Charles University in Prague

Faculty of Science

Study programme: Special Chemical and Biological Programmes

Field of study: Molecular Biology and Biochemistry of Organisms

Lucie Kulhánková

Strukturní charakterizace Prp45 a jeho interakčních partnerů

Structural characterisation of Prp45 and its interaction partners

Bachelor's thesis

Supervisor: Mgr. Marian Novotný Ph.D.

Advisor: Mgr. Kateřina Abrahámová Ph.D.

Prague, 2016

Acknowledgements

I would like express to my sincere thanks my supervisor Marian Novotný and advisor Kateřina Abrahámová for their constructive suggestions, patience and time they invested in me. I would also like to thank all member of the Laboratory of Gene Expression for giving me the opportunity to work there.

Prohlášení

Prohlašuji, že jsem závěrečnou práci zpracovala samostatně a že jsem uvedla všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze dne 12. 5. 2016

Podpis

Abstract

Prp45 is a SNW protein that is part of a spliceosome and therefore participates in splicing of pre-mRNA into mRNA. In spliceosome Prp45 joins as part of NTC complex before splicing reactions. Prp45 has several known interactional partners (such as small subunit of U2AF, Cyp2, Prp5, Prp22, Clf1 and Cwc3). These proteins serve many functions in the spliceosome from stability to regulation of splicing reactions.

Spliceosome consists of several dozens of proteins and several RNA and undergoes massive structural changes during executing its function. Until recently very little was known about the spliceosome structure; however, recent studies allow us to look at interaction network of this machine. In this work we focused on identification of both previously predicted and new interactional partners of Prp45. Using available spliceosome structures we aimed to describe these proteins structurally and functionally.

Key words: splicing, Prp45, 3D struktura, spliceosome, binding partners

Abstrakt

Prp45 je SNW protein, který je součástí spliceozomu, a tedy se účastní sestřihu pre-mRNA na mRNA. Ve spliceozomu se Prp45 přidává před sestřihovými reakcemi jako součást NTC komplexu. Prp45 má několik známých interakčních partnerů (například U2AF, Cyp2, Prp5, Prp22, Clf1 a Cwc3). Tyto proteiny mají ve spliceozomu mnoho funkcí od jeho stabilizování po regulaci sestřihových reakcí.

Spliceozom se skládá z několika desítek proteinů a několika RNA a při vykonávání své funkce prochází masivními strukturními proměnami. O struktuře spliceozomu bylo až do nedávna známo velice málo, nicméně nedávné studie nám poskytují pohled do interakční sítě toho komplexu. V této práce jsme se soustředili na identifikaci jak dříve navržených tak nových interakčních partnerů Prp45. S využitím dostupných spliceozomálních struktur jsme se pokusily tyto proteiny popsat strukturně a funkčně.

Klíčová slova: sestřih, Prp45, 3D struktura, spliceozom, vazební partneři

Content

1. Used abbreviations

HAT	half tetratricopeptide repeat
ILS	intron lariat spliceosome
Myb	myeloblastosis
NTC	nineteen complex
NTR	nineteen complex related
RRM	RNA recognition motif
SNW	SKI-interacting protein
snRNA	small nuclear ribonucleic acid
snRNP	small nuclear ribonucleoprotein
SH2-like	Src homology 2 like domain
TPR	tetratricopetide repeat
ZBD	zinc binding domain

2. Introduction

In most eukaryotes, transcription of DNA forms pre-mRNA where coding sequences (exons) are interrupted by noncoding sequences (introns). Introns are cut out of pre-mRNA and exons are ligated together and mRNA is formed in splicing. This process is catalysed by spliceosome; complex of small nuclear ribonucleic acid (snRNAs) and several dozens of proteins. Splicing is carried out co-transcriptionally where components of spliceosome are stepwise recruited while transcription is still in progress. In pre-mRNA with multiple introns alternative splicing might occur resulting in variety of mRNAs.

Study of structure of spliceosome has proven rather difficult due to its size as well as its high dynamics. Spliceosome assembly and execution of splicing requires several massive rearrangements during execution of its function. This can be demonstrated by the fact that the very same nucleotides of U6 snRNA are located almost 100 Å apart in different part of spliceosome assembly cycle (*Wan et al., 2016*). Spliceosome needs to have certain dynamics even during the splicing reactions because length of introns may vary. However recent studies put some light on spliceosome structure and in recent years structure of *Schizosaccharomyces pombe* spliceosome was published (*Yan et al. 2015*) as well as several at least partial structures of tri-snRNP for *Homo Sapiens* (*Agafonov et al., 2016*) and *Saccharomyces cerevisiae (Wan et al., 2016*)

Understanding of spliceosome structure as well as interactions among its many proteins and RNAs has its importance not only for understanding one of the core processes in eukaryotic cells and gene expression but also for understanding many diseases caused by splicing defects such as Familial dysautonomia (*Slaugenhaupt et al., 2001)* or retinis pigmentosa *(Wilkie et al., 2008)* .

In this work we focus on spliceosome protein Prpp45 which is a protein of interest in our lab and we are trying to describe its interaction network. Therefore, first aim of this work is to describe known interaction partners of Prp45 and Prp45 itself in both functional and structural way. Second aim of this work is to use new available structure containing Prp45 (*Yan et al., 2015)* and to identify new physical interaction between Prp45 and proteins in spliceosome.

3. Splicing

3.1 Spliceosome assembly and reactions

Spliceosome is a large cellular complex that consists of snRNAs and proteins. Five different snRNA molecules (U1, U2, U4, U5 and U6) are needed at some point either for assembly or for splicing reactions. These snRNAs are accompanied with several associated proteins and Sm-like proteins forming small nuclear ribonucleoprotein (snRNP). Apart from snRNPs, independent proteins and protein complexes, nineteen complex (NTC) and NTC related (NTR) proteins are also present in the spliceosome while it executes its function. While U1 and U2 snRNAs act alone U4, U5 and U6 snRNAs form preassembled tri-snRNP that is the largest particle that joins the spliceosome during its assembly.

Pre-mRNA contains several short and conserved sequences that define an intron (See Fig. 1B); 3′ splicing site (3′ss) which in higher eukaryotes is preceded by the polypyrimidine tract, 5′splicing site (5′ss) and branch point (BP) located 20-40 nucleotides upstream of 3′ss. These sequences are recognizes by parts of the spliceosome (*reviewed by Nilsen, 2003).*

The splicing is executed by two consecutive transesterification reactions (See Fig. 1A). These reactions bear similarity to group II self-splicing introns and are carried out by snRNAs of the spliceosome. Firstly, nucleophilic attack of 5′ss by the –OH group of branch point adenosine is performed leading to formation of lariat structure. In the second transesterification 5′exon′s 3′OH group attacks the 3′ss and 5′and 3′ends of exons are ligated while intron is released and spliceosome dissociates.

U1 snRNP is recruited first on the 5′ss. Then U2 snRNP binds to BP, at this point spliceosome has formed an A complex. Next preassembled U4/U6·U5 tri-snRNP binds to A complex and forms pre-catalytic B complex. Activation of the pre-catalytic B complex is accompanied by enormous structural rearrangements leading to unwinding of U4/U6 di-snRNA, release of both U4 and U1 snRNAs and recruitment of NTC and NTR protein complexes. After U4 snRNAs dissociation from the spliceosome the U6 snRNA forms new paring with U2 snRNA and also with 5′splice site. After

these changes spliceosome is termed B^{act} complex; however, B^{act} is not yet ready for splicing because it lacks its active site. Creation of active site is managed by Prp2. Prp2 by its helicase activity cause yet another structural rearrangement leading to destabilization of several SF3a and SF3b proteins and creating of B* complex (*Ohrt et al., 2012*). At this point first reaction can be catalysed yielding in C complex which catalyses the second reaction. After the second reaction spliceosome forms post-catalytic P complex and exon is released while lariat stays associated. This stage is called intron lariat spliceosome (ILS). Last step in a splicing is release of the lariat after which spliceosome disintegrates and its key components (snRNPs, NTC, and NTR) are recycled and the spliceosome can be reassembled for another splicing cycle (See Fig.1C) (*rewieved in Wahl et al., 2009; Will &* Luhrmann, *2011; and Matera et al., 2014*).

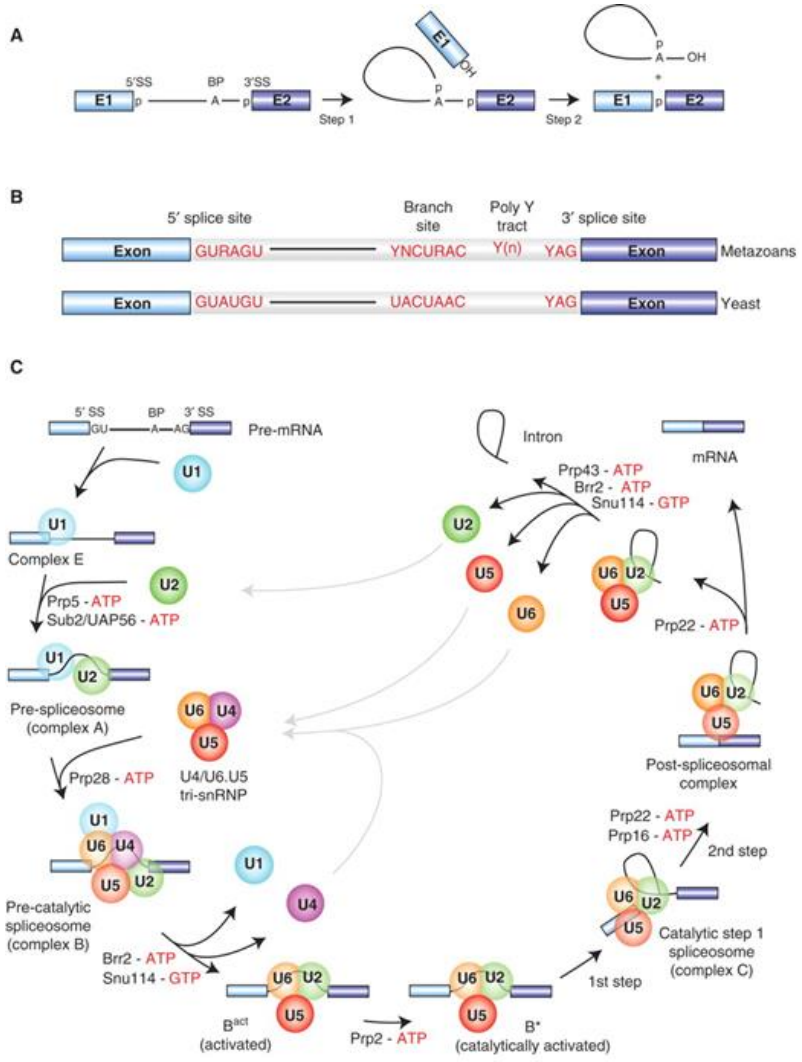

Figure 1: Splicing reactions and assembly: *A – splicing transesterification reactions leading to release of intron (line) and ligation of exon 1 and exon 2 (boxes E1 and E2 respectively,) B* –consensus sequences defining intron in yeast and metazoans respectively, C – Cycle of spliceosome assembly and function execution in regards to its snRNA content *(Will & Luhrmann, 2011, edited).*

3.2 Spliceosome

Spliceosome is a very dynamic complex. Its components are recruited to fulfil their function and then dissociate (See Fig. 1C). Therefore in different stages of spliceosome assembly we can observe very different protein and snRNA content (See Fig. 2 and Fig. 1) (*Fabrizio et. al., 2009)*. Protein contain of spliceosome is also quite different in budding yeast compared to human where the budding yeast spliceosome is less complex. In budding yeast B complex spliceosome there are about 60 detectable proteins, whereas in human B complex this number grows to about 110 proteins. Number of proteins in yeast C complex is even lower (approximately 50); however, in humans the protein content remains the same. The changes in spliceosome proteomes are massive even in the yeast itself where in transition between B and B^{act} complex more than half (*Aganofov et al., 2016*) of the 60 protein present in B complex is dissociated and not detectable in B^{act}. (*Fabrizio et al., 2009*) At whole, there are around 100 proteins that are part of spliceosome of yeast at some point of the splicing cycle. Vast majority of these proteins is highly conserved and its human counterpart can be identified (*Fabrizio et al., 2009; Herold et al., 2009).*

