Characterization of U2 snRNA associated proteins: Functional Analysis of *C. merolae* Hsh49 Protein

by

Mona Aminorroayaee

B.Sc., Tabriz University of Medical Science, 2009

Thesis Submitted in Partial Fulfillment of
The Requirements for the Degree of
Master of Science
In
Mathematical, Computer, and Physical Sciences
(Biochemistry)

The University of Northern British Columbia
January 2018

Abstract

Pre-mRNA splicing is an important process involving the removal of non-coding regions (introns) from messenger RNA. Disruption of this mechanism can cause diseases such as cancer. A particle known as the spliceosome is responsible for recognition and removal of introns. This complex machinery consists of five small nuclear RNAs, U1, U2, U4, U5, and U6, and many proteins. Each snRNA associates with proteins to form small nuclear ribonucleo-proteins or snRNPs. Due to the considerable number of proteins involved in this process in humans, the study of splicing has been a struggle. *Cyanidioschyzon merolae* (*C. merolae*) is a unicellular acidophilic red alga that we hope will be a good model for studying splicing factors due to its greatly reduced complexity (only ~70 splicing proteins, compared to ~300 in humans). Thus, we are investigating the interactions of snRNAs with their specific proteins by expressing the entire snRNP particles recombinantly and investigating their function.

I sought to use this system to investigate the interactions between the U2 snRNA and its proteins. U2 snRNA associates with two multi-protein complexes, SF3a and SF3b. The SF3b complex is required for the recognition of the intron's branch point and precise excision of introns from pre-mRNA. I attempted to express all of these proteins simultaneously as well as individually, with my efforts ultimately focused on Hsh49. This is one of five subunits of the SF3b protein complex in *C. merolae*. I was not able to detect any interaction between U2 snRNA and the Hsh49 protein by Electrophoretic mobility shift assay (EMSA), Fluorescence polarization (FP), or immunoprecipitation analysis. Nevertheless, Circular Dichroism (CD) showed that Hsh49 folded properly.

TABLE OF CONTENTS

List of Figures

List of Tables

Acknowledgments

I would like to thank first and foremost my supervisor, Dr. Stephen Rader, for giving me the opportunity to conduct my research in his lab. Stephen provided a friendly training environment that motivates me to be curious in each step of performing my research and have a deep understating of the Pre-mRNA splicing process. I would also like to thank Stephen for the time that he patiently spent to introduce and explain complicated concepts in this area. I really enjoyed my studies during these years.

I would like to acknowledge the members of my supervisory committee, Dr. Kerry Reimer, Dr. Brent Murray, and Dr. Keith Egger for their valuable questions and informative comments on this thesis.

I am also so thankful to past and current members of the Rader lab: Dr. Liz Dunn, Ambreen Siraj, Fatimat Shidi, Viktor Slat, Maya de Vos, Radu Pasca, Raliat Abioye, Kevin Huolt, and Samantha Smith for their critique questions and insightful conversation that helped me to develop my knowledge and I would like to thank Corbin Black for his assistance to find the best possible sequence alignment data. I would also like to acknowledge Sebastian J. Mackedenski for his help in circular dichroism spectropolarimetry data collection.

I would particularly like to thank Dr. Martha Stark for introducing practical techniques in the Lab and for all her helpful and informative advices through my project.

I am also grateful to Brooke Boswell and Luke Spooner for proofing and editing of this thesis.

I would like to thank UNBC for granting me the financial support that I was given to pursue my studies.

My special thanks go to my husband, Rahim Pasha Khajei, for his ongoing support and wise counsel over the years. I am so thankful to my parents, Yasamin Bozarjomehr and Javad Aminorroayaee for their love and guidance, and my brothers Emia and Aria for making me happy through passing though times.

Thank you very much, everyone!

Mona Aminorroayaee

March 22, 2018

Chapter 1
Introduction

Pre messenger RNA splicing is an important process involving the removal of non-coding regions (introns) from messenger RNA, which is carried out by a highly complex and dynamic machine called the spliceosome. This complex machinery consists of five small nuclear RNAs (snRNAs), U1, U2, U4, U5, and U6, and many proteins. Each snRNA associates with proteins to form small nuclear ribonucleoproteins, (snRNPs). Due to the considerable number of proteins involved in this process in humans, the study of splicing has been a struggle. *Cyanidioschyzon merolae* (*C. merolae*) is a unicellular acidophilic red alga that we hope will be a good model for studying splicing factors due to its greatly reduced complexity (only ~70 splicing proteins, compared to ~300 in humans; Will *et al*. 2011). Thus, we are investigating the interactions of snRNAs with their specific proteins by expressing entire snRNP particles recombinantly in bacteria.

During spliceosome assembly, U1 snRNA recognizes and base-pairs with the 5' splice site of the pre-mRNA. However, Stark *et al*. (2015) could not find any candidates for the U1-associated proteins or U1 snRNA, hinting that *C. merolae* does not have a U1 snRNP. Biochemical studies of the U2 snRNA would help to improve understanding of the initiation of the spliceosome assembly in the absence of the U1 snRNA, and provide information about the interaction of the U2 snRNA with its specific proteins.

1.1 Pre-mRNA Splicing

The 5′ splice site (5′SS), the 3′ splice site (3′SS) and the branch point site (BPS) are reactive sections on the pre-mRNA transcript and these contain small conserved sequences (Chiou *et al*. 2014). Pre-mRNA splicing involves removal of non-coding sequences known as introns and subsequent joining of adjacent protein-coding sequences called exons (Figure 1). In other words, splicing happens on precursor-messenger RNA, which is transcribed from template DNA in the nucleus, to generate mature messenger RNA (mRNA). Mature mRNA can then be translated into protein after export to the cytoplasm.

