Crystallization and biophysical characterization of spliceosomal protein complexes

Dissertation

for the award of the degree

"Doctor rerum naturalium"

of the Georg-August-Universität Göttingen

submitted by

Andreas Schmitt

from Clausthal-Zellerfeld

Göttingen, 2014

Thesis committee:

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

Prof. Dr. Kai Tittmann (Reviewer)
Department for Bioanalytics
Georg-August-Universität Göttingen

Prof. Dr. Reinhard Lührmann
Department of Cellular Biochemistry
Max-Planck-Institute for Biophysical Chemistry, Göttingen

Date of the oral examination: April 15th, 2014

I hereby declare that the PhD thesis "Crystallization and biophysical characterization of spliceosomal protein complexes" 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.

Andreas Schmitt, February 2014

FÜR SARAH (1983 - 2013)

Contents

Contents

Contents

Abstract

mRNA splicing is a highly dynamic process by which non-coding sequences are removed from a pre-mature mRNA. It is catalyzed by the spliceosome, a huge and highly dynamic ribonucleoprotein complex consisting of up to five small nuclear RNAs and more than 100 proteins. For each intron to be excised, the spliceosome has to assemble on the pre-mRNA, catalyze the two steps of the splicing reaction and is disassembled afterwards. The dynamics of the spliceosome are driven and controlled by at least eight conserved DExD/H-box ATPases which remodel RNA-RNA, protein-RNA and protein-protein interactions during assembly, catalysis and disassembly of the spliceosome and are furthermore involved in quality control.

The DEAH-box ATPase Prp2 plays a key role in the activation of the spliceosome, as it promotes the transition of the spliceosome from the Bact to the catalytically active B* complex. Prp2 exhibits an RNA-dependent NTPase activity and is active in the spliceosome after binding to its interaction partner, the G-patch conteining protein Spp2. The G-patch sequence motif consists of a conserved glycine pattern and is mostly found in proteins involved in RNA processing. So far, the structure of a G-patch motif has been unknown and the function of the G-patch motif in general and in particular of Spp2 remained elusive.

Here, the crystal structure of Spp2 from *Chaetomium thermophilum* in complex with Prp2 is shown, representing the first crystal structure of a G-patch protein. Free Spp2 is intrinsically disordered as demonstrated by CD- and NMR-spectroscopy, but at least its G-patch region adopts a well-defined fold upon binding to Prp2. The interaction with Spp2 induces major structural changes in the β-hairpin of Prp2, a region which is supposed to move upon ATP binding. However, Spp2 does not affect the ATPase activity or RNA affinity of Prp2 in vitro. Together, these findings give rise to the hypothesis that Spp2 might fulfill the function of an adaptor protein recruiting Prp2 to the spliceosome and transferring mechanical forces generated by ATP hydrolysis to other components of the spliceosome.

1 Introduction

In the eukaryotic cell, the major part of the genetic information is stored as chromosomal DNA, located in the nucleus. Upon transcription, the DNA-sequence is transcribed into precursor messenger RNA (pre-mRNA), which has to be processed before it is exported into the cytoplasm were translation into a polypeptide by ribosomes takes place (Moore and Proudfoot, 2009; Hocine *et al.*, 2010). Processing steps include the addition of an m^7Guanosin cap at the 5' end, protecting the mRNA against 5'-3' exonucleases and serving as recognition motif for proteins involved in translation or RNA export (Drummond *et al.*, 1985; Hsu and Stevens, 1993). At the 3' end of the pre-mRNA a poly(A)-tail is added which increases stability and plays an important role for translation efficiency (Tarun and Sachs, 1996).

A further necessary step is the removal of non-coding intervening sequences (introns) from the pre-mRNA which is achieved by excision of the intron, followed by ligation of the coding sequences (exons) in a process termed splicing. The process of splicing is catalyzed by a highly dynamic multi-megadalton ribonucleoprotein (RNP) complex, the spliceosome (Wahl *et al.*, 2009; Will and Lührmann, 2011). Two different modes of splicing are known today: constitutive splicing, where all exons are present in the processed mRNA and alternative splicing, where some exons are excised during mRNA maturation, alternative 5'splice sites (5'ss) or 3'splice sites (3'ss) are recognized or in rare cases introns are retained. Alternative splicing results in different species of mRNA originating from the same pre-mRNA and is found in higher eukaryotes were it increases the genomic flexibility, as different proteins can be coded by one gene (Ast, 2004; Chen and Manley, 2009).

1.1 Introns

In eukaryotes, four different classes of introns have been described so far. Besides the spliceosomal introns these are the group I & II of self-splicing introns as well

as t-RNA introns.

Among the self-splicing introns, group I introns have been discovered in 1982 and represent the first example of an RNA-catalysis (Zaug and Cech, 1980; Kruger *et al.*, 1982). Group II introns are mainly found in plants, protists and funghi (Peebles *et al.*, 1986) but they are also widely distributed in bacteria (Toro, 2003). Group II self-splicing introns are generally assumed to have diverged from a common ancestor with spliceosomal introns - a theory which is supported by crystal structure analysis of a group II intron from *Oceanobacillus iheyensis* in different states of catalysis (Toor *et al.*, 2008; Dayie and Padgett, 2008; Chan *et al.*, 2012). tRNA introns are present in bacteria, archaea and eukaryotes. While tRNA introns in bacteria are self-splicing (Biniszkiewicz *et al.*, 1994), the splicing reaction in yeast occurs in three steps requiring the activity of an endonuclease, a ligase and a phosphotransferase (Abelson *et al.*, 1998). In *Saccharomyces cerevisiae*, 59 of the 272 tRNA genes are interrupted by introns (Trotta *et al.*, 1997). Within the four classes, the major part represents spliceosomal introns which are processed by the spliceosome. Spliceosomal introns contain several characteristic sites, namely the 5'ss, the 3'ss, and the branch site (bs) located 18-40 nt upstream of the 3'splice site, which are recognized by the spliceosome via highly conserved sequence motifs at these positions (Will and Lührmann, 2011). Additional *cis*-acting elements on the pre-mRNA such as exonic or intronic splicing silencers (ESSs/ISSs) as well as exonic or intronic splicing enhancers (ESEs/ISEs) are interacting with *trans*-acting factors of the spliceosome and thereby inhibit or strengthen splice site recognition or spliceosome assembly (Burgess and Guthrie, 1993; Smith and Valcárcel, 2000). In yeast, only 3.8% of the genes contain introns, 70% of them coding for ribosomal proteins (Lopez and Séraphin, 1999). In contrast, 92% of mammalian genes contain introns (Fedorova and Fedorov, 2005).

1.2 The splicing reaction

For excision of an intron, two consecutive S_N2-type transesterification reactions have to occur (Figure 1.1). These reactions take place at the conserved 5'- and 3'ss as well as on the branch point and require no additional energy in form of ATP (Moore and Sharp, 1993).

During the first reaction, the 2'-hydroxyl group of the branch site adenosine performs a nucleophilic attack on the esterbond of the 5'ss phosphate. This results in a free 5'-exon with a free 3'-hydroxyl group and a lariat intermediate formed by a

2'-5' phosphodiester bond between the branch point adenosine and the phosphate of the 5'ss. In the second step, the 3'-hydroxyl group of the 5'-exon attacks the phosphodiester bond of the 3'ss in a nucleophilic manner resulting in ligation of the two exons by a new 3'-5' phosphodiester bond and excision of the intron-lariat (Ruskin *et al.*, 1984; Padgett *et al.*, 1984).

Excision of spliceosomal introns from pre-mRNA does not occur on its own, but requires the activity of a macromolecular machinery, the spliceosome.
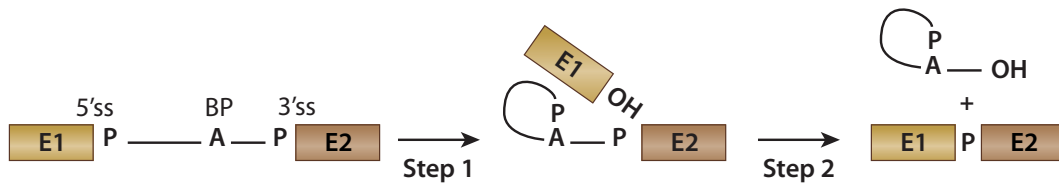
Figure 1.1: Schematic representation of the two transesterification reactions required for pre-mRNA splicing. Exons and introns are represented by boxes and lines respectively. The branchpoint (BP) adenosine is indicated by a capital A, along with the phosphates (P) at the conserved 5' and 3' splice sites (5'ss/3'ss). Figure adapted from Will and Lührmann (2011).

1.3 The spliceosome

Recognition of the splice sites of pre-mRNA and catalysis of the two transesterification reactions is performed by the spliceosome.

The spliceosome consists of up to 5 uridine-rich small nuclear RNAs (U snRNAs) and numerous associated proteins. The whole complex is assembled stepwise directly on the pre-mRNA with tremendous changes in its composition for every step. After reaching a state of catalytical activity, the spliceosome catalyzes the two splicing reactions and is subsequently disassembled. Since the spliceosome is a single-turnover machinery, the processes of assembly and disassembly have to occur for every spliced intron. (Figure 1.2

To date, two different types of spliceosomes are known up to now, the major (or U2 dependent) and the minor (or U12 dependent) spliceosome.

The major spliceosome catalyzes splicing of most pre-mRNA introns and is formed by the U1, U2, U4/U5 and U6 small nuclear ribonucleoprotein particles (snRNPs) (Will and Lührmann, 2001) as well as several associated non-splicosomal proteins (Wahl *et al.*, 2009). Each snRNP is composed of an uridine-rich RNA with a length between 100 and 200 bases and the seven Sm-proteins B/B', D1, D2, D3,

E, F and G which form a heptameric ring (Kambach *et al.*, 1999; Weber *et al.*, 2010). Additionally, each snRNP is associated with a set of several specific proteins (Wahl *et al.*, 2009).

The minor spliceosome, which in metazoan and plants excise a rare class of introns, is composed of the U11, U12, U4atac/U6atac and U5 snRNP (Tarn and Steitz, 1996; Russell *et al.*, 2006).
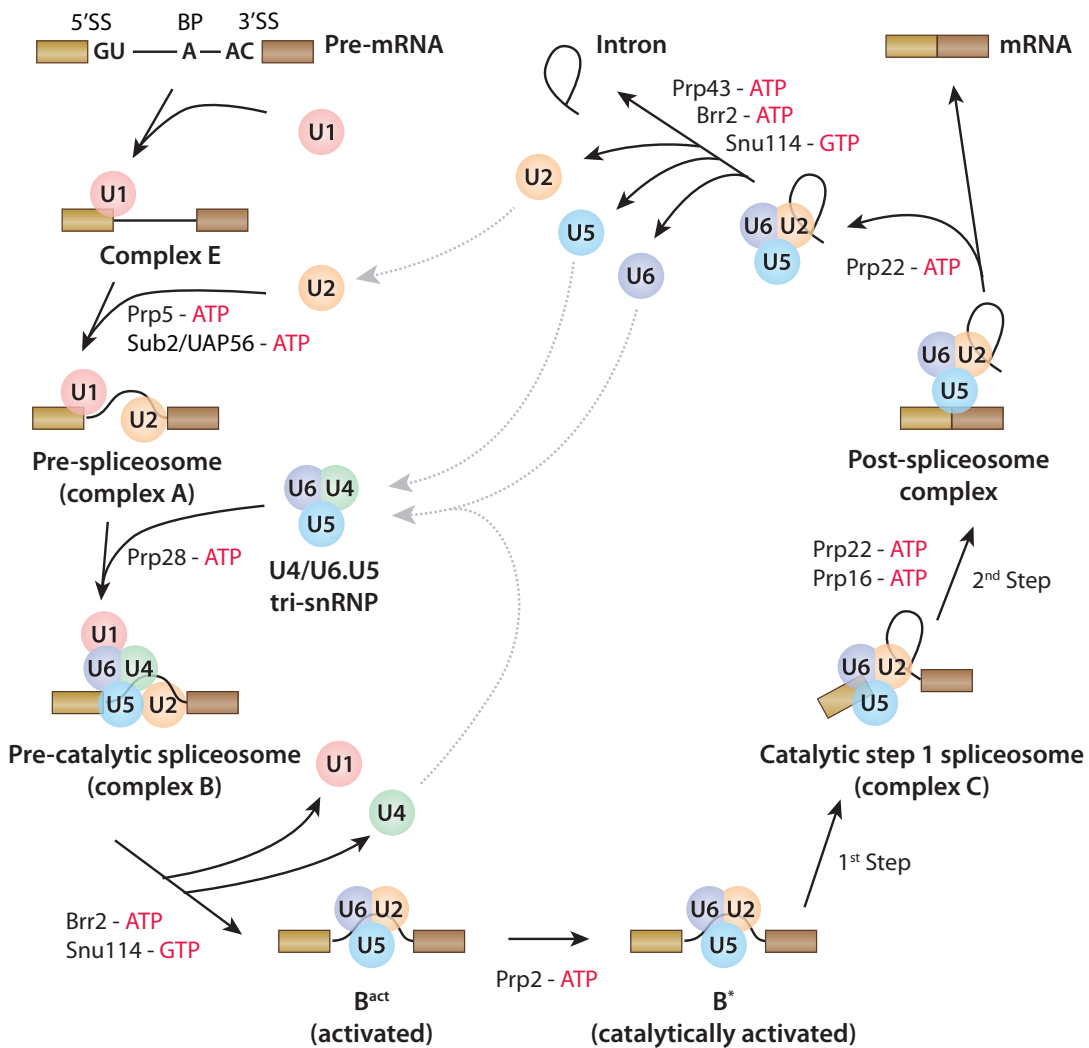
Figure 1.2: Schematic representation of assembly and disassembly steps of the U2 dependent spliceosome. The different complexes are named according to the metazoan nomenclature. The snRNPs are shown as circles, the conserved DExD/H-box proteins which facilitate conformational changes are indicated next to the arrows. Figure adapted from Will and Lührmann (2011).

1.4 The splicing cycle

The different steps of the splicing cycle can be grouped into three major parts, namely the assembly of the spliceosome, catalysis of the two transesterification reactions and disassembly of the intron-lariat spliceosome.

1.4.1 Assembly of the spliceosome

The assembly process of the spliceosome differs significantly depending on the length of the intron to be excised. If the intron does not exceed 200-250 nts the spliceosome assembles across the intron, starting with formation of complex E. In this complex, the U1 snRNP is recruited to the 5'ss and additional sequence motifs are recognized by non-snRNP factors: SF1/mBBP interacts with the branch site whereas the U2 auxiliary factor (U2AF) binds to the polypyrimidine tract (PPT) downstream of the bs (Berglund *et al.*, 1997; Fox-Walsh *et al.*, 2005).

In contrast, if an intron exceeds 200-250 nts, there are alternative ways of spliceosomal assembly including the formation of splicing complexes across an exon in a process called exon-definintion (Fox-Walsh *et al.*, 2005). This process is predominantly found in mammals whose pre-mRNAs contain multiple introns with sizes between several hundred and several thousand nucleotides (Deutsch and Long, 1999; Xiao *et al.*, 2007). Initiation of exon definition is characterized by binding of the U1 snRNP to the 5'ss downstream of an exon followed by the association of U2AF to the PPT. These events lead to the recruitment of the U2 snRNP to the branch site. The exon defined complex is then stabilized by proteins of the SR protein family, which are recruited by ESEs (Hoffman and Grabowski, 1992; Reed, 2000; Das *et al.*, 2000).

For the chemical splicing steps, the exon-defined complex has to be switched to an intron-defined splicing complex. Interaction of the U2 and U1 snRNPs bound to the 5'ss lead to formation of the cross-intron A complex (Smith and Valcárcel, 2000). The whole process is poorly understood so far and there are indications that an exon-defined complex might also be directly converted into a cross-intron B complex giving rise to the assumption that multiple pathways exist (Schneider *et al.*, 2010).

Formation of complex E is followed by binding of the U2 snRNP proteins SF3a/SF3b next to the branch site (Gozani *et al.*, 1996) and by U2AF65 which directly interacts with the bs (Valcárcel *et al.*, 1996). Formation of complex A is completed

by the ATP dependent activity of two DEAD-box proteins, Prp5 and Sub2. Prp5 acts on the U2 snRNP itself modulating intron selectivity (Xu and Query, 2007), whereas Sub2 promotes the removal of U2AF65 thus allowing stable interaction of the U2 snRNP with the pre-mRNA (Kistler and Guthrie, 2001).

In a subsequent step, the pre-assembled tri-snRNP consisting of the U4/U6 and U5 snRNPs is recruited to the spliceosome, leading to formation of the catalytically inactive B complex. In the following ATP-dependent steps, major rearrangements of RNA-RNA and RNA-protein interactions occur and give rise to the activated B^{act} complex which nevertheless is catalytically inactive. The DEAD-box helicase Prp28 catalyzes the release of the U1 snRNP (Staley and Guthrie, 1999) while the Ski2-like ATPase Brr2 disrupts the U4/U6 RNA base-pairing (Raghunathan and Guthrie, 1998), leading to U4 snRNP release. The 5' end of the U6 snRNA can now base-pair with the 5'ss and thus replaces U1 and U4 (Konforti *et al.*, 1993) while the U6 snRNP also interacts with the U2 snRNP, positioning the 5'ss near the branch site as a prerequisite for the first step of splicing (Madhani and Guthrie, 1992; Sun and Manley, 1995).

1.4.2 Catalysis of the splicing reaction

The transition into the catalytically active B* complex is promoted by the DEAH-box helicase Prp2 along with its cofactor Spp2 (Silverman *et al.*, 2004). It has been shown, that ATP-hydrolysis by Prp2 leads to a remodeling of the spliceosome, along with the destabilization of the U2 snRNP proteins SF3a and SF3b which bind near the branch site. As a consequence, the branch site adenosine becomes available for the subsequent nucleophile attack (Warkocki *et al.*, 2009). Moreover it was demonstrated, that the U2 proteins stay associated with the spliceosome at physiologic salt concentrations and that remodeling of the spliceosome by Prp2 creates a high affinity binding site for Cwc25 (Ohrt *et al.*, 2012). The activated spliceosome subsequently catalyzes the first reaction step, generating the C complex.

Prior to the second catalytic step, the DEAH-box ATPase Prp16 induces the release of the two components Yju2 and Cwc25 (Tseng *et al.*, 2011) by its interaction with an intron sequence on the pre-mRNA near the 3'ss (Umen and Guthrie, 1995). The splicing factors Slu7, Prp18 and Prp22 then identify the 3'ss prior to the second reaction step (Horowitz and Abelson, 1993; McPheeters *et al.*, 2000),

while the 3'ss seems to be dispensable for the first reaction step (Rymond and Rosbash, 1985). Alignment of the 3'- and 5'-exon for ligation is mainly mediated by the U5 snRNP whose associated protein Prp8 is suggested to stabilize the alignment by extensive interactions with the pre-mRNA (Teigelkamp *et al.*, 1995).

1.4.3 Disassembly of the spliceosome

The release of the mature mRNA from the spliceosome is induced by the DEAH-box helicase Prp22 which binds to the 3'ss of the pre-mRNA during the second transesterification reaction (McPheeters *et al.*, 2000). After ligation of the exons, it translocates in 3'-5' direction on the mRNA and thereby it is thought to disrupt the U5-exon base pairing (Schwer, 2008). In contrast to the action of Prp22 during the second reaction step, Prp22 activity during the disassembly is ATP dependent (Wagner *et al.*, 1998). In a subsequent ATP dependent step, the intron-lariat spliceosome is disassembled by the DEAH-box ATPase Prp43, which does not only include release of the intron-lariat but also separation of the U2, U5 and U6 snRNPs (Arenas and Abelson, 1997; Martin *et al.*, 2002). Prp43 acts in the so called NTC-related (NTR) complex, which also contains the two Prp43 interacting proteins Ntr1 and Ntr2 (Tsai *et al.*, 2005; Boon *et al.*, 2006). Moreover, it was shown, that the GTPase Snu144 as well as the Ski2-like helicase Brr2 are required for the disassembly process (Small *et al.*, 2006).

1.5 Components of the spliceosome

1.5.1 Biogenesis and assembly of UsnRNPs

All UsnRNAs with the exception of U6 snRNA are transcribed by RNA polymerase II as precursor snRNA in the nucleus. After co-transcription of an m^{7}G cap, they are exported to the cytoplasm by the export receptor CRM1, RanGTP, CBP20, CBP80 and the adaptor protein PHAX (Ohno *et al.*, 2000; Cougot *et al.*, 2004), where seven Sm proteins (B/B', D3, D2, D1, E, F and G) assemble as subcomplexes mediated by the SMN complex on a highly conserved so called Sm-site of the UsnRNA to a ring-like structure (Raker *et al.*, 1996). Upon Sm-core assembly, the m^{7}G cap of the snRNA is hypermethylated by the dimethyltransferase

Tgs1 to an m_3G-cap (Mouaikel *et al.*, 2002). The m_3G cap together with the Sm-core domain form a nuclear localization signal (NLS), which leads to re-import of the snRNP-core complex into the nucleus where binding of snRNP-specific proteins like 70k, A & C for the U1 snRNP occurs (see figure 1.3) (Will and Lührmann, 2001). Moreover, several modifications are introduced into the UsnR-NAs, including 2'-O-methylation and pseudouridylation (Jády and Kiss, 2001). In contrast, biogenesis of the U6 snRNP differs from the other UsnRNPs in several steps. The U6 snRNA is transcribed by RNA polymerase III (Paule and White, 2000) and does not acquire an m^7G cap but contains a unique gamma-monomethyl cap (Singh and Reddy, 1989). Moreover, the U6 snRNA lacks a canonical Sm-site, and therefore, assembly of the Sm-like proteins occurs at the 3'-end of the U6 snRNA (Achsel *et al.*, 1999).
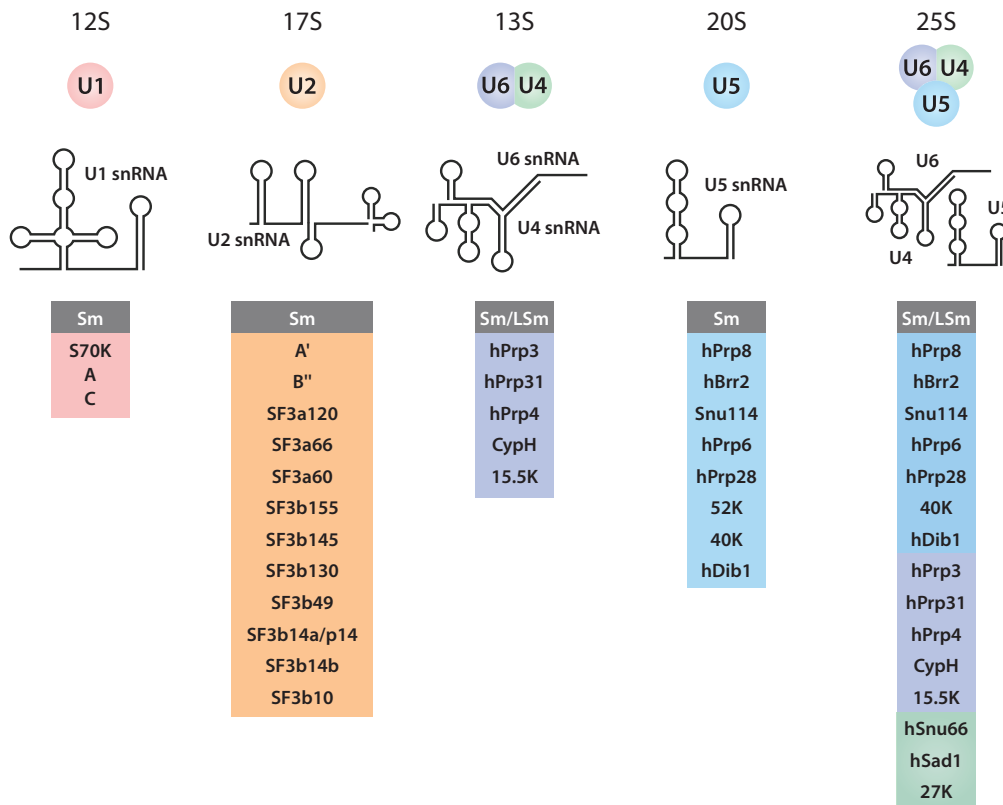

Figure 1.3: snRNA secondary structure and protein composition of the human snRNPs. The seven Sm or LSm proteins are indicated in grey at the top of the boxes listing the proteins which are associated with the individual snRNPs. The size of each snRNP is given in Svedberg units. (Figure adapted from Will and Lührmann (2011))

1.5.2 Additional protein factors of the spliceosome

In addition to the integrative U snRNP proteins, between 90 and 170 proteins associate with the spliceosome during the distinct stages (Fabrizio *et al.*, 2009). The number significantly differs between metazoans with 170 to 190 proteins and yeast, where only about 90 different proteins could be identified. For more than 85% of these, homologues could be identified in the human spliceosome, showing the conservation of the spliceosomal core design (Bessonov *et al.*, 2008; Fabrizio *et al.*, 2009). Moreover, protein factors which have been shown to be absent in yeast include SR and hnRNP proteins that are involved in regulation of alternative splicing, a process which is not observed in yeast (Fabrizio *et al.*, 2009).

The dynamics of the spliceosome protein composition has been studied on *in vitro* assembled B, B^{act} and C complexes, and the associated proteins were identified by mass spectrometry. These studies revealed, that during transition from the B to the B^{act} complex, along with the two U snRNPs U1 and U4 about 35 proteins are released from the spliceosome while 12 are recruited. Smaller changes were observed during transition from the B^{act} to the C complex, where only two proteins dissociated from the spliceosome whereas nine were recruited (Fabrizio *et al.*, 2009).

Several mechanisms have been described to regulate the dynamics of the splicing machinery. Spliceosome associated proteins are post-translationally modified by phosphorylation, methylation, hydroxylation ubiquitination and acetylation (Will and Lührmann, 2011; Korneta and Bujnicki, 2012). Remodeling of RNA/RNA, RNA/protein or protein/protein interactions is driven in an ATP dependent manner by eight helicases of the DExH/D-box family which are conserved throughout yeast and human.

1.5.3 The catalytic center

Since the splicing reaction consists of two subsequent transesterification reactions, it was originally proposed that the spliceosome has two active sites (Moore and Sharp, 1993). However, more recent studies are pointing towards a single catalytic center which changes its conformation for both reactions with the equilibrium of both states modulated by the protein factors Prp16, Isy1 and Prp8 as well as by the U6 snRNA (Konarska and Query, 2005). This hypothesis is supported by the finding, that transition between the two phases occurs in several steps and that the two splicing reactions are reversible (Liu *et al.*, 2007; Smith *et al.*, 2008;

Tseng and Cheng, 2008). Although the spliceosome contains a large number of different proteins, the splicing reaction itself is most likely catalyzed solely by the U2 and U6 snRNA, while the protein factors might act as a scaffold for the active site RNAs (Valadkhan, 2005; Wachtel and Manley, 2009; Fica *et al.*, 2013). The intramolecular stem-loop of the U6 snRNA seems to play an important role for both splicing reactions, by positioning two metal ions which are important for branching as well as for exon ligation (Yean *et al.*, 2000; Fica *et al.*, 2013). Participation in in the splicing reaction was proposed for the protein Prp8 based on crystal structures and mutation studies (Pena *et al.*, 2008; Yang *et al.*, 2008; Abelson, 2008). However, recent results suggest that Prp8 may promote formation of the catalytic core rather than being involved in the reaction itself (Fica *et al.*, 2013).

1.6 RNA helicases

RNA helicases have been found in all prokaryotic and eukaryotic organisms as well as in viruses. They are involved in various cellular processes such as translation initiation, ribosome biogenesis as well as RNA transport and splicing (Cordin *et al.*, 2006; Bleichert and Baserga, 2007). While the term RNA helicase was initially used for enzymes that couple the hydrolysis of NTPs with the unwinding of double-stranded RNA, it became clear nowadays, that they can also displace protein from RNA without RNA strand separation, anneal RNA strands or act as RNA clamps to stabilize folding intermediates (Bleichert and Baserga, 2007)

1.6.1 Helicase classification

Based on structural and mechanistic features as well as conserved sequence motifs, RNA and DNA helicases are divided into six superfamilies (SF) (Fairman-Williams *et al.*, 2010). Within these families, SF3 to SF6 comprise the hexameric or toroidal ring-forming helicases, while helicases belonging to SF1 and SF2 do not form rings (Jankowsky and Fairman, 2007). Sequence alignments of all SF1 and SF2 proteins from *H. sapiens, S. cerevisiae* and *E. coli* as well as several viral helicases revealed, that the SF1 proteins can be grouped into three families, each consisting of three or more proteins from each organism. For the SF2 proteins, nine families were identified as well as one group containing less than three proteins from one organism (for an overview see figure 1.4). Within SF1 and SF2, the

helicases share a common core which is formed by two domains, each resembling the fold of the bacterial recombinase A (RecA) and thus they have been named RecA-like domains (Subramanya *et al.*, 1996). The interface of these two domains harbors several conserved motifs and is responsible for binding of nucleotides and RNA or DNA (Bork and Koonin, 1993; Rocak and Linder, 2004). Within the SF2 helicases, the family of DEAD-box proteins contains solely RNA-helicases, whereas the other families contain both, DNA and RNA helicases. For several viral helicases the ability to unwind both, DNA as well as RNA substrates has been shown (Fairman-Williams *et al.*, 2010).
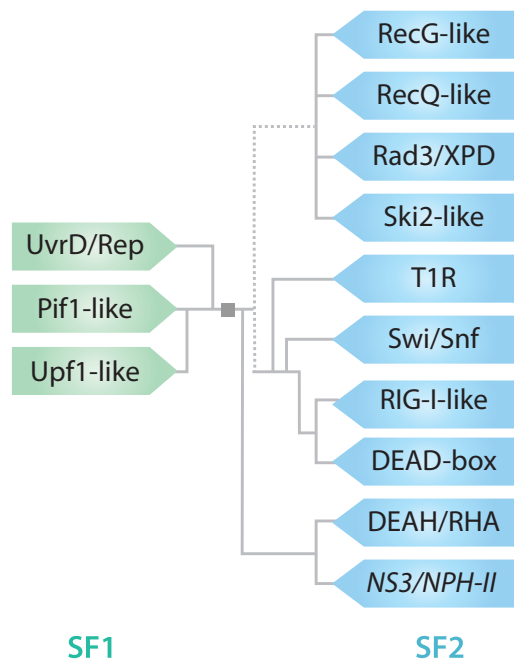
Figure 1.4: Classification of of the Superfamily 1 and 2 helicases. The cladogram shows the three identified families of the SF1 helicases on the left side and the nine families of belonging to SF2 on the right side. Families were identified by alignment of all human SF1 and SF2 helicases. Uncertain branches are indicated by dotted lines, branch length are not to scale. Figure adapted from (Fairman-Williams *et al.*, 2010)

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

The spliceosomal helicases all belong to the family of either DEAH-box, DEAD-box or Ski2-like helicases, which share up to nine conserved motifs in their respective helicase core (for overview see figure 1.5). Motif I, II, V and VI, which show the highest degree of sequence conservation, are involved in binding and hydrolysis of nucleotides, whereas RNA binding is facilitated by motifs Ia, Ib and IV. Motif II also contains the name giving DExH/D sequence motif. Motif III is suggested to play a role NTP binding as well as RNA unwinding (Walker *et al.*, 1982; Pause and Sonenberg, 1992; Cordin *et al.*, 2006). In DEAD-box as well as

Ski2-like helicases, an additional motif with a conserved glutamine residue has been identified N-terminally of motif I, the Q-motif. The conserved glutamine and aromatic residues of the Q-motif interact with the adenine and thereby, in contrast to members of the DEAH-box family, facilitate specificity for ATP (Tanner *et al.*, 2003).