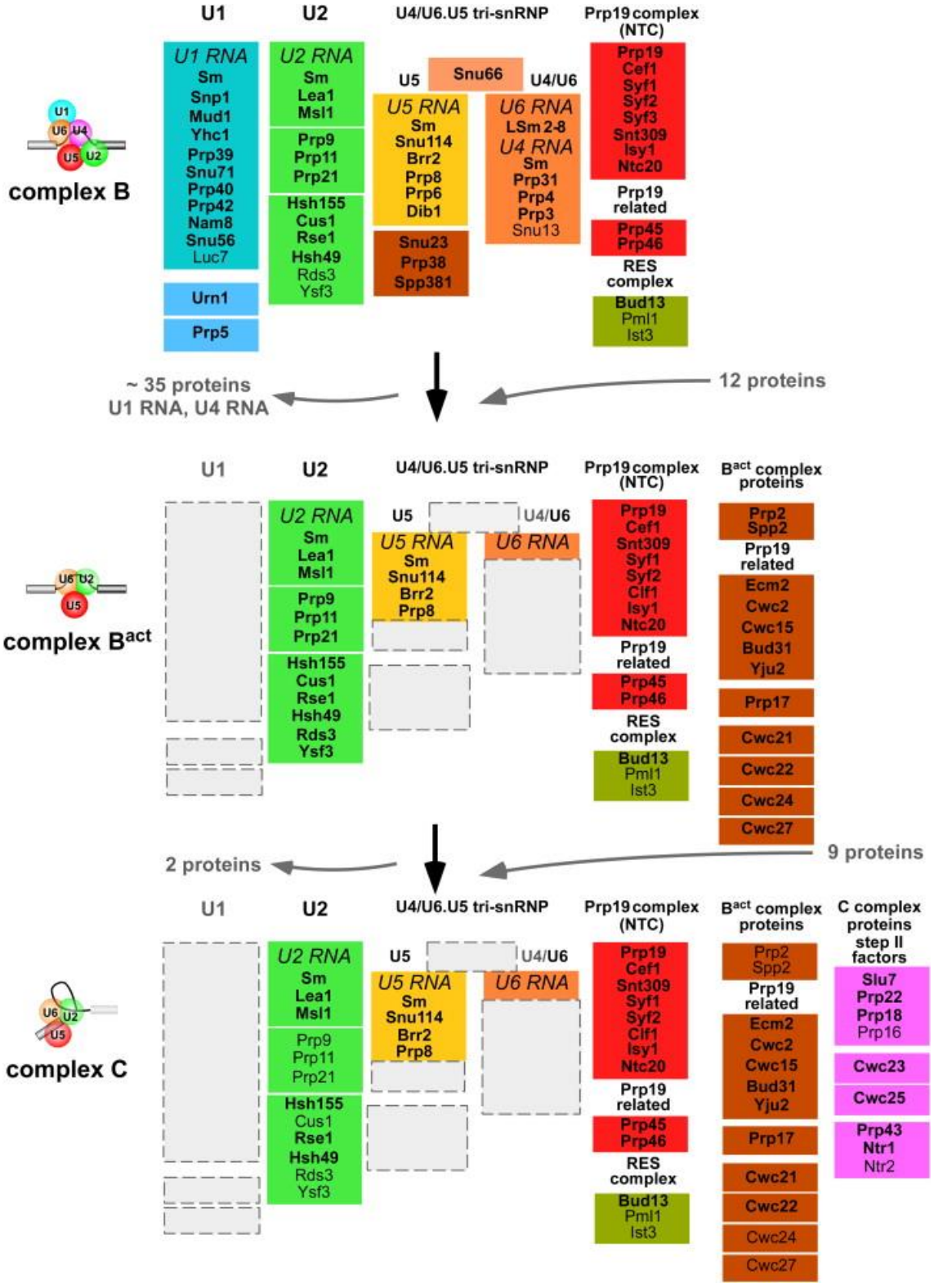

Figure 2: Comparison of protein content and dynamics of the complex in different stages of spliceosome assembly in *Saccharomyces cerevisiae (Fabrizio et al., 2009)*.

3.3 Spliceosome structure

While the mechanism of splicing is relatively well known there has been very little information about spliceosome structure and interactions. However, recent studies brought new light on the subject. A high number of proteins involved in splicing and all the massive rearrangements that occur during assembly made it difficult to look at spliceosome as a whole. By stalling the spliceosome at a certain point of its life cycle it was possible for the authors to look at the protein content of the individual stages of splicing and open door for solving the structure of the whole massive ribonucleoprotein (See Fig. 2) (*Fabrizio et al., 2009).*

First structure studies showed a rough shape of the spliceosome by using electron microscopy (*Fabrizio et al., 2009*). The overall structure of spliceosome can be viewed in Fig. 3. We can see that the B complex forms triangular shape approximately 30 Å in lengths and 7 to 10 Å in width where domains of foot and stump can be distinguished. Quite a dramatic change of this structure can be observed in activated B complex (B^{act}), where the main body of the spliceosome is more asymmetrical and compact than it was in B* complex. However, foot domain can still be observed. Yet another change of appearance of the spliceosome is visible in C complex in which the main body is still present but appears to be again less compact than in B^{act} complex. These differences in structure correlate with changes in spliceosome composition during splicing cycle (See Fig. 2, Fig. 3) (*Fabrizio et al., 2009*). More recently structures of spliceosome of *Schizosachharomyces pombe* as well as structure of the tri-snRNP were solved (*Yan et al., 2015; Wan et al., 2016).* The *S. pombe* spliceosome structure (*Yan et al., 2015)* presented more detailed view on C complex of *S. pombe* by solving the structure with 3.6 Ångstrom resolution and molecular mass of 1.3 megadaltons. Using cryo-electron microscopy they managed to obtain 3D structure of the spliceosome at resolution of 2.9 to 3.6 Å in the core region (7 to 8 Å in the periphery). Interestingly, while analysing the purified spliceosomal complex about 80 proteins were detected with approximately 50 of them in high abundance. However, these proteins are not exclusively protein that are part of C complex but are rather shared over several stages of spliceosome assembly (specifically B^{act}, B*, C, P and ILS). This means that the gained structure is not solely of C complex but rather an average of numerous stages of the spliceosome.

To assign correct components to the electron density map authors combined both homologous modelling and *de novo* atomic model building which yielded in determination of at least partial structure of total 28 proteins. They distinguish several domains: head and two arms that are connected to triangular central body (Fig 4, Fig 5) (*Yan et al., 2015*). All structures used in this work are briefly summarized in Table 1.

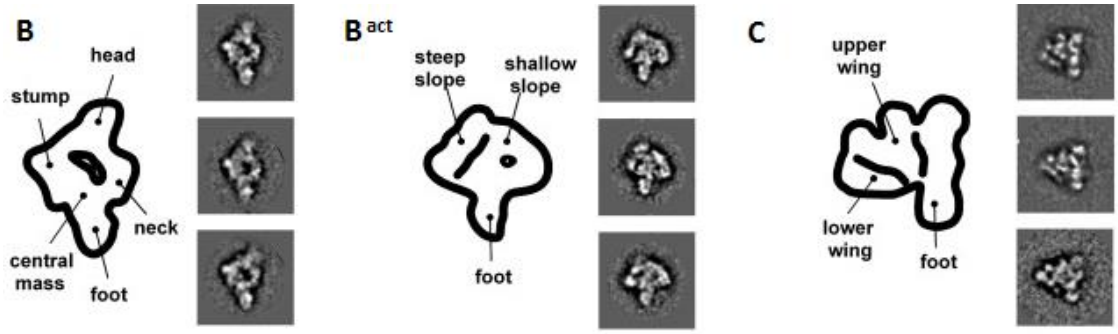

Figure 3: Overall shape of the spliceosome created by electron microscopy at different stages of spliceosome assembly. Several domains are visible as well as major rearrangements that occur between various stages of spliceosome assembly (*Fabrizio et al., 2009,* edited*).*

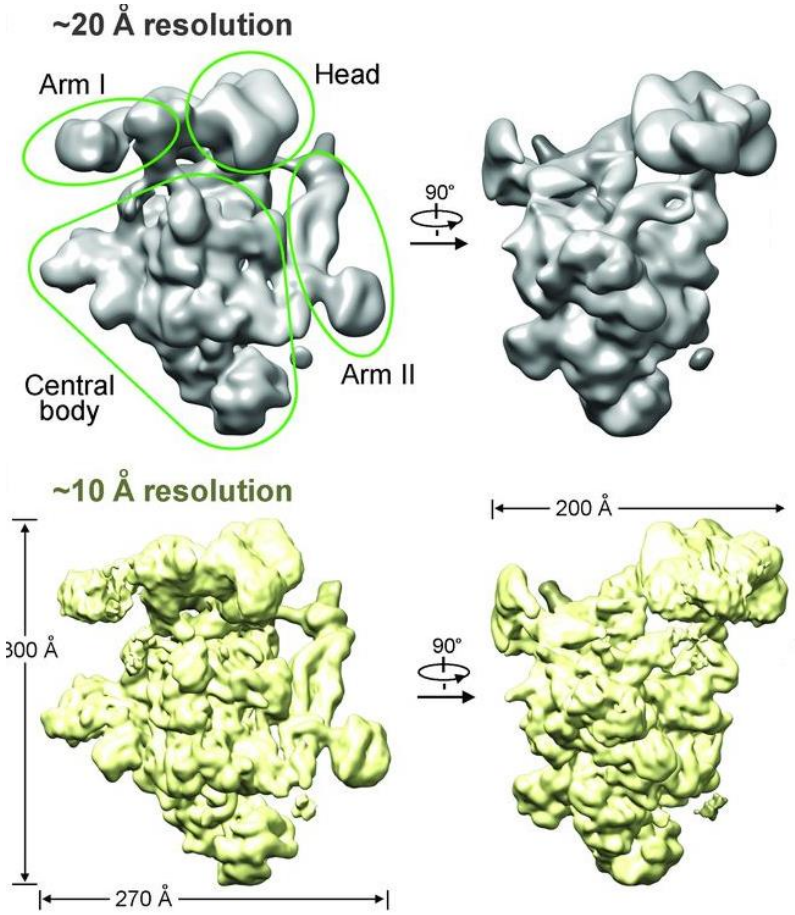

Figure 4: Structure of spliceosome with domains highlighted in 20 Å and 10 Å resolutions (*Yan et al., 2015,* edited).

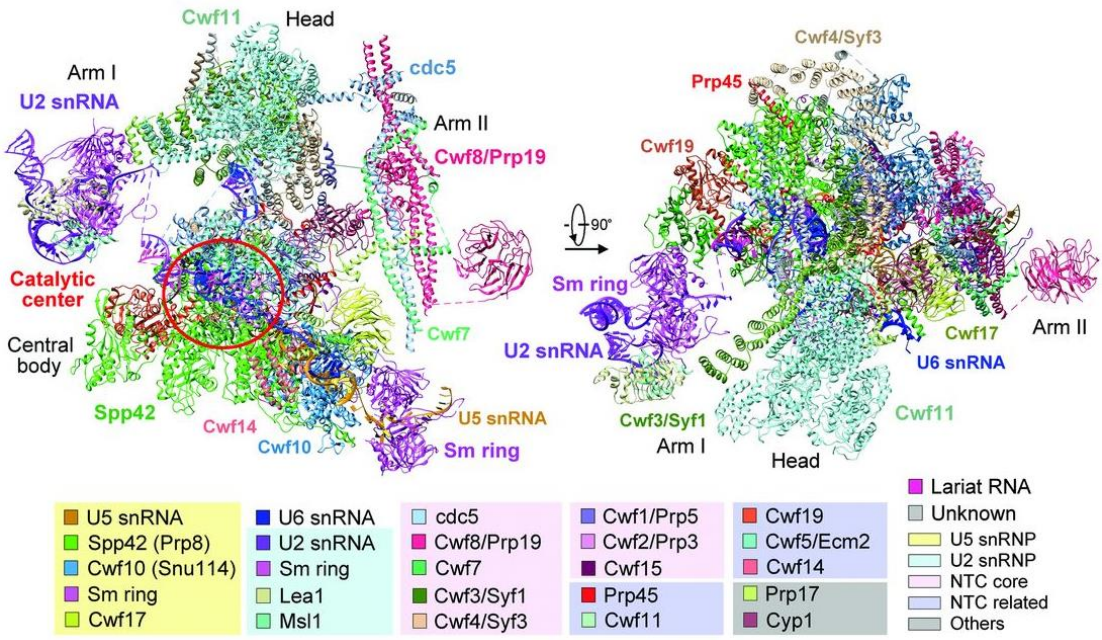

Figure 5: Protein content of structure of *S. pombe* spliceosome: With the catalytic centre in the red circle, snRNAs and proteins that are in the structure are visible here. Prp17 is missing from the list (*Yan et al., 2015*).

Table 1: Structures used in this work defined by their Protein Data Bank code and brief characterization of the structure and organism from which the structure derived.