Pre-mRNA splicing comprises two transesterification reactions. In the first reaction, the 2′ hydroxyl group of the adenosine within the branch point of the intron reacts with the phosphodiester bond of the 5′splice site (Figure 1); this reaction leads to a free 5′ exon and lariat intron. In the second reaction, the phosphodiester bond at the 3′ splice site reacts with the 3′ hydroxyl located at the 5′ exon. Through this process, the exons are joined to each other and the intron is eliminated (Wahl *et al*. 2009; Figure 1).

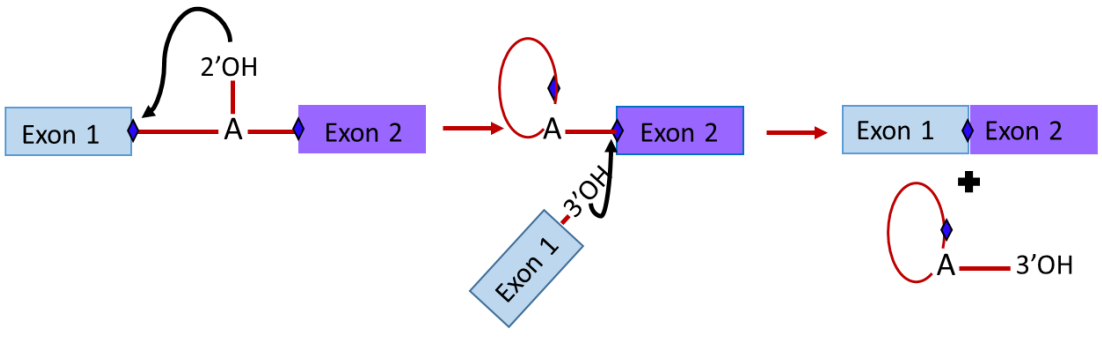


Figure 1. Pre-mRNA splicing involving removal of non-coding introns and subsequent joining of adjacent protein-coding exons. Exons and introns are indicated with box and thick line respectively. Letter (A) represents branch point Adenosine.

1.2 Spliceosome Assembly

The spliceosome is a large RNA-protein complex located in the nucleus, and it identifies and interacts with the three reactive regions of substrate transcripts (5′SS, 3′SS, BPS) in order to perform pre-mRNA splicing (Chiou *et al*. 2014). The spliceosome consists of five small nuclear ribonucleoproteins (snRNPs) including U1, U2, U4, U6, and U5, which are the major components of the spliceosome. Each snRNP contains a snRNA and associated proteins (Wahl *et al*. 2009). The human spliceosome includes five small nuclear RNAs and more than 300 proteins, while the spliceosome in *Saccharomyces cerevisiae* (*S. cerevisiae*) contains five small nuclear RNAs and almost 100 different proteins (Jurica *et al*. 2003; Wahl *et al*. 2009, Fabrizio *et al*. 2009).

The assembly of the spliceosome starts with recognition and base pairing of U1 and U2 snRNPs to the 5′ splice site and branch site of the pre-mRNA respectively. The U2 snRNA binds to the branch site, from which the catalytic adenosine at the branch point binding sequence bulges out (Figure 2, Commitment Complex). Single molecule analysis posits that either U1 or U2 can bind first to the pre-mRNA during the early steps of spliceosome assembly (Hoskins *et al*, 2016). The U6 snRNA associates with the U4 snRNA to form the U4/U6 di-snRNA; then, the binding of U5 snRNA to U4/U6 di-snRNP results in creation of the U4/U6.U5 tri-snRNP (Will *et al*. 2011). After association of tri-snRNP to form the pre-spliceosome (Figure 2, left side), the interaction between U1 snRNA and the 5′ splice site is disrupted, as well as that between the U4 and U6 snRNAs. Next, U6 snRNA associates with U2 snRNA and the 5′ splice site through base pairing interactions (Figure 2, Active Spliceosome). The order and arrangement of the spliceosome is key to the identification and excision of introns from the rest of the pre-mRNA during splicing (Wahl *et*

al. 2009). These rearrangements and changes are necessary for the spliceosome to proceed to the transesterification reaction (Brow. 2002). At the end of the splicing pathway, the spliced intron is released and the U2, U5, and U6 snRNPs, which were connected to the lariat intron, disassemble and are recycled for the next round of splicing (Chang *et al*. 2013; Figure 2, top right).