DEAD-, DEAH-box and Ski2-like helicases show remarkable differences regarding their mechanistic mode of unwinding RNA or DNA. For DNA helicases, processivity of unwinding has been predicted, since this would be a prerequisite for processes like DNA replication (Singleton *et al.*, 2007). In contrast there is evidence, that the unwinding mediated by DEAD-box helicases takes place in a non-processive manner (Hilbert *et al.*, 2009), and that some DEAH-box helicases show only a moderate processivity (Pyle, 2008). The following section will focus on specific features of the three spliceosomal helicase families, which may lead to these differences.
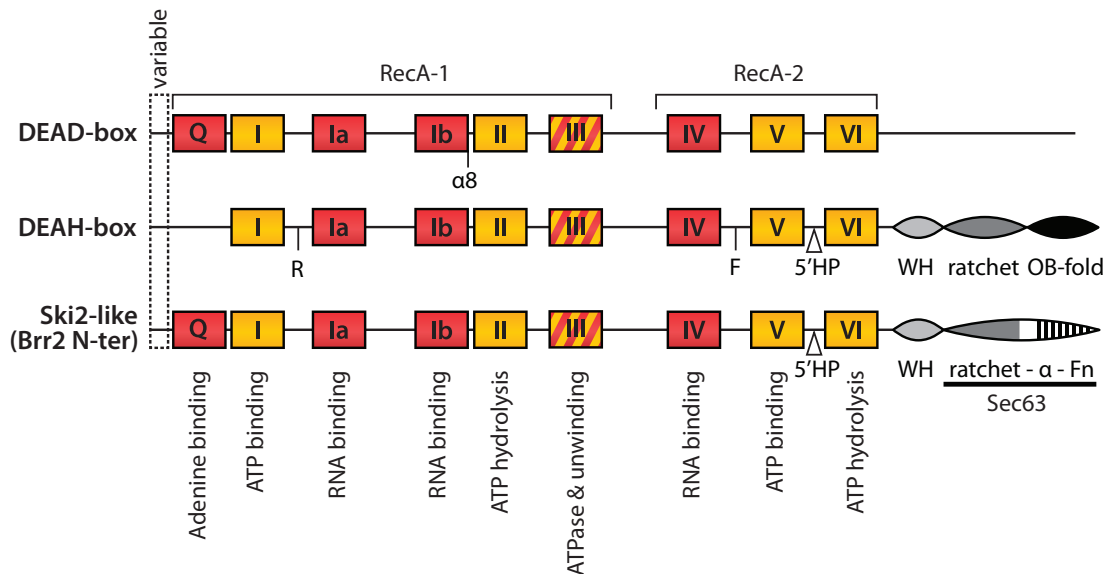
Figure 1.5: Conserved motifs of the DEAH-, DEAD-box and Ski2-like helicases. The conserved motifs located in the RecA-like domains are colored according to their interaction with ATP (yellow), RNA (red) or both (yellow/red). For the DEAD-box helicases, the location of the conserved α helix 8 is indicated by α8, R and F indicate the position of conserved arginine and phenylalanine residues of DEAH-box helicases while the position of the 5' hairpin is indicated by 5'HP. Figure adapted from Cordin *et al.* (2012)

1.6.2.1 DEAD-box helicases

As DEAD-box proteins show no processive behavior, it has been suggested that they mainly unwind short RNA duplexes or act as ATP regulated conformational switches (Pyle, 2008). It is thought, that during the cycle of ATP-binding and hydrolysis, DEAD-box proteins adopt conformations which destabilize duplex RNA and consequently lead to local strand separation (Yang *et al.*, 2007). There are different 3D crystal structures of DEAD-box proteins available, which provide an insight into the function of these proteins. In absence of ATP and RNA, the two RecA-like domains which are connected by a flexible linker adopt the so called "open conformation", which is characterized by a high variability in positioning of the RecA-domains relative to each other (Story *et al.*, 2001; Cheng *et al.*, 2005; Schütz *et al.*, 2010). In presence of RNA and ATP, the RecA-like domains bind tightly to each other and thus adopt the so called "closed conformation" which has not only been shown by crystal structures (Sengoku *et al.*, 2006; Andersen *et al.*, 2006), but was also confirmed by single-molecule FRET experiments (Theissen *et al.*, 2008). For the DEAD-box helicase eIF4AIII, crystal structures for both states are available. In comparison, these structures show the tremendous changes in positioning of the RecA-like domains. Upon binding pf ATP and RNA, both RecA-like domains tightly bind to each other, adopting the closed conformation. This conformation is stabilized by the interaction of the conserved motifs of the binding cleft with the nucleotide and the nucleic acid (Andersen *et al.*, 2006).

To date it is not clear, whether binding of ATP and RNA are two separate events or if there is a cooperative effect. Negative as well as positive cooperativity could be demonstrated for some helicases like eIF4A-III, whereas it was not observed for the DEAD-box protein DbpA (Henn *et al.*, 2008; Hilbert *et al.*, 2009). Based on the finding that in all available crystal structures helix 8 of the N-terminal RecA-like domain makes contacts to the RNA and induces a kink in the nucleic acid strand, a general mechanism for RNA unwinding by DEAD-box proteins has been proposed (Sengoku *et al.*, 2006; Hilbert *et al.*, 2009). In this mechanistic model, the unwinding cycle starts with binding of ATP and RNA, inducing the closed conformation of the helicase, and interaction of helix $\alpha 8$ with one RNA strand leads to local strand separation. Subsequent hydrolysis of ATP lowers the affinity to RNA and thus triggers the release of the RNA (Nielsen *et al.*, 2009). According to this model, ATP hydrolysis is not required for strand separation, but for release of the substrate RNA.

1.6.2.2 DExH-box helicases

The term DExH-box helicases is traditionally used for the three SF2 helicase families of DEAH-box, Ski2-like and NS3/NPHII helicases. Although their helicase core consists of two RecA-like domains, there are major differences with respect to the unwinding mechanism compared to the DEAD-box helicases. A number of studies addressing the mechanism of unwinding has been performed on members of the DExH-box helicases which resulted in two principle models describing the unwinding mechanism: the Brownian motor model and the backbone stepping motor model. Both will be described in the following.

The Brownian motor model

The Brownian motor model has been described for the DExH-box protein NS3. It is exclusively based on the thermodynamic properties of the protein while it does not take flexibility or rigidity into account. It is based on the assumption that during the cycle of ATP binding, hydrolysis and ADP release, the affinity for the nucleic acid substrate changes, leading to a switch between strong and weak binding of RNA. This allows for a movement of the protein on the RNA double strand which will induce duplex unwinding (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 nucleotide is bound (Levin *et al.*, 2003). A comparable behavior has been shown for the spliceosomal DEAH-box helicases Prp22 and Prp43, although in these cases RNA affinity is low in presence of ADP and higher when ATP is bound (Tanaka and Schwer, 2005; He *et al.*, 2010).

The backbone stepping motor model

In contrast to the Brownian motor model, the backbone stepping motor model is based on high resolution structural data of SF2 helicases in different nucleotide binding states as well as on single molecule FRET experiments on the NS3 helicase (Myong *et al.*, 2007; Hopfner and Michaelis, 2007; Büttner *et al.*, 2007; Appleby *et al.*, 2011). It proposes, that the helicase moves along the backbone of one RNA strand with a distinct step size, driven by the hydrolysis of ATP. Movement along the backbone is caused by the two RecA-like domains, which are in close contact in the ATP-bound state and increase their distance when ADP is bound. Thus, change of the affinity for RNA by the different nucleotide states of the helicase

will not result in release of the RNA substrate as in the Brownian motor model, but instead increases the flexibility of the helicase (Pyle, 2008). Crystal structures of the HCV NS3 helicase revealed, that a tryptophan residue located in the C-terminal domain of the helicase might disperse the two strands of the dsRNA and might therefore act as a physical separation device, disrupting the RNA duplex while the RecA-like domains move along the phosphate backbone (Myong *et al.*, 2007). Furthermore, single-molecule FRET experiments showed that unwinding occurs in steps of three nucleotides, and, that every unwinding step requires three translocation steps of the helicase along the RNA (Dumont *et al.*, 2006; Myong *et al.*, 2007). These steps are likely to be guided by a conserved threonine residue on each RecA-like domain, which are in a distance of three nucleotides from each other in the ATP-free state and of two nucleotides in the ATP-bound state and make contacts to the phosphate backbone of the substrate RNA (Appleby *et al.*, 2011). Therefore, each translocation event of the helicase would require the hydrolysis of one molecule ATP, also resulting in consumption of one molecule ATP per unwound base pair (Pyle, 2008; Gu and Rice, 2010; Appleby *et al.*, 2011). In the crystal structures of the Ski2-like proteins Hel308 and Ski2 (Büttner *et al.*, 2007; Halbach *et al.*, 2012) as well as of the DEAH-box helicase Prp43 (Walbott *et al.*, 2010), a β-hairpin was identified, which protrudes out of the C-terminal RecA-like domain (Figure 1.6).

It has been proposed, that this β-hairpin acts as a separation device for double stranded nucleic acids comparable to the conserved tryptophan of the NS3 helicase. This structural feature gives rise to the assumption, that the unwinding mechanism of Ski2-like and DEAH-box helicases is similar to the one proposed for NS3.
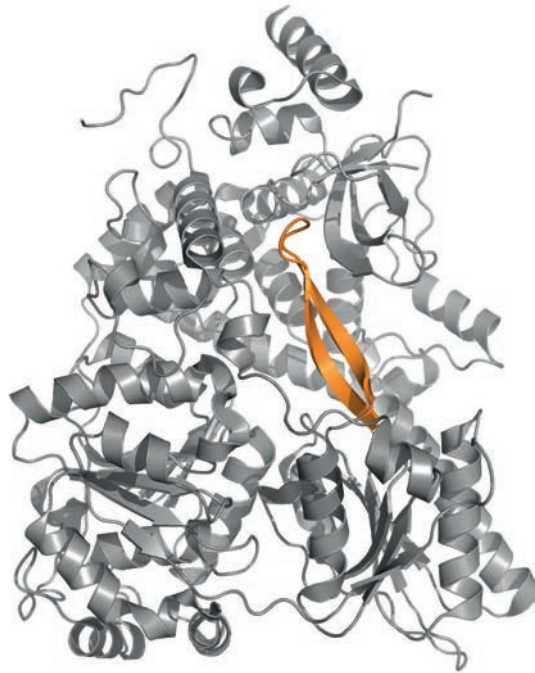
Figure 1.6: Crystal structure of the DEAH-box helicase Prp43 from *S. cerevisiae* (PDB:2XAU). The 5'β-hairpin shown in orange is protruding out of the C-terminal RecA-like domain

1.7 Spliceosomal DExH/D-box helicases

Rearrangements in the spliceosome are driven by eight conserved DExD/H-box helicases, which act as the molecular motors of the spliceosome by remodelling RNA-RNA as well as RNA-Protein interactions (Staley and Guthrie, 1998; Schwer, 2001; Wahl *et al.*, 2009). However, there is not much information about their target RNA sequence available, although some of them show RNA unwinding activity *in vitro*.

An overview of the spliceosomal helicases along with their function is shown in 1.1. Assembly of the spliceosome is dependent on the three DEAD-box helicases Sub2, Prp5 and Prp28, while catalysis and product release is promoted by the four DEAH-box helicases Prp2, Prp16, Prp22 and Prp43 (Cordin *et al.*, 2012). The Ski2-like helicase Brr2 has been shown to play a role in assembly as well as during disassembly. For most of the spliceosomal helicases, different mechanisms of regulation have been shown in biochemical studies, which might assure their specificity in the spliceosome. The ATPase activity of the DEAD-box protein Prp5 is specifically enhanced by interaction with the U2 snRNA *in vitro* (O'Day *et al.*, 1996), and it was demonstrated that phosphorylation of human Prp28 is

Table 1.1: Overview of the helicases involved in spliceosome assembly, activation and disassembly. Along with the protein name, the helicase family and the function of the protein in the spliceosome is given.

Protein	Family	Function in the spliceosome
Assembly		
Sub5	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
Catalysis		
Prp2	DEAH-box	Release of SF3a and SF3b, generation of Cwc25 binding site
Prp16	DEAH-box	Release of Yju2 and Cwc25
Prp22	DEAH-box	Identification of the 3'ss
Disassembly		
Prp22	DEAH-box	Release of mature mRNA
Prp43	DEAH-box	Disassembly of intron-lariat spliceosome

required for B complex formation (Mathew *et al.*, 2008). The Ski2-like helicase Brr2 consists of two helicase cassettes, each of it followed by a Sec63 homology domain. Remarkably it was shown, that only the N-terminal helicase cassette is required for activity *in vivo* (Kim and Rossi, 1999). However functional and crystallographic studies on human Brr2 revealed a stimulating effect of the C-terminal cassette on the helicase activity of the N-terminal cassette (Santos *et al.*, 2012). Furthermore, the RNase H-like domain of Prp8 interrupts binding of Brr2 to the U4 snRNA and in consequence inhibits the helicase activity of Brr2 (Mozaffari-Jovin *et al.*, 2012). Two of the four DEAH-box proteins in the spliceosome are dependent on G-patch containing interaction partners in order to fulfill their function in the splicing cycle. Prp43 is recruited to the spliceosome by its interaction partner Nrt1, which also stimulates its helicase activity (Tsai *et al.*, 2005; Tanaka *et al.*, 2007). Moreover, transformation of the spliceosome into the B^{act} state is facilitated by Prp2 only in presence of its interaction partner Spp2 (Roy *et al.*, 1995). In contrast to that, no regulatory interaction partners have been reported for the DEAH-box proteins Prp16 and Prp22.

1.8 The DEAH-box protein Prp2

The transition into the catalytically active B* complex is promoted by the DEAH-box helicase Prp2 along with its co-factor Spp2 (Edwalds-Gilbert *et al.*, 2004;

Ohrt *et al.*, 2012). Like all spliceosomal DExH/D-box proteins, Prp2 shows an RNA dependent ATPase activity (Kim *et al.*, 1992). However, no helicase activity could be demonstrated *in vitro*, although Prp2 seems to interact directly with pre-mRNA in the assembled spliceosome as revealed by UV-crosslink studies (Teigelkamp *et al.*, 1994). ATP-hydrolysis by Prp2 leads to a remodeling of the spliceosome, along with the destabilization of the U2 snRNP proteins SF3a and SF3b which bind near the branch site. In consequence, the branch site adenosine becomes available for the subsequent nucleophile attack (Warkocki *et al.*, 2009). Moreover, it was demonstrated that remodeling of the spliceosome by Prp2 creates a high affinity binding site for Cwc25 (Ohrt *et al.*, 2012), which seems to be important for stabilizing the pre-mRNA in a distinct conformation, suitable for the first transersterification reaction (Krishnan *et al.*, 2013). Recent data furthermore demonstrate that Prp2 might be directly or indirectly involved in destabilizing the U2/U6 helix Ia (Wlodaver and Staley, 2014). While the C-terminal region the spliceosomal DEAH-box proteins Prp2, Prp16, Prp22 and Prp43 shows a high degree of conservation and is therefore thought to be the primary interaction platform for other components of the spliceosome, the region N-terminal of RecA1 lacks significant conservation. In fact, deletion of large portions of the Prp2 N-terminus does not lead to apparent effects *in vivo* (Edwalds-Gilbert *et al.*, 2004). The function of Prp2 in the splicing process is strongly dependent on its co-factor, the G-patch protein Spp2 (supressor protein of prp2), which was originally identified as a genetic suppressor of a *prp2* mutant and is thought to recruit Prp2 to the spliceosome (Last *et al.*, 1987; Roy *et al.*, 1995). Binding and mutagenesis studies on Prp2 indicate, that Spp2 is able to interact with the C-terminal part of Prp2 (Silverman *et al.*, 2004).

1.9 Spp2 and the G-patch motif

Spp2 is a G-patch containing protein and was the first protein for which a direct interaction with a spliceosomal DExD/H box helicase was shown (Roy *et al.*, 1995).The name-giving G-patch motif consists of six conserved glycines and has first been identified in RNA-associated proteins using a bioinformatics approach (Aravind and Koonin, 1999). G-patch proteins are absent in archaea and bacteria but present in some retroviral polyproteins and widespread among eukaryotes, where the G-patch motif can be found in mRNA methyltransferases, proteins that might be involved in DNA-damage repair and DNA-binding proteins (Pang

et al., 1993; Stein *et al.*, 2003). For the spliceosomal DEAH-box helicase Prp43, interaction with three different G-patch proteins has been shown. Furthermore, its helicase activity is strongly activated by the G-patch protein Ntr1 (Tanaka *et al.*, 2007), while for Prp2 and Spp2, this effect could not be observed and the function of Spp2 remains elusive.

```
              105         115          125          135          145          155
              •           •            •            •            •            •
 S.c. Spp2    VEEFGDALLRGMGWE SDSEQDSK-GDKTQSRNKDVS-NVSQIHPDGLGIGAKLNKA---

 C.t. Spp2    VDDFGENLLRSFGWD-GKMRG----------KVKEVK-RYANLA--GLGARNVKEAED--
 H.s. Spp2    VEAYGLAMLRGMGWKPGEGIG--------RTFNQVVKPRVNSLRPKGLGLGAN-LT----
 S.c. Ntr1    TYGIGAKLLSSMGYVAGKGLGKDGSGITTPIETQS---RPMHNA--GLGMFSN-TNSS--
 H.s. Cmtr1   YNSVSQKLMAKMGFREGEGLGKYSQGRKDI------VEASSQKGRRGLGLTLRGFDQELN
```

Figure 1.7: Alignment of the G-patch motif from the G-patch containing proteins Spp2, Ntr1 and Cmtr1 of *S. cerevisiae*, *H. sapiens* and *C. thermophilum*. Conserved glycines and hydrophobic residues are colored in green and orange respectively. Numbers refer to the residues of Spp2 from *S. cerevisiae*.

The G-patch itself consists of a pattern of up to six conserved glycines along with several conserved aromatic and/or hydrophobic residues (for an overview see figure 1.7). So far, the function of the conserved glycine residues remained unclear, while the importance of the conserved leucine 109 for interaction with Prp2 has been demonstrated in mutagenesis studies (Silverman *et al.*, 2004).

1.10 Scope of the thesis

At the beginning of this work, little was known about the molecular function of the spliceosomal DEAH-box helicase Prp2. Moreover, no structural information on G-patch proteins in general and on Spp2 in particular was available and the interplay of Prp2 with Spp2 was poorly understood at a structural level. The aim of this thesis was the investigation of the Prp2-Spp2 interaction, especially regarding the influence of Spp2 on structural and functional properties of Prp2. To achieve these goals, the structure of Prp2 should be solved by means of X-ray crystallography in complex with different nucleotides, RNA and its interaction partner Spp2. Therefore, the purification of the single components as full length proteins as well as in truncated variants had to be established, followed by the reconstitution of a functional complex. In addition, the characterization of the G-patch protein Spp2 by biophysical methods such as NMR-spectroscopy and CD-spectroscopy was a major goal, as crystallization of such small uncharacterized

proteins is usually quite challenging and due to its small size, Spp2 was a good target for NMR experiments. To characterize the Prp2-Spp2 interaction on a functional level, it was planned to perform biochemical studies on the ATPase and helicase activity as well as RNA affinity of Prp2 in absence and presence of Spp2.

2 Material and methods

2.1 Material

2.1.1 Fine chemicals

All fine chemicals were and organic components were purchased from the companies AppliChem, Bio-Rad, Deutero, EURISO-Top, Fluka, Merk, MWG Biotech, Roth or Sigma-Aldrich in the purity grade *pro analysis*. Generally, the supplier with the lowest price was chosen.

2.1.2 Devices

Adjustable pipets	Eppendorf
Agarose gel electrophoresis 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
Centrifuge Avanti JA-20	Beckman Coulter
Fine balance	Sartorius
Fluorescence spectrophotometer Fluoromax III	Horiba Jobin Yvon
Microfluidizer 110 S	Microfluidics
PCR-Whatman Biometra T personal	Biometra
pH-Meter	Beckman Coulter
Real-Time PCR Detection System CFX96	Bio-Rad
Rotor JA-20	Beckman Coulter
Rotor JA-30.50 Ti	Beckman Coulter
Rotor JLA 8.1000	Beckman Coulter
SDS-PAGE chamber Hoefer miniVE	GE Healthcare
Table Top centrifuge 5417 R	Eppendorf
Thermomixer comfort	Eppendorf
Unitron shaking incubator	Infors
UV/Vis photometer	GE Healthcare

2.1.3 Supplemental material

24 well crystallisation plates	Hampton Research
96 well crystallisation plates	Hampton Research
Crystal Clear Tape	Henkel
Nylon loops for crystals	Molecular Dimensions
Sterile filter	Millipore
Reaction tubes (0.5 ml, 1.5 ml, 2.0 ml)	Eppendorf
Ultra centrifugal filter units	Millipore

2.1.4 Cell media

In this work, 2YT, TB and M9 minimal media was used for recombinant protein expression.

For 2YT and TB, the components of the media were dissolved in desalted H_2O and autoclaved at 121 °C for 20 min prior to usage. Antibiotics were added as required to a final concentration of 100 µg/ml for ampicillin and 30 µg/ml for chloramphenicol and kanamycin. The two components of TB (TB_A and TB_B) were autoclaved seperately and mixed in a 1 to 10 ratio prior to usage. To produce agar plates, 2YT media was supplemented with 1.5% (w/v) Agar prior to autoclaving. Recipies for the media are shown in the following tables.

2YT media		TB_A		TB_B	
1% (w/v)	Trypton	1.2% (w/v)	Trypton	169 mM	KH_2PO_4
1% (w/v)	Yeast Extract	2.4% (w/v)	Yeast Extract	720 mM	K_2HPO_4
1% (w/v)	NaCl	0.4% (w/v)	Glycerol		

For M9 minimal media, every stock solution was prepared an sterile filtered separately and diluted to the final working concentration with desalted and autoclaved H_2O prior to usage.

10× M9 salts		100× Trace element solution	
		13.4 mM	EDTA
		3.1 mM	$FeCl_3 \cdot 6\,H_2O$
33.7 mM	Na_2HPO_4	0.62 mM	$ZnCl_2$
22.0 mM	KH_2PO_4	76 µM	$CuCl_2 \cdot 2\,H_2O$
8.55 mM	NaCl	42 µM	$CoCl_2 \cdot 2\,H_2O$
9.35 mM	NH_4Cl	162 µM	H_3BO_3
		8.1 µM	$MnCl_2 \cdot 4\,H_2O$

M9 minimal media	
20% (w/v)	Glucose
1.0 mM	$MgSO_4$
0.3 mM	$CaCl_2$
5 mg/ml	Thiamine/HCl
1×	100× Trace element solution
1×	10× M9 salts solution

2.1.5 Chromatography systems, columns and media

All chromatographic methods were performed on Äkta Prime, Äkta Purifier and Äkta Explorer systems (all GE Healthcare). To load samples onto columns, Superloops with a volume of 150 ml and 50 ml as well as loops with a volume of 5 ml and 2 ml were used. Utilized columns are depicted in the list below.

GSTrap 5 ml	GE Healthcare
GSH Sepharose	GE Healthcare
HiPrep Desalting 26/10	GE Healthcare
HiPrep Desalting 5 ml	GE Healthcare
HisTrap 5 ml Ni-NTA Sepharose	GE Healthcare
Source 30Q	GE Healthcare
StrepTrap 5 ml	GE Healthcare
Superdex 75 26/60	GE Healthcare
Superdex 200 26/60	GE Healthcare
Superdex 200 10/300	GE Healthcare
Superose 12 10/300 GL	GE Healthcare

2.1.6 Genomic DNA

The used 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, Georg-August-Universität Göttingen.

2.1.7 Organisms

The used *E. coli* strains Rosetta II (DE3) and XL1-Blue were part of the collection of the Department for Molecular Structural Biology, Georg-August-Universität Göttingen. *E. coli* cells of the strain TOP10 were purchased from IBA-lifesciences.

2.1.8 Plasmids

Table 2.1: IBA Stargate donor vectors which have been designed and used for IBA StarGate cloning (see 2.2.6.2). In the table, the coded proteins along with the involved residues are named. The sequences of the oligonucleotides can be found in the supplemental table A.1

Plasmid	oligonucleotide
pENTRY51-ctPrp2 (1-921)	S1,S3
pENTRY51-ctPrp2Δ25 (270-921)	S2,S3
pENTRY51-ctSpp2 (1-313)	S4,S5
pENTRY51-ctSpp2_2 (172-254)	S6,S8
pENTRY51-ctSpp2_3 (211-254)	S7,S8

Table 2.2: Plasmids used for recombinant protein expression (2.2.9) during this work. The used expression vectors together with the coded proteins are named and the involved residues are given in parenthesis. Type and position (C- or N-terminal) of affinity tags are indicated, as well as the antibiotic resistance and the method used for cloning, where (RS) stands for restriction site cloning and (SG) for cloning with the IBA StarGate system. The sequences of the oligonucleotides can be found in the supplemental table A.2

Name	Affininty tag	Resistance	Cloning method	Oligonucleotide
pET21a_Prp2 (1-876)	6× His (C)	Ampicillin	RS	R1, R2
pET21a_Prp2 581ΔN (581-876)	6× His (C)	Ampicillin	RS	R4, R2
pET21a_Prp2 597ΔN (597-876)	6× His (C)	Ampicillin	RS	R3, R2
pGEX-6p-1_Spp2_2 (10-185)	GST (N)	Ampicillin	RS	R5, R6
pET21a_Spp2_2 pET (10-185)	6× His (C)	Ampicillin	RS	R5, R13
pASG25_ctPrp2 (1-921)	GST (N)	Ampicillin	SG	–
pASG25_ctPrp2Δ25 (270-921)	GST (N)	Ampicillin	SG	–
pASG3_ctSpp2 (1-313)	Strep (C)	Ampicillin	SG	–
pASG3_ctSpp2_2 (171-254)	Strep (C)	Ampicillin	SG	–
pASG3_ctSpp2_3 (211-254)	GST (N)	Ampicillin	SG	–

2.1.9 Size standards

Size standards that were used during agarose-gelelectrophoresis and SDS-PAGE are listed below

GeneRuler 1kb DNA ladder	Fermentas
Protein Molekular Weight Marker	Fermentas
PageRuler™ Unstained Protein Ladder	Fermentas
PageRuler™ Prestained Protein Ladder	Fermentas

2.1.10 RNA oligonucleotides

Table 2.3: RNA oligonucleotides used in the ATPase activity assay (2.3.1), for crystallization (2.4.1), as substrate for RNA unwinding measurements (2.3.3) and for RNA affinity measurements (2.3.4)

Name	Sequence (5'-3')
$\text{PolyA}_{(20)}$	AAAAAAAAAAAAAAAAAAAA
$\text{PolyU}_{(10)}$	UUUUUUUUUU
$\text{PolyU}_{(15)}$	UUUUUUUUUUUUUUU
$\text{PolyU}_{(20)}$	UUUUUUUUUUUUUUUUUUUU
uw	GCGCCUACGGAGCUGGUGGCGUAGGCGCAAAAAAAAAAAAAAAAAAAA
uw-l	(Cy5)-GCGCCUACGCCACCAGCUCCGUAGGCGC-(BBQ)
$\text{PolyU}_{(12)\text{fl}}$	(6-FAM)-UUUUUUUUUUUU
$\text{PolyU}_{(16)\text{fl}}$	(6-FAM)-UUUUUUUUUUUUUUUU
$\text{PolyU}_{(12)\text{fl}}$	(6-FAM)-UUUUUUUUUUUUUUUUUUUU
actin	(Cy5)-UACUAACAUCGAUUGCUUCAUUCUUUUGUUGCUAUAUUA

2.1.11 Screens for Thermal shift assay

The following screens were used thermal shift assays.

Emerald Bio Crystallography - pH Buffer screen	Jena Bioscience
Silver Bullets Bio	Hampton Research
Additive Screen	Hampton Research

2.1.12 Crystallization screens

The following screens were used during initial screening for crystallization 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 *et al.* (2005)
Morpheus	Molecular Dimensions
Natrix HT	Hampton Research
PGA	Molecular Dimensions
ProPlex	Radaev *et al.* (2006)

2.1.13 Computer programs and databases

APBS	Baker *et al.* (2001)
BLAST	Altschul *et al.* (1990)
CCP4 suite	Winn *et al.* (2011)
ClustalW	Larkin *et al.* (2007)
Coot	Emsley *et al.* (2010)
GENtle	Manske, Magnus
ESPript	Gouet *et al.* (1999)
pdb2pqr	Dolinsky *et al.* (2007)
Pfam	Finn *et al.* (2010)
Phenix	Adams *et al.* (2002)
PSIPRED	Buchan *et al.* (2010)
ProtParam	Wilkins *et al.* (1999)
Pymol	Schrödinger LLC
Sigmaplot	Systat Software
SMART	Letunic *et al.* (2012)
XDS	Kabsch (2010)

2.2 Molecular biological methods

2.2.1 Polymerase chain reaction

The Polymerase chain reaction (PCR) allows the selective amplification of a specific DNA sequence. A general PCR mixture using Phusion polymerase (Thermo-Scientific) contains 200-400 ng of DNA template, 10 pmol of the forward- as well as the reverse oligonucleotide, 1 µl of Phusion polymerase, 5 µl supplied HF-buffer and 1 µl dNTP mixture (10 mM each), added with ddH$_2$O to a total volume of 50 µl. A typical PCR program is shown below.

cycles	temperature	step	time
1×	98 °C	Initial denaturation	30 s
25×	98 °C	Denaturation	10 s
	52 - 60 °C	Annealing of the oligonucleotides	20 s
	72 °C	Elongation	15 s/1kb
1×	72 °C	Final Elongation	10 min

The PCR product was purified using the peqGOLD MicroSpin Cycle-Pure Kit (PEQLAB) and analyzed by agarose gel electrophoresis 2.2.2

2.2.2 Agarose gel electrophoresis

Agarose gel electrophoresis is a method to seperate DNA or RNA oligonucleotides by their molecular mass. In this work it was used to separate DNA fragments resulting from PCR (2.2.1) and enzymatic restriction digest (2.2.5). TAE buffer was supplied with agarose to a final concentration of 1% and boiled in a microwave oven until the agarose was completely dissolved. GelRed staining solution was added in a 1:10000 ratio according to the manufacturer. The solution was then poured into a gel mold with a comb and cooled down to room temperature. Once it was completely solid, it was transferred into an agarose gel electrophoresis chamber and covered with TAE buffer. The samples were mixed with DNA loading dye (10×) and loaded onto separate lanes. The gel was run with a voltage of 12 V/cm until a suitable separation was reached.

Recipes for 10× DNA loading dye and TAE buffer are shown below.

10× DNA loading dye		TAE buffer	
0.25% (v/v)	Bromphenolblue	40 mM	Tris
0.25% (v/v)	Xylene Cyanol	20 mM	acetic acid
30% (v/v)	Glycerol	1 mM	EDTA

2.2.3 DNA sequencing

Sequencing of DNA was performed in the Department of Developmental Biology (Prof. Pieler), Georg-August-Universität Göttingen. The sequences of oligonucleotides used for sequencing PCR can be found in supplement table A.3.

2.2.4 Determination of DNA concentration

The concentration of DNA in solution was determined by absorption measurements at 260 nm (A_{260}) using a photometer. For doublestranded DNA, 1 absorption unit at 260 nm corresponds to 50 µg/ml. The purity of the DNA was determined by the ratio of A_{260} to A_{280}, which is between 1.8 and 2.0 for pure DNA. Lower values indicate contamination with proteins.

2.2.5 Enzymatic DNA digest

In order to digest a DNA fragment obtained by PCR (2.2.1) or vector DNA (2.2.8), the DNA was incubated with the suitable DNA restriction enzyme in the appropriate buffer (according to manufacturers protocol) at 37 °C for 1 hour. For complete digest, 1 unit of restriction enzyme per 1 µg of DNA was used. The restriction enzymes were afterwards inactivated by incubation at 80 °C for 10 minutes.

A general reaction mixture for test digests contained 1 µg DNA in a total reaction volume of 20 µl while for restriction site cloning (see 2.2.6.1) 2.5 µg DNA in a total volume of 50 µl were used.

2.2.6 Molecular cloning

2.2.6.1 Restriction site cloning

The gene of interest was amplified by PCR (2.2.1), using forward and reverse oligonucleotides featuring suitable restriction sites. The purified PCR product as well as the desired vector were digested with the appropriate restriction enzymes according to the manufacturers protocol. The digested vector was dephosphorylated by incubation with Shrimp alkaline phosphatase (SAP) (0.01 unit per pmol DNA) for 30 min at 37 °C. The SAP was afterwards inactivated at 65 °C for 15 min. Ligation of vector and insert was performed over night at 16 °C by mixing 200 ng of digested and dephosphorylated vector with 5× molar excess of digested insert, 1 µl of T4 ligase (Fermentas), 1 µl of supplemented 10× reaction buffer and 1 µl of 10 mM ATP, added to 10 µl with ddH_2O. The entire reaction mixture was transformed into *E.coli* XL1-blue cells (see 2.2.7) and plated on agar plates containing the appropriate antibiotics (see 2.1.4).

2.2.6.2 IBA StarGate cloning

The IBA StarGate cloning system (IBA GmbH) is based on recombinatorial cloning, were in a first step a donor vector is created which is used in a second step to transfer the gene of interest into an expression vector.

The gene of interest was amplified by PCR (see 2.2.1), using forward and reverse oligonucleotides featuring upstream and downstream combinatorial sites. The purified PCR product was added to an reaction tube containing the pENTRY vector and mixed with specific components according to the manufacturers protocol. The reaction mixture was incubated for 1 h at 30 °C, transformed into *E.coli* TOP10 cells and plated on kanamycin containing agar plates (see 2.1.4). Plasmids from several colonies were purified using a peqGOLD Plasmid Mini Kit and subsequently sequenced to verify correct cloning.

To transfer the gene of interest into the chosen expression (acceptor) vector, the donor vector was added to a tube containing the acceptor vector and supplemented with provided specific components as described in the manual. The reaction mixture was then incubated for 30 min at 30 °C, transformed into *E.coli* TOP10 cells (see 2.2.7) and plated on ampicillin containing agar plates (see 2.1.4).

2.2.7 Transformation of chemical competent *E.coli* cells

To transform plasmid DNA into *E.coli* cells, 100-200 ng of plasmid DNA were mixed with 50 µl of competent cells, incubated on ice for 30 min followed by a heatshock of 42 °C for 1 min. The reaction vessel was placed back on ice for 5 min, then 800 µl of 2YT were added and the cells were incubated for 45 min at 37 °C. Finally, the cells were centrifuged at 3000× g for 1 min, the cell pellet was resuspended in 200 µl media and the cell suspension was plated on agarplates containing the appropriate antibiotic.

2.2.8 Preparation of plasmid DNA

To amplify plasmid DNA for further cloning steps or recombinant protein expression, the plasmid of interest was transformed (see 2.2.7) into *E. coli* XL1 blue cells (2.1.7). 5 ml of 2YT media (2.1.4) were inoculated with a single *E. coli* colony and incubated overnight at 37 °C.