Pdb code	Structure	Organism	Citation
3JB9	spliceosome	*S. pombe*	*Yan et al., 2015*
3JCM	U4/U6.U5 tri-snRNP	*S. cerevisiae*	*Wan et al., 2016*
3JCR	U4/U6.U5 tri-snRNP	*H. sapiens*	*Agafonov et. al,2016*
5GAM/5GAN/5GAO/5GAP	U4/U6.U5 tri-snRNP	*S. cerevisiae*	*Nguyen et al., 2016*
1XWN	PPIL1 and Prp45	*H. sapiens*	*Xu et al., 2006*
4YH8	U2AF	*S. pombe*	*Yoshida et al., 2015*
2OG4/3SBT	PRP8	*S. cerevisiae*	*Galej et al., 2013*
3I4U	Prp22	*H. sapiens*	*Kudlinzki et al., 2012*

4. Prp45

4.1 Function in *Saccharomyces cerevisiae*

Prp45 is an orthologue of *Homo sapiens* SNW1/SKIP protein in *Saccharomyces cerevisiae*. Its human orthologue was first identifies as a binding partner of oncogene Ski from where the name SKIP origins (Ski interacting protein) and its localization into nucleus was also established (*Dahl et al., 1998*). It was also proven to be essential in *Schizosaccharomyces pombe* (*Ambrozkova et al., 2001*). Apart of splicing function SKIP proteins carry many functions in cell and their list of interaction partners is rather diverse. SKIP interactions include proteins that are part of transcription initiation (such as Notch pathway) or repression and also cell cycle processes (rewieved by *Folk et al., 2004*). However, Prp45 in budding yeast is so far only linked to splicing (*Gahura et al., 2009*). The N-terminus was established as necessary for correct function of Prp45 because Prp45 C-terminus deletion mutant *prp45(1-190)* is able to act like wild type strain (*Martinkova et al., 2002*). While deletion mutant closer to N-terminus (*prp(119-379))* does not support grow (*Gahura et al.,2009*). In another mutant *prp45*(1-169) staling of division and termosensitivity occurred along with hypertensive phenotype for microtubule polymerization inhibitors (nicodazole, carbendazim) (*Gahura et al., 2009*).

In splicing Prp45 is part of NTR complex (nineteen complex related) and it was shown to be essential for splicing process. NTR complex recruitment happens probably at some point before first splicing reaction. After it recruitment, Prp45 is associated with spliceosome throughout the splicing reactions and it is later associated with the intron after splicing (Albers *et al., 2003*; *Fabrizio et al., 2009*). Prp45 can be already detected in B complex of spliceosome. It was showed that it does not interact directly with a region close to 3´splice site of pre-mRNA in B^{act} spliceosome (*Fabrizio et al., 2009*). Crosslink with 3´splice site region of pre-mRNA can be observed on low levels in B* complex. Prp45 was showed to interact with branch site nucleotides in first catalytic step of the splicing. In C complex crosslinking is also observable and even is about 80 % stronger than in B* complex. This finding suggests that after 1^{st} catalytic reaction Prp45 is in contact with 3´ss end of intron (*Schneider et al., 2015*; *Ajuh et al., 2000).*

4.2 Structure of Prp45

In *Saccharomyces cerevisiae* Prp45 consists of 379 amino acids. Interestingly, Prp45 protein in budding yeast is slightly different than its orthologues in other organisms because the otherwise conserved G-rich box on the N-terminus as well as polyproline track is missing from the *S. cerevisiae* Prp45. However, other known parts of this protein family such as SH2-like domain and helical segment and SNWKN sequence that gave this protein family its name, can be found in budding yeast´s orthologue as well (See Fig. 6) (*Ambrozkova et al., 2001)*.

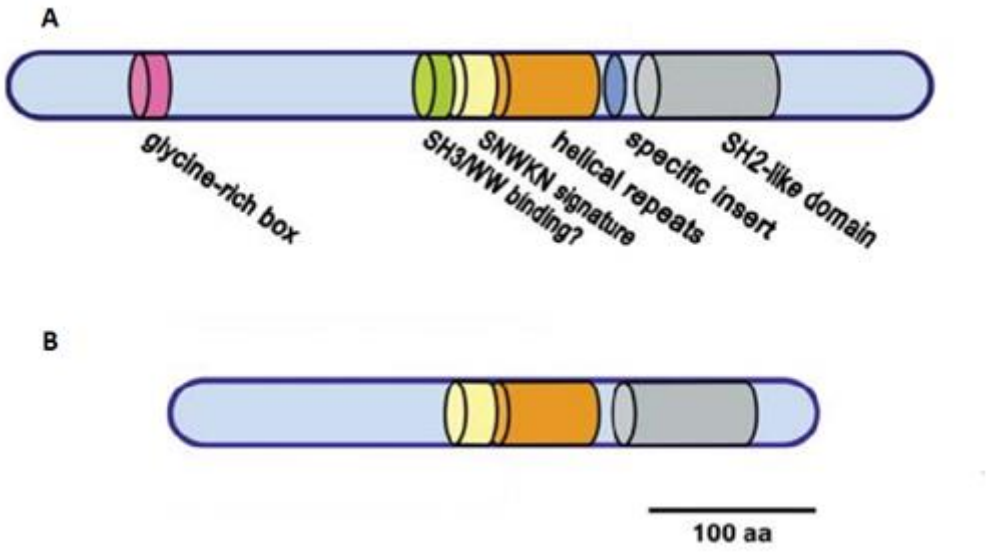

Figure 6: A) Domains in canonical SNW/SKIP protein, this domain consensus is present in *S. pombe* or *H. sapiens.* B) Domain composition of Prp45 in *S. cerevisiae* (*Folk et al., 2004, edited).*

The known structure of Prp45 is incomplete and from 547 amino acids present in the protein in *Schizosaccharomyces pombe* only positions 100-271 and 281-315 are present (they correspond to residues 25-184 and 194-224 in *Saccharomyces cerevisiae)*. Recent study shows that Prp45 is highly distorted and in fact stretches at length of over 150 Å (See Fig. 7). Besides the disordered regions two longer α-helices and one very short (only three residues long) and also two short 3_{10} helices were determined in Prp45 structure. In the middle section of the protein two short β-strands are visible. Prp45 interacts with U2 and U6 snRNAs in catalytic centre of the spliceosome. It is clear from the visualization of the catalytic centre that Prp45 is right in the centre of the spliceosome (See Fig. 5). It is likely that Prp45 helps assembly of spliceosome. This

might even permit connection between catalytic centre and periphery of the spliceosome and help to promote dynamics throughout the spliceosome (*Yan et al., 2015).*

SNW proteins were predicted to be phosphoproteins. It was predicted that the phosphorylation sites on SNW proteins were in a α-helix (reviewed in *Folk et al., 2004*). In the *S. pombe* Prp45 structure (*Yan et al., 2015*) two out of tree phosphoserines present in the protein are determined (residues 228 and 236) and neither of them is in predicted α-helical conformation, but rather in a distorted part of the protein.

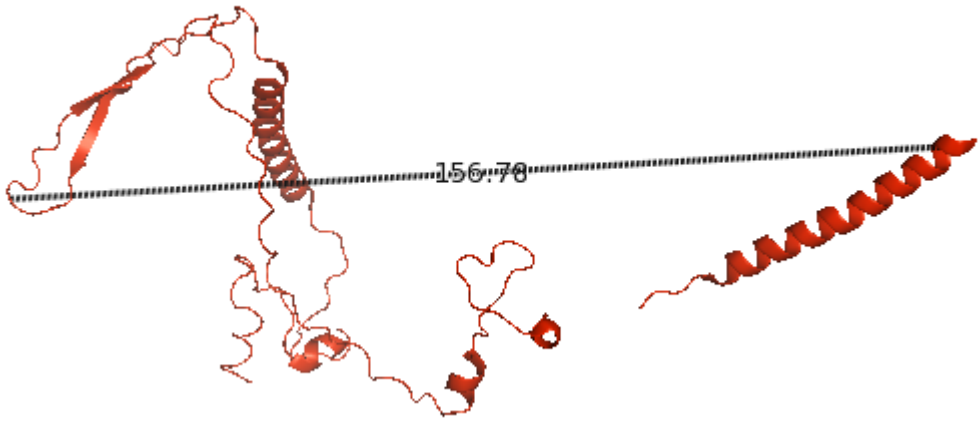

Figure 7: Visualization of Prp45 from the overall spliceosome structure. Created in PyMOL with data from 3JB9 (*Yan et al.,* 2015).

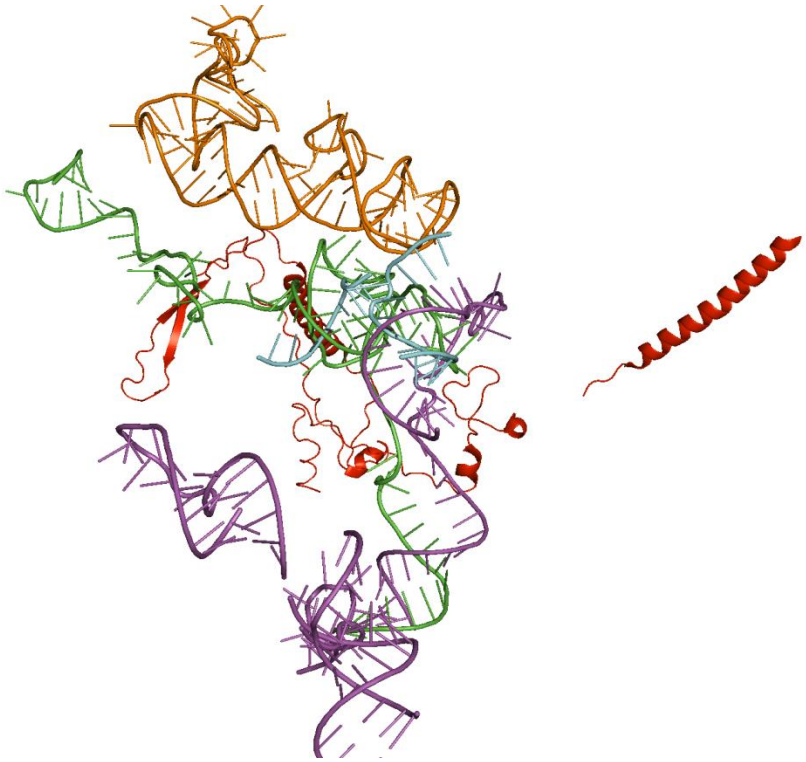

Figure 8: Prp45 displayed with RNA network. Prp45 is coloured in red, U5 snRNA in orange, U6 snRNA in green, U2 snRNA in purple and intron in light blue. Created in PyMOL with data from 3JB9 (Yan et al., 2015).

5. Interaction partners of Prp45

To determine all possible Prp45 protein interaction partners we used BioGRID database (*Chatr-aryamontri et al., 2015)* and found proteins that interact with Prp45 in either *Saccharomyces cerevisiae* or *Schizosaccahromyces pombe*. We then searched which of these proteins are involved in splicing. To our final cut we only involved interactions proposed by two-hybrid or pull down experiments. Since larger experiments (such as affinity capture or immunoprecipitation) in massive complex like spliceosome do not necessarily indicate physical interaction but rather presence in the same complex. Prp45 has several binding partners proven by two-hybrid studies that take part in different steps of splicing (*Amrozkova et al., 2001; Skruzny et al., 2001; Albers et al., 2003)*. With some of its interaction partners Prp45 also has a proven genetic interaction (Prp22, Syf1, Syf3, Cef1) (*Albers et al., 2003*; *Gahura et al., 2009*). We focused on these proposed interaction in the first part of the chapter. As you can see in Table 2 while a lot of proteins were at least partially determined in the overall structure some of the known interaction partners of Prp45 are not available (namely U2AF and Prp22) (*Yan et al., 2015*). The other part focuses on the recent structural studies of spliceosome that allow us to take better look at interaction network between spliceosomal proteins and find new physical interaction. Because of the recent work of *Yan et al.* we can for the first time visualize Prp45 and its position in regards of spliceome´s RNAs (See Fig 8.) and also its protein interaction partners in probable C state of spliceosome. We tried to determine which proteins have interaction with Prp45 using the new structure (*Yan et al., 2015).* Results of this search are summarized in Table 3.

In the oncoming chapters structures from both *Saccharomyces cerevisiae* and *Schizosaccaromyces pombe* were used. If not stated otherwise in effort to increase clarity of the text, nomenclature used in for *Schizosaccharomyces pombe* will be used primarily with corresponding equivalent for *Saccharomyces cerevisiae* in brackets in the title of the chapter.