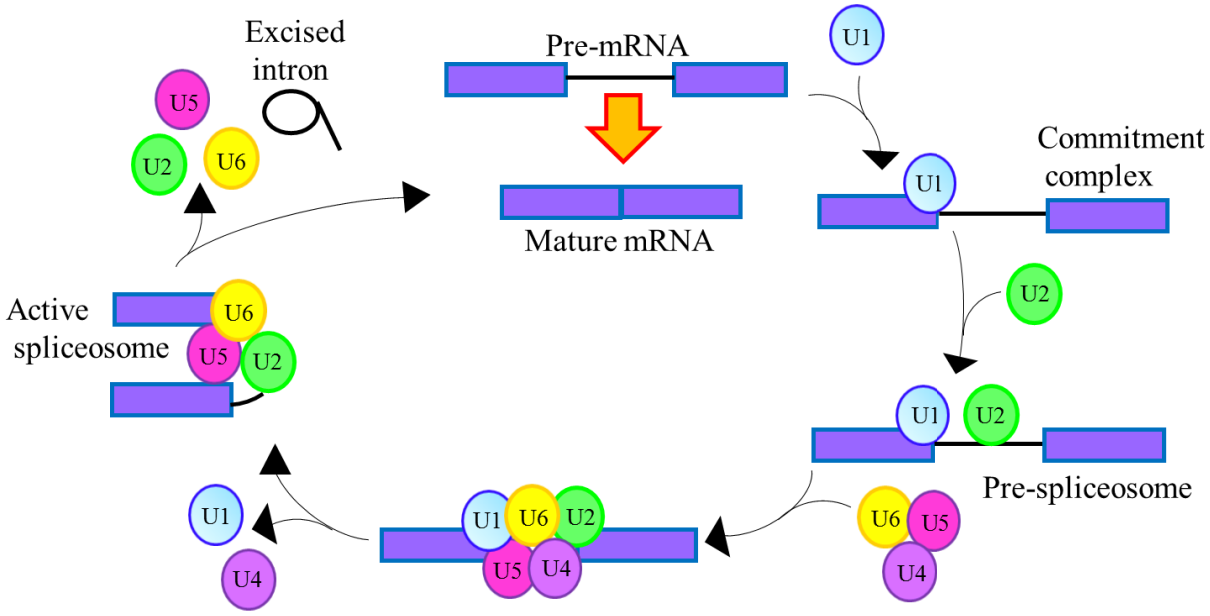


Figure 2. The pathway of spliceosome assembly during pre-mRNA splicing.

1.3 Splicing and Disease

Since pre-mRNA splicing is a vital process in gene expression, in order to create mature messenger RNA (mRNA) and then functional proteins; mis-regulation in splicing processes causes a wide variety of human diseases. Any addition or removal of nucleotides to mRNA, as a result of splicing errors, can lead to disruption of the open reading frame of mRNA and consequent truncation of the encoded protein (Faustino and Cooper 2003).

Research indicates that approximately half of all genetic diseases are due to disruptions in splicing (Lopez-Bigas *et al*. 2005). Examples of human genetic diseases caused by mutations in splicing include: Retinitis pigmentosa, which leads to blindness; spinal muscular atrophy, which causes mortality in childhood; myotonic dystrophy, the most common muscular disorder in adults; Frasier syndrome; Parkinson's disease; and Atypical cystic fibrosis (Faustino and Cooper 2003). In addition, a wide range of cancers has been linked to errors in splicing. Therefore, it seems that a better understanding of the splicing reaction is an essential step to providing more research opportunities for treatment of human genetic diseases.

1.4 Role of U2 snRNP in Yeast and Humans

Many of the interactions in the spliceosome seem to function in ensuring that activation occurs at the right time and place. As mentioned above, U2 snRNA binds to the branch point, thereby inducing the branch site A to bulge out in its active conformation. At the same time, U2 components such as SF3b appear to sequester the bulged A to prevent it from reacting prematurely (Lardelli *et al*. 2010). In addition, U6 snRNA associates with U2 snRNA and the 5′ splice site through base pairing interactions (Wahl *et al*. 2009). The association of the U2 snRNP with the pre-mRNA branch site is one of the primary stages in the pre-mRNA splicing process and plays a fundamental role during splicing assembly since they directly participate in chemical catalysis (Nilsen 1998). Moreover, the U2 snRNP undergoes several rearrangements along with the recognition of the branch point adenosine and formation of base pair interaction between theU2 snRNA and "consensus sequence within the intron" (Sashital *et al*. 2007).

1.4.1 Structure of U2 snRNA

The 5′ end of U2 snRNA, which is highly conserved, contains several stem–loops, and single-stranded regions, which interact with the U6 snRNA, the intron, and the snRNP proteins (Figure 3 and Figure 4). Donmez *et al*. (2004) investigate that stem I in the 5′ terminus of U2 snRNA has a function in early steps of spliceosome assembly. Structural studies of the U2-U6 complex show the presence of the U2 stem I as a stable structure in this complex (Sashital *et al*. 2004).