Preparation of the plasmid DNA was performed using the peqGOLD Plasmid Mini Kit (PEQLAB) based on alkaline cell lysis and binding of the DNA to a

silica matrix. After washing, DNA was eluted with 10 mM Tris/HCl pH 8.5. The DNA was stored at -20 °C until further usage.

2.2.9 Recombinant protein expression

To recombinantly express proteins in *E.coli*, the gene of interest was cloned into a suitable vector (see 2.2.6) in a way, that the coding gene is under the control of an inducable promotor. In this work, two different promotor systems were used: the lac-promotor, which was induced by addition of 0.5-1 mM IPTG to the cell culture and the tet-promotor, where Anhydrotetracycline in a final concentration of 200 µg/l was used as inducer.

For expression, a 50 ml of media containing the suitable antibiotics were inoculated with several *E.coli* colonies from an agar plate and incubated at 37 °C and 220 rpm overnight. The expression cultures were inoculated in a 1:100 ratio with the preculture and the cells were grown at 37 °C and 220 rpm until an OD_{600} of 0.8-1.0 was reached. The appropriate inducer was hen added to the cells, and the temperature was lowered to 16 °C. Expression was generally performed for 20 h.

2.2.10 Cell harvest and lysis

In order to isolate the recombinantly expressed protein, the cells had to be harvested and lysed. Cell harvest was performed by centrifugation of the cell cultures at $4800 \times$ g and 4 °C for 15 min. The cell pellet was resuspended in cold lysis buffer and either directly lysed or centrifuged again at $4800 \times$ g and 4 °C for 15 min and stored at -20 °C.

Lysis of the cells was performed pushing the cell suspension five times through a Microfluidizer 110S (Microfluidics), in which the cells pass a chamber at high pressure (0.55 MPa). The cell lysate was clarified by centrifugation at $30000 \times$ g and 4 °C for 30 min and the resulting supernatant was subsequently applied to chromatography.

2.2.11 Chromatographic methods

To separate the protein of interest from cellular *E.coli* proteins and purify it to homogeneity , several different chromatographic steps were performed which are described below. All columns were stored in 20% (v/v)ethanol, washed with 1 column volume (CV) ddH_2O and equilibrated in the appropriate buffer prior to

usage. After usage, columns were regenerated if necessary, washed with 1 CV of ddH$_2$O and 1 CV 20% (v/v) ethanol.

2.2.11.1 GST affinity chromatography

Purification of N-terminal GST fusion proteins was performed using glutathion immobilized on a sepharose matrix (Glutathion Spharose 4B) which is specifically bound by the fusion protein whereas other proteins can be washed off the column. Elution of the fusion protein was performed using a buffer supplemented with 30 mM reduced glutathion.

2.2.11.2 Histidin affinity chromatography

For purification of 6× His-tagged proteins, immobilized metal chelating affinity chromatography (IMAC) using Ni-NTA Sepharose was performed. 30 mM imidazole were added to the lysis- and loadingbuffer to reduce unspecific binding of proteins to the column. The elution of the tagged protein was performed with buffer supplemented with 500 mM imidazole.

2.2.11.3 Strep affinity chromatography

Affinity purification of Strep-tagged proteins is based on the tight and highly specific binding of an 8 amino acid long tag (WSHPQFEK) to Strep-Tactin, an engineered streptavidin. For optimal binding of the affinity tag to Strep-Tactin, a pH of 7.5 or above was used. Unbound protein was washed off and the pure protein of interest was eluted with buffer containing 3 mM Desthiobiotin.

2.2.11.4 Size exclusion chromatography

Size exclusion chromatography is a method to separate molecules based on their size using a porous material. Small molecules can enter the pores and are therefore migrating slower than large molecules. Preparative 26/60 columns were run at a flow rate of 1 ml/min whereas for analytical 10/300 columns a flow rate of 0.3 ml/min was applied.

2.2.11.5 Determination of the nucleotide loading state of proteins

For determination of the nucleotide loading state of a protein, the protein was incubated in a suitable buffer containing 10 mM EDTA to chelate all Mg^{2+} ions and

thereby strip bound nucleotides from the protein. The Protein was then desalted in an EDTA-free buffer containing 2 mM $MgCl_2$ and subsequently incubated with the nucleotide of choice. Unbound nucleotides were removed by desalting, and the protein was precipitated by incubation at 95 °C. The solution was centrifuged at $16100\times$ g and the supernatant was loaded onto a Nucleosil 4000-7 HPLC anion exchange column (Macherey-Nagel) which had been equilibrated with buffer containing 10 mM Tris pH 8.0. The nucleotides were eluted by a gradient to a buffer containing 10 mM Tris pH 8.0 and 2 M NaCl. Identification of the nucleotides was performed by comparison to column calibration runs which were performed by injection and elution of 5 nmol ADP, ATP, AMPPNP and AMPPCP.

2.2.12 Concentrating of protein solutions

Protein solutions were concentrated by transferring them into the upper chamber of a Vivaspin concentrator (Vivascience) with a vertical membrane featuring pores of a specific size range. By centrifugation at $4000 \times$ g proteins of a size dependent on the membrane were retained in the upper chamber while smaller molecules like water and ions could pass the membrane. The process was continued until the desired protein concentration was reached.

2.2.13 Determination of protein concentrations

Two different methods for determining the protein concentration were used in this work, the Bradford protein assay and UV-absorption measurements.

The Bradford assay is based on staining amino acid sidechains with the dye Coomassie Blue G-250, which binds to arginin- and aromatic sidechains and thereby undergoes a shift in its absorption maximum from 470 nm to 595 nm. Therefore, the absorption at 595 nm is directly proportional to the protein concentration were a OD_{595} equals a protein concentration of 0.1 mg/ml. UV absorption measurements were carried out in photometer at at wavelengths of 280 and 260 nm, the concentration of the protein solution was determined by Lambert-Beers law using the calculated extinction coefficient of the protein (calculated with ProtParam (see 2.1.13)). The purity of the Protein was determined by the ratio of A_{260} to A_{280}, which is between 0.5 and 0.8 for pure protein. Higher values indicate contamination with nucleic acids.

2.2.14 SDS-polyacrylamide gel electrophoresis

SDS-PAGE was used to analyze the composition of protein samples by separating the single proteins according to their mass in a process known as molecular sieving (Laemmli, 1970). Dependent on the size of the protein, separating gels with a concentration of 7.5%, 12.5%, 15% or 17.5% were chosen. The recipes of the used solutions are shown below. The gels were prepared by filling the assembled pouring chamber to about 80% with separating gel which was covered with 100% isopropanol. After complete polymerization of the gel, the isopropanol was removed and the chamber was filled with stacking gel solution to the top, followed by insertion of the combs. After polymerization of the stacking gel, the gels were stored at 4 °C until usage.

SDS-PAGE separating gel

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)	$(NH_4)_2S_2O_8$

SDS-PAGE 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)	$(NH_4)_2S_2O_8$

Prior to gel electrophoresis, the samples were mixed with Laemmli buffer, heated to 95 °C for 5 min and loaded onto the gel. Electrophoresis was run for 80-90 minutes at 400 V and 30 mA. For detection of the protein bands, the gel was boiled in staining solution in a microwave oven.

SDS-PAGE running buffer

25 mM	Tris
192 mM	Glycine
0.1 % (w/v)	SDS

2× Laemmli sample buffer

62.5 mM	Tris/HCl pH 6.8
70 mM	SDS
55% (v/v)	Glycerol
0.1% (w/v)	Bromphenol blue
5%	2-Mercaptoethanol

Staining solution	
10%	Ethanol
5%	Acetic acid
0.005% (w/v)	Coomassie brillant blue G-250
0.005% (w/v)	Coomassie brillant blue R-250

2.3 Biophysical Methods

2.3.1 Determination of ATPase activity

The ATPase assays were performed using the EnzCheck® Phosphate Assay Kit (Invitrogen, USA). The reaction mixtures with a total volume of 100 µl contained 10 mM HEPES/NaOH pH 7.5, 150 mM KCl, 10 mM MgCl$_2$ and 1 µM of the helicase as well as 1 µl purine nucleoside phosphorylase and 20 µl 2-amino-6-mercapto-7-methylpurine riboside. For the reactions in the presence of RNA, up to 10 µl poly-A 20 RNA (IBA GmbH) were added. Reaction mixtures were incubated at 25 °C for 10 min.

The reactions were started by addition of ATP in concentrations between 1 to
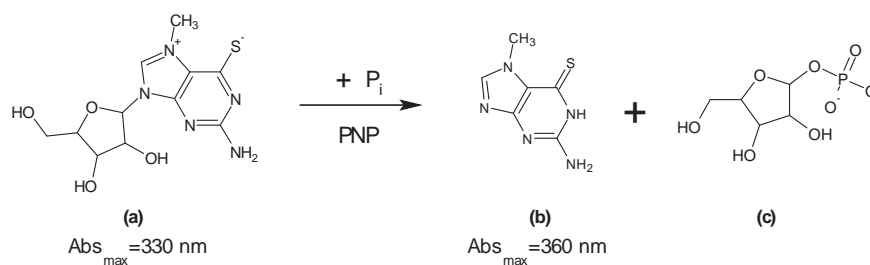
Figure 2.1: Reaction mechanism of the used EnzCheck® Phosphate assay Kit. 2-amino-6-mercapto-7-methylpurine riboside (a) is enzymatically converted to 2-amino-6-mercapto-7-methylpurine (b) and ribose 1-phosphate (c) by purine nucleoside phosphorylase (PNP). The release of orthophosphate from the ATPase reaction is thereby coupled to a shift of the absorption maximum from 330 nm to 360 nm, thus allowing for quantitation of hydrolysed ATP.

500 µM, and the absorption was measured at a wavelength of 360 nm using a Ultrospec 2100pro photometer (GE Healthcare). The starting velocity v_0 was determined as the initial slope of the change in absorption ($\Delta A/min$) three times independently for each ATP concentration. The amount of released phosphate was determined by interpolation of the measured values to a standard phosphate curve. Kinetic parameters were calculated by nonlinear regression to a hyperbola

using SigmaPlot. Since the hyperbola was expressed in a form identical to the Michaelis-Menten equation, the parameters K_{m} and v_{max} were directly obtained.

2.3.2 Circular dichroism spectroscopy

Circular dichroism spectroscopy (CD) is a method to determine the content of α helices and β sheets of a protein based on their difference in absorption of left and right circular polarized light in the far-UV range (Kelly *et al.*, 2005).

For the measurements the protein sample was desalted in a buffer containing 50 mM Sodium phosphate at pH 7.5 and diluted to 0.1 mg/ml. 300 µl of the solution were pipetted into a cuvet with a path lenght of 1 mm and placed into a Chirascan CD spectrometer (Applied Photophysics). CD spectra were recorded from 185 to 260 nm at 20 °C. Every sample was measured five times to obtain mean values and in the end a blank curve containing the CD spectra of the buffer was subtracted from the protein spectra.

2.3.3 Determination of helicase activity

The helicase activity of proteins was determined based on a method published by Belon and Frick (2008). As substrate RNA, to RNA strands were annealed of which one was labeled with the fluorophore Cyanine 5 (Cy5) at the 5'-end and with the quenching molecule BlackBerry Quencher 650 (BBQ) at the 3'-end. Cy5 was exited at a wavelength of 643 nm and emission was detected at 667 nm. Upon RNA double strand separation by the helicase, the labeled RNA strand forms a hairpin structure bringing Cy5 in close proximity to BBQ which leads to the quenching of Cy5 fluorescence. Therefore, a decrease in the Cy5 fluorescence signal is directly proportional to the rate of RNA strand separation.

Preparation of the double stranded substrate RNA was performed by mixing 4.5 µl of both RNA strands at a concentration of 100 µM each with 1 µl of 100 mM Tris pH 8.5. The reaction mixture was incubated in a thermocycler at 95 °C for 2 minutes. The temperature was then decreased from 80 °C to 20 °C at a rate of 1 °C per 20 seconds, followed by incubation of the RNA on ice. The annealed RNA was diluted to a working concentration of 1 µM.

The reaction mixtures for the helicase activity assay contained 20 nM of protein, 50 nM of annealed RNA substrate, 10 mM HEPES pH 7.5, 150 mM NaCl, and 2 mM $MgCl_2$. Prior to measurement, the samples were incubated for 15 minutes at 20 °C in the fluorescence spectrometer. The reaction was started by addition of
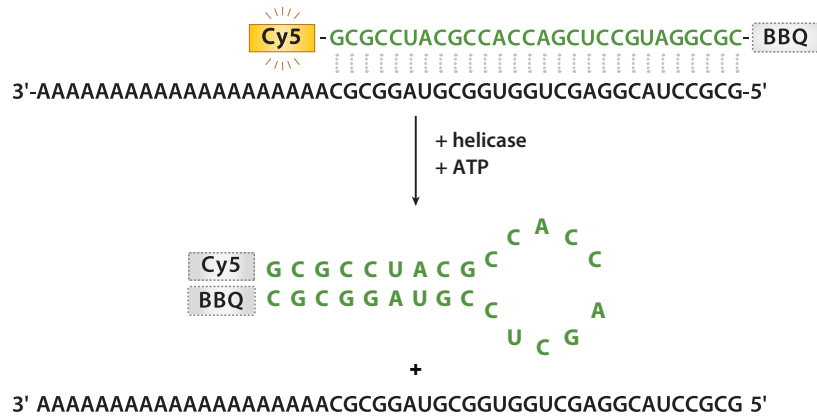
Figure 2.2: Principle of fluorescence based helicase assay. Upon addition of an active helicase and ATP to the RNA substrate the double strand is unwound, leading to hairpin formation of the labeled strand. BBQ can now quench the fluorescence of Cy5, leading to a decrease of emission at 667 nm

ATP at a final concentration of 1 mM. The samples were excited at a wavelength of 643 nm while the change in emission was monitored at 667 nm every 2 seconds for 10 minutes.

The initial slope $\Delta F/min$ representing v_0 was measured three times independently in order to determine the maximum reaction velocity (v_{max}). The amount of unwound dsRNA was calculated with respect to the signal at 667 nm for the annealed and completely unwound dsRNA, respectively.

2.3.4 Determination of binding constants for RNA

Binding constants for protein-RNA interactions were determined by MicroScale Thermophoresis, were fluorescence of molecules is used to monitor the motion of molecules along a temperature gradient. For the measurements a singlestranded RNA oligonucleotide with the sequence (5'-3'): UACUAACAUCGAUUGCUU-CAUUCUUUUGUUGCUAUAUUA was used which was labeled with Cy5 at the 5' end to allow for fluorescence detection. For the measurements, RNA was used in a concentration of 100 nM in order to obtain a good SNR for the fluorescence signal. Each sample contained 100 nM RNA, Prp2 in a concentration between 150 and 9.1 µM supplied with a 10× molar excess of ADP or AMPPNP in buffer containing 10 mM HEPES/NaOH pH 7.5, 200 mM NaCl, 5 mM MgCl$_2$ and 0.05% Tween 20. Measurements were performed in a NanoTemper Monolith NT.115 with 20% LED power and 20% MST power. Each sample was measured

$3\times$ independently. The obtained values for thermophoresis were plotted against the protein concentration, K_d values were calculated by nonlinear regression to a sigmoidal equation using SigmaPlot.

2.4 Crystallographic methods

2.4.1 High throughput screening

Initial screening for crystallization conditions was performed on a Phenix IRE pipetting robot (Rigaku) using MRC 96-3 well plates with a protein concentration between 3 and 10 mg/ml. In a general setup, two drops were pipetted for each condition in ratios of 1:1 (0.25 µl protein and 0.25 µl reservoir) and 1:2 (0.125 µl protein and 0.25 µl reservoir). As negative control for each condition, an additional drop contained only the sample buffer mixed with reservoir in a 1:1 ratio. The volume of the reservoir was 35 µl. All screens were either prepared by an Alchemist pipetting robot (Rigaku) or purchased (see 2.1.12).
After pipetting, the crystallization plates were sealed and stored in a Minstrel System (Rigaku) at 4 °C or 20 °C. Each drop was photographed automatically according to a specific schedule and photos were checked by eye using the CrystalTrak software (Rigaku).

2.4.2 Optimization of crystals

Upon identification of initial conditions which lead to crystals or promising precipitate, the concentrations of the single components in the initial condition as well as the protein concentration was varied in a finescreening process in order to obtain good diffracting single crystals of a suitable size. Finescreening was performed in 24-well sitting-drop as well as in 15-well hanging drop vapor diffusion plates with a reservoir volume of 500 µl and a drop volume between 2 and 5 µl and varying protein:reservoir ratios. The plates were sealed, stored at 4 °C or 20 °C and regularly checked for crystal growth using a binocular microscope.

2.4.3 Crystal harvest and cryoprotection

For cryocrystallization experiments, a single crystal was transferred with a nylon loop into cryo buffer containing the crystallization buffer and varying amounts

of a cryoprotectant reagents (e.g. glycerol). Immediately afterwards, the crystal was flash cooled and stored in liquid nitrogen.

2.4.4 X-ray diffraction experiments

Nylon loops with crystals were mounted on a goniometer head, cooled with a nitrogen gas stream at a temperature of 100 K and exposed to X-rays. Testing of crystals was performed on a MicroMax 007 rotating anode (Rigaku), while data collection of the ctPrp2Δ25 crystals was carried out at Beamline 14.1 BESSY, Berlin.

3 Results

3.1 Purification of Prp2 and Spp2 from *Saccharomyces cerevisiae*

One goal of this work was to establish the purification of a complex comprising the helicase Prp2 and its coactivator Spp2 from *Saccharomyces cerevisiae* (referred to as yPrp2 and ySpp2 in the following).

An overview of the purified constructs is shown in the table below.

Table 3.1: Proteins and protein fragments used for yeast Prp2-Spp2 complex formation. For truncated proteins, the boundaries are indicated in parenthesis behind the name. Mass and length of the proteins are given in kDa and number of amino acids, respectively. Type and position (C or N terminal) of affinity tags are indicated.

Name	Length (aa)	mol. Mass (kDa)	Affinity tag
Prp2	876	99.8	6× His (C)
Prp2 581ΔN (581-876)	296	33.86	6× His (C)
Prp2 597ΔN (597-876)	280	32.09	6× His (C)
Spp2_2 (10-185)	176	19.32	GST (N)
Spp2_2 pET (10-185)	176	19.32	6× His (C)

3.1.1 Purification of full length yPrp2

The DNA fragment coding for yPrp2 was amplified by PCR from a *S. cerevisiae* cDNA library and cloned into the expression vector pET21a. The protein was overexpressed in *E. coli* Rosetta 2(DE3) cells and isolated from the cell lysate by Ni^{2+} affinity chromatography. The eluted fractions were concentrated and applied on a Superdex S200 gel filtration column in a buffer containing 10 mM HEPES pH 7.5, 300 mM NaCl, 2 mM $MgCl_2$, 5% glycerol and 1 mM DTT (chromatogram shown in figure 3.1). Selected fractions were subsequently analyzed by SDS-PAGE.
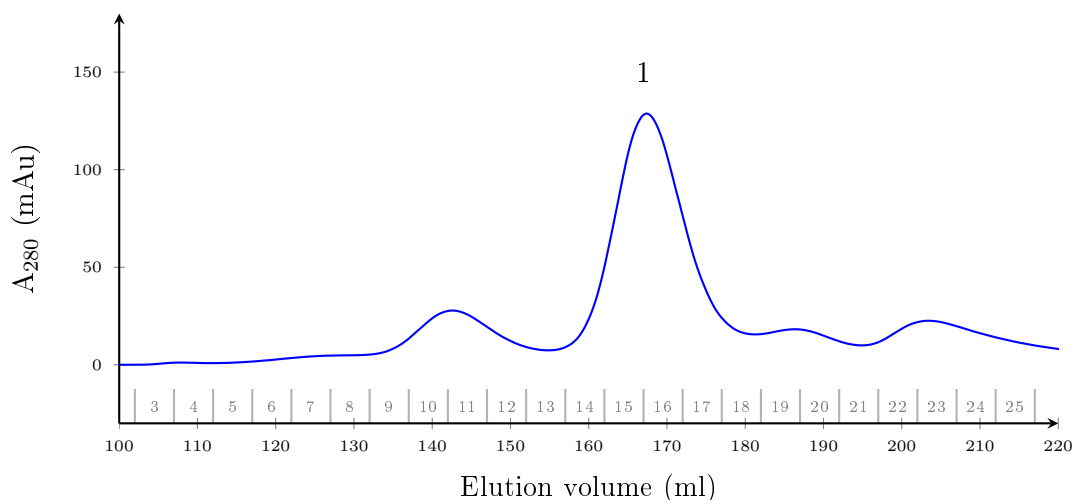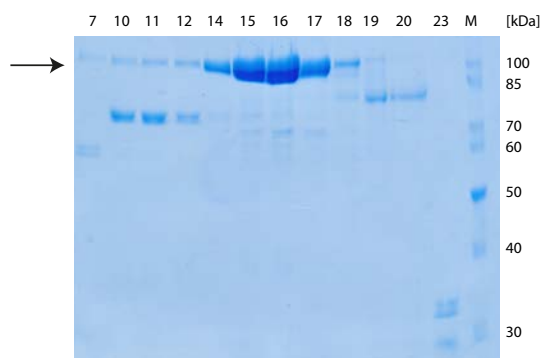
Figure 3.1: Chromatogram of the size exclusion chromatography of yPrp2. The blue line represents the UV absorption at 280 nm plotted against the elution volume. Collected fractions with a volume of 5 ml each are depicted in gray on the x-axis. The local maximum at 167 ml **(1)** corresponds to the elution of yPrp2.

Figure 3.2: Coomassie-stained 12.5% SDS-PAGE containing selected fractions of size exclusion chromatography of yPrp2 (fig. 3.1). Bands corresponding to yPrp2 are marked by an arrow.

The recombinantly expressed Prp2 eluted from the S200 size exclusion chromatography column at a volume of 167 ml, corresponding to the molecular mass of the monomeric protein (99.8 kDa). The total yield was 9.6 mg yPrp2 per liter expression culture.

3.1.2 Purification of ySpp2 (10-185)

Since expression of full length ySpp2 was unsuccessful, N-terminal truncated ySpp2 containing residues 10-185 was used in this work.

The DNA fragment coding for ySpp2 (10-185) was amplified by PCR from a *S. cerevisiae* cDNA library and cloned into the expression vector pGEX-6-P1. The protein was overexpressed in *E. coli* Rosetta 2 (DE3) cells and purified from the cell lysate by GSH-affinity chromatography. The GST-tag was cleaved off

by incubation with PreScission protease over night. Reduced glutathion was removed using a desalting column and cleaved GST was separated from ySpp2 by a second GSH-sepharose chromatography. ySpp2 containing fractions were concentrated and passed over a Superdex S75 gel filtration column in a buffer containing 10 mM HEPES pH 7.5, 150 mM NaCl and 1 mM DTT (chromatogram shown in figure 3.3). Selected fractions were subsequently analyzed by SDS-PAGE.
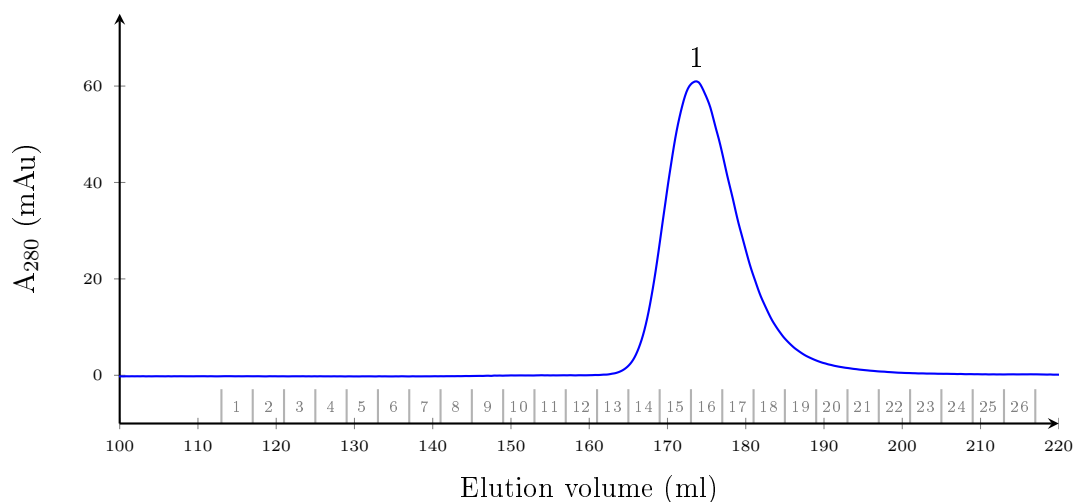

Figure 3.3: Chromatogram of the size exclusion chromatography of ySpp2 (10-185). The blue line represents the UV absorption at 280 nm plotted against the elution volume. Collected fractions with a volume of 4 ml each are depicted in gray on the x-axis. The local maximum at 174 ml **(1)** corresponds to the elution of ySpp2_2.

ySpp2 eluted from the size exclusion chromatography column in a single peak. However, the elution volume of 174 ml corresponds to a molecular mass of about 70 kDa as determined by by gel filtration calibration standard, whereas ySpp2 (10-185) has a mass of 19.32 kDa. It was at this point not clear whether this behavior of Spp2 was caused by oligomerization or if the hydrodynamic radius of the monomeric protein did not correspond to that of a globular protein of the same mass, which might be an indication for the protein to be intrinsically disordered.

The total yield was 18 mg ySpp2 per liter expression culture.

3.1.3 Purification of ^{15}N ySpp2 pET (10-185)

In order to characterize free ySpp2 structurally, NMR experiments were conducted. For that purpose ^{15}N labeled ySpp2 was expressed & purified. In order to make efficient use of the nitrogen source, a construct with a C-terminal hexahis-

Figure 3.4: Coomassie-stained 12.5 % SDS-PAGE containing selected fractions of size exclusion chromatography of ySpp2 (10-185) (fig. 3.3). Bands corresponding to ySpp2 are marked by an arrow.
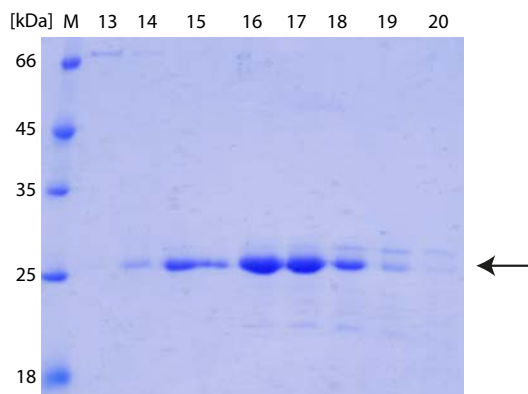
tidin tag was used. The DNA fragment coding for ySpp2 (10-185) was amplified by PCR from the ySpp2_2 (10-185) expression vector and cloned into the expression vector pET21a. The protein was overexpressed in *E. coli* Rosetta 2 (DE3) cells using M9 minimal media supplemented with ^{15}N NH$_4$Cl as nitrogen source and purified from the cell lysate by Ni^{2+} affinity chromatography. The eluted fractions were concentrated and applied onto a Superdex S200 in a buffer containing 50 mM Na$_2$HPO$_4$/NaH$_2$PO$_4$ pH 6.8, 200 mM NaCl, 1 mM DTT.
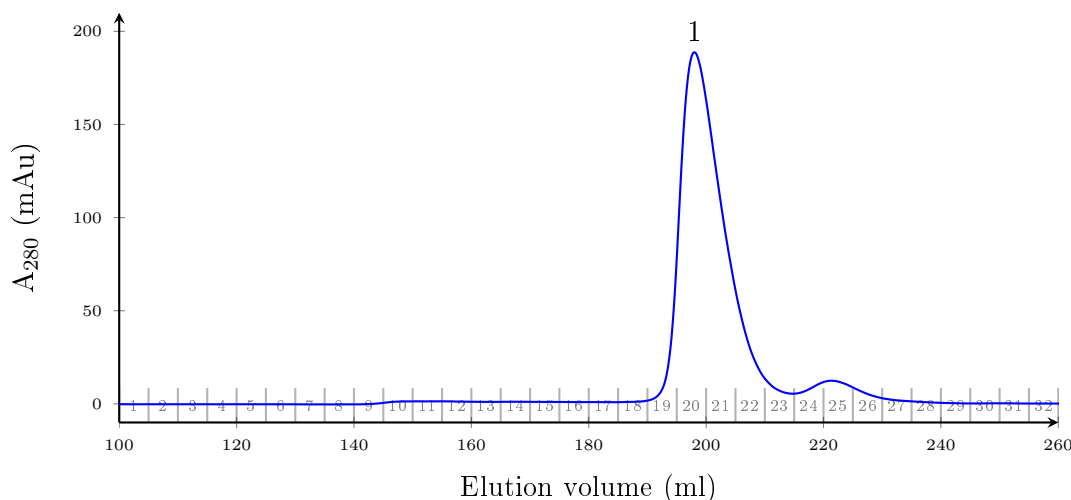
Figure 3.5: Chromatogram of the size exclusion chromatography of histidine-tagged ^{15}N labeled ySpp2 (10-185). The blue line represents the UV absorption at 280 nm plotted against the elution volume. Collected fractions with a volume of 5 ml each are depicted in gray on the x-axis. The peak at 198 ml **(1)** corresponds to the elution of ySpp2 (10-185).

ySpp2 pET eluted from the size exclusion chromatography column in a single peak at 198 ml, an elution volume which would correspond to a molecular mass of around 70 kDa. The same observation was already made during purification of ySpp2 (10-185) which was expressed as GST-fusion protein (compare 3.1.2.
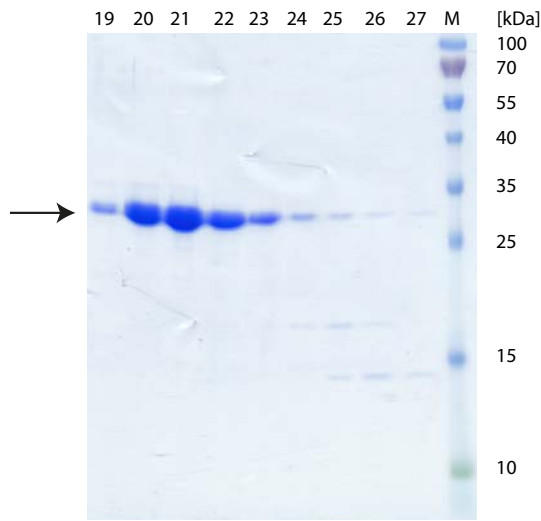
Figure 3.6: Coomassie-stained 12.5% SDS-PAGE containing selected fractions of size exclusion chromatography of histidine-tagged ^{15}N ySpp2 (10-185) (3.5). Bands corresponding to ySpp2 are marked by an arrow

ySpp2 containing fractions 19-23 were pooled and concentrated to 500 µM for use in NMR experiments.

3.2 Preparation of the yPrp2-ySpp2 complex

One major goal of this work was the preparation of a functional Prp2-Spp2 complex in order to investigate the effect of Spp2 on the NTPase and helicase activity of Prp2 as well as for crystallization trials. To see, if both proteins form a stable complex *in vitro* and to prepare a complex suitable for crystallization and biochemical studies, yPrp2 was mixed with a 2-fold molar excess of Spp2 to ensure, that Prp2 was completely saturated. The protein sample was incubated for 30 minutes on ice and afterwards applied to Superdex S200 size exclusion chromatography column in a buffer containing 10 mM HEPES pH 7.5, 200 mM NaCl, 5% glycerol and 1 mM DTT (fig. 3.7). Collected fractions were analyzed by SDS-PAGE.

SDS-PAGE analysis of the peak fractions from size exclusion chromatography showed, that both proteins eluted at the same volume in a single peak, indicating formation of a stable and homogenous complex at the given conditions. Moreover, remaining unbound Spp2 indicates that saturation of Prp2 was achieved.

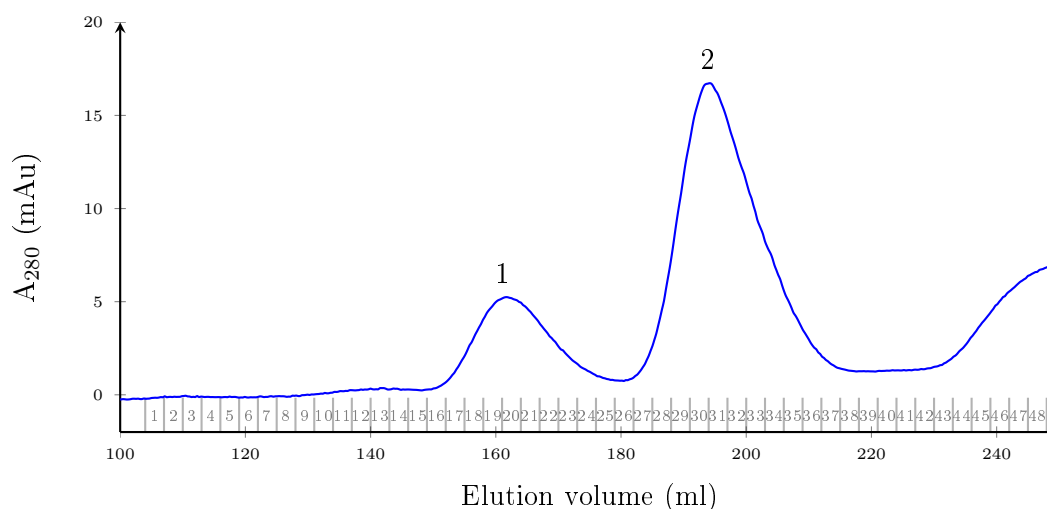The purified complex was concentrated to 4 mg/ml and used for crystallization trials.

