et al.</i> (2010)	He <i>et al.</i> (2010)
pdb-ID	2xau	2kx2
Space group	P3 ₂ 21	P3 ₂ 21
α, β, γ (°)	90, 90, 120	90, 90, 120
a, b, c (Å)	117.55, 117.55, 254.60	118.15, 118.15, 253.60
Condition	50 mM sodium cacodylate pH 6.5 100 mM ammonium acetate 8 % PEG 4000 20°C	100 mM MES pH 6.5 300 mM sodium acetate 14 % PEG 8000 4°C

Table A.2: DNA oligonucleotides used for sequencing (see section 2.2.8)

Name	Sequence (5'-3')
ENTRY-Primer-for2	GCGAAACCGATCCTCGAAG
ENTRY-Primer-rev	CCCTGATTCTGTGGATAACCG
T7-Primer-for	TAATACGACTCACTATAGGG
T7-Primer-rev	TAGTTATTGCTCAGCGGTGG
ASG-Primer-for	GAGTTATTTTACCACTCCCT
ASG-Primer-rev	CGCAGTAGCGGTAAACG

Table A.3: DNA oligonucleotides used for mutagenesis (see section 2.2.5)

Name	Sequence (5'-3')
ct43-intron-for	CGGAGTGGCTTCTTGAATCGCCCTGTCTACTACGACCTC
ct43-intron-rev	AAGAAGCCACTCCGGCCGAATGTTCC

Table A.4: DNA oligonucleotides used for IBA StarGate cloning (see section 2.2.11.2)

Nr.	Name	Sequence (5'-3')
S1	SG-y43-FL-f	AATGGGTTCCAAAAGAAGATTCTCGTCCGAAC
S2	SG-y43-FL-r	TCCCTTTCTTGGAGTGTACTCTTCTTTTTTG
S3	SG-N1-FL-f	AATGGAGGAAATCGGACTCCAACACAGATAAAAAAG
S4	SG-H-N1-FL-f	AATGCATCACCATCACCATCACAGGATTCGGACTCCAACACAGATAAAAAAG
S5	SG-N1(51)-f	AATGAATGCCCAACGATCTCAAAAATTAACGAAG
S6	SG-N1-FL-r	TCCCGAGGTCAAGGGCCATAATAAATTTGGC
S7	SG-N1(110)-r	TCCCGTTGGTATTTGAAAACATTCCTAGACCTGC
S8	SG-N1(120)-r	TCCCATCCTCATTTTCACTGTGATAGTTACTGGAG
S9	SG-N2-FL-f	AATGGCGATCAAGAAAAGAAAATAAATTAGGTTACCAAGT
S10	SG-N2-FL-r	TCCCGTTACCAATGAGCTTATTTATTTAGGCTTGC

Table A.5: DNA oligonucleotides used for restriction site based cloning (see section 2.2.11.1)

Nr.	Name	Sequence (5'-3')
R1	yPrp43-58-Nco-f	GGCCATGGGGACTTCAGCAGAAGAGGGCCCAAAAACCTGAAAGATG
R2	yPrp43-FL-S-Sal-r	TTTGTGACTCAATTTTTCGAACGTGGGGTGGCTCCATTTCTTTGGAGTGCTTACTCTTTTTTTTACCTTG
R3	ctPrp43-FL-Nco-f	GGCCATGGGGCAGATATCACGAAGGGCACGAAGCG
R4	ctPrp43-61-Nco-f	AACCATGGGGACCAACCGGAAGCAGGGCCGAGG
R5	ctPrp43-FL-S-Sal-r	CGCCGTCGACTCAATTTTTCGAACCTGCGG
R6	ctNtr1-FL-Nde-f	GGGCATATGGCGGCCCTCGACGACCTTCAATTC
R7	ctNtr1-FL-S-Sal-r	CCCGTCGACTCAATTTTTCGAACGTGGGGTGGCTCCAATACACCATATCCAGCAAGATGCCCATTC

Table A.6: DNA oligonucleotides used for In-fusion cloning (see section 2.2.11.3)

Nr.	Name	Sequence (5'-3')
I1	y43-M13-f	AGGAGATATACCATGGGGGTTCCAAAAGAAGATTCTCGTCCGAACAC
I2	y43-N1-r	GGAGTCCGAATCCTCTTTCTTGGAGTGCTTACTCTTTTGTTTTACCTTG
I3	y43(745)-N1(51)-r	GATCGTTGGGGCAATTTCCCTTGATCCCTTCCAAAAGATAAATTTGACATCAC
I4	y43-N1(51)-r	GATCGTTGGGGCAATTTTCTTGGAGTGCTTACTCTTCTTTTGTTTTACCTTG
I5	y43-N1(59)-r	ACCATATGTCCTTCGTTTTCTTGGAGTGCTTACTCTTCTTTTGTTTTACCTTG
I6	y43-N1(62)-r	CGCACCAATACCATATTTTCTTGGAGTGCTTACTCTTCTTTTGTTTTACCTTG
I7	N1(51)-y43-f	AAGCACTCCAAGAAAAATGCCCAAAGGATCTCAAAAATTAACGAAAGACATATG
I8	N1(51)-y43(745)-f	GAAAAGGATCAAGAAAAATGCCCAAACGATCTCAAAAATTAACGAAAGACATATG
I9	N1(59)-y43-f	AAGCACTCCAAGAAAAACGAAAGACATATGTTGGTCCGGAAGTTAC
I10	N1(62)-y43-f	AAGCACTCCAAGAAAAATATGTTGGTCCGAAAGTTACTTTTCGAGTATGG
I11	N1(99)-S-M13-r	CGCAAAGCTTGTTCGATCATTTTTTTCGAACTGGGGTGGCTCCAGTGCATAGGTGGCTTTGGGTTTTCTATTGG
I12	N1(102)-S-M13-r	CCGCAAGCTTGTTCGATCATTTTTTTCGAACTGGGGTGGCTCCAAAACATTCCTAGACCTGCATTTGTGCATAGG
I13	N1(110)-S-M13-r	CCGCAAGCTTGTTCGATCATTTTTTTCGAACTGGGGTGGCTCCAGTTGGTATTTGAAAAACATTCCTAGACCTGCATTG
I14	N1(120)-S-M13-r	CCGCAAGCTTGTTCGATCATTTTTTTCGAACTGGGGTGGCTCCAAATCCTCATTTTCACTGTGATAGTTACTGGGAGTTG
I15	cN2-Ex1-M13-f	AGGAGATATACCATGGCCGCTTGGCCGCGAAGCG
I16	cN2-Ex1-Ex2-r	CCGGGGGGGTAGTTCTTCTCCTGCTTGTCTCTCGCC
I17	cN2-Ex2-Ex1-f	AGCAAAGCAGGAGAAAGAACTACGGCCCGCCGGCGCGC
I18	cN2-Ex2-S-M13-r	CCGCAAGCTTGTTCGATCATTTTTTTCGAACTGGGGTGGCTCCAGCCATCATCCTCCATATCTC
I19	ct43-M13-f	AGGAGATATACCATGGCCGAGATATCAAGAAAGGCAACGAAAGCG
I20	ct43-N1(51)-r	GATCGTTGGGGCATTCGCTTGTGGCTTTCATGGCTGCTGGCG
I21	N1(51)-ct43-f	AAGGCCAGCAAGGCGAATGCCCCAACGATCTCAAAAATTAACGAAAGACATATG
I22	N11-y43-f	AAGCACTCCAAGAAAAAGAGGATTCGGACTCCCAACACACAGATAAAAAAGTTC

Appendix

Table A.7: Amino acid content of Prp43 from *Saccharomyces cerevisiae* (yPrp43) and *Chaetomium thermophilum* (ctPrp43).

	yPrp43	ctPrp43
Ala	50 (6.5 %)	62 (8.1 %)
Arg	48 (6.3 %)	52 (6.8 %)
Asn	28 (3.7 %)	25 (3.3 %)
Asp	46 (6.0 %)	41 (5.4 %)
Cys	7 (0.9 %)	9 (1.2 %)
Gln	33 (4.3 %)	33 (4.3 %)
Glu	66 (8.6 %)	65 (8.5%)
Gly	38 (5.0 %)	34 (4.5 %)
His	22 (2.9 %)	19 (2.5 %)
Ile	46 (6.0 %)	40 (5.2 %)
Leu	77 (10.0%)	78 (10.2 %)
Lys	51 (6.6 %)	50 (6.5 %)
Met	18 (2.3 %)	20 (2.6 %)
Phe	30 (3.9 %)	25 (3.3 %)
Pro	42 (5.5 %)	40 (5.2 %)
Ser	48 (6.3 %)	44 (5.8 %)
Thr	38 (5.0 %)	48 (6.3 %)
Trp	3 (0.4 %)	5 (0.7 %)
Tyr	30 (3.9 %)	30 (3.9 %)
Val	46 (6.0 %)	44 (5.8 %)

Table A.8: Abbreviations used in this work

Å	–	Ångström (1 Å = 10 ⁻¹⁰ m)
APS	–	ammoniumpersulfate
ATP	–	adenosine triphosphate
BP	–	branch point
CD	–	circular dichroism
ct	–	<i>Chaetomium thermophilum</i>
CTD	–	C-terminal domain
CV	–	column volume
Da	–	Dalton
DNA	–	deoxyribonucleic acid
dsRNA	–	double-stranded RNA
5'HP	–	5'-hairpin
InsP ₆	–	inositol hexakisphosphate
IPTG	–	isopropyl-β-D-thiogalactopyranoside
kb	–	kilobases
K_d	–	dissociation constant
kDa	–	kilodalton
K_m	–	Michaelis constant
mRNA	–	messenger RNA
NTC	–	NineTeen complex
NTE	–	N-terminal extension
NTR	–	NineTeen complex related
OB-fold	–	oligonucleotide/oligosaccharide binding-fold
OD	–	optical density
pdb-ID	–	identifier within the protein data bank
poly(A)	–	poly-adenine
poly(U)	–	poly-uridine
RHA	–	RNA helicase A
RNA	–	ribonucleic acid
RMSD	–	root mean square deviation
RNP	–	ribonucleic protein
SDS-PAGE	–	SDS polyacrylamide gel electrophoresis
SF1	–	superfamily
S _N	–	Nucleophilic substitution
ssRNA	–	single-stranded RNA
3'/5'ss	–	3'/5' splice site
tRNA	–	transfer RNA
UsnRNA	–	Uridine rich smal nuclear RNA
UTP	–	uridine triphosphate
v_{max}	–	maximal reaction velocity
WH	–	winged helix
y	–	<i>Saccharomyces cerevisiae</i> (yeast)

A.3 Supplemental figures

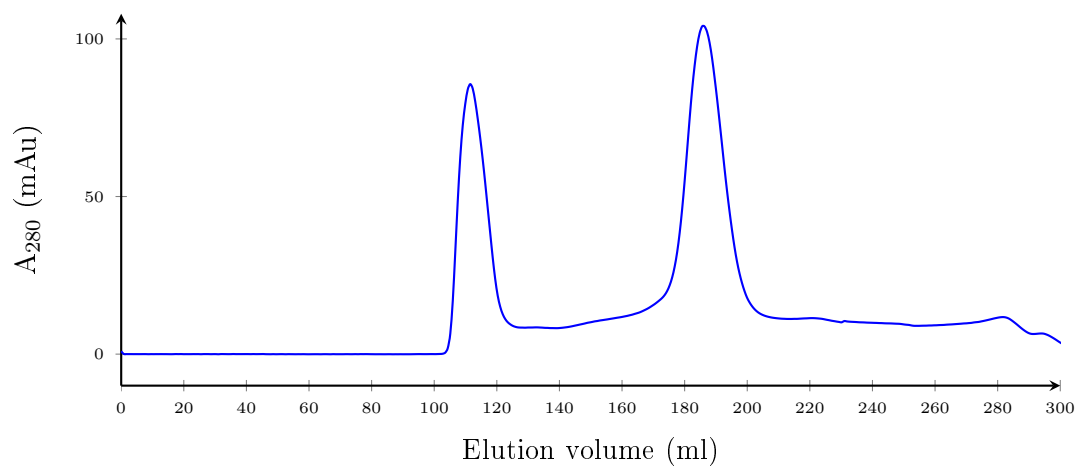


Figure A.1: Complete chromatogram of the preparative size exclusion chromatography of yPrp43-Strep, using a Superdex 200 26/60 column (see section 3.1.1). The absorbance at 280 nm in a machine-specific unit (mAu, y-axis) is shown in blue referring to the elution volume (ml, x-axis).