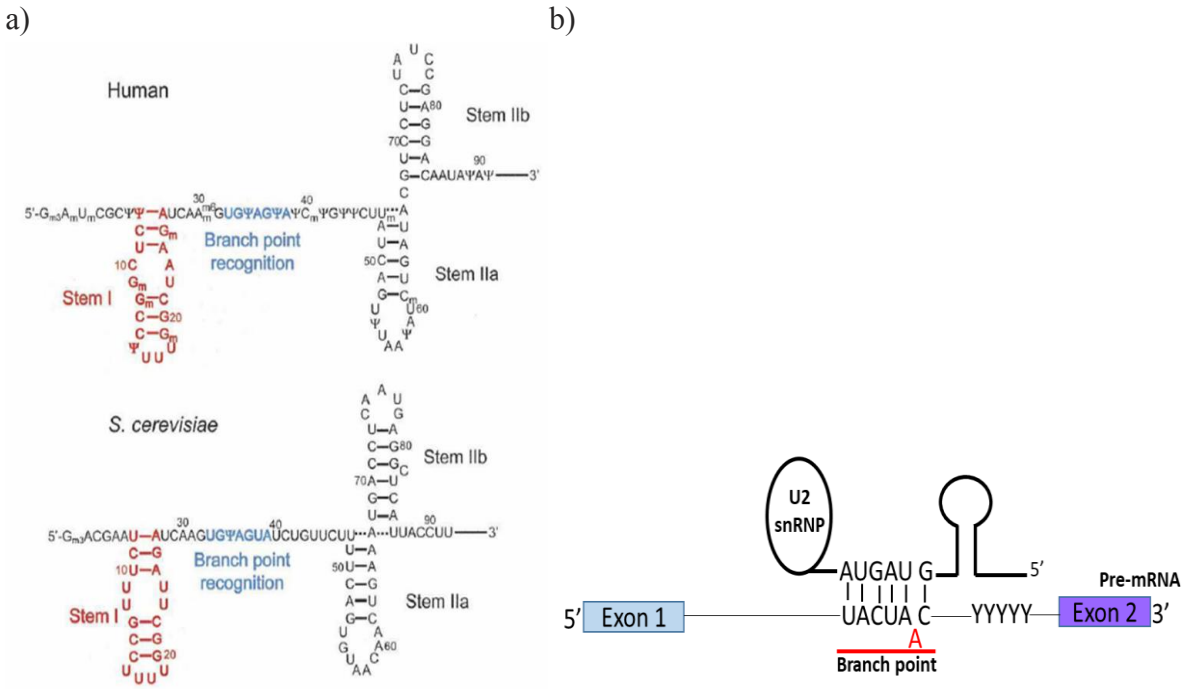


Figure 3. Secondary structures of free U2 snRNA and base pairing interaction of U2 snRNA to the branch point and a comparison of yeast and human sequences. (a) The 5′ end of U2 snRNA comprises the branch point recognition sequence (blue) and three stem–loops, including stem I (red), stem IIa and stem IIb (Sashital *et al*. 2007). (b) Base pairing interaction between branch point recognition of the U2 snRNA and branch point (A) of the pre-mRNA, Y's demonstrate the polypyrimidine tract.

The degree of the post-transcriptional modification of U2 snRNAs in yeast and humans is different. The four 2'-O-methyl guanosines and two pseudouracils are examples of

modifications that occur in Stem I of mammalian U2 snRNA (Massenet *et al*.1998), while there is no post-transcriptional modification in yeast Stem I. NMR structures of the U2 stem I from *S. cerevisiae* and the fully modified U2 stem I from humans reveal that there is a great similarity in structure. Despite the fact that they include divergent tandem wobble pairs and modified nucleotides found in human U2 stem I, this indicates that the overall folds of the stem loops of U2 snRNA are similar in yeast and humans (Sashital *et al*. 2007, Figure 3a). However, the U2 snRNA stem I is more stable in humans than yeast stem I (Sashital *et al*. 2007, Figure 3a).

Most binding of the U2 snRNP to the branch site in humans occurs through the polypyrimidine tract (Figure 3b) in an ATP-dependent manner. It seems this base pairing interaction increases the stability of the binding between U2 snRNA and the polypyrimidine tract, and helps to specify the exact location of branch formation. With this in mind, it is clear that without this interaction, the first step of the splicing pathway and spliceosome assembly would be impossible (Kramer *et al*. 1988; Wu *et al*. 1989). However, in yeast, introns do not contain a polypyrimidine tract (Rymond and Rosbash 1985; Rymond *et al*. 1987). Overall, from yeast to mammals, U2 snRNA has been conserved evolutionarily (Ares *et al*. 1986; Igel and Ares. 1988). The base paring interaction between the U2 snRNA and the branch point with the intron and the pre-mRNA splicing mechanism is similar in both yeast and mammals (Parker *et al*. 1987; Chabot *et al*. 1985; Ruskin *et al*. 1988, Zhuang *et al*. 1989).

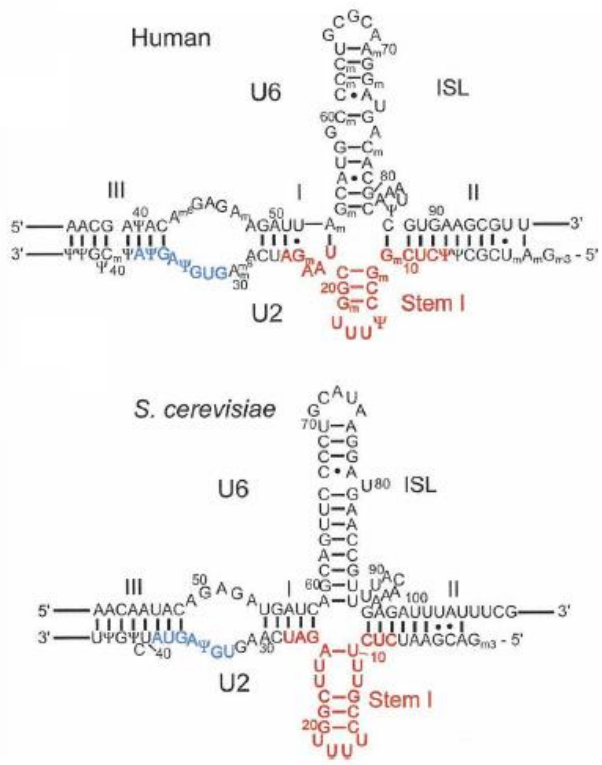


Figure 4. Secondary structure of the U2 and U6 snRNA complex in humans and yeast. The secondary structure of the U2 and U6 snRNA in humans was presented by Sun and Manley (Sun and Manley. 1995) The NMR secondary structure of the U2 and U6 snRNA in yeast was proposed by Sashital (Sashital *et al*. 2004)

1.5 U2 snRNP in *C. merolae*

Cyanidioschyzon merolae (*C. merolae*) is a unicellular, acidophilic red alga. The small size of the genome and reduced number of introns in *C. merolae* make the idea of studying splicing in this system an interesting subject. *C. merolae* contains the smallest genome size (16.5 Mb) among eukaryotes (Misumi *et al*. 2005; Suzuki *et al*. 1992). The number of genes in the *C. merolae* genome is the same as *S. cerevisiae* (yeast), but 26 genes contain introns (0.5% of the genome) in *C. merolae* (Matsuzaki *et al*. 2004) compared to 287 in *S. cerevesiae* (Juneau *et al*. 2007). It is worth mentioning that *C. merolae* grows in acidic hot springs with a pH of ~1.5 and temperature of ~ 45 C (Matsuzaki *et al*. 2004), raising the possibility that its snRNAs are intrinsically more stable. It is also the first alga in which the

genome was sequenced and the first whose sequence was 100% completed (Nozaki *et al*. 2007).