Figure 3.7: Chromatogram of the size exclusion chromatography of the yPrp2-ySpp2 complex preparation. The blue line represents the UV absorption at 280 nm plotted against the elution volume. Collected fractions with a volume of 3 ml each are depicted in gray on the x-axis. The local maximum at 161 ml **(1)** corresponds to the elution of the complex, the peak at 194 ml **(2)** contains Spp2 which was added in excess.
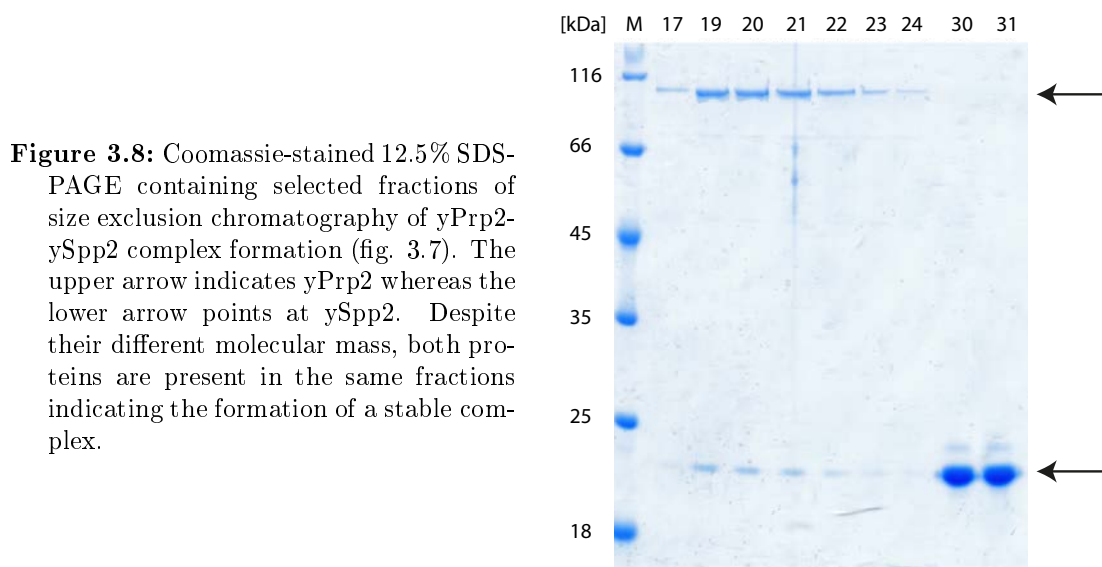

Figure 3.8: Coomassie-stained 12.5 % SDS-PAGE containing selected fractions of size exclusion chromatography of yPrp2-ySpp2 complex formation (fig. 3.7). The upper arrow indicates yPrp2 whereas the lower arrow points at ySpp2. Despite their different molecular mass, both proteins are present in the same fractions indicating the formation of a stable complex.

3.3 Characterization of Spp2

Since very little was known about the biochemical properties of Spp2 and the role and function of G-patch protein in general, beside crystallization trials, additional methods were used for structural characterization of Spp2.

3.3.1 CD spectroscopy

To get information about the secondary structure composition of ySpp2, circular dichroism spectroscopy measurements (see 2.3.2) were performed. Since Cl⁻ ions show high absorption in the far UV-range and therefore interfere with CD measurements, the protein solution was dialysed against a buffer containing only 100 mM Na_2HPO_4/NaH_2PO_4 at pH 7.5. Measurements were performed at 20 °C at wavelengths between 190 and 260 nm.
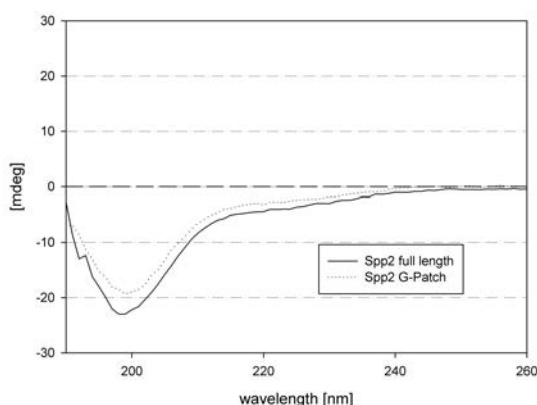

Figure 3.9: Circular dichroism spectra of Spp2 full length and an Spp2 truncation containing only the G-patch. Both spectra show no significant presence of secondary structure elements.

Comparison of the recorded spectra for ySpp2 as well as for the G-patch with reference spectra for helix, sheet and turn elements of secondary structure (see A.4) led to the conclusion, that Spp2 might not contain any secondary structure elements.

3.3.2 NMR spectroscopy

Since CD spectroscopy of Spp2 suggested that the protein does not contain secondary structure elements, further investigation on the folding state of Spp2 was carried out using NMR spectroscopy. Therefore, ^{15}N labeled ySpp2 (see 3.1.3) was concentrated to a concentration of 500 µM and an HSQC spectrum of the sample was measured. This work was performed by Shenqi Xiang from the department of NMR-based Structural Biology, Max-Plank-Institute for Biophysical

Chemistry, Göttingen. The HSQC-spectrum of ^{15}N labeled ySpp2 showed a poor

Figure 3.10: HSQC-spectrum of ySpp2. The spectra shows a very narrow chemical shift dispersion along with broad peaks, especially for the amide protons.
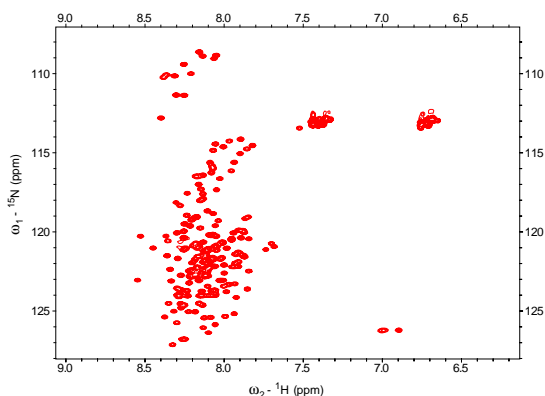
chemical shift dispersion and clustering of broad peaks around 8.3 ppm of the ^{1}H frequency which is typical for the spectra of unfolded proteins. Since there was no precipitation during sample preparation, ySpp2 was considered to be intrinsically unfolded.

3.4 Purification of Prp2 & Spp2 from *Chaetomium thermophilum*

Since crystallization of yeast Prp2 and Spp2 was not successful neither alone nor in a complex comprising both proteins, further crystallization trials were performed with orthologs of both proteins from the thermophilic fungus *Chaetomium thermophilum*.

The genome of this soil inhabiting fungus was published in 2011 (Amlacher *et al.*, 2011) and since proteins from thermophilic organisms are thought to me more rigid and therefore promising targets for crystallization trials, efforts were made to identify orthologs of Prp2 and Spp2 in this organism.

3.4.1 Identification of Prp2 & Spp2 orthologs

In order to identify the Prp2 ortholog in *C. thermophilum*, a protein-protein BLAST search (Altschul *et al.*, 1990) was performed, aligning Prp2 against all known proteins of *C. termophilum DSM 1495*. In this alignment, the highest score was obtained for a protein with the GenBank accession ID EGS18341.1, annotated as "putative mRNA splicing factor". The sequence identity between

the two proteins is 50% and the the similarity 64%, thus the protein was considered to be the ortholog of yPrp2 in *C. thermophilum*. The sequence alignment of yPrp2 and ctPrp2 shows a high degree of sequence similarity, especially within the RecA-like domains but also within the HA-domain and the OB-fold, whereas the N-terminal part shows a higher degree of variability as well as a 53 aa long insertion for ctPrp2 (for alignment see A.1). The most remarkable feature is a 502 aa long extension of ctPrp2 at the C-terminus, which also contains a G-patch motif and is separated from the helicase domain by in total 4 introns on the genomic DNA. However, amplification of the coding sequence from genomic DNA was only possible for a fragment lacking the C-terminal extension, or for the C-terminal extension alone. Therefore one has to consider, that the annotation of the genomic DNA in the database might be wrong in that point.

For Spp2, a first protein-protein BLAST search with the yeast Spp2 sequence resulted in no significant hits. Therefore, in an second trial, only amino acids 180 - 210 (resembling the last two thirds of the G-patch) of the human Spp2 ortholog GPKOW (UniProt ID Q92917) were used as search query against the group of fungi. The result with the highest score was found in *Neurospora crassa* and annotated as "hypothetical protein NCU07862" (Accession No. XP_963123). This result was used as query in a second BLAST run against all *C. thermophilum* sequences, were a protein annotated as "putative pre-mRNA splicing protein" could be identified (Accession No. EGS17798.1). In this case, the sequence identity between the two full length proteins was 41% and the the similarity 53%, the identified protein was therefore a good candidate for a *C. thermophilum* Spp2 ortholog.

Table 3.2: Proteins and truncations used for ctPrp2-ctSpp2 complex formation, for truncated proteins the boundaries are indicated in parenthesis behind the name. Mass and length of the proteins are given in kDa and number of amino acids, respectively. Type and position (C or N terminal) of affinity tags are indicated in the last column.

Name	Length (aa)	Mass (kDa)	Affinity tag
ctPrp2 (1-921)	921	104.6	GST (N)
ctPrp2 (270-921)	652	72.5	GST (N)
ctSpp2	313	36.22	GST (N) & Strep (C)
ctSpp2 (211-254)	44	4.94	GST (N)

3.4.2 Purification of ctPrp2 (1-921)

For expression, the DNA fragment coding for ctPrp2 (1-921) was amplified by PCR from *C. thermophilum* genomic DNA and cloned into the expression vector pASG-IBA25 (see 2.2.6.2). Overexpression was performed in *E. coli* Rosetta 2 (DE3) cells and the expressed GST-fusion protein was subsequently purified from the cell lysate by GSH affinity chromatography. The protein which was eluted from the column was concentrated and applied onto a Superdex S200 gel filtration column in a buffer containing 10 mM HEPES pH 7.5, 200 mM NaCl, 2 mM $MgCl_2$, 5% glycerol and 1 mM DTT. Selected fractions were subsequently analyzed by SDS-PAGE. The overexpressed protein eluted from the gel filtration column in the void volume, indicating aggregation of the protein. The A_{260}/A_{280} ratio was 1.8 (2.2.13), indicating that nucleic acids bound to the protein caused the aggregation. Several steps were performed to remove the nucleic acids from the protein, including washing of the GSH-sepharose immobilized GST-ctPrp2 fusion protein with 2 M LiCl during affinity chromatography as well as incubation of the purified protein with RNAse A and/or Benzonase. However, none of these steps led to protein free of bound nucleic acids, instead the protein still aggregated during gel filtration chromatography.

3.4.3 Purification of ctPrp2 (270-921)

Since it was not possible to purify full length ctPrp2 to homogeneity, a truncated construct of ctPrp2 lacking the first 269 amino acids was designed. The truncated residues are not part of the helicase core and have been shown to be dispensable for the function of yPrp2 *in vivo* (Edwalds-Gilbert *et al.*, 2004). The DNA fragment coding for truncated ctPrp2 (270-921) was amplified by PCR from *C. thermophilum* genomic DNA and cloned into the expression vector pASG-IBA25 (see 2.2.6.2). ctPrp2 (270-921) was overexpressed as GST-fusionprotein in *E. coli* Rosetta 2 (DE3) cells and was subsequently isolated from the cell lysate by GST affinity chromatography. After elution, the fusion-protein containing fractions were concentrated and the GST-tag was cleaved off by treatment with PreScission protease. The protein solution was then applied onto a Superdex S200 gel filtration column in a buffer containing 10 mM HEPES pH 7.5, 200 mM NaCl, 2 mM $MgCl_2$, 5% (v/v) glycerol and 1 mM DTT (chromatogram shown in figure 3.11). Selected fractions were subsequently analyzed by SDS-PAGE.

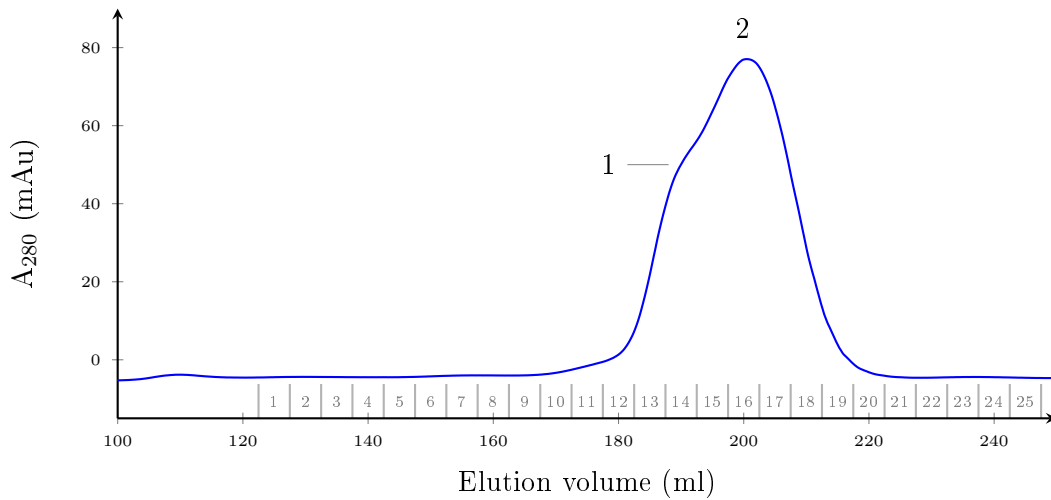ctPrp2 (270-921) eluted from the gel filtration column at a volume of 188 ml

Figure 3.11: Chromatogram of the size exclusion chromatography of ctPrp2 (270-921). The blue line represents the UV absorption at 280 nm plotted against the elution volume. Collected fractions with a volume of 5 ml each are depicted in gray on the x-axis. Elution of ctPrp2 can be observed at 188 ml **(1)** corresponding to a a molecular mass of about 70 kDa, whereas the cleaved GST dimer elutes at 200 ml **(2)**.

which agrees well with its monomeric size of 72 kDa according to the size standard calibration of the column. A second GSH affinity chromatography was performed to remove cleaved GST. The protein could be purified to homogeneity with a total yield of 25 mg per liter expression culture.

3.4.4 Purification of ctSpp2

For expression of ctSpp2, the coding DNA fragment was amplified by PCR from *C. thermophilum* genomic DNA and cloned into the expression vector pASG-IBA25. Overexpression of the GST-fusion protein was performed in *E. coli* Rosetta 2
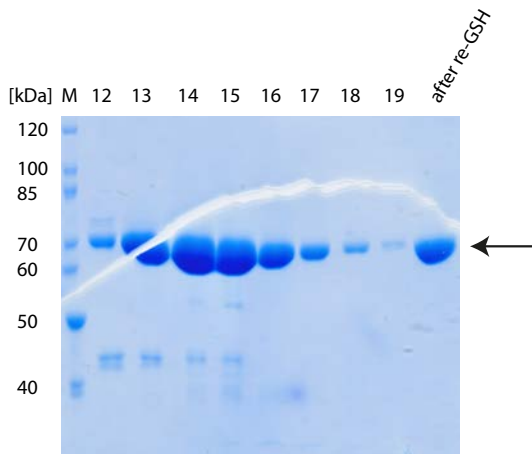

Figure 3.12: Coomassie-stained 12.5% SDS-PAGE containing selected fractions of size exclusion chromatography and of second GSH affinity chromatography of ctPrp2 (270-910). Bands corresponding to ctPrp2 are marked by an arrow.

(DE3) cells. The protein of interest was purified from the cell lysate by GST affinity chromatography, followed by treatment with PreScission protease to cleave off the GST-tag. Further purification of the protein was achieved by a Superdex S75 gel filtration column in a buffer containing 10 mM HEPES pH 7.5, 150 mM NaCl and 1 mM DTT (chromatogram shown in figure 3.13). Peak fractions of the size exclusion chromatography were subsequently analyzed by SDS-PAGE. A second GSH affinity chromatography was performed to remove cleaved GST. The protein could be purified with a total yield of 16 mg per liter expression culture.

During size exclusion chromatography, ctSpp2 showed a similar behavior as observed during the purification of ySpp2, namely that the elution volume does not correspond to the molecular mass of a monomeric protein. Since it was already shown that ySpp2 is intrinsically disordered in solution (see 3.3.2), one could suggest the same properties for ctSpp2.
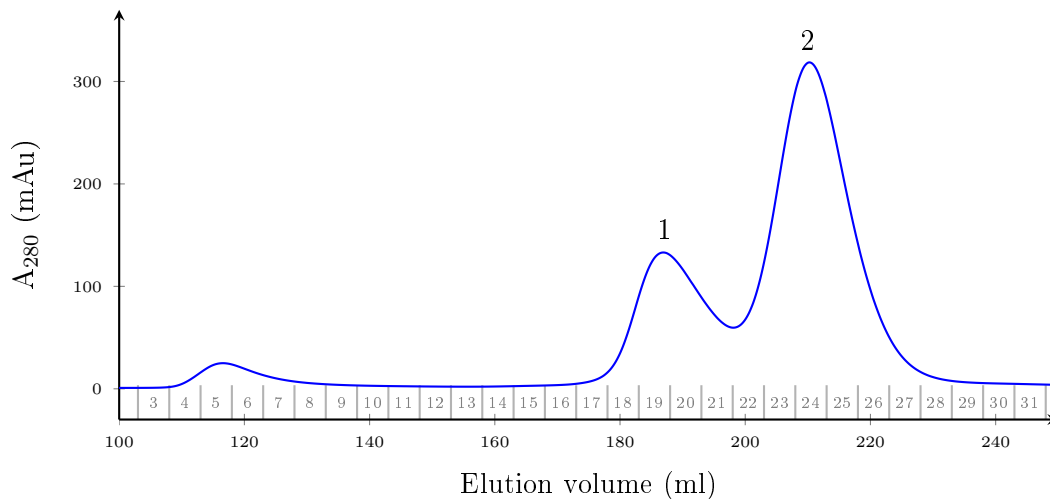

Figure 3.13: Chromatogram of the S75 size exclusion chromatography of ctSpp2. The blue line represents the UV absorption at 280 nm plotted against the elution volume. Collected fractions with a volume of 5 ml each are depicted in gray on the x-axis. Elution of ctSpp2 can be observed at 186 ml (**1**), whereas the cleaved GST dimer elutes at 210 ml (**2**).

A second GSH affinity chromatography was performed to remove cleaved GST. The protein could be purified to homogeneity with a total yield of 16 mg per liter expression culture.

3.4.5 ctPrp2-ctSpp2 complex formation

In order to test whether purified ctPrp2 (270-921) binds to purified ctSpp2, a pulldown assay with GST-tagged ctSpp2 as bait and ctPrp2 (270-921) acting
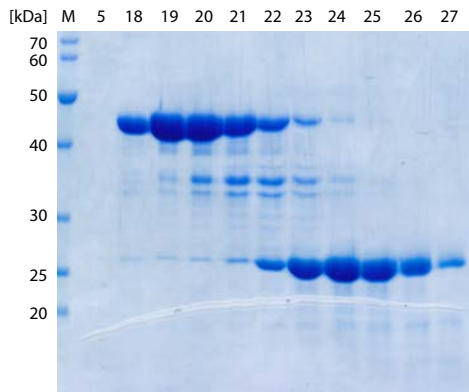
Figure 3.14: Coomassie-stained 12.5% SDS-PAGE containing selected fractions of size exclusion chromatography of ctSpp2 (fig. 3.13). Bands corresponding to ctSpp2 are marked by an arrow, the bands running at a molecular mass of 25 kDa represent GST.

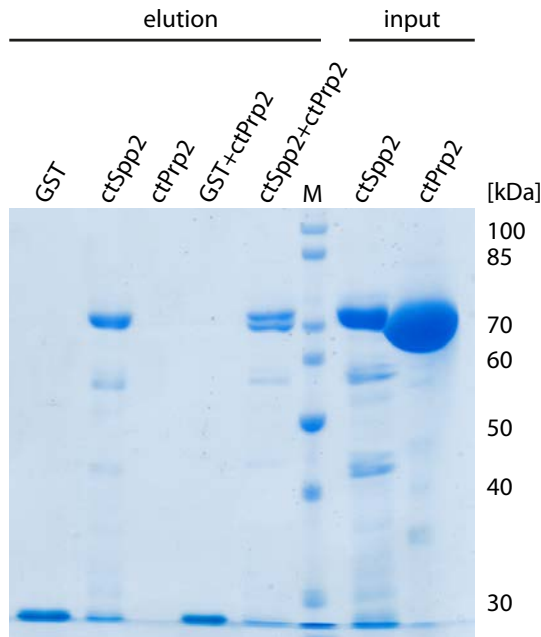as pray was performed on GSH sepharose beads. Samples were subsequently analyzed by SDS-PAGE.

Figure 3.15: SDS-PAGE of ctPrp2-ctSpp2 pulldown assay. Only in presence of ctSpp2, ctPrp2 was pulled out of solution and could be eluted from the beads (lane left of the marker), whereas no ctPrp2 could be eluted in the controls containing only the beads or immobilized GST.

In the pulldown assay (fig.3.15), an interaction between ctPrp2 and ctSpp2 is clearly visible. This also strongly indicates, that the identified and purified Spp2-candidate is in fact the Spp2 ortholog.

For following crystallization trials, the complex was prepared using a co-purification strategy which would ensure that only a completely saturated complex will be present in the sample for crystallization. In order to achieve this, the two proteins were expressed separately as GST- and Strep tagged fusion proteins, respectively. Both cell pellets were mixed prior to cell lysis and the complex was purified from the supernatant by GST affinity chromatography. The GST-tag

was cleaved using PreScission protease and the sample was further purified by Streptactin affinity chromatography. Finally, the complex was applied onto an S200 gel filtration column in a buffer containing 10 mM HEPES pH 7.5, 200 mM NaCl, 2 mM MgCl$_2$, 5% (v/v) glycerol and 1 mM DTT (fig 3.16). Fractions from SEC were subsequently analyzed by SDS-PAGE (fig 3.17).
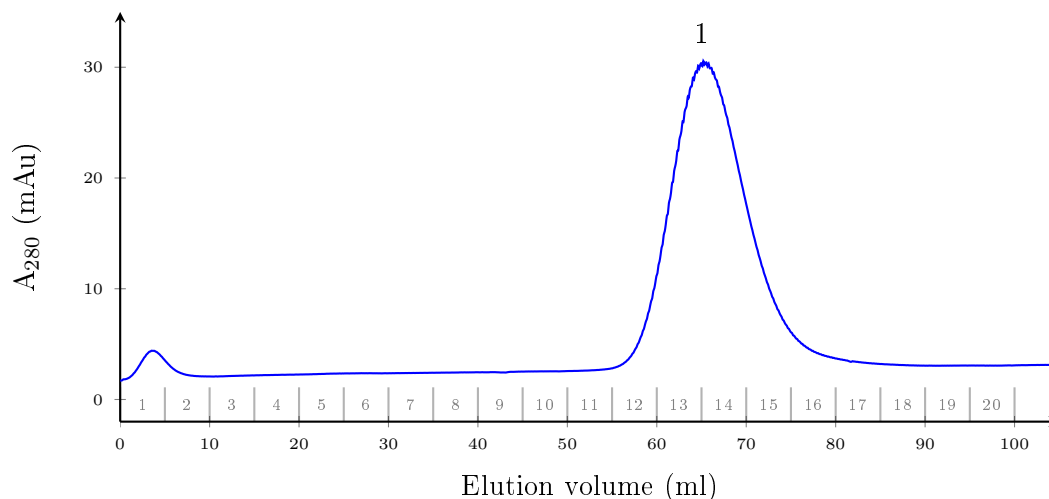

Figure 3.16: Chromatogram of the final 16/60 Superdex 75 size exclusion chromatography of the ctPrp2-ctSpp2 co-purification. The blue line represents the UV absorption at 280 nm plotted against the elution volume. Collected fractions with a volume of 5 ml each are depicted in gray on the x-axis. The local maximum at 65 ml **(1)** corresponds to the elution of the complex with a molecular mass of about 104 kDa.

Figure 3.17: SDS-PAGE of the final size exclusion chromatography step of ctPrp2-ctSpp2 co-purification. The upper band corresponds to ctPrp2 whereas the lower band represents ctSpp2.
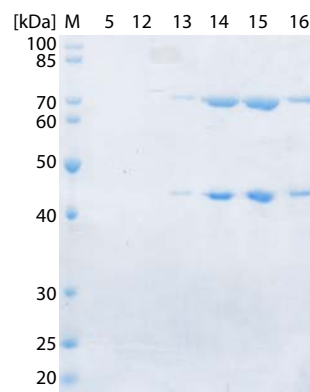
Using this approach, the complex could be purified to homogeneity with a yield of 5 mg/liter expression culture and was used for crystallization trials.

3.4.6 Purification of ctSpp2 (211-254)

Computational analysis of the ctSpp2 amino acid sequence revealed a high degree of disorder (see figure 3.18), a fact which has also been observed for ySpp2 (see section 3.3.1). Since disordered proteins might not be promising targets for crystallization, a shorter truncation was designed based on the ctSpp2 disorder prediction containing only the G-patch motif.
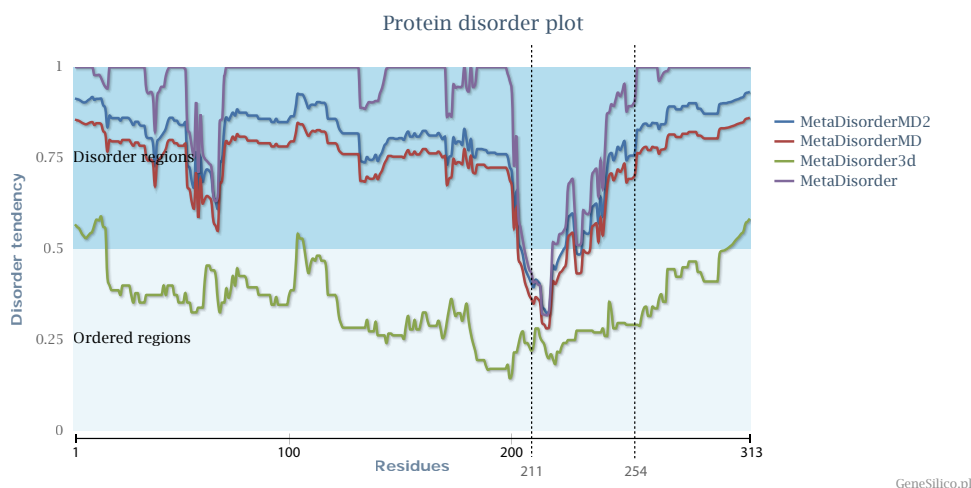

Figure 3.18: Disorder prediction of ctSpp2, calculated by GeneSilico Meta server (Kozlowski and Bujnicki, 2012). Except for a small part of the protein containing the G-patch motif, ctSpp2 is predicted to be completely disordered, the prediction is consistent throughout all used algorithms. The borders of the designed ctSpp2 truncation is indicated in the plot by dotted lines along with the number of the N- and C-terminal amino acid.

For expression of ctSpp2 (211-254), the coding DNA fragment was amplified by PCR from the full length ctSpp2 expression plasmid and cloned into the expression vector pASG-IBA25. The proteins was overexpressed in *E. coli* Rosetta 2 (DE3) cells and isolated from the cell lysate by GST affinity chromatography. The eluted fractions were concentrated, GST was cleaved using PreScission protease and the protein was further purified by a Superdex S75 gel filtration column in a buffer containing 10 mM HEPES pH 7.5, 150 mM NaCl and 1 mM DTT (chromatogram shown in figure 3.19). Selected fractions were subsequently analyzed by SDS-PAGE.

ctSpp2 eluted at a volume of 242 ml, corresponding to a molecular mass of about 10 kDa, while its calculated molecular mass is only 4.94 kDa. This effect has also been observed for full length ctSpp2 as well as of ySpp2 (compare 3.4.4 and 3.1.2). The similar behavior of the proteins show that the intrinsic disorder is not only a property of the full length protein but also of the G-patch motif itself.
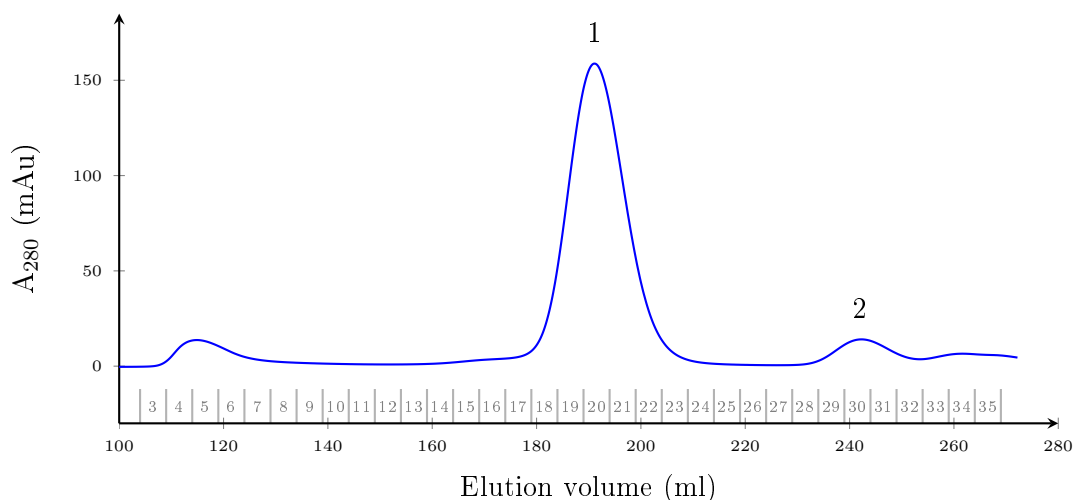
Figure 3.19: Chromatogram of the size exclusion chromatography of ctSpp2 (211-254). The blue line represents the UV absorption at 280 nm plotted against the elution volume. Collected fractions with a volume of 5 ml each are depicted in gray on the x-axis. The local maxima at 190 ml **(1)** and 240 ml **(2)** correspond to the elution of GST and ctSpp2 (211-254) respectively.

ctSpp2 (211-254) could be purified to homogeneity with a yield of 0.5 mg per liter expression culture.

3.4.7 ctPrp2-ctSpp2 (211-254) complex preparation

For preparation of the ctPrp2-ctSpp2 (211-254) complex, ctSpp2 (211-254) was added to ctPrp2 in a 2 fold molar excess, and the mixture was incubated on ice for 30 minutes. The complex was then purified by size exclusion chromatography using a Superdex S75 26/60 column in a buffer containing 10 mM HEPES pH 7.5, 200 mM NaCl and 1 mM DTT (fig. 3.21).

A stable complex consisting of ctPrp2 (270-921) and the G-patch motif of ctSpp2 (211-254) was formed and could by purified by the size exclusion chromatographic step. This is in contrast to the yeast proteins, were the isolated G-patch was not sufficient to form a stable complex with yPrp2 after the final gel filtration step. The complex was of high purity and suitable for crystallization trials.
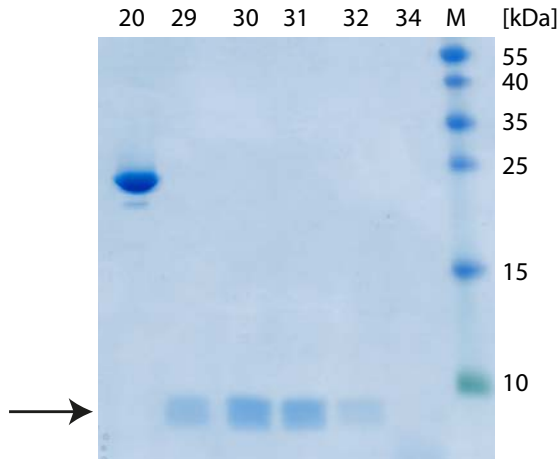
Figure 3.20: Coomassie-stained 15% SDS-PAGE containing selected fractions of size exclusion chromatography of ctSpp2 (211-254). Bands corresponding to ctSpp2 are marked by an arrow. Fractions 20 contains the GST dimer, whereas ctSpp2 elutes in fractions 29-32.
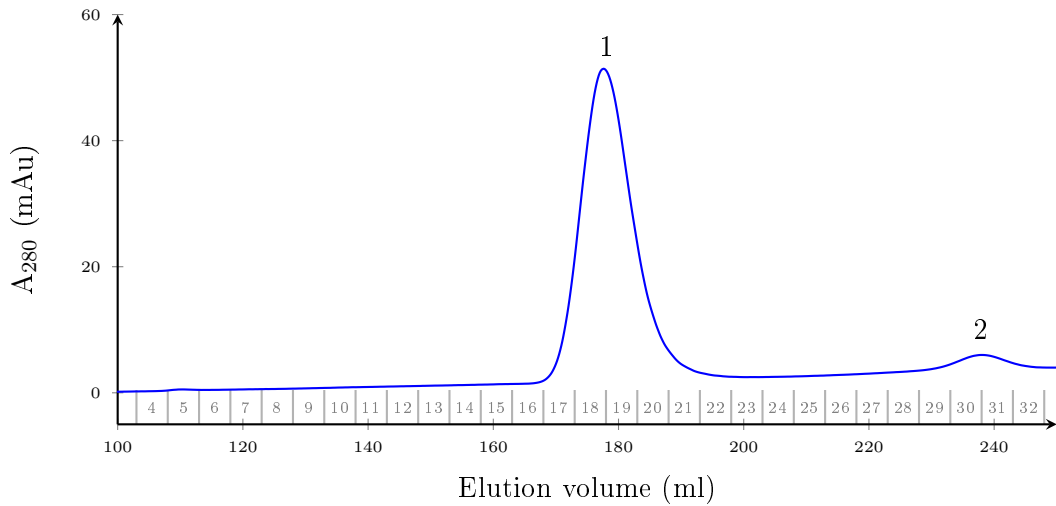
Figure 3.21: Chromatogram of the size exclusion chromatography of the ctPrp2-ctSpp2 (211-254) complex. The blue line represents the UV absorption at 280 nm plotted against the elution volume. Collected fractions with a volume of 5 ml each are depicted in gray on the x-axis. The local maximum at 178 ml **(1)** and 238 ml **(2)** correspond to the elution of the ctPrp2-ctSpp2 (211-254) complex and ctSpp2 (211-254) respectively.