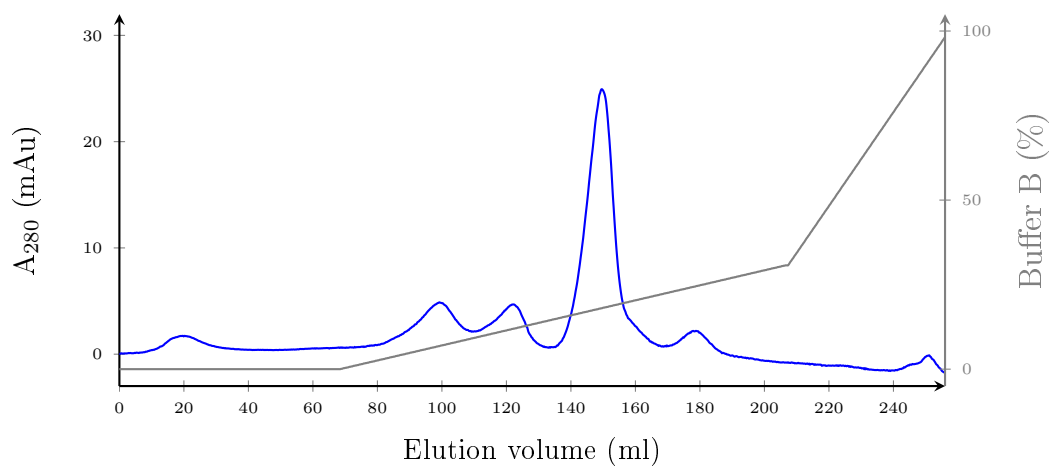


Figure A.2: Complete chromatogram of the anion exchange chromatography of yNtr2-Strep, using a Source 30Q column (see section 3.1.2). The absorption at 280 nm (mAu, left y-axis) is shown in blue with respect to the elution volume (ml, x-axis). The gradient of buffer B is shown on the right y-axis (in %) as a grey line.

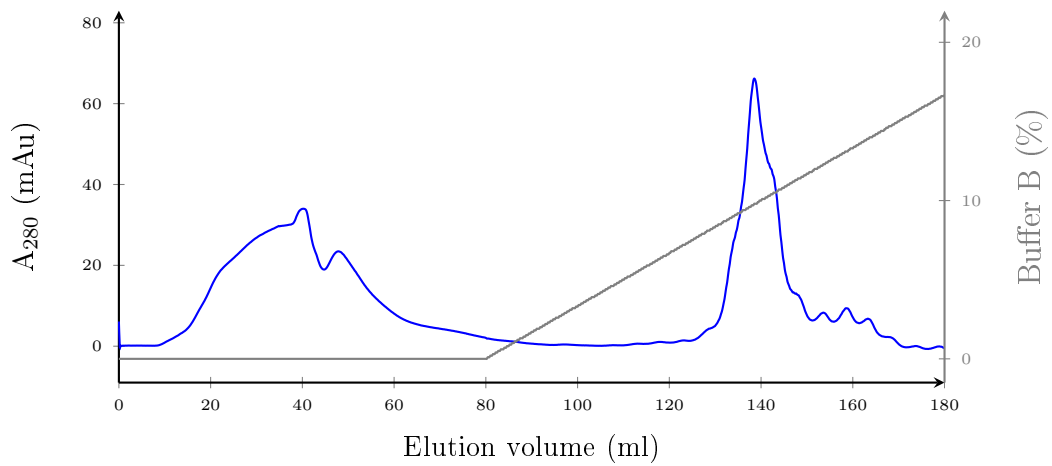


Figure A.3: Chromatogram of the anion exchange chromatography of Strep-yNtr1(1-120), using a Source 30Q column (see section 3.1.4). The absorption at 280 nm is shown in blue in a machine specific unit (mAu, left y-axis) referring to the elution volume in ml (x-axis). The gradient of buffer B is indicated on the right y-axis (in %) as a grey line.

Appendix

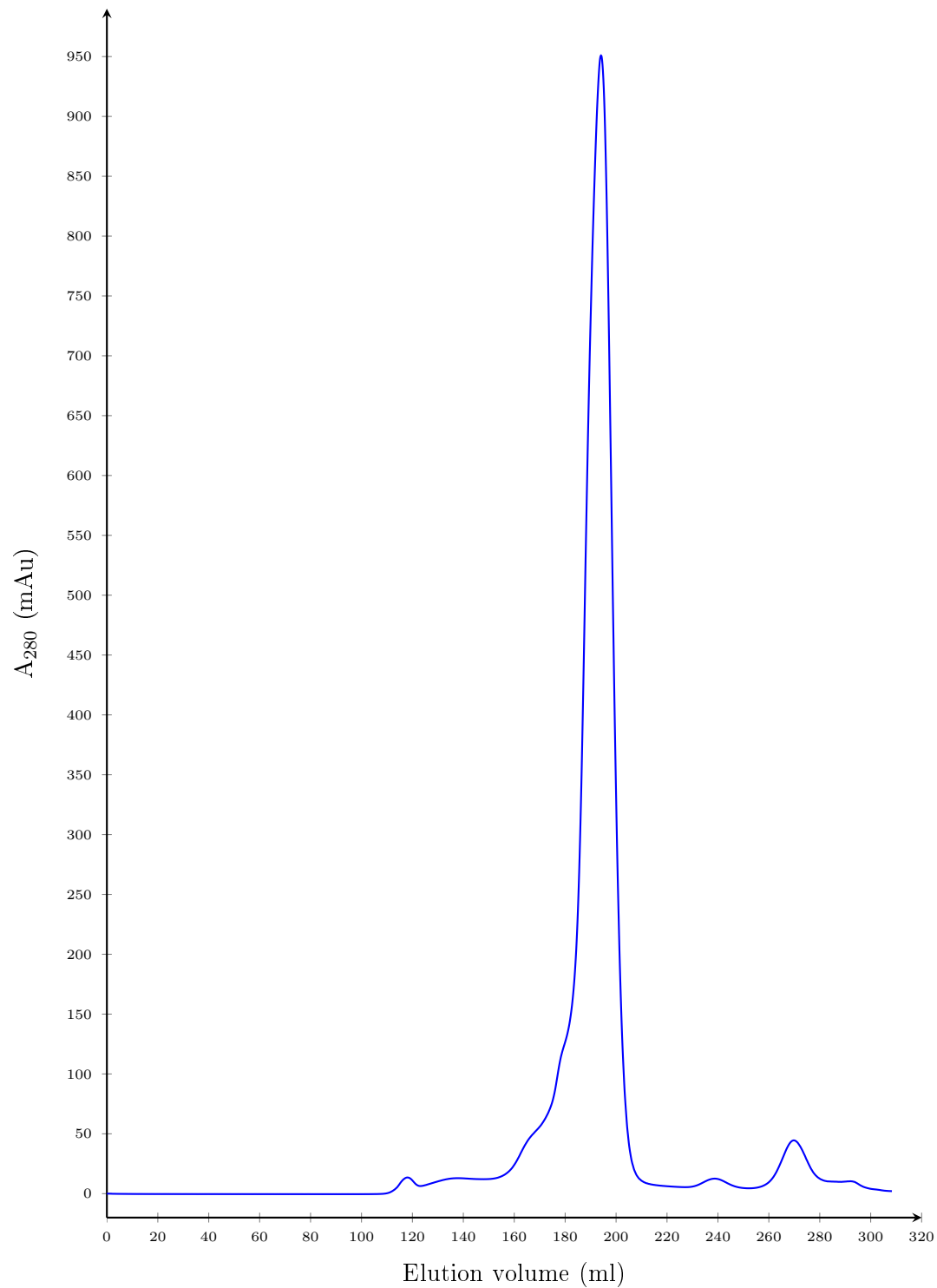


Figure A.4: Complete chromatogram of the S75 size exclusion chromatography of yNtr1 (51-110) after cleavage of the GST affinity sequence, using a Superdex 75 26/60 column (see section 3.1.5). The absorbance at 280 nm in machine specific units (mAu, y-axis) is shown as a blue line with respect to the elution volume (ml, x-axis).

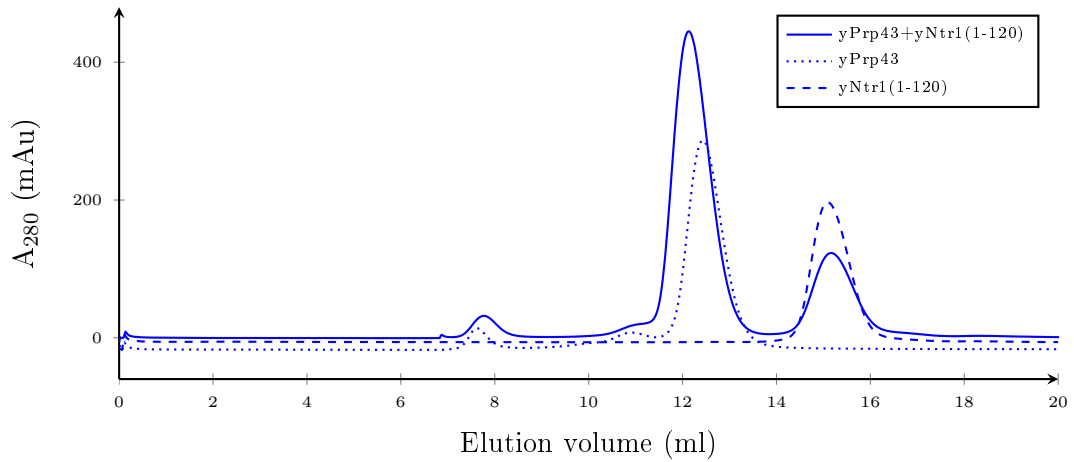


Figure A.5: Chromatogram of the analytical size exclusion chromatography of the purification of the complex yPrp43/yNtr1(1-120), using a Superdex 200 (10/300) column (see section 3.2.1). The absorption at 280 nm (machine specific unit, y-axis) is shown with respect to the elution volume (ml, x-axis) for three independent samples: yPrp43+yNtr1(1-120) (continuous line), free yPrp43 (dotted line) and free yNtr1(1-120) (dashed line).

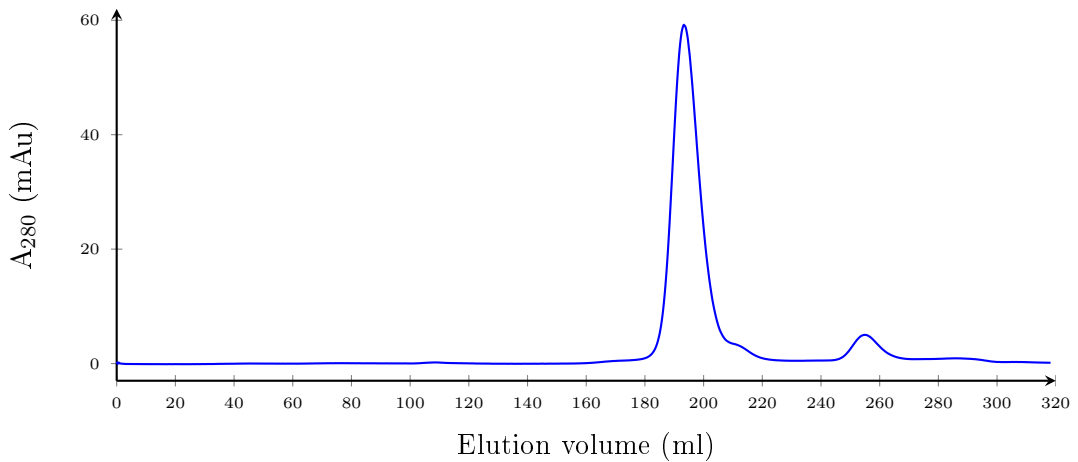


Figure A.6: Chromatogram of the size exclusion chromatography of yPrp43 and yNtr1(51-110), using a Superdex 200 26/60 column (see section 3.2.2). The blue line reflects the UV absorption at 280 nm (machine specific unit, y-axis) with respect to the elution volume (ml, x-axis).