By considering such features, some interesting questions emerge – such as whether or not proteins involved in the splicing process can remain stable in high temperatures and/or how the process of splicing occurs in this organism.

A comprehensive bioinformatics study carried out by Stark *et al.* (2015) has determined that the *C. merolae* spliceosome contains four snRNAs including U2, U4, U5, and U6. This research also did not find any candidates for the U1-associated proteins or U1 snRNA, suggesting that *C. merolae* does not have a U1 snRNP. By considering this fact, it would be interesting to know how splicing starts in the absence of the U1 snRNA and if it is possible that the U2 snRNA starts the splicing process in *C. merolae*. These questions motivated me to study the U2 snRNA in *C. merolae*. In this regard, it is imperative to know the U2 snRNA's components and understand the role and function of each particle during the splicing process.

1.5.1 U2 snRNA proteins

Here I will cover all of the functions of the U2 snRNA proteins which have been recognized either in yeast or humans. The U2 snRNA associates with ten specific proteins and seven Sm proteins in *C. merolae,* making the U2 snRNP the most complex particle in splicing (Table 1). The following are the proteins U2 snRNA associates with:

- SF3a and SF3b

These are two multiprotein complexes. SF3a contains three subunits including Prp9, Prp11, and Prp21, and SF3b consists of five subunits, Hsh155, Hsh49, Cus1, Rse1, and Rds3 proteins, all of which are present in *C. merolae*. However, SF3b in *S. cerevisiae* has one additional protein, Ysf3 (legrain and Chapon. 1993; Kramer *et al*. 1995; Brosi *et al*. 1993; Will *et al*. 2002; Wang *et al*. 2005; Stark *et al*. 2015). Both SF3a and SF3b play an important role during pre-spliceosome assembly (Gozani *et al*. 1996). The SF3a complex is crucial for initiation of the first step of the splicing reaction that occurs when SF3a dissociates from the spliceosome (Lin *et al*. 2012). The SF3a and SF3b complexes are required for recognition of the intron's branch point and precise excision of introns from pre-mRNA. These complexes exist prior to the first chemical step, when U2 snRNA base pairs with the branch point and they dissociate before the second chemical step occurs (Lardelli *et al*. 2010).

- Rds3

In yeast, Rds3 is one of the subunits of SF3b (Cus1p, Hsh49p, Hsh155p, Rse1p), as well as Ist3p/Snu17p and with Yra1p (RNA export protein). It is also an essential factor for the stability of the prespliceosome. The absence of Rds3p prevents spliceosome formation and blocks U2 snRNP addition (Wang and Rymond 2003; Figure 5).

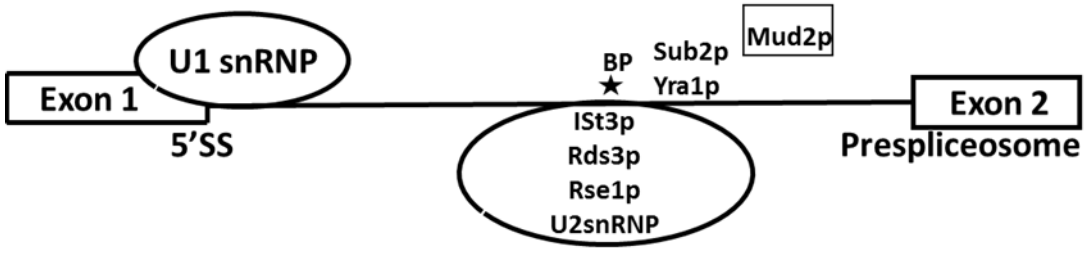


Figure 5. Model for Rds3 function. An essential factor for the stability of the spliceosome. The absence of Rds3p prevents spliceosome formation and blocks U2 snRNP addition to the commitment complex pre-spliceosome formation (Wang and Rymond. 2003)

- U2AF

U2AF is one of the splicing factors that facilitates interaction of the U2 snRNP to the branch point. It consists of a 65 kD subunit and a 35 kD subunit. The U2AF65 binds the polypyrimidine tract, while U2AF35 recognizes the 3′ AG dinucleotide in humans (Wu *et al*. 1999; Wu and Fu. 2015).

Mud2

Since there is no polypyrimidine tract in *C. merolae*, Stark *et al*. (2015) suggest that Mud2p interacts with the branch point sequences directly.

Msl5

Msl5, or branch point binding protein (BBP), binds 5′_UACUAAC sequences of the branch point and Mud2p in *S. cerevisiae.* In yeast and mammals, Msl5 recognizes the intron by interacting with the pre-mRNA branch point. Both Mud2 and Msl5 are commitment complex components and play a critical role in the stability of U1 snRNA, which occurs by interacting between these two proteins and U1 snRNP protein, Prp40 (Rutz *et al*. 1999; Abovich *et al*, 1997; Chang *et al*. 2012).

Sub2

During pre-spliceosome assembly, two ATPases, Sub2 and Prp5, are necessary for interactions between U2 snRNA and the branch point binding site (O'Day *et al*. 1996; Parker *et al*. 1987; Kistler *et al*. 2001). It is assumed that Sub2p is involved in the removal of Mud2 and Msl5 from the branch site, so that U2 snRNA is exposed to the branch site (Kistler *et al*. 2001).

Prp5

This protein contributes to the remodeling of U2 snRNA to make it more accessible to the branch point binding sequences. Prp5 activity occurs as a result of the association of the three subunits of SF3a with the U2 snRNA. In this association, Prp21 acts as a bridge between Prp9 and Prp11 (Wiest *et al*. 1996; Lin and Xu 2012; Figure 6).