Table 2: Proposed interaction partners of Prp45 list made using BioGRID database (*Chatr-aryamontri et al., 2015*) with known function in the splicing and their avalibility in the overall structure of spliceosome (*Yan et al., 2015)* and presence of physical interaction in said structure.

Protein name	Availible in structure	Interaction in 3D
Clf1	Yes	Yes
Prp22	Yes	Yes
Prp46	Yes	Yes
Cwc3	Yes	Yes
Cyp2	Yes	No
U2AF[23]	No	No

Table 3: Summarized interaction of Prp45 in the new spliceosome structure. If the interaction was observed its localization on both Prp45 and the interacting partner of the spliceosome is added. Newly identified interaction partners are highlighted in light grey (According to *Yan et al., 2015*)

Protein	Interaction in 3D	Proposed in literature	Localization in Prp45	Localization in the protein
Cdc5	Yes	No	SNW domain	Myb-like domain
Cwc3	No	No		
Cwf11	No	No		
Cwf14	No	No		
Cwf15	No	No		
Cwf17	Yes	No	SNW domain	C-terminal WD repeat
Cwf19	No	No		
Cwf2	No	No		
Cwf4	Yes	Yes	SNW domain, disordered areas	HAT3, HAT4
Cwf5	Yes	No	SNW domain	α-helices
Cwf7	No	No		
Lea1	No	No		
Msl1	No	No		
Ppi1	No	Yes		
Prp17	Yes	No	SNW domain	WD repeat 2
Prp19	No	No		
Prp5	Yes	Yes	SNW domain, disordered areas	WD repeat 4, WD repeat 5
Sm B1	Yes	No	SNW domain	coil
Sm D2		No		
Sm D3	No	No		
Sme1	No	No		
Smf1	No	No		
Smg1	No	No		
Spp42	Yes	No	SNW domain, disordered areas	Proline-rich, PROCRN domain,

5.1 Proteins with previously identified interaction with Prp45

For this section we examined proteins that were suggested as possible interaction partners of Prp45 by experiments involving two-hybrid or pull down experiments. We only chose proteins that are directly involved in splicing. We used the BioGRID database (*Chatr-aryamontri et al., 2015)* to determine proteins that meet these criteria in either *Saccharomyces cerevisiae* or *Schizosaccharomyces pombe.*

5.1.1 Prp5 (Prp46)

Prp5 is a splicing factor and part of NTR complex that is essential for splicing. Its interaction with Prp45 was demonstrated several times through two-hybrid studies (*Ajuh et al., 2000; Albers et al., 2003).* It was proven to interact with Prp45 *in vitro* and *in vivo* and was showed to crosslink with 3´splice site region of pre-mRNA from B^{act} complex to C complex. The observed crosslink in C complex was about 20 % stronger than in either B* or B_{act}. This suggests that Prp5 is in contact with 3´splice site region of pre-mRNA region throughout remodelling and activation of spliceosome. Contact site of Prp5 is immediately downstream of Prp45 contact site (*Ajuh et al., 2000*).

Prp5 is present in the overall structure of spliceosome. Pfam detected 6 WD repeats (tryptophan-aspartic acid repeats) in Prp5 in the structure. Secondary structure of Prp5 consists mostly of β-sheets with 4 antiparallel β-sheets per one WD repeat. The N-terminus consists of three α-helices. However, part of the N-terminus of the actual protein is missing from the structure (specifically residues 1-148) (*Yan et al., 2015*).

With both Prp45 and Prp5 present in the structure (*Yan et al., 2015*) we were able describe their interactions in detail. Using PyMOL, we established that there are at least two areas in which Prp45 and Prp5 make contact. First possible interaction was determined between Lys331 of Prp5 and Ser236 of Prp45. Ser236 was in the disordered area of Prp45 and Lys331 was located a part of fifth WD repeat. Second interaction that we were able to identify is located between Glu299 on Prp5 and Lys136 in Prp45. Lys146 of Prp45 was located in the first α-helix of the structure and Glu299 was in a loop between second and third β-sheet of the fourth WD repeat. Third interaction was

not far from the second one; however, it was possibly a weak interaction. It is between Val282 of Prp5 and Gln164 of Prp45, which are in the first β-strand of the fourth WD repeat and in on the C-terminal part of the first α- helix respectively (See Figure 9). These results are consistent with previous two-hybrid assay that suggested that binding region of Prp45 is in WD repeats motifs (*Albers et al., 2003*). Since large part of the N-terminus of both of these proteins is missing, there is still a possibility that other interactions between Prp45 and Prp5 exist.

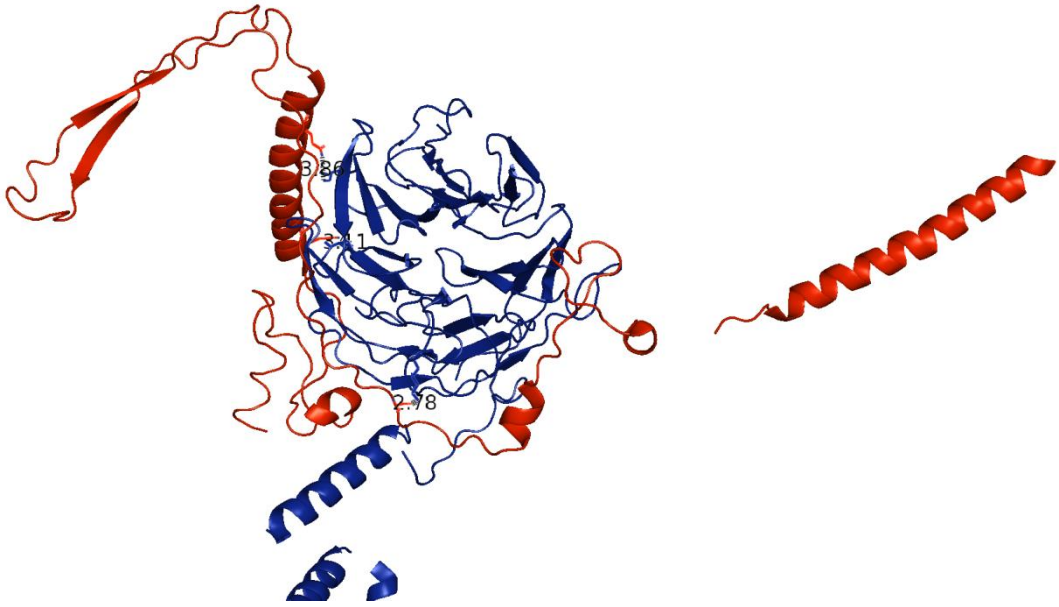

Figure 9: Interactions between Prp45 (red) and Prp5 (blue). The numbers show distance between the two proteins at the site of the interaction. Created in PyMOL with data from 3JB9 (*Yan et al., 2015*).

5.1.2 U2AF

U2AF is a heterodimeric splicing factor composed of two subunits ($U2AF^{59}$ and $U2AF^{23}$ in *S. pombe*) and is required for U2 snRNP recruitment to the spliceosome. The large subunit specifically binds to polypyrimidine track of the 3´splice site region and therefore helps to recognize binding site of U2 snRNA. U2AF65 stays in the spliceosome even after U2 snRNA binding (*Zamore &Green, 1989*). The large subunit binds around 88 % of 3´ splice sites in HeLa cells the other 12 % of spliced pre-mRNAs might not be dependent on U2AF. The U2AF also plays a role in regulation of alternative splicing where the large subunit binds into intron regions and therefore makes 3´ splice site unavailable (*Shao et al., 2014*). Interaction between *Schizosaccharomyces pombe* orthologue of Prp45 SNW1 and small subunit of U2AF was established by two-hybrid study of *S. pombe* cDNA library. Strong interaction was

observed on the C-terminus of SNW1. This interaction is stronger in deletion of U2AF (75-216) which suggests that the N-terminal region of U2AF possess inhibitory function (*Ambrozkova et al., 2001*). It was proven that $U2AF^{23}$ is highly conserved throughout evolution because $U2AF^{23}$ from fission yeast can be functionally replaced by its counterpart from human. Since the only region that differs between the two species is C-terminus, it was suggested that C-terminus is probably not needed for function of $U2AF^{23}$ (*Webb & Wise, 2004*).

Through homology searches, two Zinc binding domains (ZBD) were found in $U2AF^{23}$ (*Worthington et al. 1996*). Both of these ZBDs of $U2AF^{23}$ were mutated in the residues that coordinate Zn^{2+} ions. This mutation led to either lethality or growth defects proving that these motifs are both required for correct function. However, mutating one ZBD domain did not have same effect as mutating the other on the same residues suggesting that ZBD domains are not equal. Mutation of these motifs also lowers the RNA binding ability of the subunit suggesting that they contribute in to RNA binding of the U2AF (*Webb & Wise., 2004*).

$U2AF^{23}$ has a non-standard RNA recognition motif (RRM) termed ΨRRM (*Birney et al., 1993*). ΨRRM motif is essential for function of $U2AF^{23}$. When ΨRRM motif is not present correct contacts between the two subunits are compromised. Due to presence of aromatic resides RRM motifs can bind RNA (*Webb & Wise, 2004*). It was showed that small subunit of U2AF is *S. cerevisiae* is able to bind DNA on its own (*Yoshida et al., 2015).*

Available structure of core part of U2AF in *H. sapiens* revealed rather large interface between the two subunits of U2AF that resolves around tryptophans from both proteins (*Kielkopf et al., 2001).* More recently solved structure of U2AF heterodimer is *S. pombe* showed that $U2AF^{23}$ has RRM domain with ZBD on its side. Interaction between the two ZBD is mediated by hydrophobic interactions (*Yoshida et al., 2015).* U2AF is not present in the overall spliceosome structure (*Yan et al., 2015)*; therefore, we were not able to directly observe interaction between $U2AF^{23}$ and Prp45.

5.1.3 Cyp2

Interaction between cyclophilins and Prp45 was first discovered in *Dictyostelium discoideum* and in *Schizzosaccharomyces pombe* by a two-hybrid experiment. This

interaction was then mapped showing that the 180 N-terminal amino acids are responsible for the interaction (*Skruzny et al, 2001*).

Structure of cyclophilin along with *H. sapiens* SKIP protein was solved by NMR showing the protein interaction between the two (*Xu et al, 2006*). PPIL1 consists of eight β-sheets folded in anti-parallel order. The β-sheets are framed by two amphipathic α-helices. This secondary structure is typical for cyclophilins were described prior to this experiment (*Ottiger et al., 1997*). The last ordered part of the structure is a short 3_{10} helix that connects β-sheets β6 and β7. As the next step, authors determined details of the interactions between PPIL1 and SKIP protein. Though a pull-down experiment they discovered that of N-terminal amino acids 59-129 of SKIP are able to bind PPIL1 with sufficient strength (*Xu et al, 2006*). This result is narrowing down the previous results (*Skurzny et al, 2001).* It was suggested that after binding SKIP PPIL1 forms intramolecular bond with residues Asp-89 and Lys-91 that are located prior β5 sheet also Ile-128 and Tyr-28 probably bind to SKIP. However, the exact atom of SKIP that forms this bound is unknown. After binding of SKIP intermolecular bonds between β1, β2, β5 and β7 are weakened. This results in changes in the secondary structure of the protein. From the structure it is apparent that SKIP does not occupy PPIL1s active site or the site with which PPIL1 binds U4/U6 snRNAs (*Skruzny et al, 2001, Xu et al, 2006*).

Structure of Cyp2 is present in the structure of *Schizosaccharomyces pombe´s* spliceosome (*Yan et al., 2015*) but we were unable to find any interaction between Cyp2 and Prp45. However, the lack of the interaction is most likely caused by the fact that N-terminus of Prp45 is missing.

5.1.4 Prp22

Prp22 is a conserved DExD/H-box RNA helicase from a family of RNA-dependent ATPases that plays role in exon ligation repression of intermediates with aberrant branch site or splice site by modulating structure of the spliceosome and release of spliced mature mRNA (*Company et al, 1991*). Prp22 functions before the second catalytic step. The protein directly binds to intron in the region downstream from branch point. After the second transesterification reaction the position of Prp22 changes to the exon-exon junction region of mRNA. Prp22 also has a helicase activity which is then required for mRNA release (*Company et al., 1991*). Prp22 was not detected in

early stages of the spliceosome and becomes detectable in C complex (*Fabrizio et al, 2009*). The interaction of Prp22 and Prp45 was suggested through two-hybrid assay (*Albers et al., 2003)* and it was placed on C-terminal part of both proteins. In deletion mutant strain of *Saccharomyces cerevisiae prp45*(1-169) stoichiometry of Prp22 in spliceosome decreased. Moreover, Prp22 overexpression rescues phenotype caused by the *prp45*(1-169) mutation. It was proposed that Prp45 might be a part of recruitment and regulation of Prp22 (*Gahura et al., 2009*).