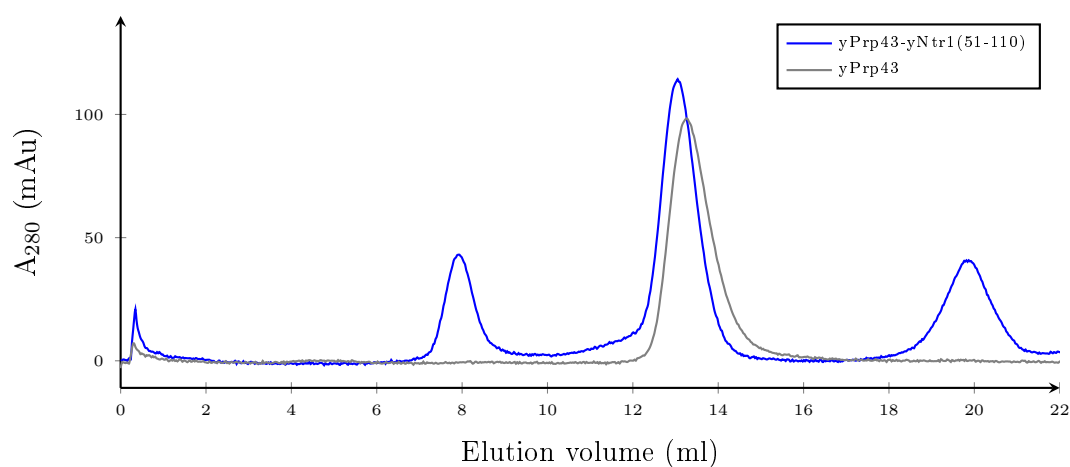


Figure A.7: Complete chromatogram of the analytical S200 size exclusion chromatography of the fusion protein yPrp43-yNtr1(51-110) in comparison to yPrp43, using a Superdex 200 (10/300) column (see section 3.3.3). The absorption at 280 nm is shown in a machine specific unit (mAu, y-axis) with respect to the elution volume (ml, x-axis). The fusion protein yPrp43-yNtr1(51-110) is shown in blue, yPrp43 in grey.

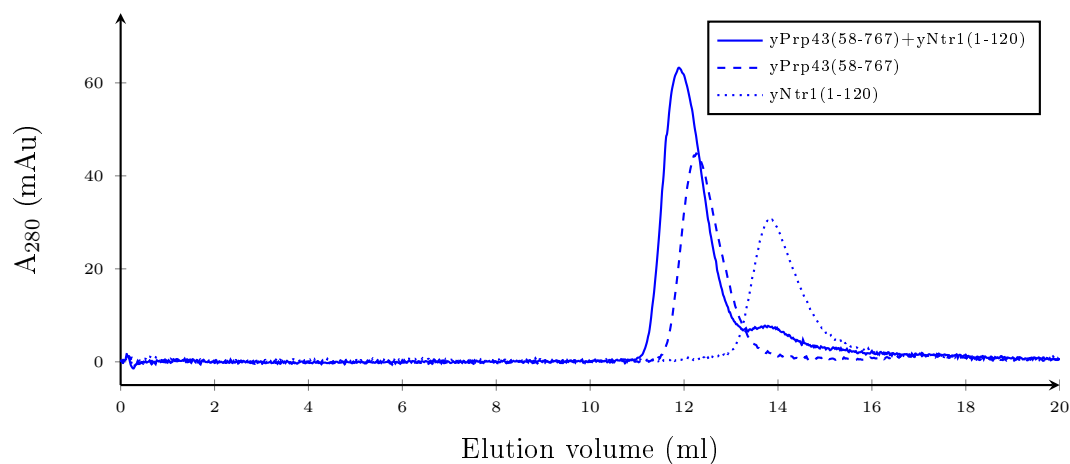


Figure A.8: Chromatogram of the analytical size exclusion chromatography of the complex yPrp43(58-767)/yNtr1(1-120), using a Superdex 200 (10/300) column (see section 3.4.3). The absorption at 280 nm (y-axis, machine-specific unit) during the three independent runs of yPrp43(58-767)+yNtr1(1-120), yPrp43(58-767) and yNtr1(1-120) is shown with respect to the elution volume (x-axis, ml).

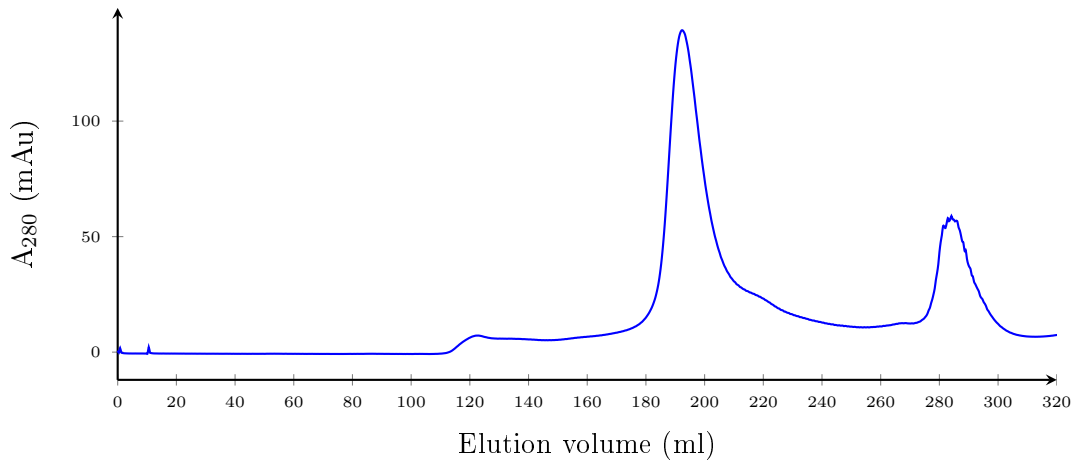


Figure A.9: Preparative size exclusion chromatography of ctPrp43-Strep, using a Superdex 200 26/60 column (see section 3.8.2). In the chromatogram, the absorption at 280 nm in a machine-specific unit (mAu, y-axis) is shown in blue referring to the elution volume (ml, x-axis).

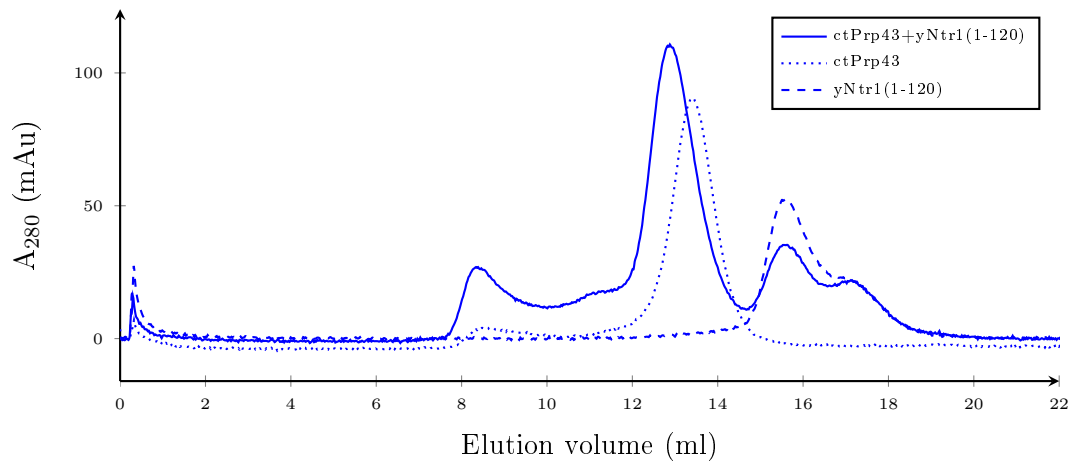


Figure A.10: Chromatogram of the analytical size exclusion chromatography of the complex ctPrp43/yNtr1(1-120), using a Superdex 200 (10/300) column (see section 3.8.2). The absorption at 280 nm (machine specific unit, y-axis) is shown with respect to the elution volume (ml, x-axis) for three independent samples: ctPrp43+yNtr1(1-120), free ctPrp43 and free yNtr1(1-120).

Appendix

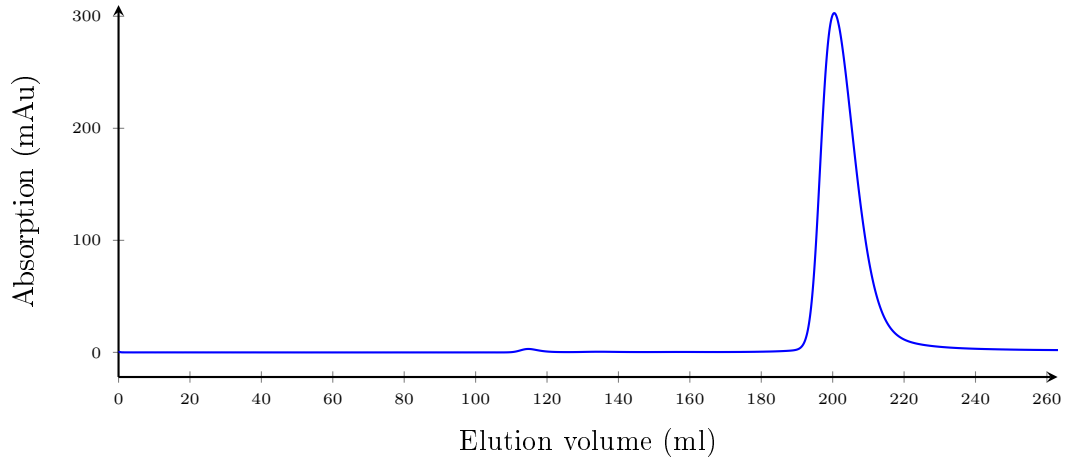
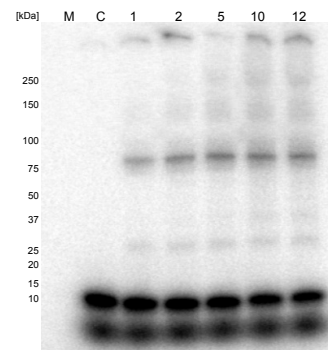


Figure A.11: Size exclusion chromatography of ctPrp43(61-764), using a Superdex 200 26/60 column (see section 3.8.3). In the chromatogram, the absorbance at 280 nm in a machine-specific unit (mAu, y-axis) is shown in blue referring to the elution volume (ml, x-axis). The collected elution fractions are indicated along the x-axis.

Figure A.12: Wet gel of protein-RNA cross-linking samples of the complex yPrp43/yNtr1(1-120). The position of the band corresponding to yPrp43 is labelled by an asterisk, of yNtr1 by a circle. Uncrosslinked RNA is present at the bottom. The position of the marker band has been adapted from the Coomassie stain of the same gel (not shown). The corresponding experiment has been performed by Romina Hofele.



A.3 Supplemental figures



Figure A.13: Sequence alignment of Prp43 from *S. cerevisiae* and *C. thermophilum*. In addition to the alignment, the secondary structure of yPrp43 is shown. The alignment has been performed using ClustalW and has been visualised using ESPrpt.