Figure 3.22: Coomassie-stained 15% SDS-PAGE of selected fractions from ctPrp2-ctSpp2 (211-254) size exclusion chromatography. The upper arrow indicates ctPrp2 whereas the lower arrow marks ctSpp2 (211-254). The complex elutes in fractions 17-20, excess of ctSpp2 (211-254) is present in fractions 30 and 31.
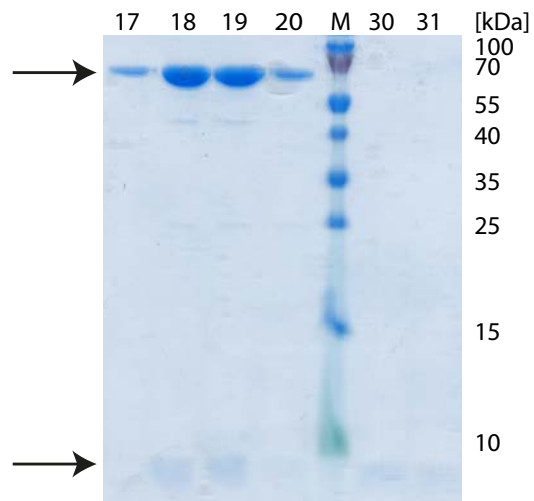
3.5 Crystallization

3.5.1 Crystallization of ctPrp2 (270-921)

3.5.1.1 ADP bound state

Initial crystallization screening was performed with a protein concentration of 5 mg/ml and a 5× molar excess of ADP. A set of nine screens (2.1.12) was tested at 20 °C and 4 °C. First crystals could be observed after 48 hours at 20 °C in various conditions containing different PEGs, salts at concentrations between 100 and 200 mM and pH values between 6 and 7.5. The initial conditions were optimized to a final condition containing 0.1 M Tris pH 7.5, 200 mM NaCl and 15% PEG 8000 (see figure 3.23).
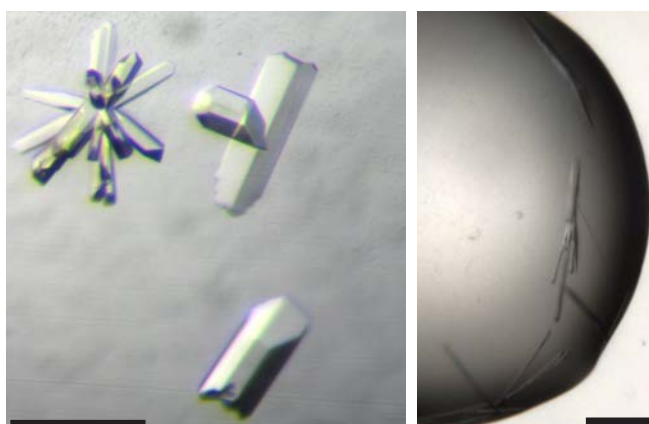

Figure 3.23: Protein crystals of ctPrp2 (270-921) in the presence of ADP. The crystals were grown in a condition containing 0.1 M Tris pH 7.5, 200 mM NaCl and 15% PEG 8000. The reference bar corresponds to a length of 200 µm.

The crystals were cryoprotected in reservoir solution supplied with 15% glycerol and 5% PEG 400 and flash cooled in liquid nitrogen. ctPrp2 (270-921) crystallized in the orthorombic spacegroup P$2_1 2_1 2_1$ and diffracted to a resolution of 2.1 Å. A complete dataset was measured at beamline 14.1 of BESSY, Berlin. The statistics of data collection and structure refinement are shown in table 3.3. Since soaking of the crystals in KBr and AuCl$_3$ for phasing resulted in no significant anomalous signal, the crystallographic phase problem was solved by molecular replacement using PHASER (McCoy et al., 2007) and the structure of yeast Prp43 (PDB ID: 2XAU) as search model.

The structure was refined by iterative cycles of CNS (Brünger et al., 1998) and manual model building in Coot. The final round of refinement was performed with PHENIX (Adams et al., 2002) to overall R$_{work}$/R$_{free}$ values of 18.74 and 22.84 % respectively (complete data statistics for refinement are shown in table 3.3). The final model of ctPrp2 (270-921) contains one molecule in the asymmetric unit and

encompasses residues 283-921. Additional density was observed in the nucleotide binding cleft were an ADP molecule as well as a magnesium ion is present.

Table 3.3: Data collection and refinement statistics of ADP bound ctPrp2 (270 - 921). Values in parentheses refer to the highest resolution shell, Ramachandran statistics were calculated with Molprobity.

Data collection		Refinement	
Spacegroup	$P2_12_12_1$	Resolution range (Å)	46.01 - 2.10 (2.15 - 2.10)
Cell dimensions a,b,c (Å)	50.26, 114.21, 117.84	No. of Reflections	39995
Wavelength (Å)	0.920	R_{work} (%)	18.78 (24.67)
X-ray source	BL 14.1 BESSY (Berlin)	R_{free} (%)	22.57 (31.34)
Resolution range (Å)	50.0 - 2.10 (2.20 - 2.10)	Total number of atoms	5431
No. of reflections	150725	Protein atoms	4990
Completeness (%)	99.7 (99.0)	Water molecules	404
R_{merge} (%)	7.6 (51.1)	*av. B-factors*	
Average I/σ	15.89 (3.04)	overall (Å^2)	32.37
Wilson B-factor (Å^2)	31.48	Protein (Å^2)	31.84
Redundancy	3.7 (3.7)	solvent (Å^2)	37.75
		RMSD	
		bond lengths (Å)	0.003
		bond angles (°)	0.735
		Ramachandran plot	
		favored (%)	97.31
		allowed (%)	2.53
		outliers (%)	0.16

3.5.1.2 Co-crystallization with ATP analoga

For co-crystallization of ctPrp2 (270-921) with different nucleotides (e.g. AMPPNP, AMPPCP), co-purified ADP had to be removed from the protein. This was achieved by addition of 10 mM EDTA to the protein followed by gel filtration chromatography in a buffer containing 10 mM HEPES pH 7.5, 200 mM NaCl, 2 mM, 5% (v/v) glycerol and 1 mM DTT supplemented with 10 mM EDTA. In a final step, the EDTA was removed by a desalting column. To check whether the ADP was removed quantitatively, the nucleotide binding state was determined using a Nucleosil column (see 2.2.11.5).

By comparison with the elution profile of ADP, the chromatogram of the Nucleosil column shows that after treatment with EDTA, ctPrp2 (270-921) is in a nucleotide-free state (see fig. 3.24).

After a nucleotide free ctPrp2 had been purified, crystallization trials were performed with a protein concentration of 4 mg/ml in presence of 5× molar excess of the non-hydrolyzable ATP analogues AMPPNP and AMPPCP respectively. Crystals could be obtained in a condition containing 0.1 M Tris pH 7.5, 200 mM
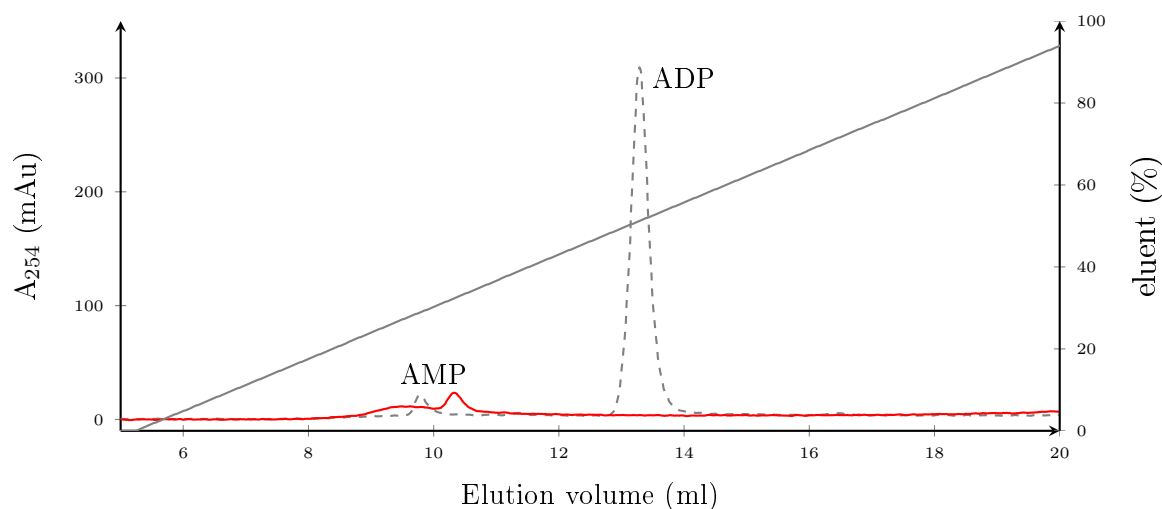
Figure 3.24: Nucleotide loading state of ctPrp2 after treatment with EDTA. In the chromatogramm, absorption at 254 nm is plotted against the concentration of the eluent. As standard, an ADP sample was loaded (gray dashed line), which shows a small peak for AMP at 9.8 ml and a major peak corresponding to ADP at 13.5 ml. For the ctPrp2 sample (red line) only a very small peak can be observed at 10.3 ml.

Li_2SO_4 and 15% PEG 8000 . The crystals were cryoprotected in reservoir solution supplied with 15% glycerol and 5% PEG 400 and flash cooled in liquid nitrogen. Several datasets were measured at beamline 14.1 of BESSY, Berlin. After refinement against the already obtained structure of ctPrp2 bound to ADP it was revealed, that the protein which was crystallized in presence of AMPPNP contained an ADP molecule in the nucleotide binding cleft. Further analysis of the nucleotide loading state of ctPrp2 incubated with AMPPNP using a Nucleosil column showed that ADP was bound to the protein, although ADP was neither present in the protein sample before incubation with AMPPNP nor in the AMPPNP solution (see Fig 3.25). It is therefore likely, that ctPrp2 (270-921) is in fact able to hydrolyze AMPPNP.

3.5.1.3 Apo state

In the refined structure obtained from datasets of ctPrp2 (270-921) crystallized in the presence of AMPPCP, no nucleotide was bound in the nucleotide binding cleft. Instead, a sulfate ion was present at a position, which harbors the β phosphate in the ADP bound state.

The apo structure of ctPrp2 (270-921) was refined to overall R_{work}/R_{free} values of 21.95 and 27.29 % (complete data statistics for refinement are shown in table 3.4). In the electron density map, the residues 283 - 921 could be traced, while
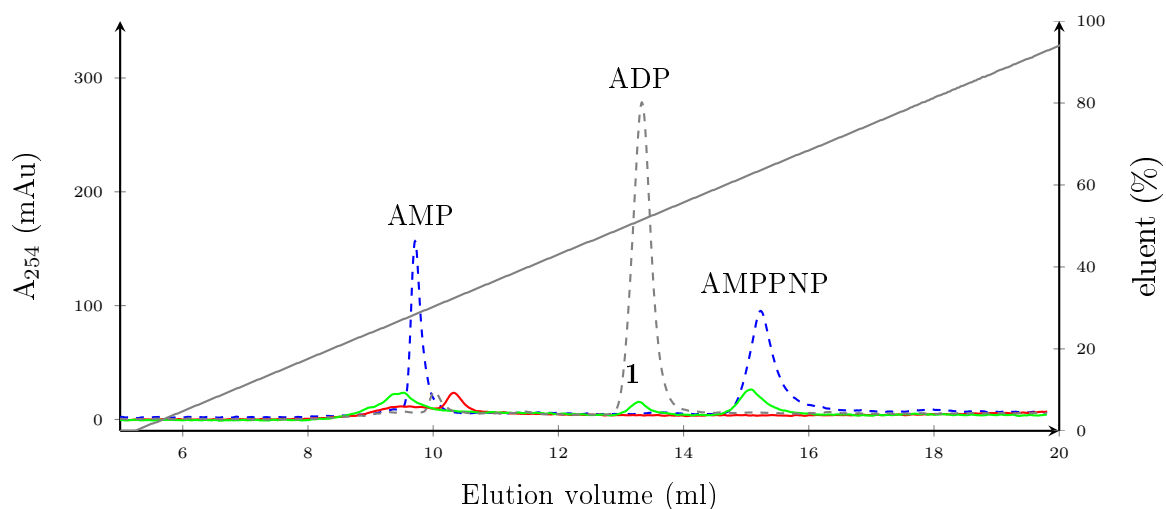
Figure 3.25: Nucleotide loading state of ctPrp2 before (red line) and after (green line) incubation with AMPPNP. In the chromatogram, absorption at 254 nm is plotted against the concentration of the eluent. As standard, an ADP (gray dashed line) as well as an AMPPNP (blue dashed line) sample was loaded. AMPPNP elutes at 15.2 ml, the peak corresponding to ADP is observed at 13.5 ml. After incubation of ctPrp2 with AMPPNP, a peak elutes at a volume corresponding to ADP (**1**, in green), indicating that AMPPNP was hydrolyzed to ADP by ctPrp2.

residues 601 - 603 located in a flexible loop region are not defined. The overall shape of the protein closely resembles that of ctPrp2 (270-921) in the ADP bound state.

Table 3.4: Data collection and refinement statistics of apo ctPrp2 (270-921). Values in parentheses refer to the highest resolution shell, Ramachandran statistics were calculated with Molprobity.

Data collection		Refinement	
Spacegroup	$P2_12_12_1$	Resolution range (Å)	41.46 - 2.7 (2.847 - 2.7)
Cell dimensions a,b,c (Å)	50.61, 114.13, 120.69	No. of Reflections	19767
Wavelength (Å)	0.920	R_{work} (%)	21.95 (30.38)
X-ray source	BL 14.1 BESSY (Berlin)	R_{free} (%)	27.29 (33.22)
Resolution range (Å)	50.0 - 2.7 (2.7 - 2.8)	Total number of atoms	5016
No. of reflections	95840	Protein atoms	4972
Completeness (%)	99.8 (99.9)	Water molecules	44
R_{merge} (%)	7.1 (64.2)	*av. B-factors*	
Average I/σ	22.66 (3.34)	overall (Å^2)	68.04
Wilson B-factor (Å^2)	56.85	Protein (Å^2)	68.21
Redundancy	4.8 (4.93)	solvent (Å^2)	42.40
		RMSD	
		bond lengths (Å)	0.003
		bond angles (°)	0.749
		Ramachandran plot	
		favored (%)	94.3
		allowed (%)	5.06
		outliers (%)	0.63

3.6 Structural analysis of ctPrp2

Analysis of the crystal structures of ctPrp2 (270-921) in the ADP-bound as well as in the apo state reveals that ctPrp2 adopts a conformation typical for DEAH-box helicases. The full length protein can be divided into six domains, five of which are present in the structure while the N-terminal domain is not part of the truncated protein used for crystallization. The structure starts at the N-terminus with the two RecA like domains (referred to as RecA1 and RecA2, blue and red in figure 3.26) which are ubiquitously found in helicase structures and are named after the *E. coli* DNA recombination protein RecA.

The RecA-like fold consists of a central five-stranded β-sheet which is flanked by α-helices. RecA1 features the so called Walker motif A or P-loop between residues 320 and 327, a conserved motif which is found in many nucleotide binding proteins. Its conserved lysine makes hydrogen bonds to the β- and γ-phosphates of nucleotides an is therefore crucial for nucleotide binding. Moreover, Walker B motif is located in the β3-strand of RecA1. RecA2 harbors an antiparallel β-hairpin, protruding out of the RecA2 domain. This structural feature is only found in DExH-box helicases and is missing in the family of DEAD-box helicases. The two RecA like domains are connected by a linker and form a cleft between each other in which the nucleotide is bound. Contacts to the nucleotide are maintained by residues within the P-loop, R362 located in RecA1 as well as T580, D582 and
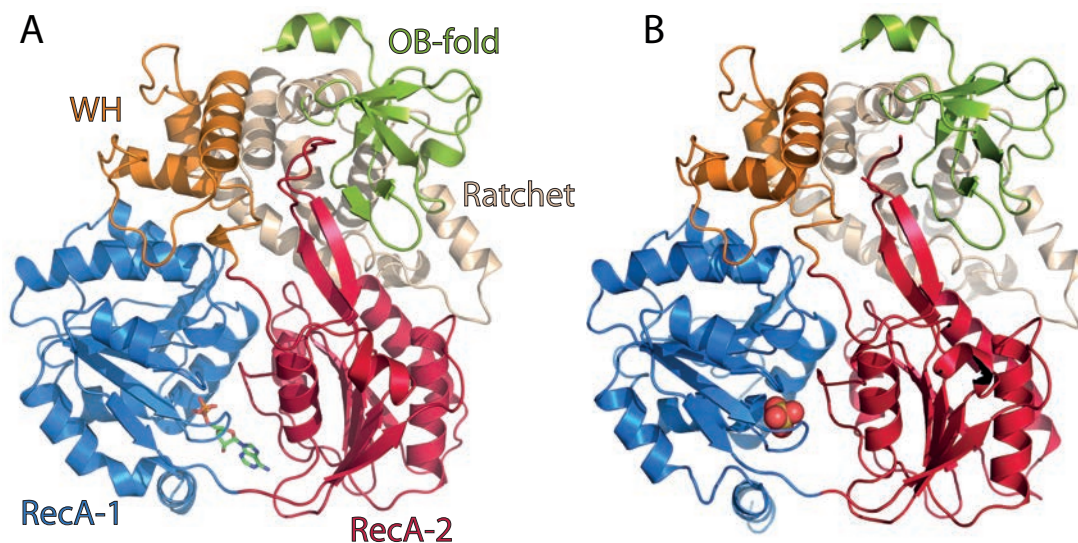
Figure 3.26: (A) Overall structure of ctPrp2 (270-921) in cartoon representation. The protein consists of two RecA-like domains (blue and red), a Winged Helix domain (orange) followed by the so called Ratchet domain (wheat) and an OB-fold domain (green) at the C-terminal end. Between the two RecA-like domains, an ADP molecule is bound. **(B)** Crystal structure of ctPrp2 (270-921) in apo-state. In the nucleotide binding cleft, a sulfate ion is bound at a position which is occupied by the β-phosphate in presence of ADP.

R628 from the RecA2 domain. The magnesium ion is coordinated by four water molecules, a β-phosphate oxygen and by Threonin 327. In the structure of apo-state ctPrp2 (270-921) additional electron density in the nucleotide binding cleft was interpreted as sulfate ion, which is located at a position that is occupied by the β-phosphate in presence of ADP. Adjacent to RecA2, a winged-helix (WH) and a so called Ratchet domain are located (orange and wheat in figure 3.26), both are homologous to the corresponding domains in the Ski2-like helicase Hel308 (Richards *et al.*, 2008). The Ratchet domain is a seven helix bundle and is named according to its proposed function of acting as a ratchet for DNA or RNA. The C-terminal domain (in green) includes a five stranded β barrel, arranged in a oligonucleotide binding fold (OB-fold) (Murzin, 1993).

Superposition of ctPrp2 with the structure of the Hel308/dsDNA complex (PDB ID: 2P6R) reveals a possible RNA binding pocket in Prp2, formed by the two RecA-like domains, the WH- and the Ratchet domain. In this model, one strand of the dsDNA binds to RecA1 as well as to a long helix in the Ratchet domain, while the other makes contacts to the RecA2- and the OB-fold domain (fig. 3.27). The superposition of the ctPrp2 apo structure with ctPrp2 in the
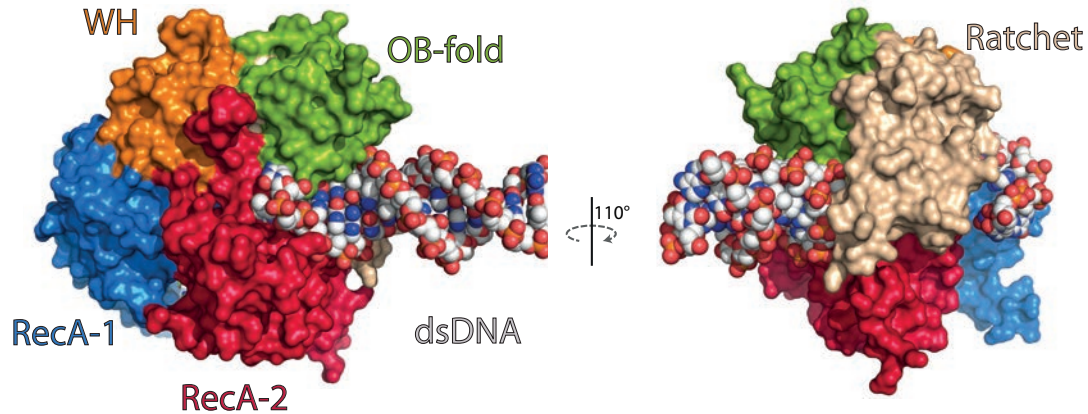
Figure 3.27: Model of a ctPrp2 (270-921)/dsDNA complex. The dsDNA was modeled into the structure by superposition with a Hel308/dsDNA complex (PDB ID: 2P6R). Coloring of the ctPrp2 domains is according to fig 3.26

ADP-bound state using LSQ-Superpose resulted in an r.m.s.d of 0.357 Å for 638 common Cα atoms indicating that no tremendous structural rearrangements of the overall shape are taking place upon binding of ADP.

3.7 Prp2-Spp2 interaction studies

In order to unravel which regions mediate the interaction between Prp2 and Spp2, crosslink interaction studies were performed were both interaction partners were crosslinked with three different chemical crosslinkers in order to probe for different distances between both proteins. While BS3 and BS2G form crosslinks between primary amines of lysines with a spacer length of 11.4 Å and 7.7 Å respectively, EDC is a zero length crosslinker that functions by converting carboxyl groups into amine-reactive isourea intermediates that bind primary amines. Therefore in proteins, EDC crosslinks glutamate and aspartate residues to lysin. The crosslinked protein complexes were separated by SDS-PAGE and subsequently analyzed by MS/MS. This work was performed by Romina Hofele from the group for Bioanalytical Mass Spectrometry, Max-Plank-Institute for Biophysical Chemistry, Göttingen.

3.7.1 *S. cerevisiae* Prp2-Spp2 crosslinks

In the yPrp2-ySpp2 complex, residues located in RecA1, RecA2 and the Ratchet-domain of Prp2 could be crosslinked to regions N- and C-terminal of the G-patch
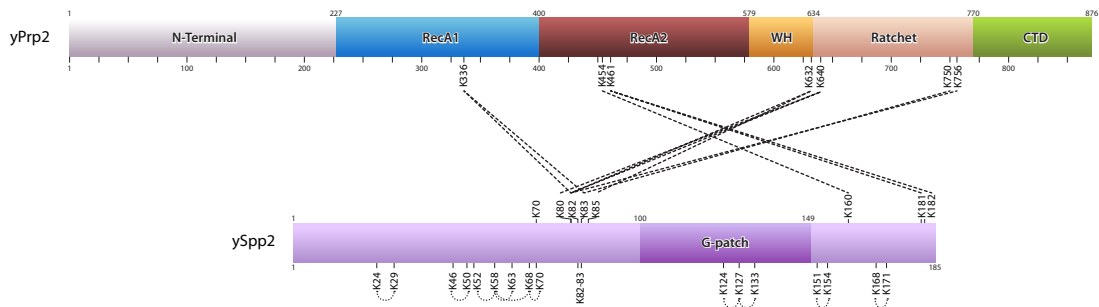
Figure 3.28: Schematic representation of yPrp2 and ySpp2. Intermolecular crosslinks are indicated by dashed lines, intramolacular crosslinks of ySpp2 are indicated by dotted lines. The G-patch motif of Spp2 is highlighted in dark purple. Most of the crosslinked residues of yPrp2 are located in RecA1, RecA2 and in the Ratchet-domain. These residues make crosslinks mostly to residues about 20 amino acids N-terminal of the G-patch motif as well as to three residues C-terminal of the G-patch.

motif. Remarkably, no crosslinks were identified between Prp2 and the G-patch, although amino acids able to form crosslinks are present in this region. Moreover, internal crosslinks in Spp2 are only formed between directly neighboring amino acids and not across the protein. This raises the question whether Spp2 adopts a folded state upon binding to Prp2, or if it is wrapped around Prp2 as an elongated chain.

To visualize and analyze the distribution of the crosslinks on the protein surface of Prp2, the corresponding residues were mapped on the structure model of ctPrp2 based on protein-protein alignment.

In the front view of fig 3.29 crosslinks can be seen in RecA1 and RecA2. In the case of RecA2, the crosslink is in close proximity to the entrance site of the RNA and to the superposed dsDNA. In the back view, crosslinks are mainly visible at positions surrounding the area were the single RNA strand (and in the figure the superposed DNA) leaves the protein core.

3.7.2 *C. thermophilum* Prp2-Spp2 crosslinks

Crosslinks between ctPrp2 and ctSpp2 could be identified within RecA2 and the C-terminal domains of Prp2. In contrast to ySpp2 were no crosslinks were formed with the G-patch, two lysine residues within the G-patch of ctSpp2, namely K236 and K250 were crosslinked to E646 and D648 located in the RecA domains of Prp2 and to K751 which is part of the WH-domain, respectively.

In context of the ctPrp2 crystal structure (3.31), a possible trace of Spp2 along the

front

180°

back
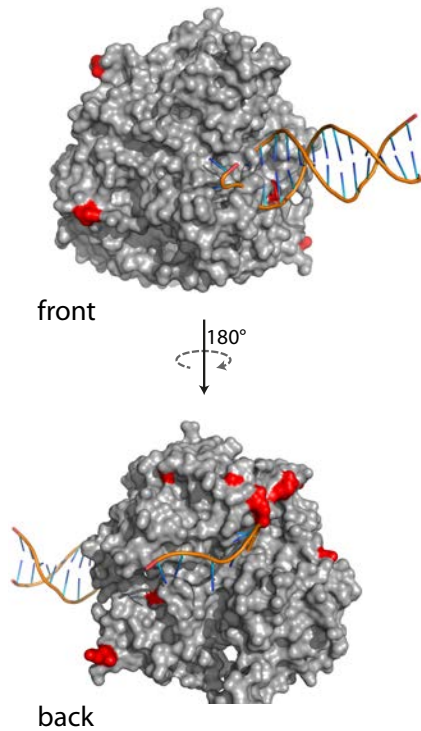
Figure 3.29: Crosslink results of the *S. cerevisiae* Prp2-Spp2 complex. Crosslinked residues of yPrp2 are mapped onto the structure of ctPrp2 and colored in red. dsDNA was superimposed from Hel308 (PDB:2P6R), orientation of front-view as in figure 3.26.

surface connecting helices from RecA1, RecA2 and proceeding to an helicase of the ratchet domain, were the superposed DNA enters the helicase can be deduced from the crosslink data. Interestingly, ctPrp2 is lacking the pronounced presence of crosslinks in the C-terminal region which was seen in the case of yPrp2.
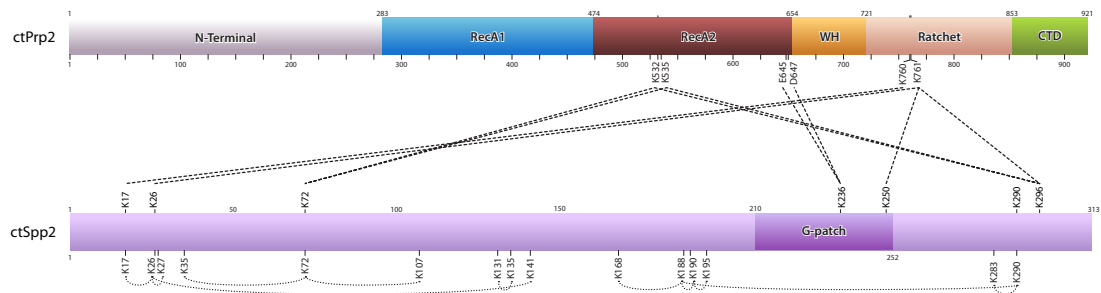
Figure 3.30: Schematic representation of identified crosslinks between ctPrp2 and ctSpp2 as well as intramolecular crosslinks of ctSpp2. Inter- and intramolecular crosslinks are indicated by dashed and dotted lines, respectively. The G-patch motif of Spp2 is highlighted in purple. Crosslinked residues of ctPrp2 are located within RecA2 and the Ratchet domain and form crosslinks to the N- and C-terminal region of ctSpp2 as well as to lysine residues located within the G-patch of Spp2.

Figure 3.31: Crosslink results of the *C. thermophilum* Prp2-Spp2 complex. Crosslinked residues of ctPrp2 are colored in red. dsDNA was superimposed from Hel308 (PDB:2P6R). A possible trace of Spp2 along the surface is indicated by a dashed line.
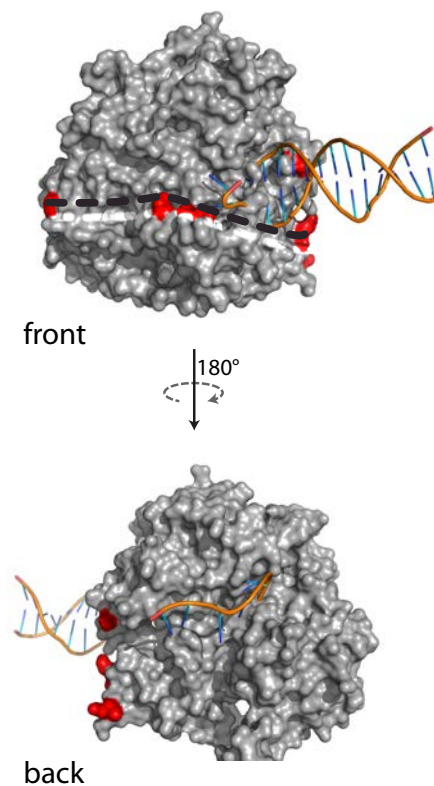
3.8 Structural studies on the Prp2-Spp2 complex

3.8.1 Crystallization of the ctPrp2-ctSpp2 (211-254) complex

Initial screening for crystallization conditions of the purified ctPrp2-ctSpp2 (211-254) complex (3.4.7) was performed with a protein concentration of 5 mg/ml and a 5× molar excess of ADP. All available screens (2.1.12) were set up at 20 °C and 4 °C, however no crystals could be obtained. The screening was therefore repeated, this time including the addition of microseeds obtained from ctPrp2 (270-921) crystals with a protein:reservoir:seed ratio of 1:1:1. Using this approach, small crystals could be obtained in a condition containing 0.1 M Tris/HCl pH 8.0, 8% (w/v) PEG 6000, 0.15 M NaCl. Crystal growth was optimized by varying the seed dilution from 1:1 to 1:100,000. Diffraction experiments showed, that the obtained crystals belonged to the same spacegroup with nearly identical cell constants compared to the ctPrp2 (270-921) crystals. After subsequent refinement of the collected data against the solved ctPrp2 structure, no additional electron density for ctSpp2 was visible in the electron density map. This indicates, that the ctPrp2-ctSpp2 complex dissociated during crystallization and as a result the crystals contained only ctPrp2.

In an alternative approach, purified ctPrp2 (270-921) was mixed with a 3-fold molar excess of ctSpp2 (211-254) and a 5× molar excess of ADP in order to prevent complex dissociation. The mixture was directly used for crystallization with a protein concentration of 4 mg/ml. Crystals were obtained in condition F10 of the Morpheus screen containing 0.1 M bicine/Trizma base pH 8.5, 10% (w/v) PEG 8000, 20% (v/v) ethylene glycol and 0.02 M of a monosaccharide mixture (for details see Gorrec (2009)). Hexagonal crystals with dimensions of 100 × 100 × 5 µm (see figure 3.32) grew within 7 days. The crystals were flash cooled in liquid nitrogen since the crystallization condition already had cryoprotecting properties.

The ctPrp2-ctSpp2 (211-254) complex crystallized in the orthorombic spacegroup $P2_12_12_1$ and the crystals diffracted to a resolution of 2.75 Å. A complete dataset was measured at beamline P13 at PETRA III, Hamburg. The statistics of the data collection are shown in table 3.5. The crystallographic phase problem was solved by molecular replacement using PHASER (McCoy *et al.*, 2007) with the already solved structure of ctPrp2 (270-921) as search model.