Prp22 is not part of the overall structure of spliceosome (*Yan et al., 2015)* and we were therefore unable to search for interaction in the structure. However, C-terminal part of Prp22 for *Homo Sapiens* was solved (*Kudlinzki et al., 2012).* This study revealed that the C-terminal part made of five β-strand that form two β-sheets and ten α-helices four of which form plane structure. The rest of the helices are placed on both sides of the plane (*Kudlinzki et al., 2012).*

5.1.5 Clf1(Syf3/Cwf4)

Clf1 (crooked neck-like factor) is an essential protein and it was discovered as an orthologue of crn protein of *Drosophila melanogaster* (*Chung et al, 1999*). Clf1 was also independently identified and named Syf3 when studying lethal mutants in cells with deletion of Prp17 gene (*Ben-Yehuda, 2000)*). Clf1 was proven to be highly conserved throughout evolution with orthologues among many other species (*Ben-Yehuda et al, 2000*). After depletion of Clf1 cells were stalled in G2/M phase of cell cycle with observable defects in splicing of pre-mRNA (*Russell et al, 2000*). This phenomenon was further studied by temperature inactivation of Clf1. This experiment showed that Clf1 deficiency causes improper recruitment and binding of tri-snRNP particle and Prp19 and therefore does not support splicing (*Chung et al, 1999, Wang et al, 2003*). In agreement with these results another study showed that Clf1 is associated with spliceosome during both catalytic reactions and might even interact with U6 and U5 snRNA (*Russell et al., 2000*).

Clf1 possess 13 tetratricopetide repeats (TPR repeats) that do not strictly correspond to classic consensus of TPR repeats (*Ben-Yehuda et al, 2000*), therefore they are termed half TPR (HAT). TPR repeat is all-helical motif made of hydrophobic residues. The only conserved residues of TPR repeats are Gly/Ala on eighth position and Ala in 20^{th} and 27^{th} positions. In structures multiple TPRs adopts helix-turn-helix secondary

structure where helices are stacked in antiparallel fashion. The motif then consists of 3 to 16 tandem-repeats of 34 residues. However, TPR does not have to be continuous but it can disperse throughout the sequence. TPRs are also known to mediate protein-protein interactions (*rewieved in D´Andrea & Regan, 2003*). Presence of these repeats in Clf1 may mean that Clf1 functions as a scaffold protein in the spliceosome (*Ben-Yehuda et al., 2000*).

It was suggested though bioinformatic analysis that Clf1 consists of two HAT superhelices and C-terminus with no HAT repeats (*Ben-Yehuda et al., 2000*). When the first two N-terminal TPR repeats of Clf1 were deleted, both snRNP and non-snRNP proteins that normally co-purify with Clf1 (Prp19, Cef1, Rse1, Hsh155) were more sensitive to dissociation. This suggests that TPR repeats are helping to stabilize their interactions (*Wang et al., 2003*).

With structure of Clf1 available in spliceosome structure we were able to determine two possible interaction of Clf1 with Prp45 (See Fig. 10). First interaction was between Tyr103 of Prp45 and Gln159. Tyr103 was on a coil at the N-terminus of determined part of Prp45, Gln159 was located is part of a α-helix. The other interaction was between His235 of Prp45 and Arg136 of Clf1. In this case His235 was a coil between β-strands and the next α-helix. Arg136 was located in a coil between two α-helices. Structurally these two interactions are in a close proximity (*according to Yan et al., 2015)*.

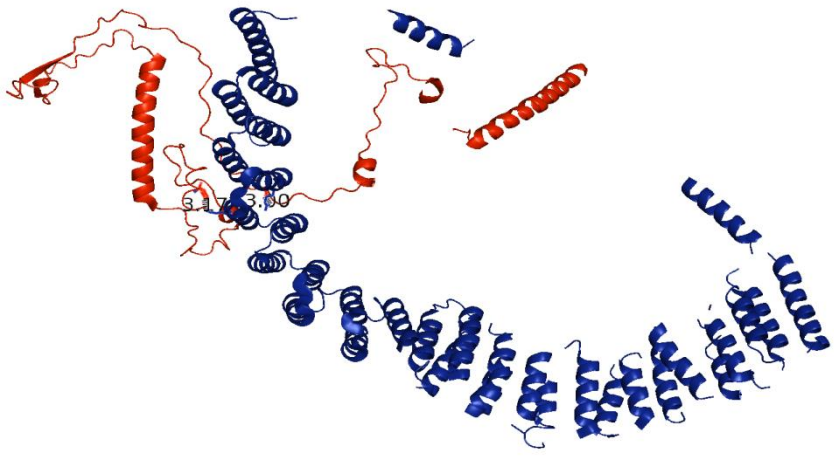

Figure 10: Interaction between Prp45 (red) and Cwf4 (blue). The numbers show distance between the two proteins at the site of the interaction. Created in PyMOL with data from 3JB9 (*Yan et al., 2015*).

5.1.6 Cwf3 (Syf1)

Cwf3 was first identified along with Clf1 in experiments that searched for genes that are lethal with mutation of Prp17. This lethality can be partially rescued by overproduction of U2 snRNA for both Cwf3 and Clf1 (*Yehuda et al., 2000*). Cwf3 is a component of spliceosome´s NTC complex and associates with spliceosome during both splicing reactions (*Russell et al., 2000*). However, several key components of NTC are able to bind to spliceosome without presence of Cwf3 (Prp19, Cef1, Clf1). Also Cwf3 in not needed for spliceosome activation via NTC (*Chang et al., 2009*). Cwf3 is required for recruitment of splicing factor Yju2 (*Chang et al., 2009*). Possibility of interaction between Cwf3 and Prp45 was established by two-hybrid experiments (*Albers et al., 2003; Yu et al., 2008*).

In *Saccharomyces cerevisiae*, Cwf3 consists of 859 amino acid residues (*Yehuda et al., 2000*). Using Prosite it was predicted that Cwf3 contains 9 TPR motifs in *Saccharomyces cerevisiae.* It was also suggested that TPR repeats TPR3 to TPR9 could form a superhelix (*Yehuda et al., 2000*). Similarly to Clf1 it is also thought that Cwf3 may function as a scaffold protein because of its TPR repeats. (*Yehuda et al., 2000*) Interaction study of Cwf3 deleted different TPR motifs from the protein. This showed that deletion of the region of predicted superhelix destroys most of Cwf3s interactions with other NTC proteins (Clf1, Cef1 and Ntc20). While deleting first two TPR motifs on the N-terminus seems to have little to no influence on interaction capabilities of Cwf3. It seems that the TPR superhelix is responsible for most of the protein function. Moreover, under normal conditions the domain is capable of executing Cwf3´s functions. TPR superhelix is also able to maintain most of Cwf3s interaction network with exception of Isy1 and Clf1 (*Chang et al., 2009*).

Structure of *Schizzosaccharomyces pombe´s* Syf1 is available in the overall structure of the spliceosome (*Yen et al., 2015*). However, we were unable to find any interaction with Prp45. The distance between Prp45 and Cwf3 was never shorter than 30 Å. This was probably due the fact that only small part of Cwf3 was determined (specifically residues 498-509, 513-581, 589-616 and 625-653 of total 653 residues) meaning that interaction between Syf1 and Prp45 was likely outside this region or in the unrepresented region of Prp45. This is consistent with previous results (*Albers et al., 2003)* that suggested that N-terminal region of Prp45 is responsible for the interaction.

5.2 Newly identified proteins with interaction with Prp45

In this chapter we took a look at proteins that are available in Cryo-EM structure of whole spliceosome including Prp45 (*Yan et al., 2015*) to examine if any previously unknown interactions between these proteins and Prp45 can be detected. In the available structure, 28 proteins and 4 nucleic acids (U2, U5, U6 snRNAs and intron) can be found. We took every one of these proteins and searched the Prp45/protein interface for possible interaction spots. It is worth noting that Prp45 sequence is incomplete in the structure only including residues 100-271 and 281-315 of 556 in *Schizosaccharomyces pombe* (these residues correspond to 25-184 and 194-224 in *Saccharomyces cerevisiae) (Yan et al., 2015).*

5.2.1 Cdc5 (Cef1)

Cdc5 is a part of NTC complex and its function in splicing is essential for splicing to occur (*Tsai et al., 1999)*. Sequences of Cdc5 are very well conserved throughout evolution. Sequence conservation is highest in the N-terminal and at it is lowest in the C-terminal third. The C-terminus is also the part of Cdc5 which binds several proteins including interaction partners of Prp45 (specifically Prp46 and Syf1) (*Ohi & Gould., 2002*). The N-terminus of Cdc5 is responsible for RNA binding and therefore Cdc5 is proposed to function as connection between NTC proteins and RNA (*Collier et al., 2014*).

Cdc5 sequence contains Myb (myeloblastosis) repeats and RRM domain (RNA recognition motive). Function of Myb repeats was not yet discovered. In the protein, there are two canonical Myb repeats and one unusual Myb-like repeat. The Myb-like repeat follows immediately after second Myb repeat (*Ohi et al., 2002*).

Myb repeats canonically contain three tryptophan residues interrupted by either 18 or 19 amino acids. Usually Myb repeats have a helical backbone consisting mainly of hydrophobic residues (*Ohi et al., 1998*). Myb repeat folds into typical structure of three α-helices (*Collier et al., 2014*).

Myb-like repeat of Cdc5 does not follow the consensus established before. Instead of tryptophan Myb-like repeat contains tyrosine. Other difference is that Myb-like repeat sequence does not seem to have hydrophobic residues that could form the helical

backbone (*Ohi et al., 1998*). This leads to Myb-like repeat's inability to fold in the three α-helices but instead fold into just two (*Collier et al., 2014*).

Part of Cdc5´s structure was recently solved (residues 8-236 of 757) (*Yan et al., 2015*). In this structure the first Myb repeat is present. With partial Cdc5´s structure available in the spliceosome structure we managed to find one direct interaction of Prp45 and Cdc5. This interaction was located between Thr265 of Prp45 and Lys23 of Cdc5 (See Fig. 11). Thr265 was part of a coil before the second 3_{10}-helix on Prp45. Lys23 was the last amino acid of the first α-helix of the protein. It is worth noting that only position 8-236 were determined in the structure while the whole length of Cdc5 in *Schizosaccharomyces pombe* is 757 amino acids. This leaves a possibility for other interactions of Cdc5 with Prp45.

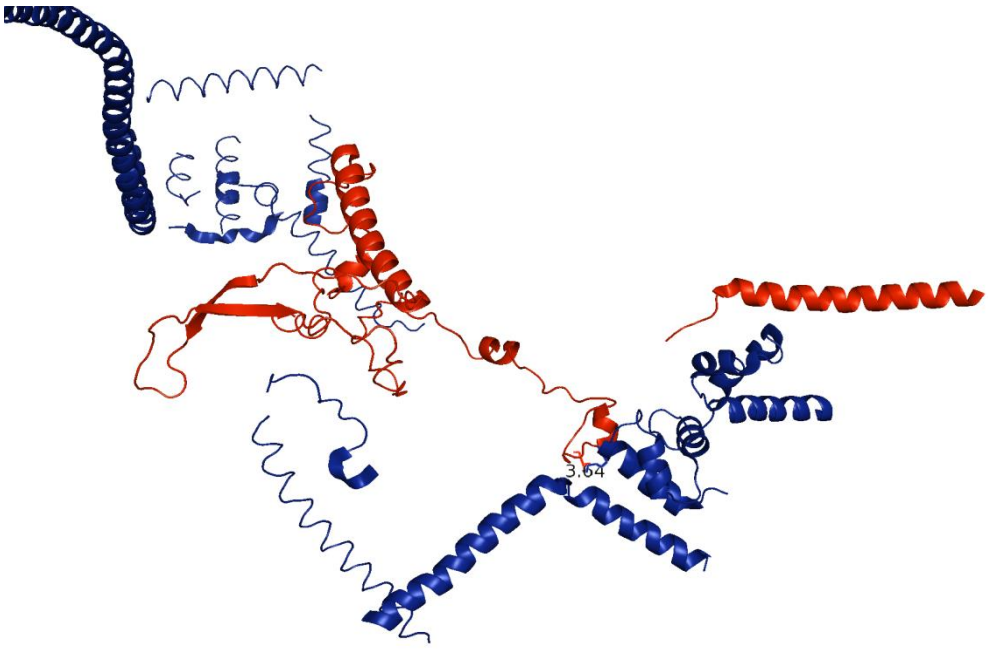

Figure 11: Interaction between Prp45 (red) and Cdc5 (blue). The numbers show distance between the two proteins at the site of the interaction. Created in PyMOL with data from 3JB9 (*Yan et al.,* 2015).