Appendix

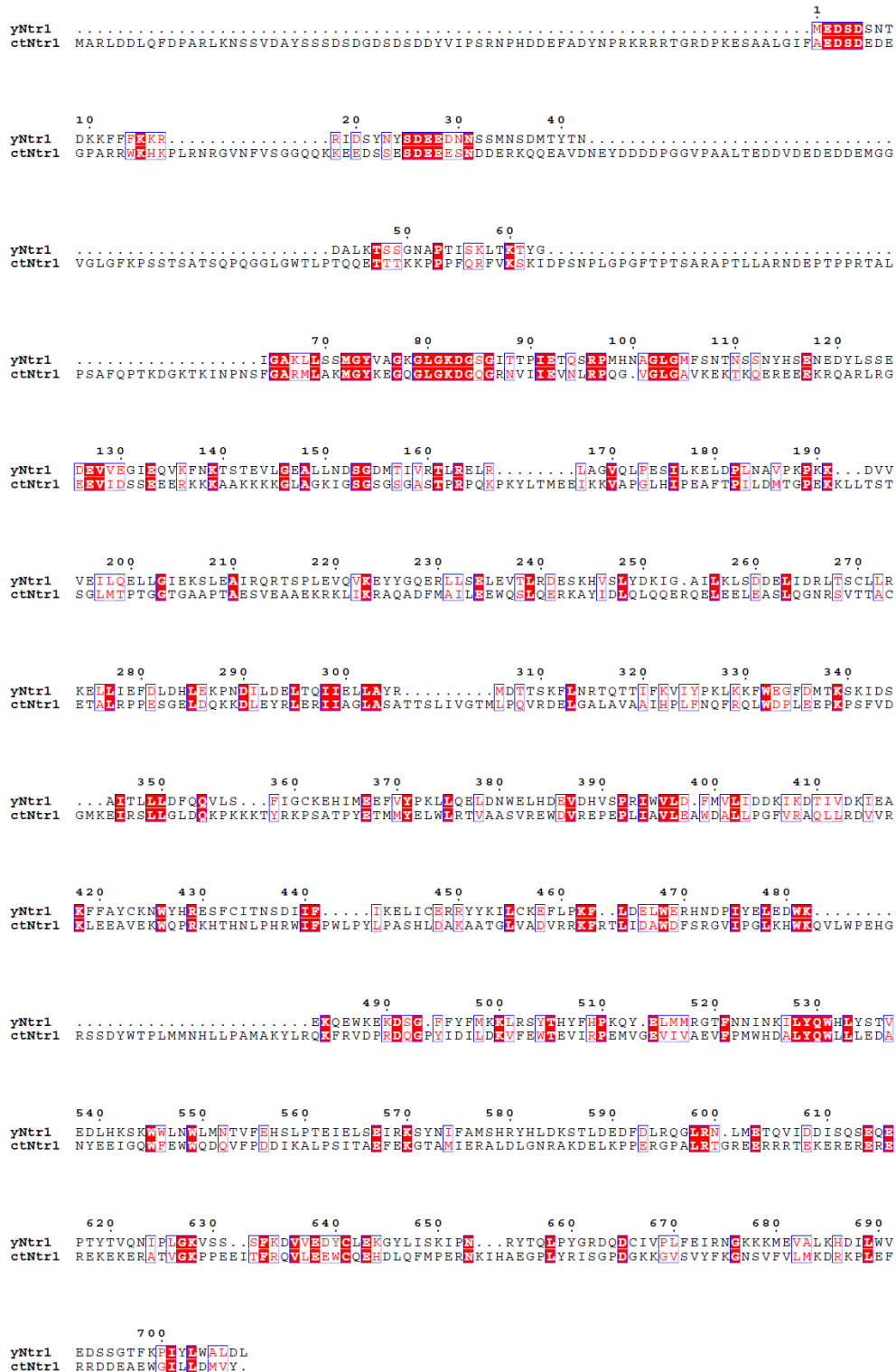


Figure A.14: Sequence alignment of Ntr1 from *S. cerevisiae* and *C. thermophilum*. The alignment has been performed using ClustalW and has been visualised using ESPript.

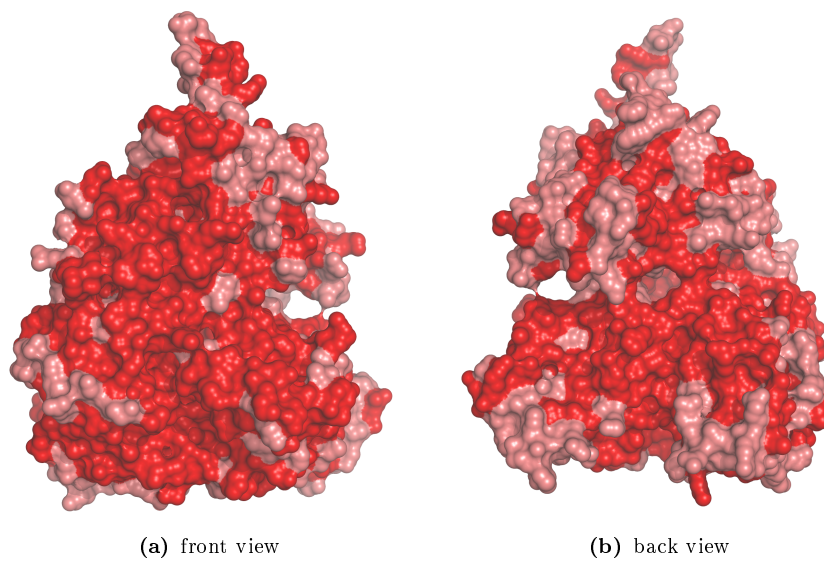


Figure A.17: Visualisation of the conserved surface residues between yPrp43 (pdb-ID 2xau) and ctPrp43(61-764) using ProtSkin in front view (a) and back view (b). Red color indicates regions of similar residues, whereas light red indicates regions of homologous residues. No region could be identified with major differences between the two orthologs. The structures have been visualised using PyMOL.

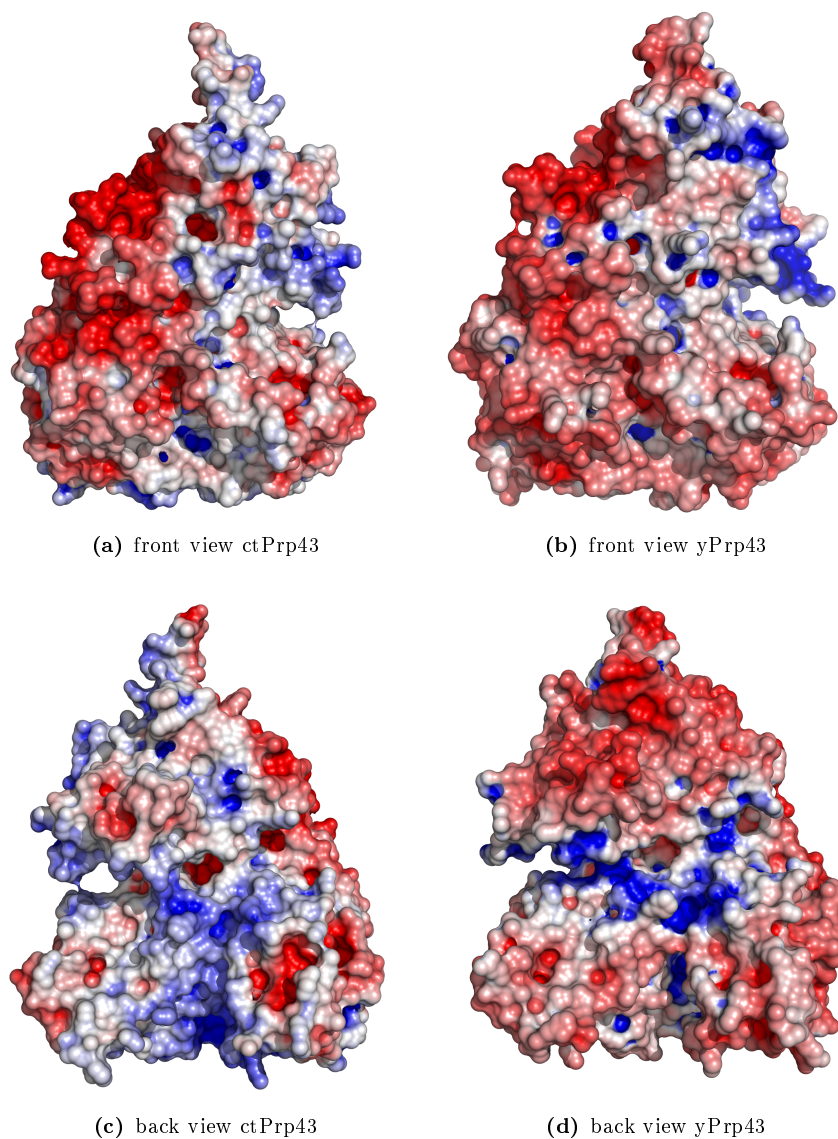


Figure A.18: Surface charge distribution in front and back view for the crystal structure of ctPrp43(61-764) in comparison to yPrp43(58-767). Blue coloring indicates positively, red coloring negatively charged regions. The contour levels were $-6/+7$ for ctPrp43 and $-8/+3$ for yPrp43. The surface charge has been calculated using pdb2pqr and has been displayed with the APBS plugin of PyMOL.

Bibliography

Abelson J, Trotta CR, and Li H.

tRNA splicing.

The Journal of biological chemistry, 273(21):12685–8, 1998.

Achsel T, Brahm H, Kastner B, Bachi A, Wilm M, and Lührmann R.

A doughnut-shaped heteromer of human Sm-like proteins binds to the 3'-end of U6 snRNA, thereby facilitating U4/U6 duplex formation in vitro.

The EMBO journal, 18(20):5789–802, 1999.

Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, and Terwilliger TC.

PHENIX : building new software for automated crystallographic structure determination.

Acta Crystallographica Section D Biological Crystallography, 58(11):1948–1954, 2002.

Altschul SF, Gish W, Miller W, Myers EW, and Lipman DJ.

Basic local alignment search tool.

Journal of molecular biology, 215(3):403–10, 1990.

Amlacher S, Sarges P, Flemming D, Noort V, Kunze R, Devos DP, Arumugam M, Bork P, and Hurt E.

Insight into structure and assembly of the nuclear pore complex by utilizing the genome of a eukaryotic thermophile.

Cell, 146(2):277–89, 2011.

Andersen CBF, Ballut L, Johansen JS, Chamieh H, Nielsen KH, Oliveira CLP, Pedersen JS, Séraphin B, Le Hir H, and Andersen GR.

Structure of the exon junction core complex with a trapped DEAD-box ATPase bound to RNA.

Science, 313(5795):1968–72, 2006.

Bibliography

Appleby TC, Anderson R, Fedorova O, Pyle AM, Wang R, Liu X, Brendza KM, and Somoza JR.

Visualizing ATP-dependent RNA translocation by the NS3 helicase from HCV. *Journal of molecular biology*, 405(5):1139–53, 2011.

Aravind L and Koonin EV.

G-patch: a new conserved domain in eukaryotic RNA-processing proteins and type D retroviral polyproteins.

Trends in biochemical sciences, 24(9):342–4, 1999.

Arenas J and Abelson J.

Prp43: An RNA helicase-like factor involved in spliceosome disassembly.

Proceedings of the National Academy of Sciences of the United States of America, 94(22):11798–802, 1997.

Ares M, Grate L, and Pauling MH.

A handful of intron-containing genes produces the lion's share of yeast mRNA.

RNA, 5(9):1138–9, 1999.

Ast G.

How did alternative splicing evolve?

Nature reviews. Genetics, 5(10):773–82, 2004.

Baker NA, Sept D, Joseph S, Holst MJ, and McCammon JA.

Electrostatics of nanosystems: application to microtubules and the ribosome.

Proceedings of the National Academy of Sciences of the United States of America, 98(18):10037–41, 2001.

Belon C and Frick D.

Monitoring helicase activity with molecular beacons.

BioTechniques, 45(4):433–40, 442, 2008.

Berget SM.

Exon recognition in vertebrate splicing.

The Journal of biological chemistry, 270(6):2411–4, 1995.

Berglund JA, Chua K, Abovich N, Reed R, and Rosbash M.

The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC.

Cell, 89(5):781–7, 1997.