Figure 3.32: Protein crystals obtained for the ctPrp2-ctSpp2 (211-254) complex in a condition containing 0.1 M bicine/Trizma base pH 8.5, 10% (w/v) PEG 8000, 20% (v/v) ethylene glycol and 0.02 M of a monosaccharide mixture. The black bar corresponds to a length of 100 µm.
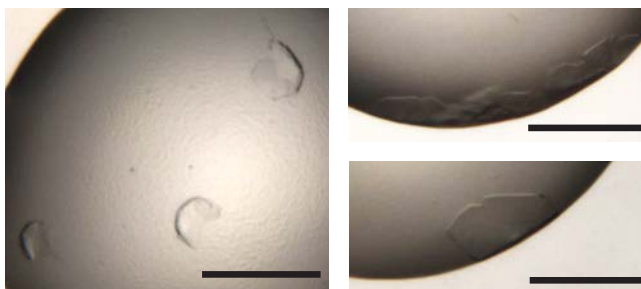
Building of the model was conducted in Coot, the coordinates were refined at a resolution of 2.75 Å using PHENIX (Adams *et al.*, 2002) and CNS (Brünger *et al.*, 1998). Alternating steps of building and refinement were performed until R-factors converged. The structure of the complex was refined to overall R_{work}/R_{free} values of 21.55 and 25.07 % (complete data statistics for refinement are shown in table 3.5). In the asymmetric unit, there are two complexes, each consisting of one molecule ctPrp2 (270-921) and one molecule ctSpp2 (211-254). With exception of residues 270-273 of one ctPrp2 molecule, all residues could be traced in the electron density. In the nucleotide binding cleft of both ctPrp2 molecules, additional density was observed which was interpreted as an ADP molecule and a magnesium ion.

Table 3.5: Data collection and refinement statistics of ADP bound ctPrp2 (270 - 921) in complex with ctSpp2 (211-254). Values in parentheses refer to the highest resolution shell, Ramachandran statistics were calculated with Molprobity.

Data collection	
Spacegroup	$P2_12_12_1$
Cell dimensions a,b,c (Å)	74.73, 113.69, 191.74
Wavelength (Å)	0.824
X-ray source	P13 DESY (Hamburg)
Resolution range (Å)	48.89 - 2.75
No. of reflections	218294
Completeness (%)	99.52 (99.50)
R_{merge} (%)	6.1 (60.1)
Average I/σ	20.16 (2.84)
Wilson B-factor (Å^2)	63.79
Redundancy	5.04 (4.97)

Refinement	
Resolution range (Å)	48.89 - 2.75 (2.81 - 2.75)
No. of Reflections	43324
R_{work} (%)	22.03 (30.60)
R_{free} (%)	25.68 (34.21)
Total number of atoms	10828
Protein atoms	10791
Water molecules	37
av. B-factors	
overall (Å^2)	79.43
Chain A/B (Å^2)	79.25
Chain C/D (Å^2)	89.50
solvent (Å^2)	38.47
RMSD	
bond lengths (Å)	0.007
bond angles (°)	1.102
Ramachandran plot	
favored (%)	97.21
allowed (%)	2.72
outliers (%)	0.07

3.8.2 Structural analysis of the ctPrp2-ctSpp2 complex

3.8.2.1 Structure and interaction interface of the G-patch motif

The 3D crystal structure of the ctPrp2-ctSpp2 complex reveals, that the N-terminus of ctSpp2 (211-254) binds to the winged-helix (WH) domain of ctPrp2. In the further course, ctSpp2 bridges the cleft between the WH and the OB-fold domain, thereby arresting the beta-hairpin of RecA2 in its position inside the cleft. The C-terminal end of ctSpp2 (211-254) is making contacts to the OB-fold as well as to the RecA2 domain. Furthermore the complex structure shows that the bound G-patch motif contains three α-helical regions, connected by flexible loops (fig. 3.33**A**). The N-terminal amphipathic helix includes residues 212-222, while the other two short helices are defined by residues 234-238 and 249-252 respectively. The G-patch motif is bound to ctPrp2 in a rather elongated shape, with all residues mediating the interaction with ctPrp2 located within the three secondary structure elements.

Stable binding of ctSpp2 to ctPrp2 is maintained by several interactions which could be identified in the crystal structure: at the N-terminus of ctSpp2, aromatic π-stacking between ctSpp2 F214 and ctPrp2 Y697 occurs along with hydrophobic interactions between L219 and W224 of ctSpp2 and L689/I690/V693 of ctPrp2. The central part of ctSpp2 which spans the cleft between the WH-domain and the OB-fold is held in position by a salt bridge formed between K236 of ctSpp2 and D648 of ctPrp2 while the C-terminus of ctSpp2 is attached to ctPrp2 mainly by a saltbridge formed between K250 of ctSpp2 and E518 of ctPrp2 (fig. 3.33**B**).

3.8.2.2 Comparison of ctPrp2 and the ctPrp2-ctSpp2 complex

Superposition of both structures reveals, that the overall fold and arrangement of the single helicase domains does not change significantly upon binding of ctSpp2 (211 - 254). However, a remarkable conformational change can be observed for the β-hairpin of the RecA2 domain. In the absence of ctSpp2, the β-hairpin possesses a high degree of flexibility and is bent slightly outward of the protein. Once ctSpp2 is bound, the tip of the β-hairpin moves by about 13 Å and the whole β-hairpin is pushed deeper and arrested into a newly formed cleft between the WH- and the OB-fold domain (fig. 3.34(**B**)). Thus, ctSpp2 acts like a seatbelt fixing the β-hairpin and reducing its flexibility.

Moreover, comparison of the two structures shows a significant difference in the
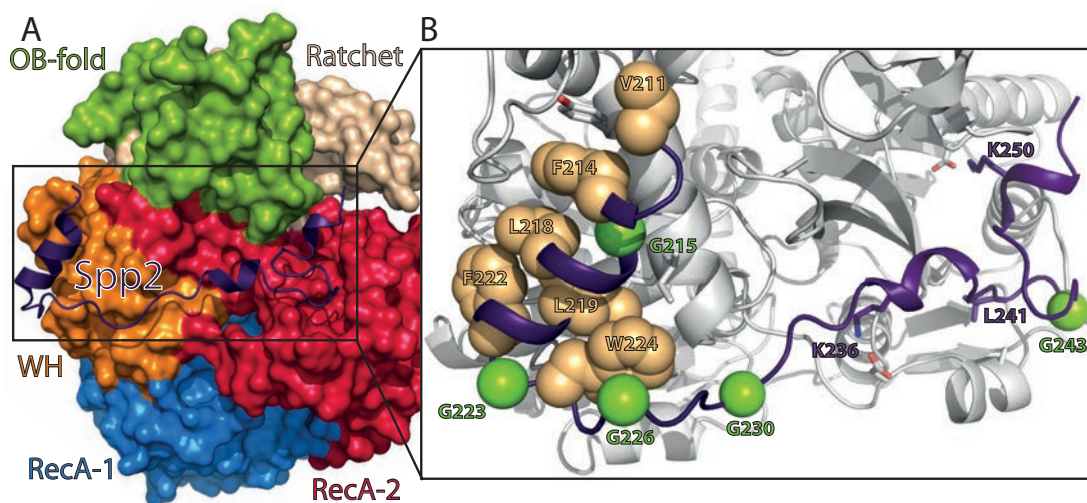
Figure 3.33: (A) Crystal structure of the ctPrp2-ctSpp2(211-254) complex with ctPrp2 shown in surface representation and the bound G-patch of ctSpp2 as ribbon (in purple). The G-patch binds to the WH-, the RecA2- and the OB-fold domain. **(B)** Interacting residues of Prp2 (gray) and Spp2 (purple). Conserved glycine residues of Spp2 are represented as green spheres, hydrophobic residues at the N-terminal helix of Spp2 are shown as spheres colored in pale orange.

conformation of the bound ADP molecule in the active site. In the absence of ctSpp2 the purine base moiety is mainly coordinated by an H-bond between the carboxyl group of D582 and the NH_2 group of the Adenine, while R362 is involved in hydrogen-bonding to the 3'-hydroxyl group of the ADP ribose. However, the situation changes upon binding of ctSpp2. In the ctPrp2-ctSpp2 complex, the purine base moiety of the ADP is sandwiched between R362 of RecA1 and F558 belonging to RecA2, both of which are involved in a cation π-stacking contact between the guanidinium group of R362 and the aromatic ring of F558. Additionally, an H-bond is formed between the 2'-hydroxyl group of the ribose and the carboxyl group of D582. The change of the ADP conformation goes along with some rearrangements in the RecA2 domain of Prp2: a loop defined by residues 625 to 632 moves further into the cleft between RecA1 and RecA2, thereby allowing the formation of a saltbridge between R625 and E321, which probably stabilizes the relative orientation of the two RecA-like domains (fig. 3.34**B**).
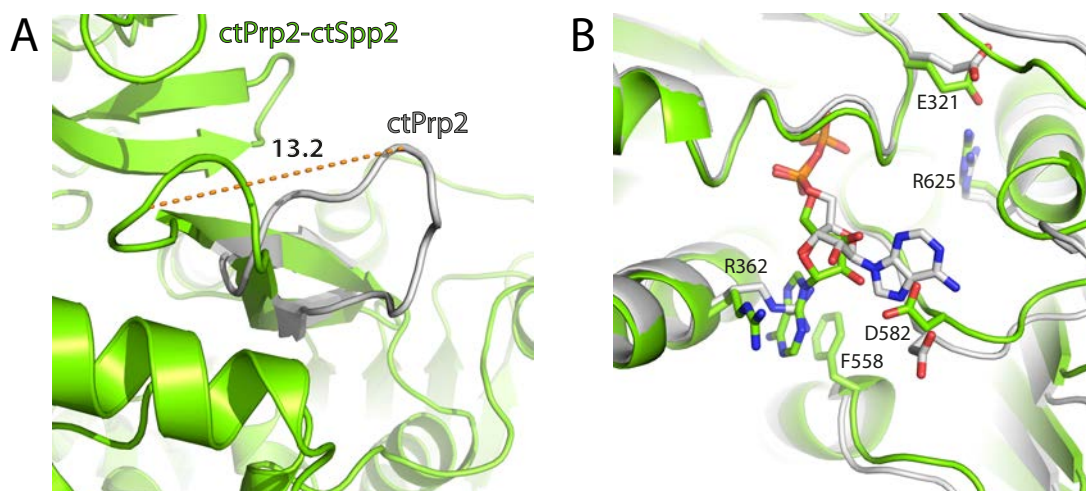
Figure 3.34: (A) Conformational change of the β-hairpin of ctPrp2 in absence (gray) and presence (green) of ctSpp2. The moving distance of the tip is given in Å. **(B)** Superposition of the nucleotide binding site of ctPrp2 (gray) and the ctPrp2-ctSpp2 complex (green). The binding mode of the Adenosin moiety changes significantly and a salt bridge between R625 and E321 is introduced upon binding of Spp2 to Prp2.

3.9 Biochemical studies on Prp2 and the Prp2-Spp2 complex

In order to investigate the role of Spp2 and its influence on the biochemical properties of Prp2, the ATPase as well as the helicase activity of yPrp2 in the presence and absence of ySpp2 were determined using real-time and fluorescence based assays. Furthermore, dissociation constants for RNA were determined by microscale thermophoresis.

3.9.1 ATPase activity

For determination of the ATPase activity of yPrp2, a commercially available photometric assay was used which measures the release of orthophosphate in real time (see 2.3.1). The assay allowed determination of the Miachaelis constant $K_{\rm m}$ as well as the turnover number $k_{\rm cat}$.

Results of the ATPase activity measurements are shown in table (3.6)

The intrinsic ATPase activity of Prp2 alone was very low with a turnover number of 4.34±0.30 min^{-1}. However, it could be significantly stimulated by RNA leading to an 10-fold increase of $k_{\rm cat}$ in the case of 15 µM RNA. At the same time $K_{\rm m}$ increased about 5 fold, which does not necessarily indicate a lower affinity
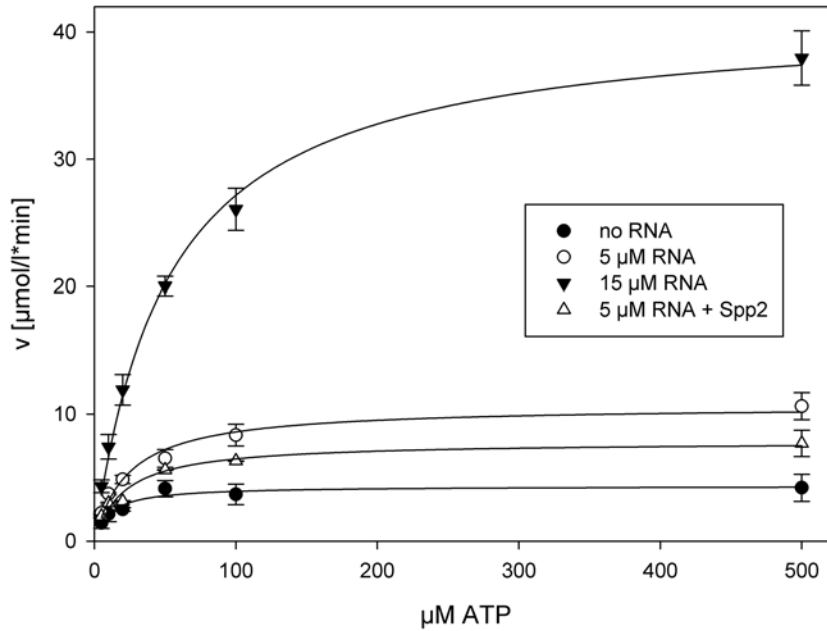
Figure 3.35: Results of the ATPase activity assay for yPrp2 in absence and presence of RNA and Spp2. The initial reaction velocity in µM * min^{-1} is plotted against ATP concentration. Poly(A)$_{20}$ RNA was used in concentrations of 5 and 15 µM.

for ATP in presence of RNA but instead might be a consequence of the increased ATP hydrolysis rate.

Interestingly, the presence of Spp2 does not stimulate the ATPase activity of Prp2, but instead led to a slight decrease of k_{cat} to 7.82±0.31 min^{-1} compared to the sample in the absence of Spp2.

3.9.2 Helicase activity

For measurements of helicase activity a fluorescence based assay was used, were decrease of measured fluorescence intensity is proportional to the amount of unwound doublestranded RNA substrate (see 2.3.3). As a positive control for the assay the spliceosomal helicase yPrp43 was used. For Prp43 it is known to posses a strong RNA helicase activity. Experiments with this protein showed a significant decrease in the fluorescence signal over time in presence of dsRNA substrate and ATP indicating, that the assay was working as expected (Christian *et al.*, 2013).

However, no helicase activity could me measured for yPrp2, neither in absence nor in presence of ySpp2 or with high ATP concentrations (2 mM), although it

Table 3.6: Kinetic parameters describing the ATPase activity of yPrp2 alone as well as in complex with ySpp2 (10-185), in absence and in presence of poly(A)$_{20}$-RNA. For each sample, the Michaelis constant K_m and the turnover number k_{cat} is provided.

Protein	K_m (µM)	k_{cat} (min^{-1})
In absence of RNA		
yPrp2	10.67±3.05	4.34±0.30
In presence of RNA		
yPrp2 + 5 µM RNA	23.81±3.21	10.64±0.42
yPrp2 + 15 µM RNA	52.03±3.77	41.29±1.02
yPrp2-ySpp2(10-185) + 5 µM RNA	20.89±2.91	7.82±0.31

was shown that yPrp2 binds RNA in presence of ATP (see 3.36).

3.9.3 RNA affinity

Binding constants for yPrp2 to RNA in the presence of ADP or AMPPNP as well as Spp2 and AMPPNP were determined by microscale thermophoresis (see 2.3.4). In this method, directed movements of particles along a microscopic temperature gradient are monitored by fluorescence detection. Any change of the hydration shell of bio molecules due to changes in their structure or conformation results in a relative change of movement along the temperature gradient and is used to determine binding affinities. The obtained data (shown in figure 3.36) demonstrate, that Prp2 is only able to bind RNA in the presence of the non-hydrolyzable ATP analogue AMPPNP (K_d 10.13±1.07 µM). The RNA affinity of the Prp2-Spp2 complex does not differ significantly from the value determined for Prp2 alone (K_d 9.36±1.31 µM), indicating that the presence of Spp2 does not influence RNA binding. In contrast, Prp2 seems not to bind RNA in the presence of ADP.
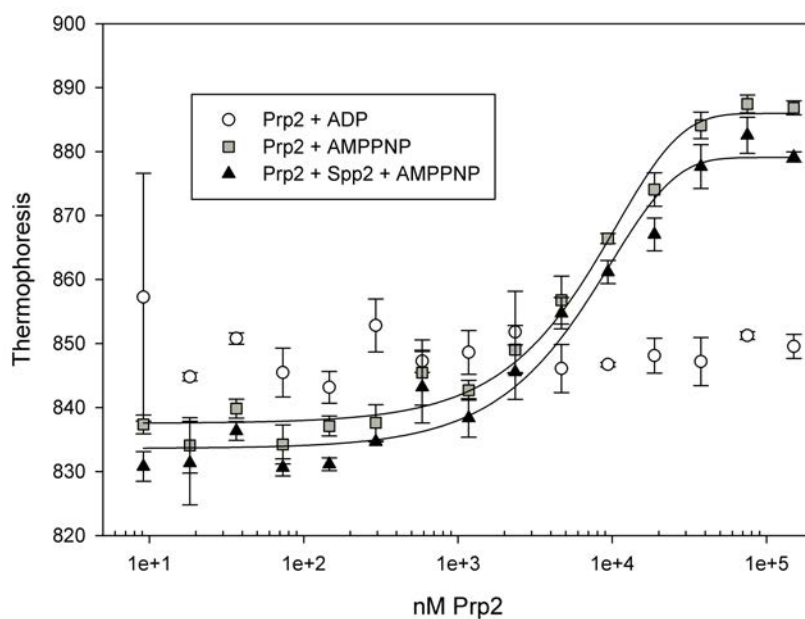
Figure 3.36: Results of the RNA affinity measurements for Prp2 in presence of ADP (white circle), AMPPNP (gray square) as well as Spp2 and AMPPNP (black triangle). Thermophoresis units are plotted against the concentration of Prp2, K_d values for the Prp2-RNA interaction were calculated by curve fitting. In the presence of AMPPNP, Prp2 binds RNA with a K_d of 10.13±1.07 µM. The affinity is not significantly changed by Spp2, while in presence of ADP no RNA binding could be measured.

4 Discussion

The molecular function of the G-patch protein Spp2 with regard to the activity of Prp2 and thus to the activation of the spliceosome has so far remained an open question. Moreover, no structural data for the G-patch motif was available up to now, neither bound to an interaction partner nor alone. The aim of this work was therefore to obtain structural and biochemical information on the Prp2-Spp2 complex in order to investigate the effect of Spp2 on Prp2.

4.1 Structural and biochemical features of Prp2

As crystallization trials with full length Prp2 as well as with different truncated Prp2 and Spp2 constructs from *S. cerevisiae* were not successful, orthologues of these proteins were identified in the thermophilic fungus *Chaetomium thermophilum*. Proteins from thermophilic organisms have been demonstrated to posses a higher rigidity compared to their mesophilic counterparts, which might make them more suitable for successful crystallization experiments (Amlacher *et al.*, 2011). Due to aggregation of the protein probably caused by binding to RNA during the purification, the preparation of full length ctPrp2 was not possible. However, since the N-terminus of yPrp2 was shown to be dispensable for its activity *in vivo* (Edwalds-Gilbert *et al.*, 2004), an analogues N-terminal truncated ctPrp2 was used for crystallization experiments and the following structure determination. The structure of ctPrp2 (270-921) could be determined in its ADP bound state, showing a high similarity to the structure of the DEAH-box helicase Prp43 (Walbott *et al.*, 2010). This similarity can be explained with the high degree of sequence conservation throughout the family of DEAH-box helicases, especially within the helicase core and the C-terminal domains. However, such similarity is nevertheless surprising, since the biochemical features of both proteins differ in many ways. The biochemical properties of Prp43 have been assessed in several studies. Throughout these it was demonstrated, that Prp43 which plays a role in splicing and in the biogenesis of ribosomes, possesses AT-

Pase as well as helicase activity *in vitro*. Furthermore it was shown, that these activities are stimulated by the G-patch protein Ntr1 in the spliceosomal context or by Pfa1 during ribosome biogenesis (Tanaka *et al.*, 2007; Lebaron *et al.*, 2009; Christian *et al.*, 2013). Another extensively studied spliceosomal DEAH-box helicase is Prp22 which is required during disassembly of the spliceosome. This helicase shows high ATP hydrolysis and RNA unwinding rates. For example, the ATP turnover rate equals 56 min^{-1} in the absence of RNA without requiring an activating interaction partner (Tanaka and Schwer, 2005). In contrast, the ATP hydrolysis rate of Prp2 is significantly lower, with the k_{cat} value averaging 4.34±0.30 min^{-1} in the absence of RNA and 41.29±1.02 min^{-1} in presence of a 15× molar excess of Poly(A)$_{20}$-RNA. While for Prp43 a three to five fold stimulation of its ATPase activity by Ntr1 was shown (Christian *et al.*, 2013), Spp2 doesn't seem to have such stimulating effects on the ATP hydrolysis rate of Prp2 under the assayed conditions (see table 3.6 for details). Moreover, the results of the helicase assay indicate, that Prp2 doesn't seem to unwind dsRNA (see 3.9.2 for details), whereas Ntr1 dependent RNA unwinding activity could be demonstrated *in vitro* for Prp43 (Christian *et al.*, 2013). Nevertheless, the RecA2 domain of Prp2 contains a 5'β-hairpin, a structure which is also found in Prp43 as well as as in the viral NS3 helicase and is thought to act as a separation device for dsRNA (Walbott *et al.*, 2010). Since dsRNA unwinding activity was not observed for Prp2, the question arises, whether the β-hairpin might fulfill a different function in Prp2.

4.1.1 Non-canonical binding of ADP in absence of Spp2

The helicase core of all SF2 helicases is formed by two RecA-like domains which contain several conserved motifs that are involved in Nucleotide and/or RNA binding. The Nucleotide is bound in a cleft between both domains and coordinated by residues belonging to the conserved motifs I, II, V and VI. In the available crystal structures of Prp43 from *S. cerevisiae* and *C. thermophilum* the adenine moiety of the ADP is found sandwiched between a phenylalanine of RecA2 and an arginine of RecA1 (Walbott *et al.*, 2010; He *et al.*, 2010). This binding mode is referred to as "canonical binding mode" in the following. In the ctPrp2-ctSpp2 complex structure, the ADP molecule adopts a conformation corresponding to the canonical binding mode, with the adenine base stacked between F558 and R362 while D582 forms an H-bond to the 2'OH of the ribose. However, in the absence

of ctSpp2, the conformation of the ADP molecule differs significantly from the one observed in the canonical binding mode and so do the coordinating residues of the RecA-like domains. While the phosphate moieties are bound in the same position by the P-loop, the adenine base is no longer stacked between F558 and R362 but instead bound by D582 while R362 forms an H-bond to the ribose. This leads to a rotation of about 110 ° around the ADP α angle (O2α-Pα-O5'-C5') and is accompanied by a conformational change of the ADP from the anti- into the syn-conformation. Moreover, changes in a loop region defined by residues 626 - 632 can be observed, which lead to the disruption of a salt bridge between E321 and R625 (fig. 3.34). Since this salt bridge is build between RecA1 and RecA2, it might help to stabilize the helicase core in a defined state once Spp2 is bound, although no direct influence on the ATP hydrolysis rate was observed in the activity assays.

The non-canonical ADP conformation has never been observed in other DEAH-box helicases and also differs from the conformation observed in DEAD-box helicases (fig. 4.1), where the adenine moiety is bound by the Q-motif located in the RecA1 domain (Sengoku *et al.*, 2006; Oyama *et al.*, 2009). The comparison with the ctPrp2 structure in the nucleotide-free state reveals, that the positions of the residues involved in nucleotide binding as well as of the loop region 626 - 632 are similar to the positions seen in the ADP bound state. This finding indicates, that the nucleotide binding site of ctPrp2 can accommodate ADP exclusively in the non-canonical mode as long as ctSpp2 is absent and that the switch to the canonical binding mode is induced by the interaction with the G-patch motif of ctSpp2. It should, however, be noted that the interactions of the nucleotides with the protein residues of the different motifs are likely to be influenced by binding of the protein to nucleic acids, which has been observed for different structures of NS3 (Luo *et al.*, 2008). Therefore, the observed interaction might change again upon binding of RNA.
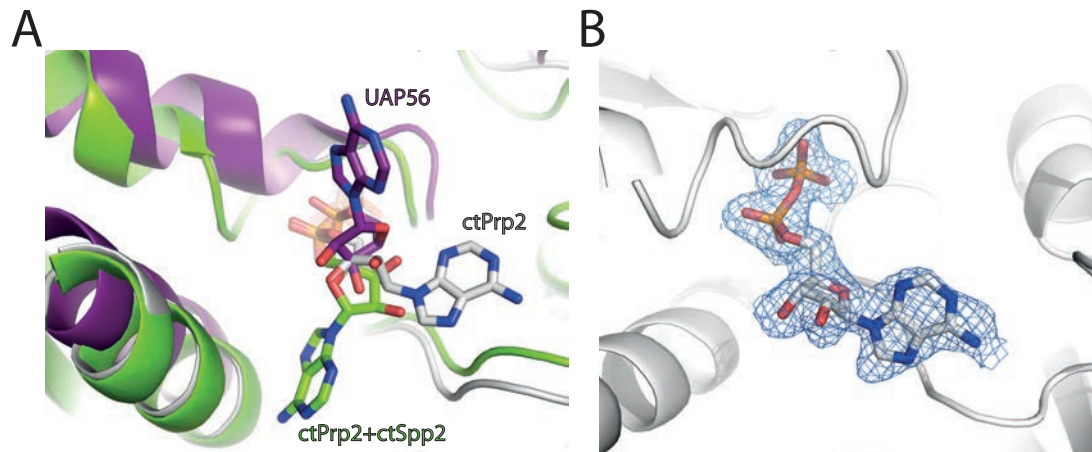
Figure 4.1: (A) Superposition of different ADP conformations in the structures of ctPrp2 (gray), the ctPrp2-ctSpp2 complex (green) and the DEAD-box helicase UAP56 (purple), the corresponding ADP molecules are shown as sticks. The ADP conformation observed in ctPrp2 differs significantly from the canonical conformation observed in the ctPrp2-ctSpp2 complex as well as from the conformation observed in crystal structures of other DEAD-box helicases. **(B)** Simulated annealing mFo-DFc omit map contoured at 2.0 σ (blue mesh) for the ADP molecule in the ctPrp2 (270-921) structure. The ADP molecule is shown in stick representation, the protein is shown as ribbon.

4.2 Intrinsically disordered proteins

Intrinsically disordered proteins (IDPs) are widespread throughout all spliceosome associated proteins, in fact it has been predicted that about 45% of the human spliceosomal proteins are disordered and about 80% contain disordered regions with a length of at least 30 residues (Korneta and Bujnicki, 2012). Disorder of proteins goes along with some functional advantages like the adaptability to different binding partners and the ability to bind with high specificity but at the same time relatively low affinity, which enables the IDPs to associate and dissociate fast in dynamic complexes (Dyson and Wright, 2005; Hegyi *et al.*, 2007). In the context of the highly dynamic environment of the spliceosome which contains a huge number of proteins, these properties would enable IDPs to specifically recognize their interaction partners in order to recruit them to the spliceosomal core. Moreover, disordered parts of the protein might play a role in a mechanism known as moonlighting. Moonlighting proteins are able to fulfill more than one function in unrelated processes. Examples for moonlighting yeast spliceosomal proteins are Clf1p, which has been shown to additionally play a role in DNA replication, as well as U5-52K, which besides its function in the spliceosome is a binding protein for the CD2 receptor in the immune response (Jeffery, 2009). The switch

in function is originally explained by change in expression, cellular localization or variations in oligomerization (Jeffery, 2003a,b). However, structural flexibility induced by disordered elements might also increase the structural adaptability to different binding partners (Tompa *et al.*, 2005).

4.2.1 The IDP Spp2

It has been demonstrated that the presence of Prp2 and ATP is not sufficient for spliceosome activation, but that this step is also dependent on the presence of an additional factor, the G-patch protein Spp2 that binds to Prp2 (Last *et al.*, 1987; Roy *et al.*, 1995). A similar mechanism has been described for the helicase Prp43, which fulfills its function in spliceosome disassembly after it has been activated by its interaction partner Ntr1, which possesses a G-patch motif as well (Tanaka *et al.*, 2007). While a significant stimulating effect on the ATPase as well as on the helicase activity of Prp43 by Ntr1 has been shown (Tanaka *et al.*, 2007; Christian *et al.*, 2013), Spp2 does not seem to enhance the ATPase activity of Prp2 nor does it stimulate its helicase activity (see 3.9.1 & 3.9.2). Thus, Prp2 might either require an additional factor for activation or the function of Spp2 might not be the activation but the recruitment of Prp2.

The CD-spectroscopy results for ySpp2 (10-185) as well as for the truncated ySpp2 (100-150) containing only the G-patch motif clearly indicate the absence of any secondary structure elements (for details see section 3.3.1). These results were confirmed by the HSQC NMR spectrum of ^{15}N labeled ySpp2 (10-185) demonstrating that Spp2 is indeed disordered in its free form (see section 3.3.2). In contrast, the ctPrp2-ctSpp2 crystal structure reveals that upon binding to Prp2 the G-patch of Spp2 adopts a distinct fold, comprising three α-helices connected by flexible linker loops (see section 3.8.2.1. This suggests that Spp2 has the ability to fold upon binding to Prp2.

4.3 Interaction of Prp2 with Spp2

In the ctPrp2-ctSpp2 crystal structure presented in this work, the G-patch motif of ctSpp2 adopts an elongated shape that is held in position by hydrophobic interactions between its amphipathic N-terminal helix and the WH-domain of ctPrp2. Additionally, two salt bridges are formed between the ctSpp2/ctPrp2 residues K236/D648 and K250/E518. While the residues involved in inter-protein con-

tacts are located in α-helical regions, the parts of the protein connecting this anchoring points contain no secondary structure elements. A sequence alignment of Prp2 orthologues from different organisms (fig. A.5) revealed patches of conserved residues located on the surface of ctPrp2 (fig. 4.2). Interestingly, two of these patches coincide with the contact areas of the N- and C-terminal helices of Spp2, thus suggesting a similar mode of Prp2/Spp2 interaction in different organisms.
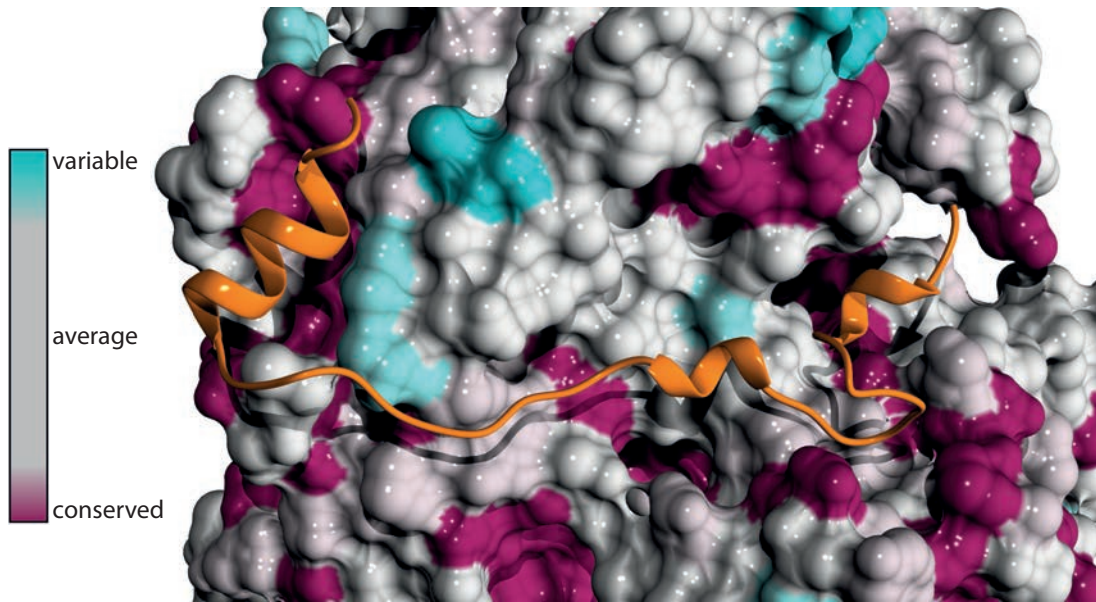

Figure 4.2: Surface representation of ctPrp2 with ctSpp2 (211-254) shown as ribbon (orange). The surface of ctPrp2 in colored according to the degree of sequence conservation (blue: variable, gray: average, purple: conserved). The N- and C-terminal helices of ctSpp2 are found to make contacts to conserved regions of ctPrp2.