Cus1

Assembly of U2 snRNP into the spliceosome to form the pre-spliceosome is the earliest function of Cus1p. After this protein binds to U2 snRNA, an interaction with pre-mRNA results. During the splicing process, Cus1p forms a complex with Hsh49 and Hsh155 and remains associated with the spliceosome (Pauling *et al*. 2000).

Prp43

Prp43 is one of the U2 snRNA proteins related to the DEXH/D-box RNA helicase family. It functions in the early steps of spliceosome disassembly. Prp43 removes U2, U5, and U6 snRNPs from the post-splicing lariat intron ribonucleoprotein complex in order to release mature RNA (Arenas and Abelson. 1997; Martin *et al*. 2002, Combs *et al*. 2006). The function of the Prp43 protein before spliceosome disassembly is not well understood.

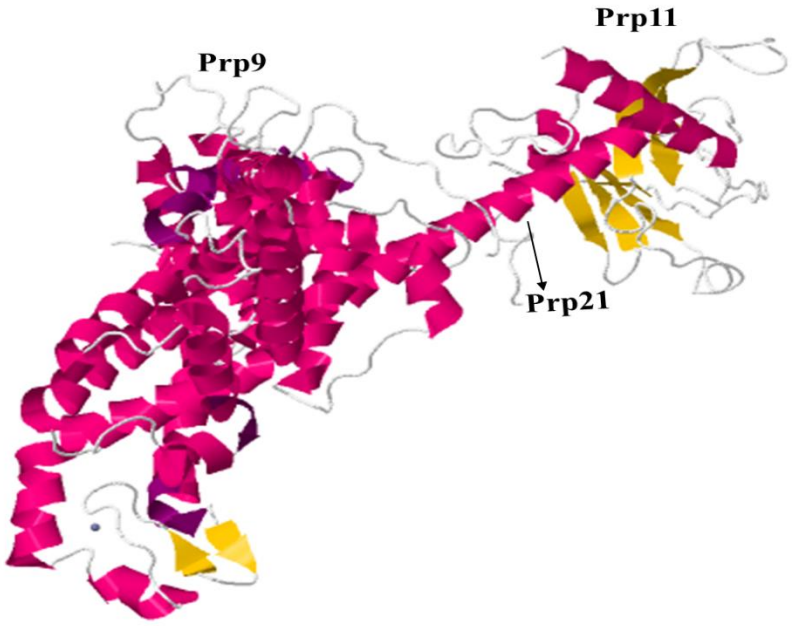


Figure 6. The structure of the SF3a core. Prp11 and Prp9 do not bind each other; however, Prp21 acts as a bridge between these two proteins (PDB file 4DGW; Lin & Xu. 2012).

Lea1, Msl1 and Cus2

Lea1, Msl1, and Cus2 are non-essential proteins in yeast. The Lea1 protein is a specific particle in yeast U2 snRNP and interacts with Msl1. The association of Msl1 with U2 snRNA is possible only in the presence of Lea1 (Caspary *et al*. 1998). Cus2 interacts with Prp11, which is one of the subunits of SF3a, and there is evidence that Cus2 might has a role in recruiting SF3a to the U2 snRNP (Yan *et al*. 1998).

Table 1. Comparison of all U2 snRNA proteins in humans, yeast, and *C. merolae* [a]

All of the proteins in the table have been found by mass spectrometry except for MSL5. Msl5 and Mud2 are the commitment complex proteins that are identified with 10 U2 specific proteins in *C. merolae* (Stark *et al*. 2015).

C. merolae	*Saccharomyces cerevisiae*	Human	Aliases of human U2 snRNA protein
CmSmB/B'	SmB/B'	SmB/B'	
CmSmD1	SmD1	SmD1	HsT2456, SMD1, SNRPD, Sm-D1
CmSmD2	SmD2	SmD2	
CmSmD3	SmD3	SmD3	SMD3, Sm-D3
CmSmE	SmE	SmE	B-raf, HYPT11, SME, Sm-E
CmSmF	SmF	SmF	
CmSmG	SmG	SmG	SMG, Sm-G
CmPrp9	Prp9	SF3a3	PRP9, PRPF9, SAP61, SF3a60
CmPrp11	Prp11	SF3a2	PRP11, PRPF11, SAP62, SF3a66
CmPrp21	Prp21	SF3a1	PRP21, PRPF21, SAP11420, SF3A1
CmHsh155	Hsh155	SF3b1	Hsh155, MDS, PRP10, PRPF10, SAP155, SF3b155
CmCus1	Cus1	SF3b2	Cus1, SAP145, SF3B145, SF3b1, SF3b150
CmRse1	Rse1	SF3b3	RSE1, SAP130, SF3b130, STAF130
CmHsh49	Hsh49	SF3b4	AFD1, Hsh49, SAP49, SF3b49
CmRds3	Rds3	PHF5	SPOC1
		SF3b6	CGI-110, HSPC175, Ht006, P14, SAP14, SAP14a, SF3B14, SF3B14a
CmPrp43	Prp43	hPrp43	YGL120C
CmPrp5	Prp5	hPRP5	YBR237WRNA5
CmMud2	Mud2	U2AF2	U2AF65
CmMsl5	MSL5/BBP	SF1	BBP, D11S636, MBBP, ZCCHC25, ZFM1, ZNF162
		U2AF1	FP793, RN, RNU2AF1, U2AF35, U2AFBP
	Msl1	SNRPB2	SPBC8D2.09c
	Ysf3	SF3B5	SF3b10, Ysf3
	Lea1	SNRPA1	Lea1
	RES Complex: Bud13, Pml1, Ist3	RES Complex: MGC13125, SNIP, CGI-79	
		U2 related: U2AF, PUF60, SPF30, SPF31, SPF45, CHERP, SR140	

[a] Cvitkovic & Jurica; (2013) NCBI Gene.