- Bessonov S, Anokhina M, Will CL, Urlaub H, and Lührmann R.
Isolation of an active step I spliceosome and composition of its RNP core.
Nature, 452(7189):846–50, 2008.
- Bleichert F and Baserga SJ.
The long unwinding road of RNA helicases.
Molecular cell, 27(3):339–52, 2007.
- Bohnsack MT, Martin R, Granneman S, Ruprecht M, Schleiff E, and Tollervey D.
Prp43 bound at different sites on the pre-rRNA performs distinct functions in ribosome synthesis.
Molecular cell, 36(4):583–92, 2009.
- Bono F, Ebert J, Lorentzen E, and Conti E.
The crystal structure of the exon junction complex reveals how it maintains a stable grip on mRNA.
Cell, 126(4):713–25, 2006.
- Boon KL, Auchynnikava T, Edwalds-Gilbert G, Barrass JD, Droop AP, Dez C, and Beggs JD.
Yeast Ntr1/Spp382 mediates Prp43 function in postspliceosomes.
Molecular and cellular biology, 26(16):6016–23, 2006.
- Bork P and Koonin EV.
An expanding family of helicases within the 'DEAD/H' superfamily.
Nucleic acids research, 21(3):751–2, 1993.
- Box JA, Bunch JT, Tang W, and Baumann P.
Spliceosomal cleavage generates the 3' end of telomerase RNA.
Nature, 456(7224):910–4, 2008.
- Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, and Warren GL.
Crystallography & NMR system: A new software suite for macromolecular structure determination.
Acta crystallographica. Section D, Biological crystallography, 54(Pt 5):905–21, 1998.

Bibliography

- Buchan DWA, Ward SM, Lobley AE, Nugent TCO, Bryson K, and Jones DT.
Protein annotation and modelling servers at University College London.
Nucleic acids research, 38(Web Server issue):W563–8, 2010.
- Burgess S and Guthrie C.
Beat the clock: paradigms for NTPases in the maintenance of biological fidelity.
Trends in biochemical sciences, 18(10):381–4, 1993.
- Büttner K, Nehring S, and Hopfner KP.
Structural basis for DNA duplex separation by a superfamily-2 helicase.
Nature structural & molecular biology, 14(7):647–52, 2007.
- Caruthers JM, Johnson ER, and McKay DB.
Crystal structure of yeast initiation factor 4A, a DEAD-box RNA helicase.
Proceedings of the National Academy of Sciences of the United States of America, 97(24):13080–5, 2000.
- Chakrabarti S, Jayachandran U, Bonneau F, Fiorini F, Basquin C, Domcke S, Le Hir H, and Conti E.
Molecular mechanisms for the RNA-dependent ATPase activity of Upf1 and its regulation by Upf2.
Molecular cell, 41(6):693–703, 2011.
- Chamieh H, Ballut L, Bonneau F, and Le Hir H.
NMD factors UPF2 and UPF3 bridge UPF1 to the exon junction complex and stimulate its RNA helicase activity.
Nature structural & molecular biology, 15(1):85–93, 2008.
- Chan RT, Robart AR, Rajashankar KR, Pyle AM, and Toor N.
Crystal structure of a group II intron in the pre-catalytic state.
Nature structural & molecular biology, 19(5):555–7, 2012.
- Chan SP and Cheng SC.
The Prp19-associated complex is required for specifying interactions of U5 and U6 with pre-mRNA during spliceosome activation.
The Journal of biological chemistry, 280(35):31190–9, 2005.
- Chan SP, Kao DI, Tsai WY, and Cheng SC.
The Prp19p-associated complex in spliceosome activation.
Science, 302(5643):279–82, 2003.

- Chapman KB and Boeke JD.
Isolation and characterization of the gene encoding yeast debranching enzyme.
Cell, 65(3):483–92, 1991.
- Chen HC and Cheng SC.
Functional roles of protein splicing factors.
Bioscience reports, 32(4):345–59, 2012.
- Chen HC, Tseng CK, Tsai RT, Chung CS, and Cheng SC.
Link of NTR-Mediated Spliceosome Disassembly with DEAH-Box ATPases Prp2, Prp16, and Prp22.
Molecular and cellular biology, 33(3):514–25, 2013.
- Chen M and Manley JL.
Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches.
Nature reviews. Molecular cell biology, 10(11):741–54, 2009.
- Chen VB, Arendall WB, Headd J, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, and Richardson DC.
MolProbity: all-atom structure validation for macromolecular crystallography.
Acta crystallographica. Section D, Biological crystallography, 66(Pt 1):12–21, 2010.
- Cheng Z, Coller J, Parker R, and Song H.
Crystal structure and functional analysis of DEAD-box protein Dhh1p.
RNA, 11(8):1258–70, 2005.
- Chiu YF, Liu YC, Chiang TW, Yeh TC, Tseng CK, Wu NY, and Cheng SC.
Cwc25 is a novel splicing factor required after Prp2 and Yju2 to facilitate the first catalytic reaction.
Molecular and cellular biology, 29(21):5671–8, 2009.
- Clerici M, Mourão A, Gutsche I, Gehring NH, Hentze MW, Kulozik A, Kadlec J, Sattler M, and Cusack S.
Unusual bipartite mode of interaction between the nonsense-mediated decay factors, UPF1 and UPF2.
The EMBO journal, 28(15):2293–306, 2009.

Bibliography

- Company M, Arenas J, and Abelson J.
Requirement of the RNA helicase-like protein PRP22 for release of messenger RNA from spliceosomes.
Nature, 349(6309):487–93, 1991.
- Cordin O, Banroques J, Tanner NK, and Linder P.
The DEAD-box protein family of RNA helicases.
Gene, 367:17–37, 2006.
- Cordin O, Hahn D, and Beggs JD.
Structure, function and regulation of spliceosomal RNA helicases.
Current opinion in cell biology, 24(3):431–8, 2012.
- Das R, Zhou Z, and Reed R.
Functional association of U2 snRNP with the ATP-independent spliceosomal complex E.
Molecular cell, 5(5):779–87, 2000.
- Dayie KT and Padgett RA.
A glimpse into the active site of a group II intron and maybe the spliceosome, too.
RNA, 14(9):1697–703, 2008.
- Deckert J, Hartmuth K, Boehringer D, Behzadnia N, Will CL, Kastner B, Stark H, Urlaub H, and Lührmann R.
Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions.
Molecular and cellular biology, 26(14):5528–43, 2006.
- Deprez C, Llobès R, Gavioli M, Marion D, Guerlesquin F, and Blanchard L.
Solution structure of the E.coli TolA C-terminal domain reveals conformational changes upon binding to the phage g3p N-terminal domain.
Journal of molecular biology, 346(4):1047–57, 2005.
- Deutsch M and Long M.
Intron-exon structures of eukaryotic model organisms.
Nucleic acids research, 27(15):3219–28, 1999.
- Dolinsky TJ, Czodrowski P, Li Y, Nielsen JE, Jensen JH, Klebe G, and Baker NA.
PDB2PQR: expanding and upgrading automated preparation of biomolecular

structures for molecular simulations.

Nucleic acids research, 35(Web Server issue):W522–5, 2007.

Dosztányi Z, Mészáros B, and Simon I.

ANCHOR: web server for predicting protein binding regions in disordered proteins.

Bioinformatics, 25(20):2745–6, 2009.

Drummond DR, Armstrong J, and Colman A.

The effect of capping and polyadenylation on the stability, movement and translation of synthetic messenger RNAs in *Xenopus* oocytes.

Nucleic acids research, 13(20):7375–94, 1985.

Dumont S, Cheng W, Serebrov V, Beran RK, Tinoco I, Pyle AM, and Bustamante C.

RNA translocation and unwinding mechanism of HCV NS3 helicase and its coordination by ATP.

Nature, 439(7072):105–8, 2006.

Edwards-Gilbert G, Kim DH, Silverman E, and Lin RJ.

Definition of a spliceosome interaction domain in yeast Prp2 ATPase.

RNA, 10(2):210–20, 2004.

Egecioglu DE and Chanfreau G.

Proofreading and spellchecking: a two-tier strategy for pre-mRNA splicing quality control.

RNA, 17(3):383–9, 2011.

Emsley P, Lohkamp B, Scott WG, and Cowtan K.

Features and development of Coot.

Acta crystallographica. Section D, Biological crystallography, 66(Pt 4):486–501, 2010.

Fabrizio P, Dannenberg J, Dube P, Kastner B, Stark H, Urlaub H, and Lührmann R.

The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome.

Molecular cell, 36(4):593–608, 2009.

Bibliography

- Fairman-Williams ME, Guenther UP, and Jankowsky E.
SF1 and SF2 helicases: family matters.
Current opinion in structural biology, 20(3):313–24, 2010.
- Fedorova L and Fedorov A.
Puzzles of the Human Genome : Why Do We Need Our Introns ?
Current genomics, 6(8):589–595, 2005.
- Finn RD, Mistry J, Tate J, Coggill P, Heger A, Pollington JE, Gavin OL, Gunasekaran P, Ceric G, Forslund K, Holm L, Sonnhammer ELL, Eddy SR, and Bateman A.
The Pfam protein families database.
Nucleic acids research, 38(Database issue):D211–22, 2010.
- Fiorini F, Boudvillain M, and Le Hir H.
Tight intramolecular regulation of the human Upf1 helicase by its N- and C-terminal domains.
Nucleic acids research, 41(4):2404–15, 2013.
- Fourmann JB, Schmitzová J, Christian H, Urlaub H, Ficner R, Boon KL, Fabrizio P, and Lührmann R.
Dissection of the factor requirements for spliceosome disassembly and the elucidation of its dissociation products using a purified splicing system.
Genes & development, 27(4):413–28, 2013.
- Fox-Walsh KL, Dou Y, Lam BJ, Hung SP, Baldi PF, and Hertel KJ.
The architecture of pre-mRNAs affects mechanisms of splice-site pairing.
Proceedings of the National Academy of Sciences of the United States of America, 102(45):16176–81, 2005.
- Frénel K, Callebaut I, Wecker K, Prochnicka-Chalufour A, Dendouga N, Zinn-Justin S, Delepierre M, Tomavo S, and Wolff N.
Structural and functional characterization of the TgDRE multidomain protein, a DNA repair enzyme from *Toxoplasma gondii*.
Biochemistry, 45(15):4867–74, 2006.
- Galej WP, Oubridge C, Newman AJ, and Nagai K.
Crystal structure of Prp8 reveals active site cavity of the spliceosome.
Nature, 493(7434):638–43, 2013.

- Gouet P, Courcelle E, Stuart DI, and Métoz F.
ESPrpt: analysis of multiple sequence alignments in PostScript.
Bioinformatics, 15(4):305–8, 1999.
- Gozani O, Feld R, and Reed R.
Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A.
Genes & development, 10(2):233–43, 1996.
- Grabowski PJ, Padgett RA, and Sharp PA.
Messenger RNA splicing in vitro: an excised intervening sequence and a potential intermediate.
Cell, 37(2):415–27, 1984.
- Gu M and Rice CM.
Three conformational snapshots of the hepatitis C virus NS3 helicase reveal a ratchet translocation mechanism.
Proceedings of the National Academy of Sciences of the United States of America, 107(2):521–8, 2010.
- Guerrier-Takada C, Gardiner K, Marsh T, Pace N, and Altman S.
The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme.
Cell, 35(3 Pt 2):849–57, 1983.
- Guglielmi B and Werner M.
The yeast homolog of human PinX1 is involved in rRNA and small nucleolar RNA maturation, not in telomere elongation inhibition.
The Journal of biological chemistry, 277(38):35712–9, 2002.
- Halbach F, Rode M, and Conti E.
The crystal structure of *S. cerevisiae* Ski2, a DExH helicase associated with the cytoplasmic functions of the exosome.
RNA, 18(1):124–34, 2012.
- Hamm J and Mattaj IW.
Monomethylated cap structures facilitate RNA export from the nucleus.
Cell, 63(1):109–18, 1990.