Upon binding of the G-patch motif, the position of the β-hairpin undergoes a remarkable change of about 13 Å while it is pushed into a cleft between the WH and the OB-fold domain (fig. 3.34). The β-hairpin is thought to act like a wedge on RNA double strand substrates, which are then separated by the translocation of the helicase onto the RNA (Walbott *et al.*, 2010). Comparison of several crystal structures of the DEAD-box helicases eIF4A-III, Vasa, and Prp28 bound to different nucleotides as well as in the apo state revealed major conformational changes that occur during the cycle of ATP binding, hydrolysis and release of ADP (Andersen *et al.*, 2006; Bono *et al.*, 2006; Schütz *et al.*, 2010). Based on these findings and on superpositions with the crystal structure of eIF4A-III bound to AMPPNP, the β-hairpin of Prp43 was predicted to perform a movement outward of its binding cleft upon binding of ATP, which would support the separation of dsRNA

(He *et al.*, 2010). However, the ctPrp2-ctSpp2 complex structure shows, that the β-hairpin can not perform this movement once Spp2 is bound, because the G-patch motif acts like a seatbelt and seems to arrest the β-hairpin in its position. Additionally, in the ctPrp2-ctSpp2 complex structure, the two C-terminal helices of the G-patch motif are located in a small cleft formed between RecA2 and the OB-fold on the surface of ctPrp2. Based on a superposition with the structure of the Hel308/DNA complex (PDB 2P6R) (Büttner *et al.*, 2007), this region was supposed to interact with the 5'-end of the modeled ssDNA after separation of the double strand by the β-hairpin (Walbott *et al.*, 2010). This is in conflict with the structural data presented in this work (fig. 4.3).
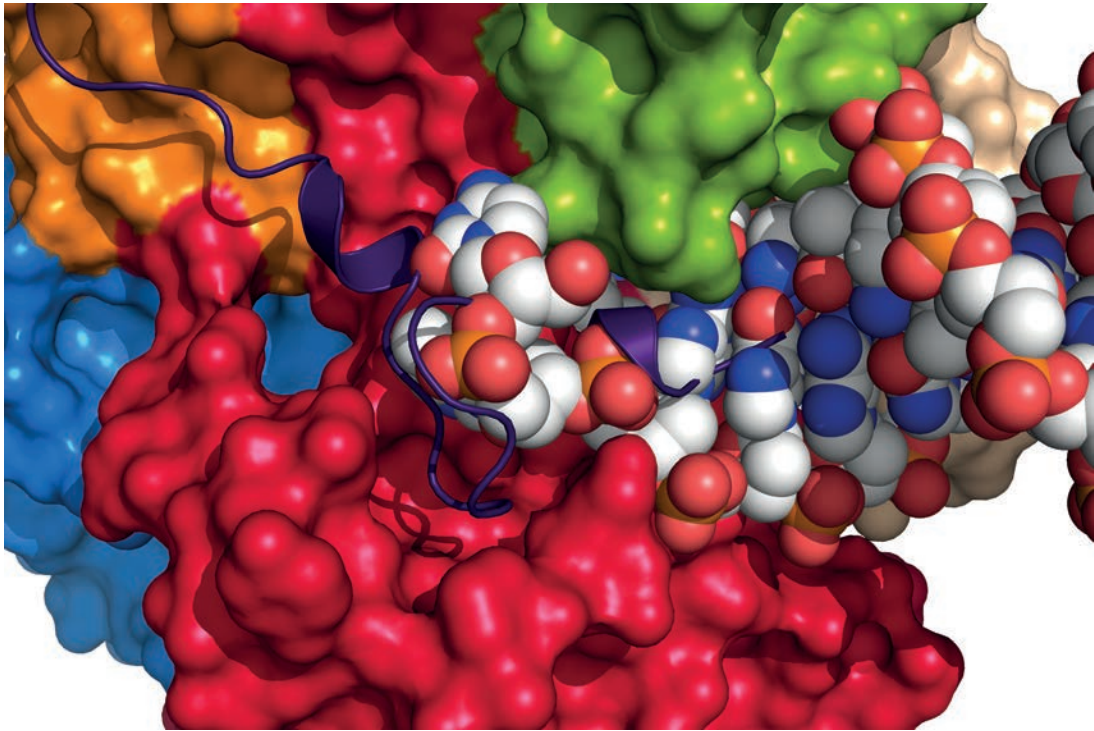
Figure 4.3: Superposition of the ctPrp2-ctSpp2 crystal structure with dsDNA from the structure of Hel308 (PDB:2P6R). Coloring of the protein is according to fig. **3.33(A)**, the backbone carbons of the DNA molecule are shown as gray spheres, ctPrp2 and ctSpp2 are shown in surface and ribbon representation, respectively. The C-terminus of ctSpp2 (211-254) clashes with the 5'-end of one strand the modeled DNA.

In contrast, binding of the second strand which is thought to pass through a cavity between the two RecA-like domains and the C-terminal domains where it contacts the long ratchet helix and thereby might allow the helicase to translocate along the RNA (Büttner *et al.*, 2007), seems not to be opposed as the entrance as well as the cavity itself are not blocked by the truncated ctSpp2 molecule in

the crystal structure. It is not clear whether ctSpp2 would prevent unwinding of double stranded RNA due to steric clashes with the RNA strand, since no dsRNA unwinding activity could be observed for Prp2 *in vitro*, neither in absence nor in presence of Spp2 (section 3.9.2; Reinhard Lührmann, personal communication). In the asymmetric unit of the crystal, the two ctPrp2-ctSpp2 complex molecules are oriented in a way with the two ctSpp2 molecules facing each other. Therefore, the observed conformation and binding position of ctSpp2 might likely be influenced by the crystallization conditions and not represent the situation in solution. To assess this problem, *in solution* crosslink studies were performed with ctPrp2 (270-921) and full length ctSpp2. In these experiments, ten crosslinks between ctPrp2 and ctSpp2 could be identified. Of these, two were located within the crystallized portion of ctSpp2, namely the lysine residues K236 and K250 which were found to form crosslinks to the ctPrp2 residues E646, D648 and K761 respectively. These crosslinks are perfectly fitting the crosslinker distances in the crystal structure and could thereby confirm that the position of the G-patch as observed in the crystal structure is in good agreement with the situation in solution. Moreover, crosslink studies on the interaction of Prp43 with the G-patch protein Ntr1 have recently been published (Christian *et al.*, 2013). The binding site of Ntr1 proposed in this study is very similar to the one observed for ctSpp2 in the crystal structure. In fact, by superposition based modeling of the ctSpp2 G-patch onto the Prp43 crystal structure, several of the identified crosslinks could be confirmed to be in good agreement with respect to the observed distances in the model (fig. 4.4). In detail, K67 which is located N-terminal of the G-patch crosslinks to the residues K733, K737 and K746 that are located at the C-terminal helix of Prp43. In the model, the measured Cα-Cα distances for these residues are 14.7, 15.4. and 26.6 Å respectively. Thus the distances are in a good range to fit the crosslinking distance of the used reagent BS3 and BS2G if flexibility of the proteins C-terminus is taken into account. The same applies for a crosslink between Ntr1 K78 and Prp43 K4 with a measured distance of 23.1 Å in the model. These findings further support the hypothesis, that the observed interaction in the crystal structure is not unique to the ctPrp2-ctSpp2 complex but my also be present in other helicase/G-patch protein complexes and might therefore be considered as a molecular basis to explain the mechanistic function of the G-patch/helicase interaction in general.

The question remains, whether folding of Spp2 is induced after the binding event or if folding into a defined shape is a prerequisite for association and thus
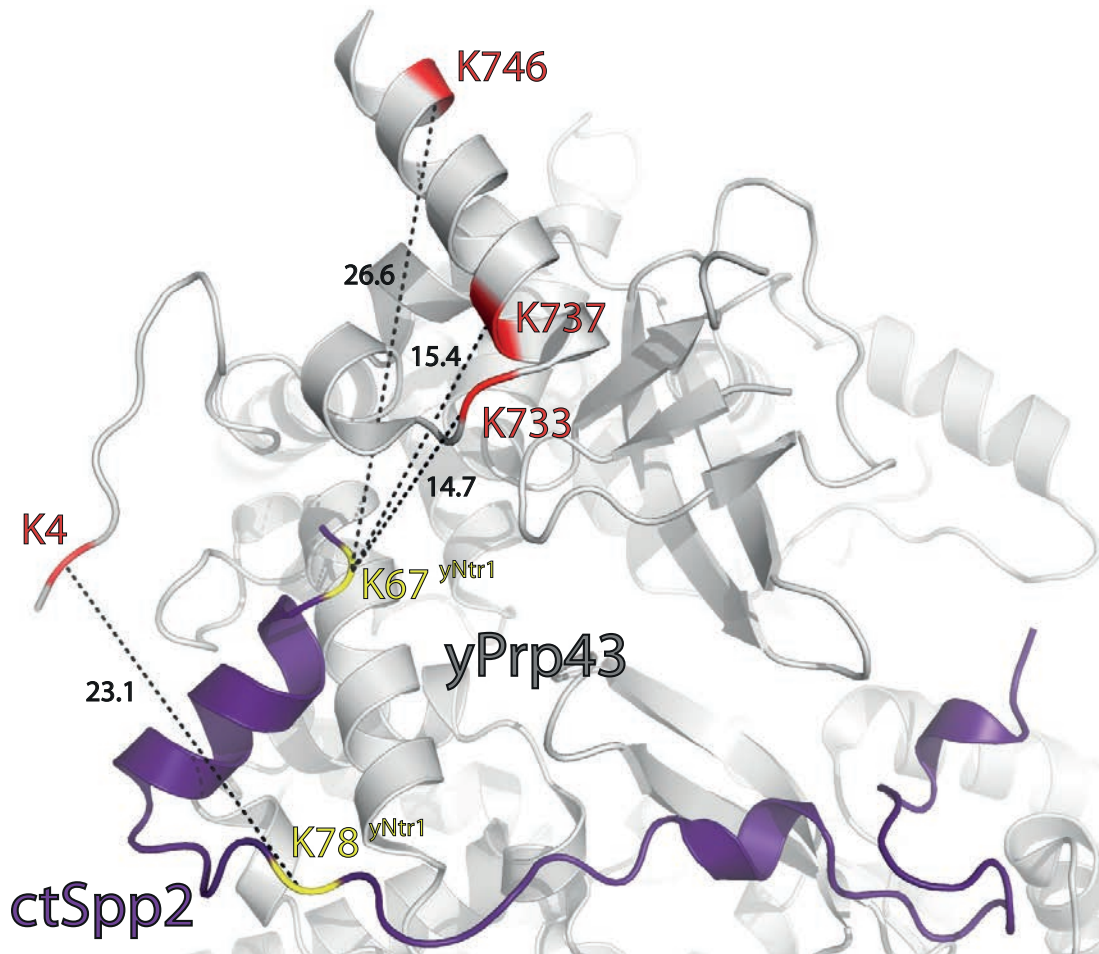
Figure 4.4: Model of the Prp43/G-patch interaction. For the model, the structure of yPrp43 (PDB: 2XAU) was superimposed onto the ctPrp2-ctSpp2 complex structure and the crosslinked residues of yNtr1 were mapped on the ctSpp2 molecule based on protein sequence alignment. Prp43 is shown in gray, Spp2 is colored in purple. Crosslinked residues of yPrp43 and yNtr1 are highlighted in red and yellow respectively, crosslinks are indicated by dashed lines and the Cα-Cα crosslink distances are given in Å.

has to happen first. However, since the interaction between the G-patch of Spp2 and Prp2 is mainly facilitated by anchoring points located in the three α-helical parts of the G-patch which are connected by long and more flexible linkers, one could speculate that these parts are able to initially bind and fold independent of each other. The N-terminus for example might initially be positioned close to its binding site by single interactions to a hydrophobic patch on the surface of Prp2. Subsequently, formation of the amphipathic α-helix might be induced by the hydrophobic environment, leading to an increase in the number of hydrophobic interactions which in the end results in a more stable binding of Spp2 to Prp2.

4.3.1 Structural properties of the bound G-patch motif

While it was previously suggested, that the G-patch has a specific function in RNA processing (Aravind and Koonin, 1999), there are no studies on the function of the conserved residues of the G-patch available. Since structural information about the G-patch in a bound state was lacking, the functional relevance of the characteristic six conserved glycine residues as well as of the conserved aromatic and hydrophobic residues remained unknown so far. With the results obtained during this work, it was possible for the first time to elucidate the necessity of glycine residues at certain positions on a structural basis. The structure shows, that for the conserved residues 223, 226 and 230 of Spp2, the observed combination of ϕ and ψ torsion angles are only possible for glycine residues at these positions according to the Ramachandran plot (Ramachandran and Sasisekharan, 1968). Additionally, the residue at position 226 has to be small because the side chain of the conserved W224 blocks the space needed for a larger side chain. The structure strongly emphasizes that beside the glycines, the conserved hydrophobic residues L218 and L219 are important for the function of Spp2 since they are part of an amphipathic helix and involved in hydrophobic interactions to the interaction partner Prp2. This suggestion is supported by mutational studies on Spp2 from *S. cerevisiae* showing that a mutation of L109 (which corresponds to L218 in *C. thermophilum*) to arginine or glutamate eliminated the interaction with Prp2. Interestingly an L109V mutant showed even higher binding affinity compared to the wild type which strongly supports the idea, that L109 is involved in hydrophobic contacts with Prp2.

Sequence alignments of Spp2 orthologues from *Homo sapiens*, *Saccharomyces cerevisiae* and *Chaetomium thermophilum* show a variable length of the linker region between third conserved glycine of the G-patch and the second α-helix (see A.3). This region is not involved in protein-protein interactions and the difference in length might therefore originate from evolutionary adaptation to different binding targets. However, more structural data on different G-patch protein complexes are required to confirm this hypothesis.

4.4 A model for the molecular function of Prp2 & Spp2

Although Spp2 has been demonstrated to be essential for the activation of the spliceosome (Silverman *et al.*, 2004), its working mechanism remained elusive so far. It was demonstrated for several SF1 and SF2 helicases, that their ATPase and helicase activities are stimulated to a certain extend by activator proteins which bind to the helicases. For example, the two RecA-like domains of eIF4A come closer together upon binding of its activator eIF4G, leading to an increase of ATP and RNA affinity (Oberer *et al.*, 2005; Schütz *et al.*, 2008). The SF1 helicase Upf1 is stimulated by the intrinsically unstructured protein Upf2, which decreases the RNA affinity of Upf1 and thus increases RNA unwinding activity (Clerici *et al.*, 2009; Chakrabarti *et al.*, 2011). Recent data for the DEAH-box helicase Prp43 show an activation of ATPase and RNA unwinding activity by the intrinsically disordered G-patch protein Ntr1 (Tanaka *et al.*, 2007; Christian *et al.*, 2013). Despite the fact, that the high sequence conservation and structural similarity of Prp2 and Prp43 suggests a similar functional mechanism for both proteins, there seem to be major differences regarding their mechanistic function in the spliceosome, and the role of their interaction partners Spp2 and Ntr1. Additionally, several results of this work are not consistent with the current assumption, that Prp2 is a classical RNA unwinding helicase. First, dsRNA unwinding activity could not be shown for Prp2, neither in this work (3.9.2) nor by other groups (Reinhard Lührmann, personal communication), although recent data suggest the possibility of at least an indirect destabilization of snRNA by Prp2 (Wlodaver and Staley, 2014). Second, the structural data of the ctPrp2-ctSpp2 complex combined with a model for dsRNA binding by DEAH-box helicases implicate steric clashes between ctSpp2 and the ssRNA after strand separation. Together, these findings support the theory that Prp2 does not unwind dsRNA *in vivo*. Nevertheless, the structure of Prp2 shows the typical feature of DEAH-box helicases - the 5'-β-hairpin which is thought to act as dsRNA separation device and thus to be a key element for RNA unwinding. This secondary structure element is supposed to move outward of its binding cleft upon binding and hydrolysis of ATP and possibly participates in RNA strand separation (Büttner *et al.*, 2007; He *et al.*, 2010). However, in the crystal structure of the ctPrp2-ctSpp2 complex, the 5'-HP seems to be arrested in its position by the G-patch of Spp2 making its function in RNA separation

questionable. Instead, it might act as cantilever pushing against Spp2 upon ATP binding or hydrolysis, with Spp2 serving as adapter molecule, which transfers mechanical energy to other components of the spliceosome. Since no structural information about the Prp2-Spp2 complex in its ATP-bound state is available, this hypothesis is however speculative and more structural and functional data on the interaction of Spp2 with the 5'-HP are required to investigate the physiological relevance of this model.

Appendix

A.1 Supplemental tables

Table A.1: DNA oligonucleotides used for IBA StarGate cloning (see 2.2.6.2)

Nr.	Name	Sequence (5'-3')
S1	ctPrp2 Fwd	AGCGGCTCTTCAATGTCGTCAAAAACGGTACGCCTTCGTC
S2	ctPrp2 270 Fw	AGCGGCTCTTCAATGACGCCGAGCAGGCAGGCTGCTGAAG
S3	ctPrp2 921 Rev	AGCGGCTCTTCTCCCTGCCCCAAACTCACTCAACCACCTCGG
S4	ctSpp2 1 Fwd	AGCGGCTCTTCAATGTCAGATCACGATACAAAATCTTC
S5	ctSpp2 313 Rev	AGCGGCTCTTCTCCCCGTCACTCCCACGCTC
S6	ctSpp2 172 Fwd	AGCGGCTCTTCAATGACCCTCGAAGAGGAAGCC
S7	ctSpp2 211 Fwd	AGCGGCTCTTCAATGGTTGACGACTTGGTGAAAACC
S8	ctSpp2 254 Rev	AGCGGCTCTTCTCCCATCTTCGGCCTCTTTAACGTTTC

Table A.2: DNA oligonucleotides used for restriction site cloning (see 2.2.6.1)

Nr.	Name	Sequence (5'-3')
R1	Prp2 Fwd	GGAATTCCATATGATGTCAAGTATTACATCTGAAACCGG
R2	Prp2 Rev	CCCTCGAGCCGCCTCCCCTATTTGT
R3	Prp2 597 Fwd	CGGGATCCCTTGGCGTTACTGACCTAATAAAGTTTC
R4	Prp2 581 Fwd	CGGGATCCGAGATCACAAGAACAAATCTTTCT
R5	Spp2 Rev	CCGCTCGAGTTATGACTTATTTTTTGTCGTCATC
R6	Spp2 Δ9 F	CGGGATCCAGTAAGACCTTGAAAAAAATATTTCCAA
R7	Spp2 Δ33 F	CGGGATCCGATTGGGATGATGCGGAAACTGCAA
R8	Spp2 Δ91 F	CGGGATCCGTCACCGAAAAAAGAATATAATGAGGTA
R9	Spp2 Δ99 F	CGGGATCCGTACCTGTCGAGGAATTTGGAGATGC
R10	Spp2 Δc171 R	CCGCTCGAGTTAATCGATTTTTACGACGGGCATAAATGA
R11	Spp2 Δc183 R	CCGCTCGAGTTATTTTTGTCGTCATCAACTTTAGTGCCC
R12	Spp2 Δc150 R	CCGCTCGAGTTAACCAATTCCCAGTCCATCAGGATGTAT
R13	Spp2 Δ9 F Nde	GGAATTCCATATGAGTAAGACCTTGAAAAAAATATTTCCAA

Table A.3: DNA oligonucleotides used for sequencing (see 2.2.3)

Nr.	Name	Sequence (5'-3')
Seq1	ENTRY-Primer-for2	GCGAAACGATCCTCGAAG
Seq2	ENTRY-Primer-rev	CCCCTGATTCTGTGGATAACCG
Seq3	T7 fwd	TAATACGACTCACTATAGGG
Seq4	T7 term	CTAGTTATTGCTCAGCGGTG
Seq5	pGEX fwd	GGGCTGGCAAGCCACGTTTGGTG
Seq6	pGEX rev	CCGGGAGCTGCATGTGTCAGAGG

Appendix

Table A.4: Abbreviations used in this work

Å	–	Ångström (1 Å = 10^{-10} m)
APS	–	ammoniumpersulfate
ATP	–	adenosine triphosphate
BP	–	branch point
CD	–	circular dichroism
ct	–	*Chaetomium thermophilum*
CTD	–	C-terminal domain
CV	–	column volume
Da	–	Dalton
DNA	–	deoxyribonucleic acid
dsDNA	–	double-stranded DNA
dsRNA	–	double-stranded RNA
5'HP	–	5'-hairpin
IPTG	–	isopropyl-β-D-thiogalactopyranoside
kb	–	kilobases
K_cat	–	Turnover number
K_d	–	dissociation constant
kDa	-	kilodalton
K_m	–	Michaelis constant
mRNA	–	messenger RNA
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
RNA	–	ribonucleic acid
RMSD	–	root mean square deviation
RNP	–	ribonucleic protein
SDS-PAGE	–	SDS polyacrylamide gel electrophoresis
SF	–	superfamily
ssDNA	–	single-stranded DNA
ssRNA	–	single-stranded RNA
3'/5'ss	–	3'/5' splice site
tRNA	–	transfer RNA
UsnRNA	–	Uridine rich small nuclear RNA
UTP	–	uridine triphosphate
v_{max}	–	maximal reaction velocity
WH	–	winged helix
y	–	*Saccharomyces cerevisiae* (yeast)

A.2 Supplemental figures

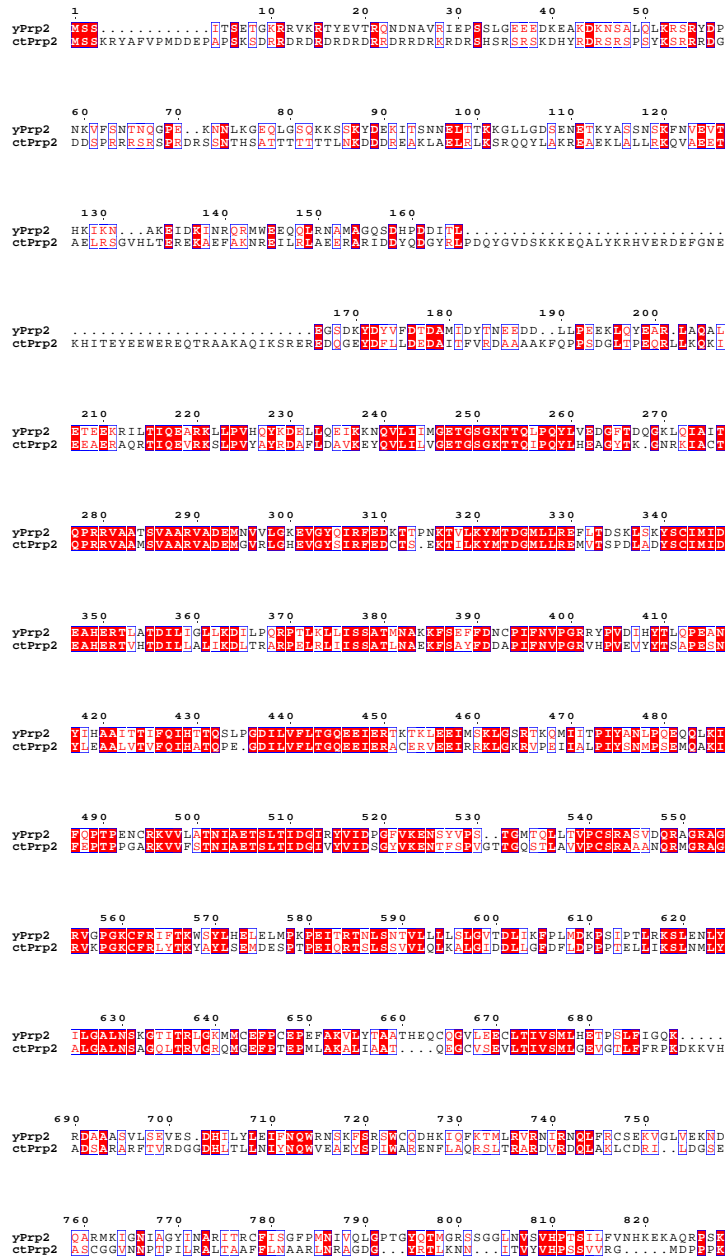**Figure A.1:** Alignment of the aminoacid sequence of yPrp2 and ctPrp2. The alignment was generated using ClustalW

```
ySpp2     1  .......MSKFSLKLG.GKTLKKNISKKTK..KKNSLQKANLFDWDDAETASLSHKPQSK
ctSpp2    1  MSDHDTKSSRINIQLGKSSRSSLDTKKRSRPVYGKQYRAHALHDESESDDGDFEQEYGRA


ySpp2    51  IKTQSIDKFDLDEE.........................SSSKKKLVIKLSENADTKKN
ctSpp2   61  EATTSYDIYDTKSDGGDRRRARDRSSSSDRDRSRTRNGDTASDRRDKARRDDRDRSRSRN


ySpp2    85  DAPLVEYVTEK.......................................................
ctSpp2  121  SNGVEDNETEKNPIKYGLTIKTKSREDIGERTSRDSRSQSEDFGKEGKGSKTLEEEALEA


ySpp2    96  ........................EYNEVPVEEFGDALLRGMGWESDSEQDSKGDKTQSR
ctSpp2  181  LMGSGVPKRKQPVFDEPDREPQLEDYKAVPVDDFGENLLRSFGWDGKMRGKVKEVKRYAN


ySpp2   132  NKDVSNVSQIHPDGLGIG......AKLNKAINVEEASFMPVVKIDKITGTKVPDDKKNKS
ctSpp2  241  LAGLGARNVKEAEDLGAWDPKAGKSSSSRPVRLDDYRREEERKRRRLEDKYGDSYKRERE


ySpp2        ..............
ctSpp2  301  RERERERERGSDR
```

Figure A.2: Alignment of the aminoacid sequence of ySpp2 and ctSpp2. The alignment was generated using ClustalW

```
                105          115          125          135          145          155
                 •            •            •            •            •            •
S.c. Spp2   VEEFGDALLRGMGWE SDSEQDSK-GDKTQSRNKDVS-NVSQIHPDGLGIGAKLNKA---

C.t. Spp2   VDDFGENLLRSFGWD-GKMRG----------KVKEVK-RYANLA--GLGARNVKEAED--
H.s. Spp2   VEAYGLAMLRGMGWKPGEGIG--------RTFNQVVKPRVNSLRPKGLGLGAN-LT----
S.c. Ntr1   TYGIGAKLLSSMGYVAGKGLGKDGSGITTPIETQS---RPMHNA--GLGMFSN-TNSS--
H.s. Cmtr1  YNSVSQKLMAKMGFREGEGLGKYSQGRKDI------VEASSQKGRRGLGLTLRGFDQELN
```

Figure A.3: Alignment of the aminoacid sequence of G-patch motifs from different proteins. The alignment was generated using ClustalW

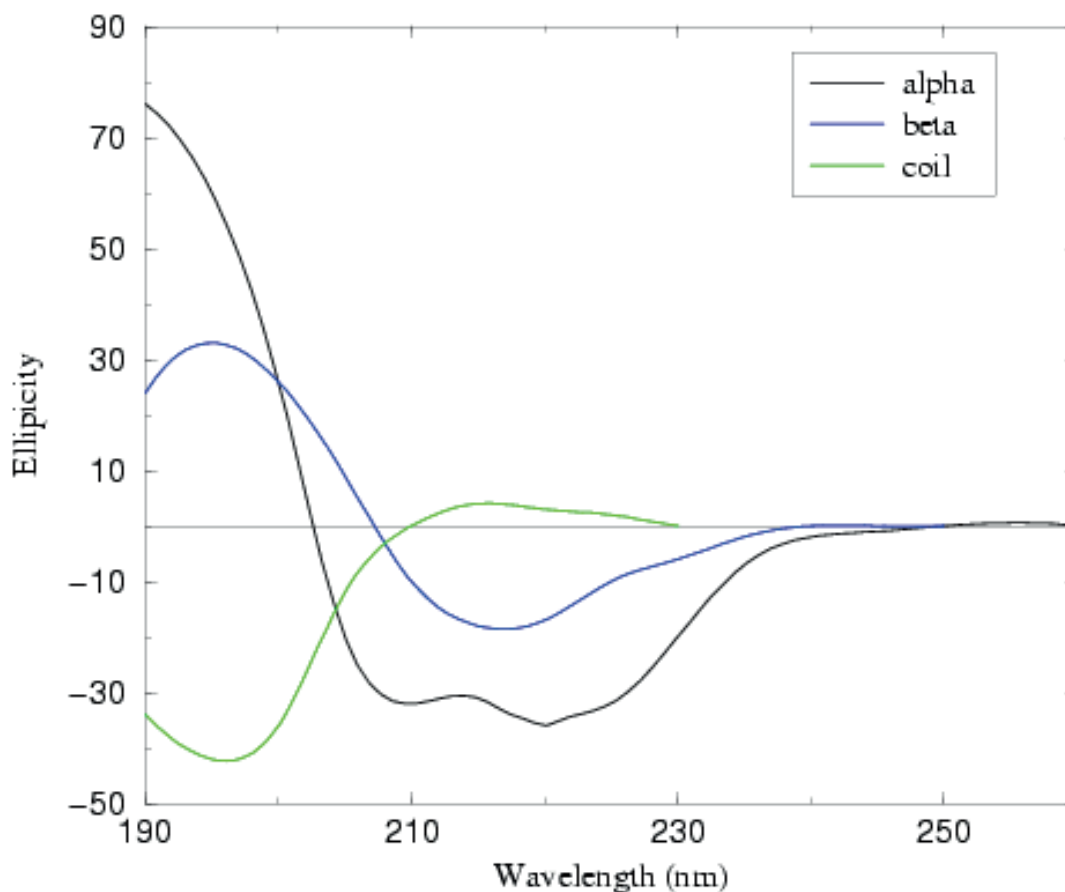**Figure A.4:** Characteristic circular dichroism spectra for the helix, sheet and turn elements of secondary structure *(from: http://besley.chem.nottingham.ac.uk/research/research-prospec.html)*
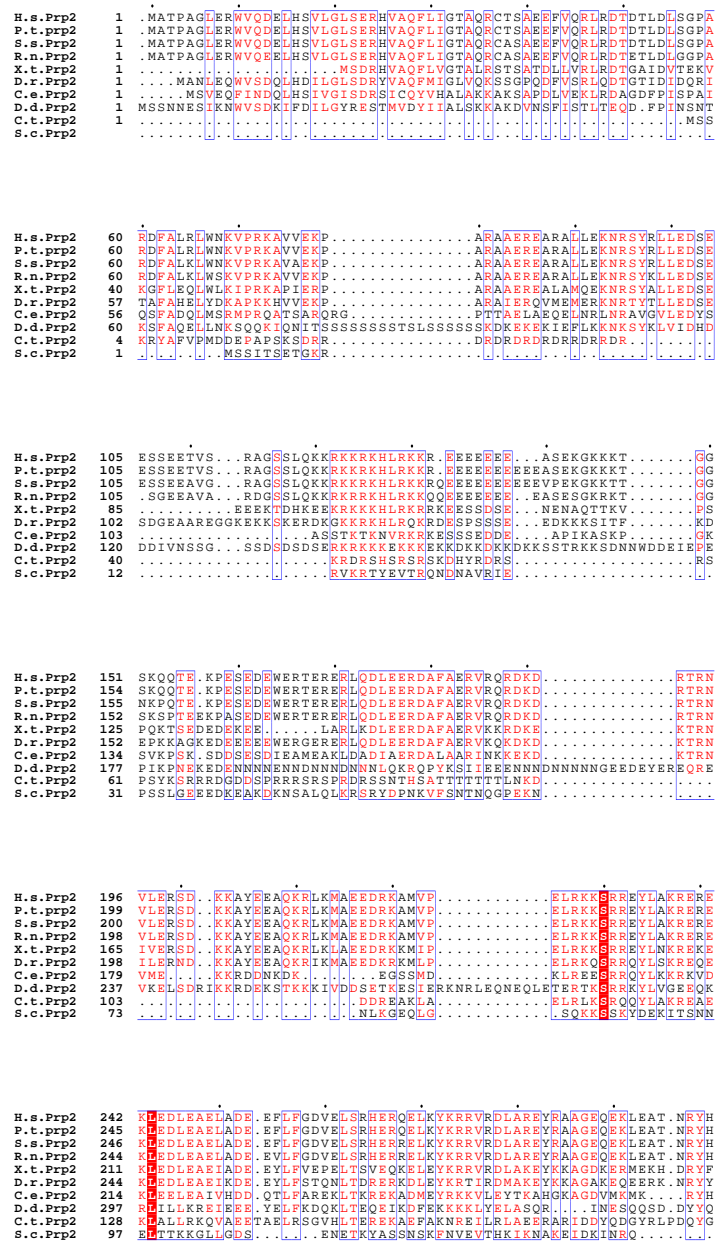
Figure A.5: Alignment of the aminoacid sequence of Prp2 orthologues from different organisms. The alignment was generated using ClustalW

Bibliography

Abelson John.
  Is the spliceosome a ribonucleoprotein enzyme?
  Nature structural & molecular biology, 15(12):1235–7, 2008.

Abelson John, Trotta C R, and Li H.
  tRNA splicing.
  The Journal of biological chemistry, 273(21):12685–8, 1998.

Achsel T, Brahms 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 Paul D., Grosse-Kunstleve Ralf W., Hung Li-Wei, Ioerger Thomas R.,
  McCoy Airlie J., Moriarty Nigel W., Read Randy J., Sacchettini James C.,
  Sauter Nicholas K., and Terwilliger Thomas C.
  PHENIX : building new software for automated crystallographic structure de-
  termination.
  Acta Crystallographica Section D Biological Crystallography, 58(11):1948–1954,
  2002.

Altschul S F, Gish Warren, Miller W, Myers E W, and Lipman D J.
  Basic local alignment search tool.
  Journal of molecular biology, 215(3):403–10, 1990.

Amlacher Stefan, Sarges Phillip, Flemming Dirk, Noort Vera, Kunze Ruth, De-
  vos Damien P., Arumugam Manimozhiyan, Bork Peer, and Hurt Ed.
  Insight into structure and assembly of the nuclear pore complex by utilizing
  the genome of a eukaryotic thermophile.
  Cell, 146(2):277–89, 2011.

Bibliography

Andersen Christian B F, Ballut Lionel, Johansen Jesper S, Chamieh Hala, Nielsen Klaus H, Oliveira Cristiano L P, Pedersen Jan Skov, Séraphin Bertrand, Le Hir Hervé, and Andersen Gregers Rom.
Structure of the exon junction core complex with a trapped DEAD-box ATPase bound to RNA.
Science (New York, N.Y.), 313(5795):1968–72, 2006.