5.2.2 Prp17/Cdc40

In spliceosome Prp17 is part of the second transesterification. Prp17 binds to Prp16 and helps to remodel the spliceosome. However, Prp17 is a part of the spliceosome before both transesterifications and stays connected till post-catalytic P complex (*Sapra et al., 2008)*. Interestingly Prp17 is non-essencial for the second transesterification (*Vijayraghavan et al., 1989*) but its absence from spliceosome lowers the kinetics of

second splicing step (*Sapra et al., 2008*). Without Prp17 cells also show temperature sensitive phenotype (*Ben-Yehuda et al., 1998*).

By sequence analysis of Prp17, five to seven WD repeats were found in the protein. The repeats probably function as carrier of protein-protein interactions of Prp17. These interactions likely require C-terminus of the protein. However, the N-terminus of the protein is functionally important (*Ben-Yehuda et al., 1998*).

By looking at the spliceosome structure (*Yan et al., 2015*) we were able to find three probable interactions of Prp17 and Prp45 (See Fig. 12). First interaction that was visible in the structure is between Gln191 of Prp45 and Thr86 of Prp17. Thr86 was on a coil between fourth α-helix and the only β-sheet of Cdc40 and Gln191 was in a coil between the two β-strands of Prp45. Second interaction was found between residues Glu92 on Cdc40 and Arg185 on Prp45. Glu92 was the last amino acid of the β-sheet on Cdc40 and Arg185 was part of the first β-sheet on Prp45. Last interaction we found was between Lys76 of Cdc40 and Met192 of Prp45 where Lys76 was part of the third α-helix of Cdc40 and Met192 was between the two β-sheets of Prp45. It is clear form Fig.12 that interaction interface between the two proteins is established with two β-sheets one from each protein running in antiparallel matter.

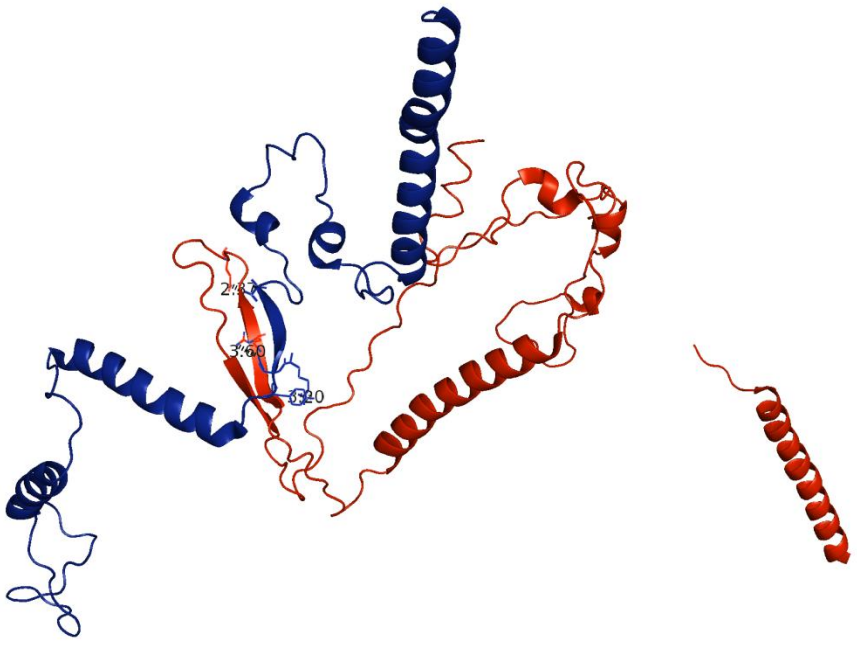

Figure 12: Interactions between Prp45 (red) and Prp17 (blue). The numbers show distance between the two proteins at the site of the interaction. Created in PyMOL with data from 3JB9 (*Yan et al.,* 2015)

5.2.3 Spp42 (Prp8)

Spp42 is U5 snRNP protein that was showed to crosslink with 3´splice site and recognizes the uridine tract (*Umen & Guthrie et al, 1995a*). I was also observed that Spp42 helps to recognize 3´splice site in second step of splicing. Spp42 binds to 3´ss after Prp16 and Prp17 binding (*Umen & Guthrie, 1995b).* Through biochemical study it was determined that Spp42 crosslinks with intron since B* complex with increase of the crosslink after the first transesterification. In the C complex crosslink with branch point was also observed (*Schneider et al., 2015).*Through a pull-down study a direct interaction of Spp42 and helicase Brr2 was established. It was also proven that Prp8 stimulates Brr2s helicase activity and therefore promotes unwinding of U4/U6 snRNA duplex (Maeder et al., 2009).

Spp42 orthologue in *S. cerevisiae* was previously partially crystalized (residues 885-2413) in complex with Aar2 (*Galej et al., 2013)*. This study observed new domain on Prp8 that consists of finger, palm and thumb regions. The finger/palm region of the protein folds into reverse transcriptase like fold. Whole domain is then directly connected via linker to endonuclease domain of the protein. The authors of the study found that vast majority of known crosslinks of Prp8 and snRNAs are located in the RT/En domain. Two more domains are located on the C-terminal part of the protein: RNaseH-like domain (*Pena et al., 2008)* and Jab1/MPN domain. RNaseH-like domain is located on the very N-terminus and is connected to Jab1/MPN only by a short linker (*Pena et al., 2008; Galej et al., 2013).* It was suggested that RNaseH-like domain might help with positioning of the RNAs (*Pena et al., 2008).*

Since Spp42 is present it the overall spliceosome structure we were able to determine at least three independent interaction regions between Spp42 and Prp45. Spp42 is in the very core of the spliceosome most of its sequence was assigned in the structure (residues 47-302, 314-1532, 1536-1780, 1784-2030 of total 2362 residues) (*Yan et al.,2015)*. The interaction interface of Prp45 and Spp42 is slightly more complex compared to other proteins because Prp45 to some degree copies the shape of Spp42.

First interaction that we found was on Ser172 on Prp45 and on Lys114 on Spp42. Ser172 was located in the coil between the first α-helix and the first β-sheet, whereas Lys114 was part of Spp42s α-helix. Second interaction was in close proximity of the first one and was located between Arg218 on Prp45 which was the last residue of

the firs α-helix and Glu686 which was part of a coil between two α-helices. Next two interactions were fairly close to each other. First one was between Lys728 (coil between α-helices) on Spp42 and Trp249 on Prp45 (first residue after the second α-helix). The other one was between Glu738 on Spp42 (α-helix) and Asn257 on Prp45 (coil after second α-helix). Lastly we found one interaction between Glu841 on Prp8 (α-helix) and Arg309 on Prp45 (second α-helix) (See Fig 13).

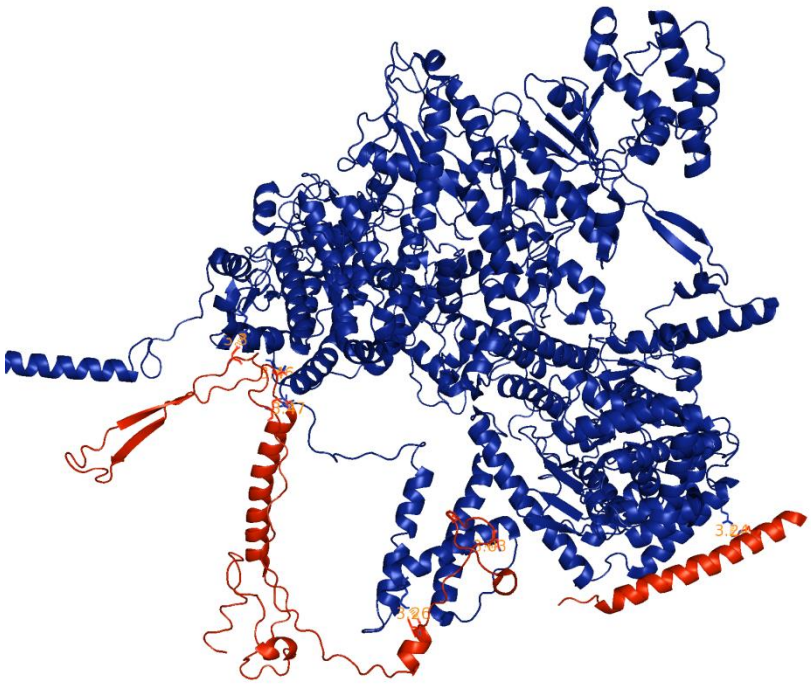

Figure 13: Interactions between Prp45 (red) and Prp8 (blue). The numbers show distance between the two proteins at the site of the interaction. Created in PyMOL with data from 3JB9 (*Yan et al., 2015*)

5.2.4 Smb1

Smb1 in *S. cerevisiae* was discovered as orthologue of *H. sapiens* Sm proteins (*Salgado-Garrido et al., 1999*). Sm proteins have two motifs called Sm1 and Sm2 that are separated by eight amino acid long linker. Sm1 motif is on C-terminal region of the protein and consists of 32 amino acid residues of which at least 16 are highly conserved. Sm2 is a shorter sequence closer to N-terminus. It consists of 14 amino acids with consensus IRGXNI sequence (*Hermann et al., 1995)*. The very C-terminal tails of Sm proteins have function in stabilizing of the spliceosome (*Zhang et al., 2001)*. In spliceosome, seven Sm proteins make one complex bound to snRNA. Sm proteins of

S. cerevisiae co-precipitate with all snRNAs with exception of SmB proteins, which do not co-precipitate with U1 snRNA (*Salgado-Garrido et al., 1999*).

Structures of different Sm or Lsm protein were crystalized before usually as a part of larger complex (*Nguyen et al., 2016, Wan et al., 2016*). Most of the residues of Smb1 were assigned in the overall structure of spliceosome (2-47, 60-86, 94-118 of total 146), revealing that it mostly consists of antiparallel β-sheets with exception of one α-helix on the N-terminus of the protein. The determined C-terminal part appeared to be disordered (*Yan et al., 2015)*. In this region we managed to find one possible weak interaction with Prp45 between Val115 on Smb1 and Arg185 on Prp45 (See Fig 14). Val115 was located on the coil. Arg185 was part of the first β-strand of Prp45.

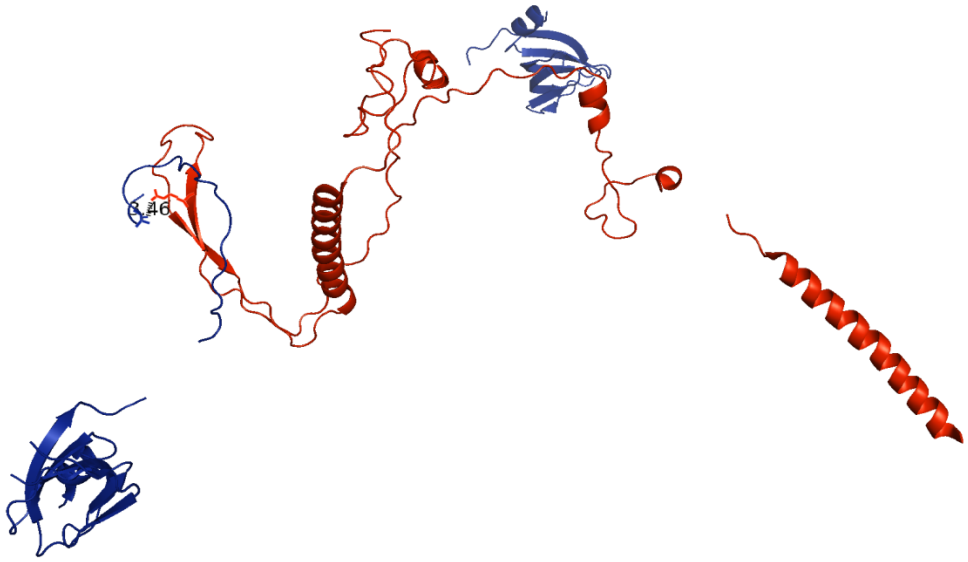

Figure 14: Interaction between Prp45 (red) and Smb1 (blue). The numbers show distance between the two proteins at the site of the interaction. Created in PyMOL with data from 3JB9 (*Yan et al., 2015).*

5.2.5 Cwf5 (Ecm2/Slt11)

Cwf5 in *Saccharomyces cerevisiae* is nonessential 364 amino acids long protein. It was first discovered as a splicing-related protein that functions before the first transesterification for which Cwf5 is required (*Xu et al., 1998)*. From the sequence three segments are detectable. First 150 residues contain two zinc fingers, RNA binding domain and lysine rich region on C-terminus (*Xu et al., 1998)*. In *Schizosaccharomyces pombe's* Cwf5 the lysine rich area is not present (*Xu et al., 1998; Xu* & Friesen*., 2001*). Mutating Cwf5 together with either mutation in U5 snRNA loop 1 or 11-nt substitution on U2snRNA loop 1 is synthetically lethal (*Xu et al., 1998*). In

mutation studies it was suggested that Cwf5 is involved in pairing of U2/U6 snRNA´s in helix II region. (*Xu &* Friesen, *2001*)

The new spliceosome structure (*Yan et al., 2015*) gave us a first insight into how Cwf5´s structure looks like. Most of the protein structure of Cwf5 was determined (residues 18-151, 178-184, 224-272 and 279-285 of total 285 residues). The structure revealed that this part of the protein is globular and can be divided into two parts. First of which is made out of four β-strands folded in antiparallel matter enclosed by 5 α-helices. Another part of the protein consists of two antiparallel β-strands with three α-helices and one α-helix that link these two parts together (See Fig 15).