Bibliography

- He Y, Andersen GR, and Nielsen KH.
Structural basis for the function of DEAH helicases.
EMBO reports, 11(3):180–6, 2010.
- Hegyí H, Schad E, and Tompa P.
Structural disorder promotes assembly of protein complexes.
BMC structural biology, 7:65, 2007.
- Herold N, Will CL, Wolf E, Kastner B, Urlaub H, and Lührmann R.
Conservation of the protein composition and electron microscopy structure of *Drosophila melanogaster* and human spliceosomal complexes.
Molecular and cellular biology, 29(1):281–301, 2009.
- Herrmann G, Kais S, Hoffbauer J, Shah-Hosseini K, Brüggelolte N, Schober H, Fäsi M, and Schär P.
Conserved interactions of the splicing factor Ntr1/Spp382 with proteins involved in DNA double-strand break repair and telomere metabolism.
Nucleic acids research, 35(7):2321–32, 2007.
- Hilbert M, Karow AR, and Klostermeier D.
The mechanism of ATP-dependent RNA unwinding by DEAD box proteins.
Biological chemistry, 390(12):1237–50, 2009.
- Hocine S, Singer R, and Grünwald D.
RNA processing and export.
Cold Spring Harbor perspectives in biology, 2(12):a000752, 2010.
- Hoffman BE and Grabowski PJ.
U1 snRNP targets an essential splicing factor, U2AF65, to the 3' splice site by a network of interactions spanning the exon.
Genes & development, 6(12B):2554–68, 1992.
- Hopfner KP and Michaelis J.
Mechanisms of nucleic acid translocases: lessons from structural biology and single-molecule biophysics.
Current opinion in structural biology, 17(1):87–95, 2007.
- Horowitz DS and Abelson J.
Stages in the second reaction of pre-mRNA splicing: the final step is ATP

independent.

Genes & development, 7(2):320–9, 1993.

Hotz HR and Schwer B.

Mutational analysis of the yeast DEAH-box splicing factor Prp16.

Genetics, 149(2):807–15, 1998.

House AE and Lynch KW.

An exonic splicing silencer represses spliceosome assembly after ATP-dependent exon recognition.

Nature structural & molecular biology, 13(10):937–44, 2006.

Hsu CL and Stevens A.

Yeast cells lacking 5'→3' exoribonuclease 1 contain mRNA species that are poly(A) deficient and partially lack the 5' cap structure.

Molecular and cellular biology, 13(8):4826–35, 1993.

Ismail S, Vetter IR, Sot B, and Wittinghofer A.

The structure of an Arf-ArfGAP complex reveals a Ca²⁺ regulatory mechanism.

Cell, 141(5):812–21, 2010.

Jacquier A.

Self-splicing group II and nuclear pre-mRNA introns: how similar are they?

Trends in biochemical sciences, 15(9):351–4, 1990.

Jády BE, Darzacq X, Tucker KE, Matera AG, Bertrand E, and Kiss T.

Modification of Sm small nuclear RNAs occurs in the nucleoplasmic Cajal body following import from the cytoplasm.

The EMBO journal, 22(8):1878–88, 2003.

Jurica MS and Moore MJ.

Pre-mRNA splicing: awash in a sea of proteins.

Molecular cell, 12(1):5–14, 2003.

Karow AR and Klostermeier D.

A conformational change in the helicase core is necessary but not sufficient for RNA unwinding by the DEAD box helicase YxiN.

Nucleic acids research, 37(13):4464–71, 2009.

Bibliography

Karow AR, Theissen B, and Klostermeier D.

Authentic interdomain communication in an RNA helicase reconstituted by expressed protein ligation of two helicase domains.

The FEBS journal, 274(2):463–73, 2007.

Keating KS, Toor N, Perlman PS, and Pyle AM.

A structural analysis of the group II intron active site and implications for the spliceosome.

RNA, 16(1):1–9, 2010.

Kelly SM, Jess TJ, and Price NC.

How to study proteins by circular dichroism.

Biochimica et biophysica acta, 1751(2):119–39, 2005.

Kim DH and Rossi JJ.

The first ATPase domain of the yeast 246-kDa protein is required for in vivo unwinding of the U4/U6 duplex.

RNA, 5(7):959–71, 1999.

Kistler AL and Guthrie C.

Deletion of MUD2, the yeast homolog of U2AF65, can bypass the requirement for sub2, an essential spliceosomal ATPase.

Genes & development, 15(1):42–9, 2001.

Konarska MM and Query CC.

Insights into the mechanisms of splicing: more lessons from the ribosome.

Genes & development, 19(19):2255–60, 2005.

Konforti BB, Koziolkiewicz MJ, and Konarska MM.

Disruption of base pairing between the 5' splice site and the 5' end of U1 snRNA is required for spliceosome assembly.

Cell, 75(5):863–73, 1993.

Koodathingal P, Novak T, Piccirilli JA, and Staley JP.

The DEAH box ATPases Prp16 and Prp43 cooperate to proofread 5' splice site cleavage during pre-mRNA splicing.

Molecular cell, 39(3):385–95, 2010.

- Korneta I and Bujnicki JM.
Intrinsic disorder in the human spliceosomal proteome.
PLoS computational biology, 8(8):e1002641, 2012.
- Korneta I, Magnus M, and Bujnicki JM.
Structural bioinformatics of the human spliceosomal proteome.
Nucleic acids research, 40(15):7046–65, 2012.
- Kruger K, Grabowski PJ, Zaug J, Sands J, Gottschling DE, and Cech TR.
Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*.
Cell, 31(1):147–57, 1982.
- Kudlinzki D, Schmitt A, Christian H, and Ficner R.
Structural analysis of the C-terminal domain of the spliceosomal helicase Prp22.
Biological chemistry, 393(10):1131–40, 2012.
- Lardelli RM, Thompson JX, Yates JR, and Stevens SW.
Release of SF3 from the intron branchpoint activates the first step of pre-mRNA splicing.
RNA, 16(3):516–28, 2010.
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, and Higgins DG.
Clustal W and Clustal X version 2.0.
Bioinformatics, 23(21):2947–8, 2007.
- Le Hir H and Andersen GR.
Structural insights into the exon junction complex.
Current opinion in structural biology, 18(1):112–9, 2008.
- Lebaron S, Froment C, Fromont-Racine M, Rain JC, Monsarrat B, Caizergues-Ferrer M, and Henry Y.
The splicing ATPase prp43p is a component of multiple preribosomal particles.
Molecular and cellular biology, 25(21):9269–82, 2005.
- Lebaron S, Papin C, Capeyrou R, Chen YL, Froment C, Monsarrat B, Caizergues-Ferrer M, Grigoriev M, and Henry Y.

Bibliography

- The ATPase and helicase activities of Prp43p are stimulated by the G-patch protein Pfa1p during yeast ribosome biogenesis.
The EMBO journal, 28(24):3808–19, 2009.
- Leidig C, Bange G, Kopp J, Amlacher S, Aravind A, Wickles S, Witte G, Hurt E, Beckmann R, and Sinning I.
Structural characterization of a eukaryotic chaperone-the ribosome-associated complex.
Nature structural & molecular biology, 20(1):23–8, 2013.
- Letunic I, Doerks T, and Bork P.
SMART 7: recent updates to the protein domain annotation resource.
Nucleic acids research, 40(Database issue):D302–5, 2012.
- Levin MK, Gurjar MM, and Patel SS.
ATP binding modulates the nucleic acid affinity of hepatitis C virus helicase.
The Journal of biological chemistry, 278(26):23311–6, 2003.
- Lin ML, Fukukawa C, Park JH, Naito K, Kijima K, Shimo A, Ajiro M, Nishidate T, Nakamura Y, and Katagiri T.
Involvement of G-patch domain containing 2 overexpression in breast carcinogenesis.
Cancer science, 100(8):1443–50, 2009.
- Liu YC, Chen HC, Wu NY, and Cheng SC.
A novel splicing factor, Yju2, is associated with NTC and acts after Prp2 in promoting the first catalytic reaction of pre-mRNA splicing.
Molecular and cellular biology, 27(15):5403–13, 2007.
- Lopez PJ and Séraphin B.
Genomic-scale quantitative analysis of yeast pre-mRNA splicing: implications for splice-site recognition.
RNA, 5(9):1135–7, 1999.
- Lorsch JR and Herschlag D.
The DEAD box protein eIF4A. 1. A minimal kinetic and thermodynamic framework reveals coupled binding of RNA and nucleotide.
Biochemistry, 37(8):2180–93, 1998.

Luo P and Baldwin RL.

Mechanism of helix induction by trifluoroethanol: a framework for extrapolating the helix-forming properties of peptides from trifluoroethanol/water mixtures back to water.

Biochemistry, 36(27):8413–21, 1997.

Madhani HD and Guthrie C.

A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome.

Cell, 71(5):803–17, 1992.

Mallam AL, Del Campo M, Gilman B, Sidote DJ, and Lambowitz AM.

Structural basis for RNA-duplex recognition and unwinding by the DEAD-box helicase Mss116p.

Nature, 490(7418):121–5, 2012.

Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Lu F, Marchler GH, Mullokandov M, Omelchenko MV, Robertson CL, Song JS, Thanki N, Yamashita RA, Zhang D, Zhang N, Zheng C, and Bryant SH.

CDD: a Conserved Domain Database for the functional annotation of proteins.

Nucleic acids research, 39(Database issue):D225–9, 2011.

Marcia M and Pyle AM.

Visualizing Group II Intron Catalysis through the Stages of Splicing.

Cell, 151(3):497–507, 2012.

Martin A, Schneider S, and Schwer B.

Prp43 is an essential RNA-dependent ATPase required for release of lariatintron from the spliceosome.

The Journal of biological chemistry, 277(20):17743–50, 2002.

Mathew R, Hartmuth K, Möhlmann S, Urlaub H, Ficner R, and Lührmann R.

Phosphorylation of human PRP28 by SRPK2 is required for integration of the U4/U6-U5 tri-snRNP into the spliceosome.

Nature structural & molecular biology, 15(5):435–43, 2008.

Bibliography

Mattaj IW.

Cap trimethylation of U snRNA is cytoplasmic and dependent on U snRNP protein binding.

Cell, 46(6):905–11, 1986.

Mayas RM, Maita H, Semlow DR, and Staley JP.

Spliceosome discards intermediates via the DEAH box ATPase Prp43p.

Proceedings of the National Academy of Sciences of the United States of America, 107(22):10020–5, 2010.

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, and Read RJ.

Phaser crystallographic software.

Journal of applied crystallography, 40(4):658–674, 2007.

McPheeters DS, Schwer B, and Muhlenkamp P.

Interaction of the yeast DExH-box RNA helicase prp22p with the 3' splice site during the second step of nuclear pre-mRNA splicing.

Nucleic acids research, 28(6):1313–21, 2000.

Monecke T, Haselbach D, Voß B, Russek A, Neumann P, Thomson E, Hurt E, Zachariae U, Stark H, Grubmüller H, Dickmanns A, and Ficner R.

Structural basis for cooperativity of CRM1 export complex formation.

Proceedings of the National Academy of Sciences of the United States of America, 110(3):960–5, 2013.

Montpetit B, Thomsen ND, Helmke KJ, Seeliger M, Berger JM, and Weis K.

A conserved mechanism of DEAD-box ATPase activation by nucleoporins and InsP6 in mRNA export.

Nature, 472(7342):238–42, 2011.

Moore MJ and Proudfoot NJ.

Pre-mRNA processing reaches back to transcription and ahead to translation.

Cell, 136(4):688–700, 2009.

Moore MJ and Sharp PA.

Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing.

Nature, 365(6444):364–8, 1993.