Appleby Todd C, Anderson Robert, Fedorova Olga, Pyle Anna M, Wang Ruth, Liu Xiaohong, Brendza Katherine M, and Somoza John R.
Visualizing ATP-dependent RNA translocation by the NS3 helicase from HCV.
Journal of molecular biology, 405(5):1139–53, 2011.

Aravind L and Koonin E V.
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 E and Abelson J N.
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.

Ast Gil.
How did alternative splicing evolve?
Nature reviews. Genetics, 5(10):773–82, 2004.

Baker N A, Sept D, Joseph S, Holst M J, and McCammon J a.
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 Craig a and Frick David N.
Monitoring helicase activity with molecular beacons.
BioTechniques, 45(4):433–40, 442, 2008.

Berglund J a, 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 Sergey, Anokhina Maria, Will Cindy L, Urlaub Henning, and Lührmann Reinhard.
Isolation of an active step I spliceosome and composition of its RNP core.
Nature, 452(7189):846–50, 2008.

Biniszkiewicz Detlev, Cesnaviciene E, and Shub David A.
Self-splicing group I intron in cyanobacterial initiator methionine tRNA: evidence for lateral transfer of introns in bacteria.
The EMBO journal, 13(19):4629–35, 1994.

Bleichert Franziska and Baserga Susan J.
The long unwinding road of RNA helicases.
Molecular cell, 27(3):339–52, 2007.

Bono Fulvia, Ebert Judith, Lorentzen Esben, and Conti Elena.
The crystal structure of the exon junction complex reveals how it maintains a stable grip on mRNA.
Cell, 126(4):713–25, 2006.

Boon Kum-Loong, Auchynnikava Tatsiana, Edwalds-Gilbert Gretchen, Barrass J David, Droop Alastair P, Dez Christophe, and Beggs Jean D.
Yeast ntr1/spp382 mediates prp43 function in postspliceosomes.
Molecular and cellular biology, 26(16):6016–23, 2006.

Bork P and Koonin E V.
An expanding family of helicases within the 'DEAD/H' superfamily.
Nucleic acids research, 21(3):751–2, 1993.

Brünger A T, Adams Paul D, Clore G Marius, DeLano W L, Gros Piet, Grosse-Kunstleve R W, Jiang J S, Kuszewski John, Nilges Michael, Pannu Navraj S, Read Randy J, Rice Luke M, Simonson Thomas, and Warren G L.
Crystallography & NMR system: A new software suite for macromolecular structure determination.
Acta crystallographica. Section D, Biological crystallography, 54(Pt 5):905–21, 1998.

Buchan D W a, Ward S M, Lobley a E, Nugent T C O, Bryson K, and Jones D T.
Protein annotation and modelling servers at University College London.
Nucleic acids research, 38(Web Server issue):W563–8, 2010.

Bibliography

Burgess S M and Guthrie C.
A mechanism to enhance mRNA splicing fidelity: the RNA-dependent ATPase Prp16 governs usage of a discard pathway for aberrant lariat intermediates.
Cell, 73(7):1377–91, 1993.

Büttner Katharina, Nehring Sebastian, and Hopfner Karl-Peter.
Structural basis for DNA duplex separation by a superfamily-2 helicase.
Nature structural & molecular biology, 14(7):647–52, 2007.

Chakrabarti Sutapa, Jayachandran Uma, Bonneau Fabien, Fiorini Francesca, Basquin Claire, Domcke Silvia, Le Hir Hervé, and Conti Elena.
Molecular mechanisms for the RNA-dependent ATPase activity of Upf1 and its regulation by Upf2.
Molecular cell, 41(6):693–703, 2011.

Chan Russell T, Robart Aaron R, Rajashankar Kanagalaghatta R, Pyle Anna Marie, and Toor Navtej.
Crystal structure of a group II intron in the pre-catalytic state.
Nature structural & molecular biology, 19(5):555–7, 2012.

Chen Mo and Manley James L.
Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches.
Nature reviews. Molecular cell biology, 10(11):741–54, 2009.

Cheng Zhihong, Coller Jeff, Parker Roy, and Song Haiwei.
Crystal structure and functional analysis of DEAD-box protein Dhh1p.
RNA (New York, N.Y.), 11(8):1258–70, 2005.

Christian Henning, Hofele Romina V., Urlaub Henning, and Ficner Ralf.
Insights into the activation of the helicase Prp43 by biochemical studies and structural mass spectrometry.
Nucleic acids research, 2013.

Clerici Marcello, Mourão André, Gutsche Irina, Gehring Niels H, Hentze Matthias W, Kulozik Andreas, Kadlec Jan, Sattler Michael, and Cusack Stephen.
Unusual bipartite mode of interaction between the nonsense-mediated decay

factors, UPF1 and UPF2.
The EMBO journal, 28(15):2293–306, 2009.

Cordin Olivier, Banroques Josette, Tanner N Kyle, and Linder Patrick.
The DEAD-box protein family of RNA helicases.
Gene, 367:17–37, 2006.

Cordin Olivier, Hahn Daniela, and Beggs Jean D.
Structure, function and regulation of spliceosomal RNA helicases.
Current opinion in cell biology, 24(3):431–8, 2012.

Cougot Nicolas, Dijk Erwin, Babajko Sylvie, and Séraphin Bertrand.
'Cap-tabolism'.
Trends in biochemical sciences, 29(8):436–44, 2004.

Das Rita, Zhou Zhaolan, and Reed Robin.
Functional Association of U2 snRNP with the ATP-Independent Spliceosomal Complex E.
Molecular Cell, 5(5):779–787, 2000.

Dayie Kwaku T and Padgett Richard a.
A glimpse into the active site of a group II intron and maybe the spliceosome, too.
RNA (New York, N.Y.), 14(9):1697–703, 2008.

Deutsch M and Long M.
Intron-exon structures of eukaryotic model organisms.
Nucleic acids research, 27(15):3219–28, 1999.

Dolinsky Todd J, Czodrowski Paul, Li Hui, Nielsen Jens E, Jensen Jan H, Klebe Gerhard, and Baker Nathan a.
PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations.
Nucleic acids research, 35(Web Server issue):W522–5, 2007.

Drummond Douglas R, Armstrong John, and Colman Alan.
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.

Bibliography

Dumont Sophie, Cheng Wei, Serebrov Victor, Beran Rudolf K, Tinoco Ignacio, Pyle Anna Marie, and Bustamante Carlos.
RNA translocation and unwinding mechanism of HCV NS3 helicase and its coordination by ATP.
Nature, 439(7072):105–8, 2006.

Dyson H Jane and Wright Peter E.
Intrinsically unstructured proteins and their functions.
Nature reviews. Molecular cell biology, 6(3):197–208, 2005.

Edwalds-Gilbert Gretchen, Kim Dong-Ho, Silverman Edward, and Lin Ren-Jang.
Definition of a spliceosome interaction domain in yeast Prp2 ATPase.
RNA, 10(2):210–220, 2004.

Emsley P, Lohkamp B, Scott W G, and Cowtan K.
Features and development of Coot.
Acta crystallographica. Section D, Biological crystallography, 66(Pt 4):486–501, 2010.

Fabrizio Patrizia, Dannenberg Julia, Dube Prakash, Kastner Berthold, Stark Holger, Urlaub Henning, and Lührmann Reinhard.
The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome.
Molecular cell, 36(4):593–608, 2009.

Fairman-Williams Margaret E, Guenther Ulf-Peter, and Jankowsky Eckhard.
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(December 2004):589–595, 2005.

Fica Sebastian M., Tuttle Nicole, Novak Thaddeus, Li Nan-Sheng, Lu Jun, Koodathingal Prakash, Dai Qing, Staley Jonathan P., and Piccirilli Joseph a.
RNA catalyses nuclear pre-mRNA splicing.
Nature, 503(7475):229–234, 2013.

Finn Robert D, Mistry Jaina, Tate John, Coggill Penny, Heger Andreas, Pollington Joanne E, Gavin O Luke, Gunasekaran Prasad, Ceric Goran,

Forslund Kristoffer, Holm Liisa, Sonnhammer Erik L L, Eddy Sean R, and Bateman Alex.
The Pfam protein families database.
Nucleic acids research, 38(Database issue):D211–22, 2010.

Fox-Walsh Kristi L, Dou Yimeng, Lam Bianca J, Hung She-Pin, Baldi Pierre F, and Hertel Klemens J.
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.

Gorrec Fabrice.
The MORPHEUS protein crystallization screen.
Journal of applied crystallography, 42(Pt 6):1035–1042, 2009.

Gouet Patrice, Courcelle Emmanuel, Stuart David I, and Métoz Frédéric.
ESPript: analysis of multiple sequence alignments in PostScript.
Bioinformatics (Oxford, England), 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.

Gu Meigang and Rice Charles M.
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.

Halbach Felix, Rode Michaela, and Conti Elena.
The crystal structure of S. cerevisiae Ski2, a DExH helicase associated with the cytoplasmic functions of the exosome.
RNA (New York, N.Y.), 18(1):124–34, 2012.

He Yangzi, Andersen Gregers R, and Nielsen Klaus H.
Structural basis for the function of DEAH helicases.
EMBO reports, 11(3):180–6, 2010.

Bibliography

Hegyi Hedi, Schad Eva, and Tompa Peter.
  Structural disorder promotes assembly of protein complexes.
  BMC structural biology, 7:65, 2007.

Henn Arnon, Cao Wenxiang, Hackney David D, and De La Cruz Enrique M.
  The ATPase cycle mechanism of the DEAD-box rRNA helicase, DbpA.
  Journal of molecular biology, 377(1):193–205, 2008.

Hilbert Manuel, Karow Anne R, and Klostermeier Dagmar.
  The mechanism of ATP-dependent RNA unwinding by DEAD box proteins.
  Biological chemistry, 390(12):1237–50, 2009.

Hocine Sami, Singer Robert H, and Grünwald David.
  RNA processing and export.
  Cold Spring Harbor perspectives in biology, 2(12):a000752, 2010.

Hoffman B E and Grabowski P J.
  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 Karl-Peter and Michaelis Jens.
  Mechanisms of nucleic acid translocases: lessons from structural biology and
  single-molecule biophysics.
  Current opinion in structural biology, 17(1):87–95, 2007.

Horowitz D S 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.

Hsu Cecilia L and Stevens Audrey.
  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.

Jády B E and Kiss T.
  A small nucleolar guide RNA functions both in 2'-O-ribose methylation and
  pseudouridylation of the U5 spliceosomal RNA.
  The EMBO journal, 20(3):541–51, 2001.

Jankowsky Eckhard and Fairman Margaret E.
  RNA helicases–one fold for many functions.
  Current opinion in structural biology, 17(3):316–24, 2007.

Jeffery Constance J.
  Multifunctional proteins: examples of gene sharing.
  Annals of medicine, 35(1):28–35, 2003a.

Jeffery Constance J.
  Moonlighting proteins: old proteins learning new tricks.
  Trends in genetics : TIG, 19(8):415–7, 2003b.

Jeffery Constance J.
  Moonlighting proteins–an update.
  Molecular bioSystems, 5(4):345–50, 2009.

Kabsch Wolfgang.
  XDS.
  Acta crystallographica. Section D, Biological crystallography, 66(Pt 2):125–32, 2010.

Kambach C, Walke S, Young R, Avis J M, Fortelle E, Raker V a, Lührmann R, Li J, and Nagai K.
  Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs.
  Cell, 96(3):375–87, 1999.

Kelly Sharon M, Jess Thomas J, and Price Nicholas C.
  How to study proteins by circular dichroism.
  Biochimica et biophysica acta, 1751(2):119–39, 2005.

Kim D H and Rossi John J.
  The first ATPase domain of the yeast 246-kDa protein is required for in vivo unwinding of the U4/U6 duplex.
  RNA (New York, N.Y.), 5(7):959–71, 1999.

Kim SH, Smith Joseph, Claude Alejandro, and Lin RJ.
  The purified yeast pre-mRNA splicing factor PRP2 is an RNA-dependent NT-Pase.
  The EMBO journal, 1(6):2319–2326, 1992.

Bibliography

Kistler Amy L and Guthrie Christine.
  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 Maria M and Query Charles C.
  Insights into the mechanisms of splicing: more lessons from the ribosome.
  Genes & development, 19(19):2255–60, 2005.

Konforti B B, Koziolkiewicz M J, and Konarska M M.
  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.

Korneta Iga and Bujnicki Janusz M.
  Intrinsic disorder in the human spliceosomal proteome.
  PLoS computational biology, 8(8):e1002641, 2012.

Kozlowski Lukasz P and Bujnicki Janusz M.
  MetaDisorder: a meta-server for the prediction of intrinsic disorder in proteins.
  BMC bioinformatics, 13(1):111, 2012.

Krishnan Ramya, Blanco Mario R, Kahlscheuer Matthew L, Abelson John,
  Guthrie Christine, and Walter Nils G.
  Biased Brownian ratcheting leads to pre-mRNA remodeling and capture prior
  to first-step splicing.
  Nature structural & molecular biology, 20(12):1450–1457, 2013.

Kruger K, Grabowski P J, Zaug a J, Sands J, Gottschling D E, and Cech T R.
  Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA in-
  tervening sequence of Tetrahymena.
  Cell, 31(1):147–57, 1982.

Laemmli U K.
  Cleavage of structural proteins during the assembly of the head of bacteriophage
  T4.
  Nature, 227(5259):680–5, 1970.

Larkin M a, Blackshields G, Brown N P, Chenna R, McGettigan P a,
  McWilliam H, Valentin F, Wallace I M, Wilm A, Lopez R, Thompson J D,

Gibson T J, and Higgins D G.
Clustal W and Clustal X version 2.0.
Bioinformatics (Oxford, England), 23(21):2947–8, 2007.

Last R L, Maddock J R, and Woolford J L.
Evidence for related functions of the RNA genes of Saccharomyces cerevisiae.
Genetics, 117(4):619–31, 1987.

Lebaron Simon, Papin Christophe, Capeyrou Régine, Chen Yan-Ling, Froment Carine, Monsarrat Bernard, Caizergues-Ferrer Michèle, Grigoriev Mikhail, and Henry Yves.
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.

Letunic Ivica, Doerks Tobias, and Bork Peer.
SMART 7: recent updates to the protein domain annotation resource.
Nucleic acids research, 40(Database issue):D302–5, 2012.

Levin Mikhail K, Gurjar Madhura M, and Patel Smita S.
ATP binding modulates the nucleic acid affinity of hepatitis C virus helicase.
The Journal of biological chemistry, 278(26):23311–6, 2003.

Liu Yen-chi, Chen Hsin-chou, Wu Nan-ying, Cheng Soo-chen, and Iol M O L C E L L B.
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 Pascal J and Séraphin Bertrand.
Genomic-scale quantitative analysis of yeast pre-mRNA splicing: implications for splice-site recognition.
RNA (New York, N.Y.), 5(9):1135–7, 1999.

Luo Dahai, Xu Ting, Watson Randall P, Scherer-Becker Daniella, Sampath Aruna, Jahnke Wolfgang, Yeong Sui Sum, Wang Chern Hoe, Lim Siew Pheng, Strongin Alex, Vasudevan Subhash G, and Lescar Julien.
Insights into RNA unwinding and ATP hydrolysis by the flavivirus NS3 protein.
The EMBO journal, 27(23):3209–19, 2008.

Bibliography

Madhani H D 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.

Martin Arnold, Schneider Susanne, and Schwer Beate.
Prp43 is an essential RNA-dependent ATPase required for release of lariat-intron from the spliceosome.
The Journal of biological chemistry, 277(20):17743–50, 2002.

Mathew Rebecca, Hartmuth Klaus, Möhlmann Sina, Urlaub Henning, Ficner Ralf, and Lührmann Reinhard.
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.

McCoy Airlie J, Grosse-Kunstleve Ralf W, Adams Paul D, Winn Martyn D, Storoni Laurent C, and Read Randy J.
Phaser crystallographic software.
Journal of applied crystallography, 40(Pt 4):658–674, 2007.

McPheeters D S, 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.

Moore M J and Sharp P A.
Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing.
Nature, 365(6444):364–8, 1993.

Moore Melissa J and Proudfoot Nick J.
Pre-mRNA processing reaches back to transcription and ahead to translation.
Cell, 136(4):688–700, 2009.

Mouaikel John, Verheggen Céline, Bertrand Edouard, Tazi Jamal, and Bordonné Rémy.
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 Sina, Santos Karine F., Hsiao He-Hsuan, Will Cindy L., Urlaub Henning, Wahl Markus C., Lührmann Reinhard, and Lu Reinhard.
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 Alexey G.
OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences.
The EMBO journal, 12(3):861–7, 1993.

Myong Sua, Bruno Michael M, Pyle Anna M, and Ha Taekjip.
Spring-loaded mechanism of DNA unwinding by hepatitis C virus NS3 helicase.
Science (New York, N.Y.), 317(5837):513–6, 2007.

Newman Janet, Egan David, Walter Thomas S, Meged Ran, Berry Ian, Ben Jelloul Marouane, Sussman Joel L, Stuart David I, and Perrakis Anastassis.
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 Klaus H, Chamieh Hala, Andersen Christian B F, Fredslund Folmer, Hamborg Kristiane, Le Hir Hervé, and Andersen Gregers R.
Mechanism of ATP turnover inhibition in the EJC.
RNA (New York, N.Y.), 15(1):67–75, 2009.

Oberer Monika, Marintchev Assen, and Wagner Gerhard.
Structural basis for the enhancement of eIF4A helicase activity by eIF4G.
Genes & development, 19(18):2212–23, 2005.

O'Day C L, Dalbadie-McFarland G, and Abelson John.
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.

Bibliography

Ohno M, Segref A, Bachi A, Wilm M, and Mattaj I W.
PHAX, a mediator of U snRNA nuclear export whose activity is regulated by phosphorylation.
Cell, 101(2):187–98, 2000.

Ohrt Thomas, Prior Mira, Dannenberg Julia, Odenwälder Peter, Dybkov Olexandr, Rasche Nicolas, Schmitzová Jana, Gregor Ingo, Fabrizio Patrizia, Enderlein Jörg, and Lührmann Reinhard.
Prp2-mediated protein rearrangements at the catalytic core of the spliceosome as revealed by dcFCCS.
RNA (New York, N.Y.), 18(6):1244–56, 2012.

Oyama Takuji, Oka Hayato, Mayanagi Kouta, Shirai Tsuyoshi, Matoba Kyoko, Fujikane Ryosuke, Ishino Yoshizumi, and Morikawa Kosuke.
Atomic structures and functional implications of the archaeal RecQ-like helicase Hjm.
BMC structural biology, 9:2, 2009.

Padgett R a, Konarska M M, Grabowski P J, Hardy S F, and Sharp P a.
Lariat RNA's as intermediates and products in the splicing of messenger RNA precursors.
Science (New York, N.Y.), 225(4665):898–903, 1984.

Pang Q, Hays J B, and Rajagopal I.
Two cDNAs from the plant Arabidopsis thaliana that partially restore recombination proficiency and DNA-damage resistance to E. coli mutants lacking recombination-intermediate-resolution activities.
Nucleic acids research, 21(7):1647–53, 1993.

Paule Marvin R and White Robert J.
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.

Peebles C L, Perlman P S, Mecklenburg K L, Petrillo M L, Tabor J H, Jarrell K a, and Cheng H L.
  A self-splicing RNA excises an intron lariat.
  Cell, 44(2):213–23, 1986.

Pena Vladimir, Rozov Alexey, Fabrizio Patrizia, Lührmann Reinhard, and Wahl Markus C.
  Structure and function of an RNase H domain at the heart of the spliceosome.
  The EMBO journal, 27(21):2929–40, 2008.

Pyle Anna Marie.
  Translocation and unwinding mechanisms of RNA and DNA helicases.
  Annual review of biophysics, 37:317–36, 2008.

Radaev Sergei, Li Sean, and Sun Peter D.
  A survey of protein-protein complex crystallizations.
  Acta crystallographica. Section D, Biological crystallography, 62(Pt 6):605–12, 2006.

Raghunathan P L and Guthrie C.
  RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2.
  Current biology : CB, 8(15):847–55, 1998.

Raker V a, 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.

Ramachandran G N and Sasisekharan V.
  Conformation of polypeptides and proteins.
  Advances in protein chemistry, 23:283–438, 1968.

Reed R.
  Mechanisms of fidelity in pre-mRNA splicing.
  Current opinion in cell biology, 12(3):340–5, 2000.

Richards Jodi D, Johnson Kenneth a, Liu Huanting, McRobbie Anne-Marie, McMahon Stephen, Oke Muse, Carter Lester, Naismith James H, and

Bibliography

White Malcolm F.
Structure of the DNA repair helicase hel308 reveals DNA binding and autoinhibitory domains.
The Journal of biological chemistry, 283(8):5118–26, 2008.

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

Roy J, Kim K, Maddock J R, Anthony J G, and Woolford J L.
The final stages of spliceosome maturation require Spp2p that can interact with the DEAH box protein Prp2p and promote step 1 of splicing.
RNA (New York, N.Y.), 1(4):375–90, 1995.

Ruskin B, Krainer a R, Maniatis T, and Green M R.
Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro.
Cell, 38(1):317–31, 1984.

Russell Anthony G, Charette J Michael, Spencer David F, and Gray Michael W.
An early evolutionary origin for the minor spliceosome.
Nature, 443(7113):863–6, 2006.

Rymond B C 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 Karine F, Jovin Sina Mozaffari, Weber Gert, Pena Vladimir, Lührmann Reinhard, and Wahl Markus C.
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 Marc, Will Cindy L, Anokhina Maria, Tazi Jamal, Urlaub Henning, and Lührmann Reinhard.
Exon definition complexes contain the tri-snRNP and can be directly converted

into B-like precatalytic splicing complexes.
Molecular cell, 38(2):223–35, 2010.

Schütz Patrick, Bumann Mario, Oberholzer Anselm Erich, Bieniossek Christoph, Trachsel Hans, Altmann Michael, and Baumann Ulrich.
Crystal structure of the yeast eIF4A-eIF4G complex: an RNA-helicase controlled by protein-protein interactions.
Proceedings of the National Academy of Sciences of the United States of America, 105(28):9564–9, 2008.

Schütz Patrick, Karlberg Tobias, Berg Susanne, Collins Ruairi, Lehtiö Lari, Högbom Martin, Holmberg-Schiavone Lovisa, Tempel Wolfram, Park Hee-Won, Hammarström Martin, Moche Martin, Thorsell Ann-Gerd, and Schüler Herwig.
Comparative structural analysis of human DEAD-box RNA helicases.
PloS one, 5(9), 2010.

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

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

Sengoku Toru, Nureki Osamu, Nakamura Akira, Kobayashi Satoru, and Yokoyama Shigeyuki.
Structural basis for RNA unwinding by the DEAD-box protein Drosophila Vasa.
Cell, 125(2):287–300, 2006.

Silverman E.J., Maeda Ayaka, Wei Janet, Smith Paul, Beggs J.D., and Lin R.J.
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, 2004.

Singh R and Reddy R.
Gamma-monomethyl phosphate: a cap structure in spliceosomal U6 small nuclear RNA.

Bibliography

Proceedings of the National Academy of Sciences of the United States of America, 86(21):8280–3, 1989.

Singleton Martin R, Dillingham Mark S, and Wigley Dale B.
Structure and mechanism of helicases and nucleic acid translocases.
Annual review of biochemistry, 76:23–50, 2007.

Small Eliza C, Leggett Stephanie R, Winans Adrienne a, and Staley Jonathan P.
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 W and Valcárcel J.
Alternative pre-mRNA splicing: the logic of combinatorial control.
Trends in biochemical sciences, 25(8):381–8, 2000.

Smith Duncan J, Query Charles C, and Konarska Maria M.
"Nought may endure but mutability": spliceosome dynamics and the regulation of splicing.
Molecular cell, 30(6):657–66, 2008.

Staley J P and Guthrie C.
Mechanical devices of the spliceosome: motors, clocks, springs, and things.
Cell, 92(3):315–26, 1998.

Staley J P and Guthrie C.
An RNA switch at the 5' splice site requires ATP and the DEAD box protein Prp28p.
Molecular cell, 3(1):55–64, 1999.

Stein Lincoln D, Bao Zhirong, Blasiar Darin, Blumenthal Thomas, Brent Michael R, Chen Nansheng, Chinwalla Asif, Clarke Laura, Clee Chris, Coghlan Avril, Coulson Alan, D'Eustachio Peter, Fitch David H A, Fulton Lucinda A, Fulton Robert E, Griffiths-Jones Sam, Harris Todd W, Hillier LaDeana W, Kamath Ravi, Kuwabara Patricia E, Mardis Elaine R, Marra Marco A, Miner Tracie L, Minx Patrick, Mullikin James C, Plumb Robert W, Rogers Jane, Schein Jacqueline E, Sohrmann Marc, Spieth John, Stajich Jason E, Wei C, Willey David, Wilson Richard K, Durbin Richard, and Waterston Robert H.

The genome sequence of Caenorhabditis briggsae: a platform for comparative genomics.
PLoS biology, 1(2):E45, 2003.

Story R M, Li H, and Abelson J N.
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 H S, Bird L E, Brannigan J A, and Wigley D B.
Crystal structure of a DExx box DNA helicase.
Nature, 384(6607):379–83, 1996.

Sun J S and Manley J L.
A novel U2-U6 snRNA structure is necessary for mammalian mRNA splicing.
Genes & development, 9(7):843–54, 1995.

Tanaka Naoko and Schwer Beate.
Characterization of the NTPase, RNA-binding, and RNA helicase activities of the DEAH-box splicing factor Prp22.
Biochemistry, 44(28):9795–803, 2005.

Tanaka Naoko, Aronova Anna, and Schwer Beate.
Ntr1 activates the Prp43 helicase to trigger release of lariat-intron from the spliceosome.
Genes & development, 21(18):2312–25, 2007.

Tanner N Kyle, Cordin Olivier, Banroques Josette, Doère Monique, and Linder Patrick.
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 W Y and Steitz J a.
A novel spliceosome containing U11, U12, and U5 snRNPs excises a minor class (AT-AC) intron in vitro.
Cell, 84(5):801–11, 1996.

Bibliography

Tarun S Z and Sachs B.
  Association of the yeast poly(A) tail binding protein with translation initiation
  factor eIF-4G.
  The EMBO journal, 15(24):7168–77, 1996.

Teigelkamp S, McGarvey M, Plumpton M, and Beggs J D.
  The splicing factor PRP2, a putative RNA helicase, interacts directly with pre-
  mRNA.
  The EMBO journal, 13(4):888–97, 1994.

Teigelkamp S, Newman a J, and Beggs J D.
  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 Bettina, Karow Anne R, Köhler Jürgen, Gubaev Airat, and Kloster-
  meier Dagmar.
  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 Amer-
  ica*, 105(2):548–53, 2008.

Tompa Peter, Szász Csilla, and Buday László.
  Structural disorder throws new light on moonlighting.
  Trends in biochemical sciences, 30(9):484–9, 2005.

Toor Navtej, Keating Kevin S, Taylor Sean D, and Pyle Anna Marie.
  Crystal structure of a self-spliced group II intron.
  Science (New York, N.Y.), 320(5872):77–82, 2008.

Toro Nicolás.
  Bacteria and Archaea Group II introns: additional mobile genetic elements in
  the environment.
  Environmental microbiology, 5(3):143–51, 2003.

Trotta C R, Miao F, Arn E a, Stevens S W, Ho C K, Rauhut R, and Abelson J N.
  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 Rong-tzong, Fu Ru-huei, Yeh Fu-lung, Tseng Chi-kang, Lin Yu-chieh, Huang Yu-Hsin, and Cheng Soo-Chen.
Spliceosome disassembly catalyzed by Prp43 and its associated components Ntr1 and Ntr2.
Genes & development, 19(24):2991–3003, 2005.

Tseng Chi-Kang and Cheng Soo-Chen.
Both catalytic steps of nuclear pre-mRNA splicing are reversible.
Science (New York, N.Y.), 320(5884):1782–4, 2008.

Tseng Chi-Kang, Liu Hsueh-lien, and Cheng Soo-chen.
DEAH-box ATPase Prp16 has dual roles in remodeling of the spliceosome in catalytic steps.
RNA (New York, N.Y.), 17(1):145–54, 2011.

Umen J G 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 (New York, N.Y.), 1(6):584–97, 1995.

Valadkhan Saba.
snRNAs as the catalysts of pre-mRNA splicing.
Current opinion in chemical biology, 9(6):603–8, 2005.

Valcárcel J, Gaur R K, Singh R, and Green M R.
Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA [corrected].
Science (New York, N.Y.), 273(5282):1706–9, 1996.

Wachtel Chaim and Manley James L.
Splicing of mRNA precursors: the role of RNAs and proteins in catalysis.
Molecular bioSystems, 5(4):311–6, 2009.

Wagner J D, Jankowsky E, Company M, Pyle a M, and Abelson J N.
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.

Bibliography

Wahl Markus C, Will Cindy L, and Lührmann Reinhard.
  The spliceosome: design principles of a dynamic RNP machine.
  Cell, 136(4):701–18, 2009.

Walbott Hélène, Mouffok Saïda, Capeyrou Régine, Lebaron Simon, Humbert Odile, Tilbeurgh Herman, Henry Yves, and Leulliot Nicolas.
  Prp43p contains a processive helicase structural architecture with a specific regulatory domain.
  The EMBO journal, 29(13):2194–204, 2010.

Walker John E, Saraste Matti, Runswick Michael J, and Gay Nicholas J.
  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.

Warkocki Zbigniew, Odenwälder Peter, Schmitzová Jana, Platzmann Florian, Stark Holger, Urlaub Henning, Ficner Ralf, Fabrizio Patrizia, and Lührmann Reinhard.
  Reconstitution of both steps of Saccharomyces cerevisiae splicing with purified spliceosomal components.
  Nature structural & molecular biology, 16(12):1237–43, 2009.

Weber Gert, Trowitzsch Simon, Kastner Berthold, Lührmann Reinhard, and Wahl Markus C.
  Functional organization of the Sm core in the crystal structure of human U1 snRNP.
  The EMBO journal, 29(24):4172–84, 2010.

Wilkins M R, Gasteiger E, Bairoch A, Sanchez J C, Williams K L, Appel R D, and Hochstrasser D F.
  Protein identification and analysis tools in the ExPASy server.
  Methods in molecular biology (Clifton, N.J.), 112:531–52, 1999.

Will C L and Lührmann R.
  Spliceosomal UsnRNP biogenesis, structure and function.
  Current opinion in cell biology, 13(3):290–301, 2001.

Will Cindy L and Lührmann Reinhard.
Spliceosome structure and function.
Cold Spring Harbor perspectives in biology, 3(7), 2011.

Winn Martyn D, Ballard Charles C, Cowtan Kevin D, Dodson Eleanor J, Emsley Paul, Evans Phil R, Keegan Ronan M, Krissinel Eugene B, Leslie Andrew G W, McCoy Airlie, McNicholas Stuart J, Murshudov Garib N, Pannu Navraj S, Potterton Elizabeth a, Powell Harold R, Read Randy J, Vagin Alexei, and Wilson Keith S.
Overview of the CCP4 suite and current developments.
Acta crystallographica. Section D, Biological crystallography, 67(Pt 4):235–42, 2011.

Wlodaver Alissa M and Staley Jonathan P.
The DExD/H-box ATPase Prp2p destabilizes and proofreads the catalytic RNA core of the spliceosome.
RNA (New York, N.Y.), 2014.

Xiao Xinshu, Wang Zefeng, Jang Minyoung, and Burge Christopher B.
Coevolutionary networks of splicing cis-regulatory elements.
Proceedings of the National Academy of Sciences of the United States of America, 104(47):18583–8, 2007.

Xu Yong-Zhen and Query Charles C.
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 Kui, Zhang Lingdi, Xu Tao, Heroux Annie, and Zhao Rui.
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 Quansheng, Del Campo Mark, Lambowitz Alan M, and Jankowsky Eckhard.
DEAD-box proteins unwind duplexes by local strand separation.
Molecular cell, 28(2):253–63, 2007.

Bibliography

Yean S L, Wuenschell G, Termini J, and Lin R J.
  Metal-ion coordination by U6 small nuclear RNA contributes to catalysis in
  the spliceosome.
  Nature, 408(6814):881–4, 2000.

Zaug A J and Cech T R.
  In vitro splicing of the ribosomal RNA precursor in nuclei of Tetrahymena.
  Cell, 19(2):331–8, 1980.

Acknowledgements

Curriculum vitae

Personal data

Name	Andreas Schmitt
Birthday	December 16th, 1982 in Clausthal-Zellerfeld (Germany)
Nationality	German
Adress	Im Hassel 11, 37077 Göttingen
Contact	aschmit1@uni-goettingen.de

Education

2009 - 2014	Göttingen Graduate School for Neurosciences, Biophysics, and Molecular Biosciences (GGNB) - University of Göttingen
	Doctoral thesis: *"Crystallization and biophysical characterization of spliceosomal protein complexes"*; in the department of Molecular Structural Biology under the supervision of Prof. Dr. Ralf Ficner
2008	Diploma in Biology - University of Göttingen
	Diploma thesis: *"Charakterisierung und Kristallisation des spleißosomalen DEAD-Box Proteins Prp28"*; in the department of Molecular Structural Biology under the supervision of Prof. Dr. Ralf Ficner
2005	Vordiplom in Biology - University of Göttingen
2003	University of Göttingen
2002	Allgemeine Hochschulreife (Abitur) Modellschule Obersberg, Bad Hersfeld