In this protein we found two possible areas of interactions between Cwf5 and Prp45. In these areas we found three interactions (See Fig. 15). First interaction was between Lys198 on Prp45 which was located in coil between the two β-strands and Glu24 in the first α-helix of Cwf5. The second one was located on Arg200 on Prp45 and on Gly28 on Cwf5. Arg200 was in the coil between the two β-strands and Gly28 was the first residues after the first α-helix. The first two interactions were remarkably close with only two amino-acid residues apart in Prp45 and four residues apart in Cwf5. Last found interaction were between residues Glu114 on Cwf5 (α-helix) and Lys220 which was located in a coil after the two β-strands on Prp45.

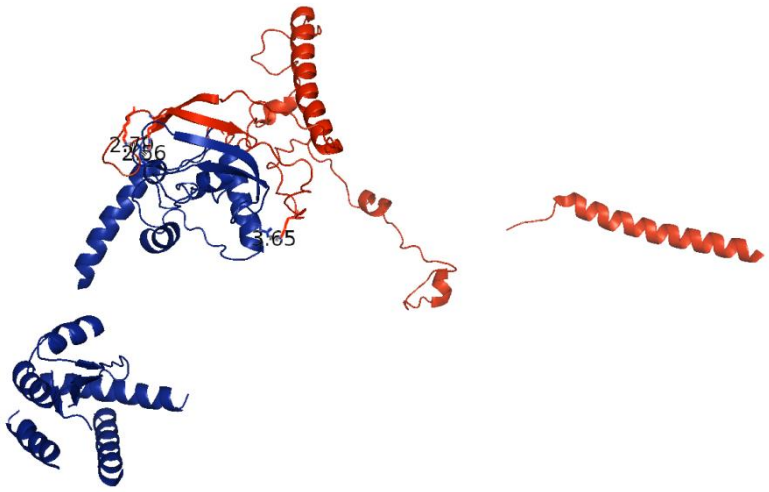

Figure 15: Interactions between Prp45 (red) and Cwf5 (blue). The numbers show distance between the two proteins at the site of the interaction. Created in PyMOL with data from 3JB9 (*Yan et al., 2015).*

5.2.6 Cwf17

Cwf17 was found to be a splicing factor of *Schizosaccharomyces pombe* but it is not present in *Saccharomyces cerevisiae.* It was also found that in vivo Cwf17 associates with Cdc5 (*Ohi et al., 2002*). Using Prosite it was determined that Cwf17 consists of 7 WD repeats.

Almost whole Cwf17 was determined in the overall structure of spliceosome (residues 42-80, 82-146, 148-249 and 254-340 of total 340 residues) revealing a globular protein of β-sheets and linking coils. There are six WD repeats present in the structure with very little residues in sequence between them. The structure folds into a circle with space in the middle (*Yan et al., 2015).*

In case of Cwf17 we have been able to determine one interaction with Prp45, since the interface was rather small one. The interaction was located between residues Arg332 on Cwf17 and Asp180 on Prp45. The residues were located in a coil right before the last β-strand and in a coil right before the first β-strand, respectively (See Fig. 16).

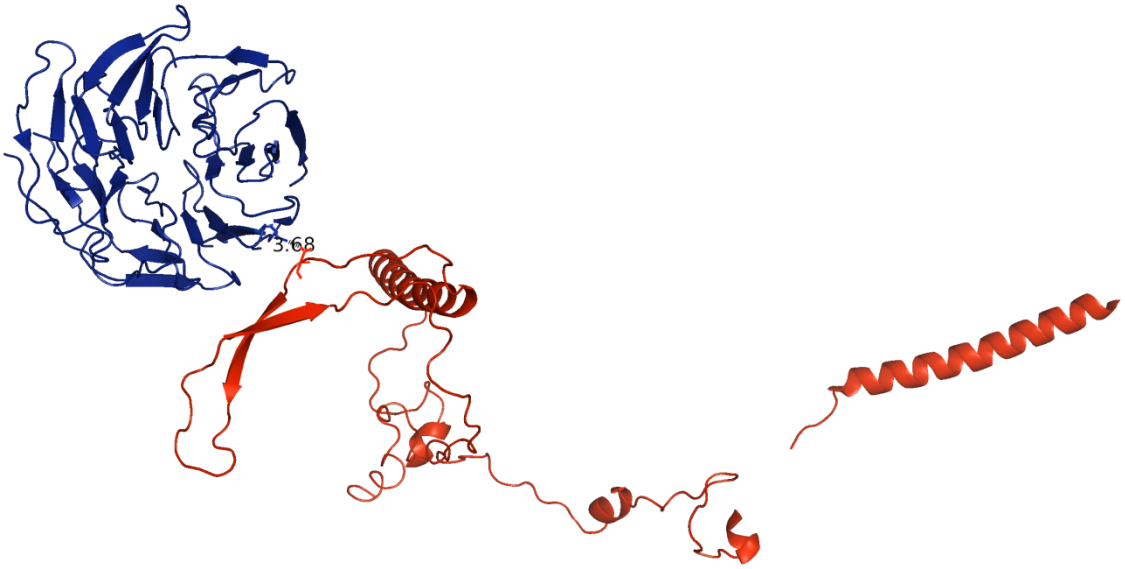

Figure 16: Interaction between Prp45 (red) and Cwf17 (blue). The numbers show distance between the two proteins at the site of the interaction. Created in PyMOL with data from 3JB9 (*Yan et al., 2015).*

6. Conclusion

Prp45 is a known part of NTC complex of the spliceosome. Recent structural studies give us the unique opportunity to observe Prp45s structure and explore its interactions. In this work we tried to map predicted interactions of Prp45 with its partners and to identify new interaction partners of Prp45.

Prp45 is located in the very core of spliceosome and has very unique extended structure. Aside of interactions that were previously established by two-hybrid assay, novel structure of spliceosome uncovered six new protein interaction partners of Prp45. However, one of this newly identified interaction partner (Smb1) has only a weak interaction with Prp45. It is apparent that most of the interactions between Prp45 and other proteins happen via conserved domains such as WD repeats, Myb repeat domains or HAT domains. This thesis support the notion that studying of interactions in spliceosome may help with understanding of processes and rearrangements that spliceosome undergoes.

Spliceosome is a very intricate complex with many components that go through extensive changes in their structure and composition. It is therefore not easy to study protein-protein or protein-RNA interaction in it. We expect that more structures of spliceosome will soon follow. These studies could shine new light on interaction network of spliceosome and also uncover new interaction partners of Prp45.

7. References

7.1 Literature

Agafonov, D. E., Kastner, B., Dybkov, O., Hofele, R. V., Liu, W. T., Urlaub, H. Stark, H. (2016). Molecular architecture of the human U4/U6.U5 tri-snRNP. *Science, 351*(6280), 1416-1420. doi:10.1126/science.aad2085

Ajuh, P., Kuster, B., Panov, K., Zomerdijk, J., Mann, M., & Lamond, A. I. (2000). Functional analysis of the human CDC5L complex and identification of its components by mass spectrometry. *Embo Journal, 19*(23), 6569-6581. doi:10.1093/emboj/19.23.6569

Albers, M., Diment, A., Muraru, M., Russell, C. S., & Beggs, J. D. (2003). Identification and characterization of Prp45p and Prp46p, essential pre-mRNA splicing factors. *Rna-a Publication of the Rna Society, 9*(1), 138-150. doi:10.1261/rna.2119903

Ambrozkova, M., Puta, F., Fukova, I., Skruzny, M., Brabek, J., & Folk, P. (2001). The fission yeast ortholog of the coregulator SKIP interacts with the small subunit of U2AF. *Biochemical and Biophysical Research Communications, 284*(5), 1148-1154. doi:10.1006/bbrc.2001.5108

Ben Yehuda, S., Dix, I., Russell, C. S., Levy, S., Beggs, J. D., & Kupiec, M. (1998). Identification and functional analysis of hPRP17, the human homologue of the PRP17/CDC40 yeast gene involved in splicing and cell cycle control. *Rna-a Publication of the Rna Society, 4*(10), 1304-1312.

Ben-Yehuda, S., Dix, I., Russell, C. S., McGarvey, M., Beggs, J. D., & Kupiec, M. (2000). Genetic and physical interactions between factors involved in both cell cycle progression and pre-mRNA splicing in Saccharomyces cerevisiae. *Genetics, 156*(4), 1503-1517.

Birney, E., Kumar, S., & Krainer, A. R. (1993). Analysis of the RNA-recognition motif and RS and RGG domains-conservation in metazoan pre-messenger-RNA splicing factors. *Nucleic Acids Research, 21*(25), 5803-5816. doi:10.1093/nar/21.25.5803

Chang, K. J., Chen, H. C., & Cheng, S. C. (2009). Ntc90 is required for recruiting first step factor Yju2 but not for spliceosome activation. *Rna-a Publication of the Rna Society, 15*(9), 1729-1739. doi:10.1261/rna.1625309

Chung, S. Y., McLean, M. R., & Rymond, B. C. (1999). Yeast ortholog of the Drosophila crooked neck protein promotes spliceosome assembly through stable U4/U6.U5 snRNP addition. *Rna-a Publication of the Rna Society, 5*(8), 1042-1054. doi:10.1017/s1355838299990635

Collier, S. E., Voehler, M., Peng, D. G., Ohi, R., Gould, K. L., Reiter, N. J., & Ohi, M. D. (2014). Structural and Functional Insights into the N-Terminus of Schizosaccharomyces pombe Cdc5. *Biochemistry, 53*(41), 6439-6451. doi:10.1021/bi5008639

Company, M., Arenas, J., & Abelson, J. (1991). Requirement of the RNA helicase-like protein PRP22 for release of messenger RNA from spliceosomes. *Nature, 349*(6309), 487-493. doi:10.1038/349487a0

D'Andrea, L. D., & Regan, L. (2003). TPR proteins: the versatile helix. *Trends in Biochemical Sciences, 28*(12), 655-662. doi:10.1016/j.tibs.2003.10.007

Dahl, R., Wani, B., & Hayman, M. J. (1998). The Ski oncoprotein interacts with Skip, the human homolog of Drosophila Bx42. *Oncogene, 16*(12), 1579-1586. doi:10.1038/sj.onc.1201687

Fabrizio, P., Dannenberg, J., Dube, P., Kastner, B., Stark, H., Urlaub, H., & Luhrmann, R. (2009). The Evolutionarily Conserved Core Design of the Catalytic Activation Step of the Yeast Spliceosome. *Molecular Cell, 36*(4), 593-608. doi:10.1016/j.molcel.2009.09.040

Folk, P., Puta, F., & Skruzny, M. (2004). Transcriptional coregulator SNW/SKIP: the concealed tie of dissimilar pathways. *Cellular and Molecular Life Sciences, 61*(6), 629-640. doi:10.1007/s00018-003-3215-4

Gahura, O., Abrhamova, K., Skruzny, M., Valentova, A., Munzarova, V., Folk, P., & Puta, F. (2009). Prp45 Affects Prp22 Partition in Spliceosomal Complexes and Splicing Efficiency of Non-Consensus Substrates. *Journal of Cellular Biochemistry, 106*(1), 139-151. doi:10.1002/jcb.21989

Galej, W. P., Oubridge, C., Newman, A. J., & Nagai, K. (2013). Crystal structure of Prp8 reveals active site cavity of the spliceosome. *Nature, 493*(7434), 638-+. doi:10.1038/nature11843

.