- Mouaikel J, Verheggen C, Bertrand E, Tazi J, and Bordonné R.
Hypermethylation of the cap structure of both yeast snRNAs and snoRNAs requires a conserved methyltransferase that is localized to the nucleolus.
Molecular cell, 9(4):891–901, 2002.
- Mozaffari-Jovin S, Santos K, Hsiao HH, Will CL, Urlaub H, Wahl MC, and Lührmann R.
The Prp8 RNase H-like domain inhibits Brr2-mediated U4/U6 snRNA unwinding by blocking Brr2 loading onto the U4 snRNA.
Genes & development, 26(21):2422–34, 2012.
- Murzin AG.
OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences.
The EMBO journal, 12(3):861–7, 1993.
- Myong S, Bruno MM, Pyle AM, and Ha T.
Spring-loaded mechanism of DNA unwinding by hepatitis C virus NS3 helicase.
Science, 317(5837):513–6, 2007.
- Newman AJ and Norman C.
U5 snRNA interacts with exon sequences at 5' and 3' splice sites.
Cell, 68(4):743–54, 1992.
- Newman J, Egan D, Walter TS, Meged R, Berry I, Ben Jelloul M, Sussman JL, Stuart DI, and Perrakis A.
Towards rationalization of crystallization screening for small- to medium-sized academic laboratories: the PACT/JCSG+ strategy.
Acta crystallographica. Section D, Biological crystallography, 61(Pt 10):1426–31, 2005.
- Nielsen KH, Chamieh H, Andersen CBF, Fredslund F, Hamborg K, Le Hir H, and Andersen GR.
Mechanism of ATP turnover inhibition in the EJC.
RNA, 15(1):67–75, 2009.
- Niu Z, Jin W, Zhang L, and Li X.
Tumor suppressor RBM5 directly interacts with the DExD/H-box protein

Bibliography

- DHX15 and stimulates its helicase activity.
FEBS letters, 586(7):977–83, 2012.
- Oberer M, Marintchev A, and Wagner G.
Structural basis for the enhancement of eIF4A helicase activity by eIF4G.
Genes & development, 19(18):2212–23, 2005.
- O’Day CL, Dalbadie-McFarland G, and Abelson J.
The *Saccharomyces cerevisiae* Prp5 protein has RNA-dependent ATPase activity with specificity for U2 small nuclear RNA.
The Journal of biological chemistry, 271(52):33261–7, 1996.
- Ogden RC, Lee MC, and Knapp G.
Transfer RNA splicing in *Saccharomyces cerevisiae*: defining the substrates.
Nucleic acids research, 12(24):9367–82, 1984.
- Ohr T, Prior M, Dannenberg J, Odenwalder P, Dybkov O, Rasche N, Schmitzova J, Gregor I, Fabrizio P, Enderlein J, and Luhrmann R.
Prp2-mediated protein rearrangements at the catalytic core of the spliceosome as revealed by dcFCCS.
RNA, 18(6):1244–56, 2012.
- Okada-Katsuhata Y, Yamashita A, Kutsuzawa K, Izumi N, Hirahara F, and Ohno S.
N- and C-terminal Upf1 phosphorylations create binding platforms for SMG-6 and SMG-5:SMG-7 during NMD.
Nucleic acids research, 40(3):1251–66, 2012.
- Padgett RA, Konarska MM, Grabowski PJ, Hardy SF, and Sharp PA.
Lariat RNA’s as intermediates and products in the splicing of messenger RNA precursors.
Science, 225(4665):898–903, 1984.
- Pandit S, Lynn B, and Rymond BC.
Inhibition of a spliceosome turnover pathway suppresses splicing defects.
Proceedings of the National Academy of Sciences of the United States of America, 103(37):13700–5, 2006.

- Pandit S, Paul S, Zhang L, Chen M, Durbin N, Harrison SMW, and Rymond BC.
Spp382p interacts with multiple yeast splicing factors, including possible regulators of Prp43 DExD/H-Box protein function.
Genetics, 183(1):195–206, 2009.
- Patel SB and Bellini M.
The assembly of a spliceosomal small nuclear ribonucleoprotein particle.
Nucleic acids research, 36(20):6482–93, 2008.
- Paule MR and White RJ.
Survey and summary: transcription by RNA polymerases I and III.
Nucleic acids research, 28(6):1283–98, 2000.
- Pause A and Sonenberg N.
Mutational analysis of a DEAD box RNA helicase: the mammalian translation initiation factor eIF-4A.
The EMBO journal, 11(7):2643–54, 1992.
- Pause A, Méthot N, and Sonenberg N.
The HRIGRXXXR region of the DEAD box RNA helicase eukaryotic translation initiation factor 4A is required for RNA binding and ATP hydrolysis.
Molecular and cellular biology, 13(11):6789–98, 1993.
- Peebles CL, Perlman PS, Mecklenburg KL, Petrillo ML, Tabor JH, Jarrell KA, and Cheng HL.
A self-splicing RNA excises an intron lariat.
Cell, 44(2):213–23, 1986.
- Pena V, Rozov A, Fabrizio P, Lührmann R, and Wahl MC.
Structure and function of an RNase H domain at the heart of the spliceosome.
The EMBO journal, 27(21):2929–40, 2008.
- Pertschy B, Schneider C, Gnädig M, Schäfer T, Tollervey D, and Hurt E.
RNA helicase Prp43 and its co-factor Pfa1 promote 20 to 18 S rRNA processing catalyzed by the endonuclease Nob1.
The Journal of biological chemistry, 284(50):35079–91, 2009.
- Pyle AM.
Translocation and unwinding mechanisms of RNA and DNA helicases.
Annual review of biophysics, 37:317–36, 2008.

Bibliography

Query CC and Konarska MM.

Suppression of multiple substrate mutations by spliceosomal prp8 alleles suggests functional correlations with ribosomal ambiguity mutants.

Molecular cell, 14(3):343–54, 2004.

Radaev S, Li S, and Sun PD.

A survey of protein-protein complex crystallizations.

Acta crystallographica. Section D, Biological crystallography, 62(Pt 6):605–12, 2006.

Raghunathan PL and Guthrie C.

RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2.

Current biology, 8(15):847–55, 1998.

Raker V, Plessel G, and Lührmann R.

The snRNP core assembly pathway: identification of stable core protein heteromeric complexes and an snRNP subcore particle in vitro.

The EMBO journal, 15(9):2256–69, 1996.

Ramanagoudr-Bhojappa R, Blair LP, Tackett AJ, and Raney KD.

Physical and functional interaction between yeast Pif1 helicase and Rim1 single-stranded DNA binding protein.

Nucleic acids research, 41(2):1029–46, 2013.

Rice P, Longden I, and Bleasby A.

EMBOSS: the European Molecular Biology Open Software Suite.

Trends in genetics, 16(6):276–7, 2000.

Richter FM, Hsiao HH, Plessmann U, and Urlaub H.

Enrichment of protein-RNA crosslinks from crude UV-irradiated mixtures for MS analysis by on-line chromatography using titanium dioxide columns.

Biopolymers, 91(4):297–309, 2009.

Ritchie DB, Schellenberg MJ, Gesner EM, Raithatha SA, Stuart DT, and Macmillan AM.

Structural elucidation of a PRP8 core domain from the heart of the spliceosome.

Nature structural & molecular biology, 15(11):1199–205, 2008.

Rocak S and Linder P.

DEAD-box proteins: the driving forces behind RNA metabolism.
Nature reviews. Molecular cell biology, 5(3):232–41, 2004.

Rudolph MG, Heissmann R, Wittmann JG, and Klostermeier D.

Crystal structure and nucleotide binding of the *Thermus thermophilus* RNA helicase Hera N-terminal domain.
Journal of molecular biology, 361(4):731–43, 2006.

Ruskin B, Krainer R, Maniatis T, and Green MR.

Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro.
Cell, 38(1):317–31, 1984.

Rymond BC and Rosbash M.

Cleavage of 5' splice site and lariat formation are independent of 3' splice site in yeast mRNA splicing.
Nature, 317(6039):735–7, 1985.

Santos K, Jovin SM, Weber G, Pena V, Lührmann R, and Wahl MC.

Structural basis for functional cooperation between tandem helicase cassettes in Brr2-mediated remodeling of the spliceosome.
Proceedings of the National Academy of Sciences of the United States of America, 109(43):17418–23, 2012.

Schneider M, Will CL, Anokhina M, Tazi J, Urlaub H, and Lührmann R.

Exon definition complexes contain the tri-snRNP and can be directly converted into B-like pre-catalytic splicing complexes.
Molecular cell, 38(2):223–35, 2010.

Schneider S and Schwer B.

Functional domains of the yeast splicing factor Prp22p.
The Journal of biological chemistry, 276(24):21184–91, 2001.

Schütz P, Bumann M, Oberholzer AE, Bieniossek C, Trachsel H, Altmann M, and Baumann U.

Crystal structure of the yeast eIF4A-eIF4G complex: an RNA-helicase controlled by protein-protein interactions.

Bibliography

Proceedings of the National Academy of Sciences of the United States of America, 105(28):9564–9, 2008.

Schwer B.

A new twist on RNA helicases: DExH/D box proteins as RNPsases.
Nature structural biology, 8(2):113–6, 2001.

Schwer B.

A conformational rearrangement in the spliceosome sets the stage for Prp22-dependent mRNA release.
Molecular cell, 30(6):743–54, 2008.

Schwer B and Gross CH.

Prp22, a DExH-box RNA helicase, plays two distinct roles in yeast pre-mRNA splicing.
The EMBO journal, 17(7):2086–94, 1998.

Semlow DR and Staley JP.

Staying on message: ensuring fidelity in pre-mRNA splicing.
Trends in biochemical sciences, 37(7):263–73, 2012.

Sengoku T, Nureki O, Nakamura A, Kobayashi S, and Yokoyama S.

Structural basis for RNA unwinding by the DEAD-box protein Drosophila Vasa.
Cell, 125(2):287–300, 2006.

Sharma S, Kohlstaedt LA, Damianov A, Rio DC, and Black DL.

Polypyrimidine tract binding protein controls the transition from exon definition to an intron defined spliceosome.
Nature structural & molecular biology, 15(2):183–91, 2008.

Shen H.

UAP56- a key player with surprisingly diverse roles in pre-mRNA splicing and nuclear export.
BMB reports, 42(4):185–8, 2009.

Shi H, Cordin O, Minder CM, Linder P, and Xu RM.

Crystal structure of the human ATP-dependent splicing and export factor UAP56.
Proceedings of the National Academy of Sciences of the United States of America, 101(51):17628–33, 2004.

- Silverman E, Maeda A, Wei J, Smith P, Beggs JD, and Lin RJ.
Interaction between a G-patch protein and a spliceosomal DEXD/H-box ATPase that is critical for splicing.
Molecular and cellular biology, 24(23):10101–10, 2004.
- Singh R and Reddy R.
Gamma-monomethyl phosphate: a cap structure in spliceosomal U6 small nuclear RNA.
Proceedings of the National Academy of Sciences of the United States of America, 86(21):8280–3, 1989.
- Singleton MR, Dillingham MS, and Wigley DB.
Structure and mechanism of helicases and nucleic acid translocases.
Annual review of biochemistry, 76:23–50, 2007.
- Sinz A.
Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions.
Mass spectrometry reviews, 25(4):663–82, 2006.
- Small EC, Leggett SR, Winans AA, and Staley JP.
The EF-G-like GTPase Snu114p regulates spliceosome dynamics mediated by Brr2p, a DEXD/H box ATPase.
Molecular cell, 23(3):389–99, 2006.
- Smith C and Valcárcel J.
Alternative pre-mRNA splicing: the logic of combinatorial control.
Trends in biochemical sciences, 25(8):381–8, 2000.
- Smith DJ, Query CC, and Konarska MM.
"Nought may endure but mutability": spliceosome dynamics and the regulation of splicing.
Molecular cell, 30(6):657–66, 2008.
- Sontheimer EJ, Sun S, and Piccirilli JA.
Metal ion catalysis during splicing of premessenger RNA.
Nature, 388(6644):801–5, 1997.
- Spiller MP, Boon KL, Reijns MM, and Beggs JD.
The Lsm2-8 complex determines nuclear localization of the spliceosomal U6

Bibliography

snRNA.

Nucleic acids research, 35(3):923–9, 2007.

Staley JP and Guthrie C.

Mechanical devices of the spliceosome: motors, clocks, springs, and things.

Cell, 92(3):315–26, 1998.

Story RM, Li H, and Abelson JN.

Crystal structure of a DEAD box protein from the hyperthermophile *Methanococcus jannaschii*.

Proceedings of the National Academy of Sciences of the United States of America, 98(4):1465–70, 2001.

Subramanya HS, Bird LE, Brannigan JA, and Wigley DB.

Crystal structure of a DExx box DNA helicase.

Nature, 384(6607):379–83, 1996.

Sun JS and Manley JL.

A novel U2-U6 snRNA structure is necessary for mammalian mRNA splicing.

Genes & development, 9(7):843–54, 1995.

Svec M, Bauerová H, Pichová I, Konvalinka J, and Stríšovský K.

Proteinases of betaretroviruses bind single-stranded nucleic acids through a novel interaction module, the G-patch.

FEBS letters, 576(1-2):271–6, 2004.

Tanaka N and Schwer B.

Characterization of the NTPase, RNA-binding, and RNA helicase activities of the DEAH-box splicing factor Prp22.