Almost all U2 snRNA proteins interact with each other during pre-mRNA splicing. In the SF3a complex, Prp21 has more interactions with other U2 snRNA proteins compared with Prp9 and Prp11. Among the SF3b complex, Hsh155 mostly interacts with other U2 proteins while other proteins of the SF3b complex interact with each other rather than U2 snRNA proteins. Moreover, Sms interact with each other and other U2 snRNA proteins (Figure 7). The diagram shows the physical interactions between U2 snRNA proteins. These studies are curated by BioGRID and are listed in the Saccharomyces Genome Database.

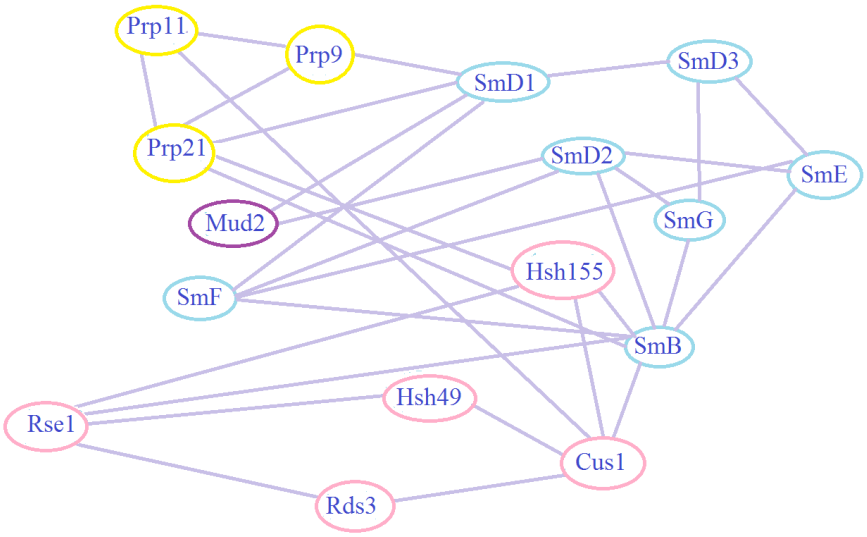


Figure 7. Interaction network between U2 snRNA proteins. Physical interactions or protein-protein interactions are based on the purification of proteins from *Saccharomyces cerevisiae* genome. The number of lines connecting to each protein show how many other proteins it is able to interact with. The Prp5 and Prp43 do not have any interaction with other U2 snRNA proteins in *S. cerevisiae* (*Saccharomyces* Genome Database). Proteins labeled with the same colors form a complex.

1.6 Previous expression of U2-associated proteins

Production of recombinant proteins is a powerful technique to study the biochemical characterization of a desired protein. The expression or co-expression of the U2

snRNA proteins has been done in different organisms either for functional or structural analyses.

In one study, the human SF3b complex was co-expressed in insect cells to investigate the structure of these proteins (Cretu *et al*. 2016). In another study, two subunits of SF3b complex, yeast Hsh49p, and Cus1protein, were expressed individually in a strain of *Escherichia coli*, BL21 (Igel *et al*. 1998; Pauling *et al*. 2000). Also, Rds3p was expressed separately in yeast cells (Wang and Rymond. 2003). Yeast SF3a complex with full length proteins has been co-expressed in BL21 (Lin *et al*. 2012). Rosetta cells, another strain of *E.coli,* were used for expression of yeast Prp43 and Prp5 proteins (Tauchert *et al*. 2016; Liang *et al*. 2015).

1.7 Overall Research Objective

Since the presence of the U1 snRNA or the U1-associated proteins have not been recognized in *C. merolae,* a question that arises is: Does the U2 snRNA help initiate spliceosome assembly? Additionally, how do U2 snRNA specific proteins act upon each other during pre-mRNA splicing process?

Understanding the function and structure of the U2 snRNPs helps to address these questions about initiation and regulation of spliceosome assembly, as well as the role of each U2 snRNA protein in the whole process of pre-mRNA splicing. My overall research goal was to generate plasmids for recombinant expression of the U2 snRNA proteins, assembling them into a 35 kbp plasmid containing the genes for the U2 snRNA, seven Sm proteins and ten U2 associated proteins. I investigated expression conditions for the entire snRNP, as well as tested expression of sub-complexes and individual proteins. This allowed me to use the

recombinantly expressed proteins to investigate *C. merolae*'s recognition of the branch-point sequence, and to further understand the association of the U6 snRNA to the U2 snRNA.

Chapter 2
Co-expression of all U2 snRNPs and SF3a proteins

U2 snRNA in *C. merolae* is associated with ten snRNP-specific proteins and seven Sm proteins, making U2 snRNP the most complex particle for splicing (Table 2). SF3a is one of the multi-protein complexes that associates with the U2 snRNA and contains three subunits including Prp9, Prp11, and Prp21. The SF3a complex is crucial for initiation of the first step of the splicing reaction that occurs when SF3a dissociates from the spliceosome. Moreover, SF3a is required for the spliceosome assembly. X-ray crystallography of SF3a revealed that there is no interaction between Prp9 and Prp11. In fact, Prp21 acts as a bridge to connect Prp11 and Prp9 in *Saccharomyces cerevisiae* (Lin *et al*. 2012).