Hermann, H., Fabrizio, P., Raker, V. A., Foulaki, K., Hornig, H., Brahms, H., & Luhrmann, R. (1995). SnRNP Sm proteins share two evolutionarily conserved sequence motifs which are involved in Sm protein-protein interactions. *Embo Journal, 14*(9), 2076-2088.

Herold, N., Will, C. L., Wolf, E., Kastner, B., Urlaub, H., & Luhrmann, R. (2009). Conservation of the Protein Composition and Electron Microscopy Structure of Drosophila melanogaster and Human Spliceosomal Complexes. *Molecular and Cellular Biology, 29*(1), 281-301. doi:10.1128/mcb.01415-08

Kielkopf, C. L., Rodionova, N. A., Green, M. R., & Burley, S. K. (2001). A novel peptide recognition mode revealed by the X-ray structure of a core U2AF-(35)/U2AF(65) heterodimer. *Cell, 106*(5), 595-605. doi:10.1016/s0092-8674(01)00480-9

Kudlinzki, D., Schmitt, A., Christian, H., & Ficner, R. (2012). Structural analysis of the C-terminal domain of the spliceosomal helicase Prp22. *Biological Chemistry, 393*(10), 1131-1140. doi:10.1515/hsz-2012-0158

Maeder, C., Kutach, A. K., & Guthrie, C. (2009). ATP-dependent unwinding of U4/U6 snRNAs by the Brr2 helicase requires the C terminus of Prp8. *Nature Structural & Molecular Biology, 16*(1), 42-48. doi:10.1038/nsmb.1535

Martinkova, K., Lebduska, P., Skruzny, M., Folk, P., & Puta, F. (2002). Functional mapping of Saccharomyces cerevisiae Prp45 identifies the SNW domain as essential for viability. *Journal of Biochemistry, 132*(4), 557-563.

Matera, A. G., & Wang, Z. F. (2014). A day in the life of the spliceosome. *Nature Reviews Molecular Cell Biology, 15*(2), 108-121. doi:10.1038/nrm3742

Nguyen, T. H. D., Galej, W. P., Bai, X. C., Oubridge, C., Newman, A. J., Scheres, S. H. W., & Nagai, K. (2016). Cryo-EM structure of the yeast U4/U6.U5 tri-snRNP at 3.7 angstrom resolution. *Nature, 530*(7590), 298-+. doi:10.1038/nature16940

Nilsen, T. W. (2003). The spliceosome: the most complex macromolecular machine in the cell? *Bioessays, 25*(12), 1147-1149. doi:10.1002/bies.10394

Ohi, M. D., & Gould, K. L. (2002). Characterization of interactions among the Cef1 p-Prp19p-associated splicing complex. *Rna-a Publication of the Rna Society, 8*(6), 798-815. doi:10.1017/s1355838202025050

Ohi, M. D., Link, A. J., Ren, L. P., Jennings, J. L., McDonald, W. H., & Gould, K. L. (2002). Proteomics analysis reveals stable multiprotein complexes in both fission and budding yeasts containing Myb-related Cdc5p/Cef1p, novel pre-mRNA splicing factors, and snRNAs. *Molecular and Cellular Biology, 22*(7), 2011-2024. doi:10.1128/mcb.22.7.2011-2024.2002

Ohi, R., Feoktistova, A., McCann, S., Valentine, V., Look, A. T., Lipsick, J. S., & Gould, K. L. (1998). Myb-related Schizosaccharomyces pombe cdc5p is structurally and functionally conserved in eukaryotes. *Molecular and Cellular Biology, 18*(7), 4097-4108.

Ohrt, T., Prior, M., Dannenberg, J., Odenwalder, P., Dybkov, O., Rasche, N. Luhrmann, R. (2012). Prp2-mediated protein rearrangements at the catalytic core of the spliceosome as revealed by dcFCCS. *Rna-a Publication of the Rna Society, 18*(6), 1244-1256. doi:10.1261/rna.033316.112

Ottiger, M., Zerbe, O., Guntert, P., & Wuthrich, K. (1997). The NMR solution conformation of unligated human cyclophilin A. *Journal of Molecular Biology, 272*(1), 64-81. doi:10.1006/jmbi.1997.1220

Pena, V., Rozov, A., Fabrizio, P., Luhrmann, R., & Wahl, M. C. (2008). Structure and function of an RNase H domain at the heart of the spliceosome. *Embo Journal, 27*(21), 2929-2940. doi:10.1038/emboj.2008.209

Russell, C. S., Ben-Yehuda, S., Dix, I., Kupiec, M., & Beggs, J. D. (2000). Functional analyses of interacting factors involved in both pre-mRNA splicing and cell cycle progression in Saccharomyces cerevisiae. *Rna-a Publication of the Rna Society, 6*(11), 1565-1572. doi:10.1017/s1355838200000984

Salgado-Garrido, J., Bragado-Nilsson, E., Kandels-Lewis, S., & Seraphin, B. (1999). Sm and Sm-like proteins assemble in two related complexes of deep evolutionary origin. *Embo Journal, 18*(12), 3451-3462. doi:10.1093/emboj/18.12.3451

Sapra, A. K., Khandelia, P., & Vijayraghavan, U. (2008). The splicing factor Prp17 interacts with the U2, U5 and U6 snRNPs and associates with the spliceosome pre- and post-catalysis. *Biochemical Journal, 416*, 365-374. doi:10.1042/bj20081195

Schneider, C., Agafonov, D. E., Schmitzova, J., Hartmuth, K., Fabrizio, P., & Luhrmann, R. (2015). Dynamic Contacts of U2, RES, Cwc25, Prp8 and Prp45 Proteins with the Pre-mRNA Branch-Site and 3' Splice Site during Catalytic Activation and Step 1 Catalysis in Yeast Spliceosomes. *Plos Genetics, 11*(9), 27. doi:10.1371/journal.pgen.1005539

Shao, C. W., Yang, B., Wu, T. B., Huang, J., Tang, P., Zhou, Y., . . . Fu, X. D. (2014). Mechanisms for U2AF to define 3 ' splice sites and regulate alternative splicing in the human genome. *Nature Structural & Molecular Biology, 21*(11), 997-1005. doi:10.1038/nsmb.2906

Skruzny, M., Ambrozkova, M., Fukova, I., Martinkova, K., Blahuskova, A., Hamplova, L. Folk, P. (2001). Cyclophilins of a novel subfamily interact with SNW/SKIP coregulator in Dictyostelium discoideum and Schizosaccharomyces pombe. *Biochimica Et Biophysica Acta-Gene Structure and Expression, 1521*(1-3), 146-151. doi:10.1016/s0167-4781(01)00301-3

Slaugenhaupt, S. A., Blumenfeld, A., Gill, S. P., Leyne, M., Mull, J., Cuajungco, M. P., Gusella, J. F. (2001). Tissue-specific expression of a splicing mutation in the IKBKAP gene causes familial dysautonomia. *American Journal of Human Genetics, 68*(3), 598-605. doi:10.1086/318810

Tsai, W. Y., Chow, Y. T., Chen, H. R., Huang, K. T., Hong, R. I., Jan, S. P.Cheng, S. C. (1999). Cef1p is a component of the Prp19p-associated complex and essential for pre-mRNA splicing. *Journal of Biological Chemistry, 274*(14), 9455-9462. doi:10.1074/jbc.274.14.9455

Umen, J. G., & Guthrie, C. (1995a). A novel role for a U5 snRNP protein in 3' splice site selection. *Genes & Development, 9*(7), 855-868. doi:10.1101/gad.9.7.855

Umen, J. G., & Guthrie, C. (1995b) Prp16p, Slu7p, and Prp8p interact with the 3' splice site in two distinct stages during the second catalytic step of pre-mRNA splicing. *RNA 1, 1*(6), 584-597.

Vijayraghavan, U., Company, M., & Abelson, J. (1989). Isolation and characterization of pre-messenger RNA splicing mutants of Saccharomyces cerevisiae. *Genes & Development, 3*(8), 1206-1216. doi:10.1101/gad.3.8.1206

Wahl, M. C., Will, C. L., & Luhrmann, R. (2009). The Spliceosome: Design Principles of a Dynamic RNP Machine. *Cell, 136*(4), 701-718. doi:10.1016/j.cell.2009.02.009

Wan, R. X., Yan, C. Y., Bai, R., Wang, L., Huang, M., Wong, C. C. L., & Shi, Y. G. (2016). The 3.8 angstrom structure of the U4/U6.U5 tri-snRNP: Insights into spliceosome assembly and catalysis. *Science, 351*(6272), 466-475. doi:10.1126/science.aad6466

Wang, Q., Hobbs, K., Lynn, B., & Rymond, B. C. (2003). The Clf1p splicing factor promotes spliceosome assembly through N-terminal tetratricopeptide repeat contacts. *Journal of Biological Chemistry, 278*(10), 7875-7883. doi:10.1074/jbc.M210839200

Webb, C. J., & Wise, J. A. (2004). The splicing factor U2AF small subunit is functionally conserved between, fission yeast and, humans. *Molecular and Cellular Biology, 24*(10), 4229-4240. doi:10.1128/mcb.24.10.4229-4240.2004

Wilkie, S. E., Vaclavik, V., Wu, H. M., Bujakowska, K., Chakarova, C. F., Bhattacharya, S. S., . . . Hunt, D. M. (2008). Disease mechanism for retinitis pigmentosa (RP11) caused by missense mutations in the splicing factor gene PRPF31. *Molecular Vision, 14*(81-83), 683-690.

Will, C. L., & Luhrmann, R. (2011). Spliceosome Structure and Function. *Cold Spring Harbor Perspectives in Biology, 3*(7), 23. doi:10.1101/cshperspect.a003707

Worthington, M. T., Amann, B. T., Nathans, D., & Berg, J. M. (1996). Metal binding properties and secondary structure of the zinc-binding domain of Nup475. *Proceedings of the National Academy of Sciences of the United States of America, 93*(24), 13754-13759. doi:10.1073/pnas.93.24.13754

Xu, C., Zhang, J. H., Huang, X. J., Sun, J. P., Xu, Y. Q., Tang, Y. J. Zhang, Q. H. (2006). Solution structure of human peptidyl prolyl isomerase-like protein 1 and insights into its interaction with SKIP. *Journal of Biological Chemistry, 281*(23), 15900-15908. doi:10.1074/jbc.M511155200

Xu, D. M., Field, D. J., Tang, S. J., Moris, A., Bobechko, B. P., & Friesen, J. D. (1998). Synthetic lethality of yeast sit mutations with U2 small nuclear RNA mutations suggests functional interactions between U2 and U5 snRNPs that are important for both steps of pre-mRNA splicing. *Molecular and Cellular Biology, 18*(4), 2055-2066.

Xu, D. M., & Friesen, J. D. (2001). Splicing factor Slt11p and its involvement in formation of U2/U6 helix II in activation of the yeast spliceosome. *Molecular and Cellular Biology, 21*(4), 1011-1023. doi:10.1128/mcb.21.4.1011-1023.2001

Yan, C. Y., Hang, J., Wan, R. X., Huang, M., Wong, C. C. L., & Shi, Y. G. (2015). Structure of a yeast spliceosome at 3.6-angstrom resolution. *Science, 349*(6253), 1182-1191. doi:10.1126/science.aac7629

Yoshida, H., Park, S. Y., Oda, T., Akiyoshi, T., Sato, M., Shirouzu, M., . . . Obayashi, E. (2015). A novel 3 ' splice site recognition by the two zinc fingers in the U2AF small subunit. *Genes & Development, 29*(15), 1649-1660. doi:10.1101/gad.267104.115

Yu, H. Y., Braun, P., Yildirim, M. A., Lemmens, I., Venkatesan, K., Sahalie, J.Vidal, M. (2008). High-quality binary protein interaction map of the yeast interactome network. *Science, 322*(5898), 104-110. doi:10.1126/science.1158684

Zamore, P. D., & Green, M. R. (1989). Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor.*Proceedings of the National Academy of Sciences of the United States of America, 86*(23), 9243-9247. doi:10.1073/pnas.86.23.9243

Zhang, D., Abovich, N., & Rosbash, M. (2001). A biochemical function for the Sm complex. *Molecular Cell, 7*(2), 319-329. doi:10.1016/s1097-2765(01)00180-0

7.2 Used software

Chatr-aryamontri, A., Breitkreutz, B. J., Oughtred, R., Boucher, L., Heinicke, S., Chen, D. C., Tyers, M. (2015). The BioGRID interaction database: 2015 update. *Nucleic Acids Research, 43*(D1), D470-D478. doi:10.1093/nar/gku1204

The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.