Biochemistry, 44(28):9795–803, 2005.

Tanaka N and Schwer B.

Mutations in PRP43 that uncouple RNA-dependent NTPase activity and pre-mRNA splicing function.

Biochemistry, 45(20):6510–21, 2006.

Tanaka N, Aronova A, and Schwer B.

Ntr1 activates the Prp43 helicase to trigger release of lariat-intron from the spliceosome.

Genes & development, 21(18):2312–25, 2007.

- Tanner NK, Cordin O, Banroques J, Doère M, and Linder P.
The Q motif: a newly identified motif in DEAD box helicases may regulate ATP binding and hydrolysis.
Molecular cell, 11(1):127–38, 2003.
- Tarn WY and Steitz J.
A novel spliceosome containing U11, U12, and U5 snRNPs excises a minor class (AT-AC) intron in vitro.
Cell, 84(5):801–11, 1996.
- Tarun SZ and Sachs AB.
Association of the yeast poly(A) tail binding protein with translation initiation factor eIF-4G.
The EMBO journal, 15(24):7168–77, 1996.
- Tazi J, Durand S, and Jeanteur P.
The spliceosome: a novel multi-faceted target for therapy.
Trends in biochemical sciences, 30(8):469–78, 2005.
- Teigelkamp S, McGarvey M, Plumpton M, and Beggs JD.
The splicing factor PRP2, a putative RNA helicase, interacts directly with pre-mRNA.
The EMBO journal, 13(4):888–97, 1994.
- Teigelkamp S, Newman J, and Beggs JD.
Extensive interactions of PRP8 protein with the 5' and 3' splice sites during splicing suggest a role in stabilization of exon alignment by U5 snRNA.
The EMBO journal, 14(11):2602–12, 1995.
- Theissen B, Karow AR, Köhler J, Gubaev A, and Klostermeier D.
Cooperative binding of ATP and RNA induces a closed conformation in a DEAD box RNA helicase.
Proceedings of the National Academy of Sciences of the United States of America, 105(2):548–53, 2008.
- Theobald DL, Mitton-Fry RM, and Wuttke DS.
Nucleic acid recognition by OB-fold proteins.
Annual review of biophysics and biomolecular structure, 32:115–33, 2003.

Bibliography

- Tompa P, Fuxreiter M, Oldfield CJ, Simon I, Dunker AK, and Uversky VN.
Close encounters of the third kind: disordered domains and the interactions of proteins.
BioEssays : news and reviews in molecular, cellular and developmental biology, 31(3):328–35, 2009.
- Toor N, Keating KS, Taylor SD, and Pyle AM.
Crystal structure of a self-spliced group II intron.
Science, 320(5872):77–82, 2008.
- Toor N, Keating KS, and Pyle AM.
Structural insights into RNA splicing.
Current opinion in structural biology, 19(3):260–6, 2009.
- Trotta CR, Miao F, Arn E, Stevens SW, Ho CK, Rauhut R, and Abelson J.
The yeast tRNA splicing endonuclease: a tetrameric enzyme with two active site subunits homologous to the archaeal tRNA endonucleases.
Cell, 89(6):849–58, 1997.
- Tsai RT, Fu RH, Yeh FL, Tseng CK, Lin YC, Huang YH, and Cheng SC.
Spliceosome disassembly catalyzed by Prp43 and its associated components Ntr1 and Ntr2.
Genes & development, 19(24):2991–3003, 2005.
- Tsai RT, Tseng CK, Lee PJ, Chen HC, Fu RH, Chang KJ, Yeh FL, and Cheng SC.
Dynamic interactions of Ntr1-Ntr2 with Prp43 and with U5 govern the recruitment of Prp43 to mediate spliceosome disassembly.
Molecular and cellular biology, 27(23):8027–37, 2007.
- Tseng CK and Cheng SC.
Both catalytic steps of nuclear pre-mRNA splicing are reversible.
Science, 320(5884):1782–1784, 2008.
- Tseng CK, Liu HL, and Cheng SC.
DEAH-box ATPase Prp16 has dual roles in remodeling of the spliceosome in catalytic steps.
RNA, 17(1):145–54, 2011.

- Turunen JJ, Niemelä EH, Verma B, and Frilander MJ.
The significant other: splicing by the minor spliceosome.
Wiley interdisciplinary reviews. RNA, 4(1):61–76, 2012.
- Umen JG and Guthrie C.
Prp16p, Slu7p, and Prp8p interact with the 3' splice site in two distinct stages during the second catalytic step of pre-mRNA splicing.
RNA, 1(6):584–97, 1995.
- Valadkhan S.
snRNAs as the catalysts of pre-mRNA splicing.
Current opinion in chemical biology, 9(6):603–8, 2005.
- Valcárcel J, Gaur RK, Singh R, and Green MR.
Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA [corrected].
Science, 273(5282):1706–9, 1996.
- Noort V, Bradatsch B, Arumugam M, Amlacher S, Bange G, Creevey C, Falk S, Mende DR, Sinning I, Hurt E, and Bork P.
Consistent mutational paths predict eukaryotic thermostability.
BMC evolutionary biology, 13(1):7, 2013.
- Moeller H, Basquin C, and Conti E.
The mRNA export protein DBP5 binds RNA and the cytoplasmic nucleoporin NUP214 in a mutually exclusive manner.
Nature structural & molecular biology, 16(3):247–54, 2009.
- Wachtel C and Manley JL.
Splicing of mRNA precursors: the role of RNAs and proteins in catalysis.
Molecular bioSystems, 5(4):311–6, 2009.
- Wagner JD, Jankowsky E, Company M, Pyle AM, and Abelson J.
The DEAH-box protein PRP22 is an ATPase that mediates ATP-dependent mRNA release from the spliceosome and unwinds RNA duplexes.
The EMBO journal, 17(10):2926–37, 1998.
- Wahl MC, Will CL, and Lührmann R.
The spliceosome: design principles of a dynamic RNP machine.
Cell, 136(4):701–18, 2009.

Bibliography

Walbott H, Mouffok S, Capeyrou R, Lebaron S, Humbert O, Tilbeurgh H, Henry Y, and Leulliot N.

Prp43p contains a processive helicase structural architecture with a specific regulatory domain.

The EMBO journal, 29(13):2194–204, 2010.

Walker JE, Saraste M, Runswick MJ, and Gay NJ.

Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold.

The EMBO journal, 1(8):945–51, 1982.

Wang F, Yang Y, Singh TR, Busygina V, Guo R, Wan K, Wang W, Sung P, Meetei AR, and Lei M.

Crystal structures of RMI1 and RMI2, two OB-fold regulatory subunits of the BLM complex.

Structure, 18(9):1159–70, 2010.

Wang L and Chance MR.

Structural mass spectrometry of proteins using hydroxyl radical based protein footprinting.

Analytical chemistry, 83(19):7234–41, 2011.

Wang Z and Burge CB.

Splicing regulation: from a parts list of regulatory elements to an integrated splicing code.

RNA, 14(5):802–13, 2008.

Ward AJ and Cooper TA.

The pathobiology of splicing.

The Journal of pathology, 220(2):152–63, 2010.

Warkocki Z, Odenwalder P, Schmitzova J, Platzmann F, Stark H, Urlaub H, Ficner R, Fabrizio P, and Luhrmann R.

Reconstitution of both steps of *Saccharomyces cerevisiae* splicing with purified spliceosomal components.

Nature structural & molecular biology, 16(12):1237–43, 2009.

Webb MR.

A continuous spectrophotometric assay for inorganic phosphate and for measuring phosphate release kinetics in biological systems.

Proceedings of the National Academy of Sciences of the United States of America, 89(11):4884–7, 1992.

Weir JR, Bonneau F, Hentschel J, and Conti E.

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 of the United States of America, 107(27):12139–44, 2010.

Weirich CS, Erzberger JP, Flick JS, Berger JM, Thorner J, and Weis K.

Activation of the DExD/H-box protein Dbp5 by the nuclear-pore protein Gle1 and its coactivator InsP6 is required for mRNA export.

Nature cell biology, 8(7):668–76, 2006.

Will CL and Lührmann R.

Spliceosomal UsnRNP biogenesis, structure and function.

Current opinion in cell biology, 13(3):290–301, 2001.

Will CL and Lührmann R.

Spliceosome structure and function.

Cold Spring Harbor perspectives in biology, 3(7), 2011.

Will CL, Schneider C, Hossbach M, Urlaub H, Rauhut R, Elbashir S, Tuschl T, and Lührmann R.

The human 18S U11/U12 snRNP contains a set of novel proteins not found in the U2-dependent spliceosome.

RNA, 10(6):929–41, 2004.

Wolff T and Bindereif A.

Conformational changes of U6 RNA during the spliceosome cycle: an intramolecular helix is essential both for initiating the U4-U6 interaction and for the first step of splicing.

Genes & development, 7(7B):1377–89, 1993.

Woodman IL and Bolt EL.

Winged helix domains with unknown function in Hel308 and related helicases.

Biochemical Society transactions, 39(1):140–4, 2011.

Bibliography

Xu YZ and Query CC.

Competition between the ATPase Prp5 and branch region-U2 snRNA pairing modulates the fidelity of spliceosome assembly.

Molecular cell, 28(5):838–49, 2007.

Yang K, Zhang L, Xu T, Heroux A, and Zhao R.

Crystal structure of the beta-finger domain of Prp8 reveals analogy to ribosomal proteins.

Proceedings of the National Academy of Sciences of the United States of America, 105(37):13817–22, 2008.

Yang Q, Del Campo M, Lambowitz AM, and Jankowsky E.

DEAD-box proteins unwind duplexes by local strand separation.

Molecular cell, 28(2):253–63, 2007.

Zamore PD and Green MR.

Biochemical characterization of U2 snRNP auxiliary factor: an essential pre-mRNA splicing factor with a novel intranuclear distribution.

The EMBO journal, 10(1):207–14, 1991.

Zhang Y, Stec B, and Godzik A.

Between order and disorder in protein structures: analysis of "dual personality" fragments in proteins.

Structure, 15(9):1141–7, 2007.

Zhao R, Shen J, Green MR, MacMorris M, and Blumenthal T.

Crystal structure of UAP56, a DExD/H-box protein involved in pre-mRNA splicing and mRNA export.

Structure, 12(8):1373–81, 2004.

Zhu B, Cai G, Hall EO, and Freeman GJ.

In-fusion assembly: seamless engineering of multidomain fusion proteins, modular vectors, and mutations.

BioTechniques, 43(3):354–9, 2007.

Acknowledgements

Foremost, I am very grateful to Prof. Dr. Ralf Ficner for his substantial guidance and support during the past four years as well as for his permanent help and encouragement.

I want to acknowledge Prof. Dr. Reinhard Lührmann and Prof. Dr. Holger Stark for the helpful discussions and for sharing their expertise during the thesis committee meetings.

Furthermore, I want to acknowledge Prof. Dr. Henning Urlaub and Romina Hofele for the productive collaboration regarding the mass-spectrometric analysis of the yPrp43/yNtr1 complexes. I am also grateful to Prof. Dr. Reinhard Lührmann and Dr. Jean-Baptiste Fourmann for the fruitful collaboration and for testing different samples in functional assays. I want to thank PD Dr. Reinhard Rauhut his help during the identification of the ctNTR orthologs and Dr. Bernhard Schmidt for mass spectrometric analysis.

Moreover, I want to acknowledge all former and current members of the Department of Molecular Structural Biology for the friendly and cooperative atmosphere. My special thanks go to Andreas Schmitt and Dr. Yasar Luqman Ahmed for helpful clues and intensive discussions. I am very grateful to Dr. Piotr Neumann for his help during diffraction data collection, structure solution and structure interpretation. Furthermore, I want to acknowledge Dr. Achim Dickmanns, Dr. Thomas Monecke, Dr. Jean-Baptiste Fourmann, Dr. Piotr Neumann, Dr. Yasar Luqman Ahmed, Andreas Schmitt and Stephanie Schell for useful comments on parts of this thesis and Lin-Ta Hsu for creating beautiful figures.

In addition, I want to acknowledge the "Göttingen Graduate School for Neurosciences and Molecular Biology" (GGNB) for the financial support.