To characterize the function, structure, and interaction of the U2 snRNA proteins and the SF3a protein complex in *C. merolae*, it is necessary to develop an *in vitro* reconstitution system. A challenging but key step towards this approach is to express seventeen recombinant proteins and purify all of them. No-one has ever biochemically expressed or purified the U2 snRNPs from *C. merolae* with the purpose of reconstituting the U2 snRNPs. In order to carry out biochemical experiments on the U2 snRNPs, I attempted to express U2 snRNPs and SF3a recombinantly. Romier *et al*. (2006) state that co-expression of protein complexes seems more stable and prevents proteins from degrading or aggregating when compared to the expression of a single protein. Moreover, some proteins rely on each other to function properly. For instance, in *Saccharomyces cerevisiae*, Cus2 protein, which interacts with Prp11, has a role in recruiting SF3a to the U2 snRNP (Yan *et al*. 1998).

Two plasmids were generated; the first plasmid (35 kbp) contains the gene for the U2 snRNA, ten U2-associated proteins and seven Sm proteins. The second plasmid (8.2 kbp) is for recombinant expression of the SF3a multi-protein complexes. In order to do *in vitro* reconstitution of the U2 snRNP proteins, their respective genes were obtained through a polymerase chain reaction (PCR) using the *C. merolae* genomic DNA. Those genes were introduced into an expression vector by a ligation independent cloning technique. I expect these recombinant snRNPs to be useful for the mechanistic and structural studies of splicing in *C. merolae*.

Table 2. The U2 snRNA and all U2 snRNP genes with their associated accession numbers, masses, and lengths (Stark *et al*. 2015).

U2 snRNP	Accession number	Mass (kDa)	Length (base pair)
Prp9	CMQ406C	60.28	1617
Prp11	CMN095C	19.09	504
Prp21	CMJ300C	50.36	1361
Hsh155	CMB002C	104.04	2864
Cus1	CMT357C	29.09	771
Rse1	CML103C	179.36	5036
Hsh49	CME063C	14.2	399
Rds3	CMS014C	14.04	393
Prp5	CMR433C	112.57	3084
Prp43	CMM048C	83.04	2178
U2 snRNA	AP006493		131
SmB/B’	CMK022C	8.85	243
SmD1	CMF084C	15.22	405
SmD2	CMN302C	36.29	999
SmD3	CMM065C	18.83	513
*SmE	CMM109C	11.93	318
	CMH215C		
SmF	CMQ171C	10	273
SmG	CMO342C	10.68	303

*SmE proteins are encoded by CMM109C and CMH215C genes and the SmE proteins will vary by only one amino acid.

2.1 Materials and Methods

2.1.1 Preparation of *C. merolae* genomic DNA

I used *C. merolae* 10D strain (NIES-1332) provided by the Microbial Culture Collection at the National Institute for Environmental Studies in Tsukuba, Japan (mcc.nies.go.jp/). *C. merolae* genomic DNA was prepared by Martha Stark as previously described (Stark *et al*. 2015).

2.1.2 Construction of Co-expression Vectors containing U2 snRNP genes and SF3a genes

Amplification of all U2 snRNP genes was performed by using a polymerase chain reaction from one microgram of *C. merolae* genomic DNA. Gene specific primers (synthesized by Eurofins Genomics), which were employed for amplification of these genes, had a BamHI restriction site in the forward primer and a NotI restriction site in the reverse primer (Table 3).

The ligation independent cloning technique was used for directional cloning of amplified U2 snRNP (cloning of the Sms genes in the pQlink vector was performed by Fatimat Shidi) and SF3a gene products. This method creates long cohesive sticky ends on both the vector and the insert that results in the generation of desired molecules after annealing the insert and the vector. All U2 snRNP and SF3a genes were combined into a polycistronic system, which means each gene contains its own promoter, ribosome binding site and stop codon.

After the target was amplified by PCR, the PCR products were treated with T4 DNA polymerase. First, 50 ng of an amplified gene were treated with 1 µL of T4 DNA polymerase (3′ to 5′ exonuclease activity) in the presence of 25 mM dCTP (T4 DNA polymerase degrades the DNA from the 3′ end until it encounters to the first dC residue). The pQlinkN^{mod} was digested with the EcoRI and HindIII site by using 10 U PmlI /µL (three hours digestion without gel purification) and 60 ng of this linearized co-expression vector. Next, the pQlinkN^{mod} was treated with 1 µL of T4 DNA polymerase in the presence of 25 mM dGTP (T4 DNA polymerase removes nucleotides until it encounters to the first dG residue) (Kim. 2007), 100 mM DTT, and 2 µL of 10 x T4 DNA polymerase buffer. The pQlinkN^{mod} was produced by Liz Dunn (fragment between EcoRI and BamHI was removed from original pQlink) (Addgene plasmid 13670; Scheich *et al*. 2007) and cloned this part with an oligo duplex, oSDR1084/1085. Each of the T4 DNA polymerase treatment reactions was incubated at room temperature for 30 minutes separately. These two reactions were combined together and heated at 70 C for 20 minutes, then allowed to sit at room temperature for five minutes to anneal each U2 snRNP and SF3a gene into the co-expression pQlinkN^{mod} vector. Each of the U2 snRNP and SF3a genes was sequenced by the UNBC genetic facility by using oSDR1084 and oSDR1085 primers. For better sequencing of extensive genes (more than 800 nucleotides), in addition to using oSDR1084 and oSDR1085 primers, gene specific primers were also used (Table 3). The 4Peaks software was used to confirm the results of sequencing (http://nucleobytes.com/index.php/4peaks).

Table 3. DNA oligonucleotides which were used to amplify and sequence the desired gene. The first DNA oligonucleotide for each gene is the forward and the second DNA oligonucleotide in the reverse primer. TACTTCCAATCCCACGAGGAGAAATTAACT is the LIC site for the forward primer and TTATCCACTTCCCACG is the LIC site for the reverse primer.

oSDR#	**U2 snRNA Genes**	**Sequence**
oSDR1105	Cm Prp9	TACTTCCAATCCCACGAGGAGAAATTAACTATGAACGCGCTCACC
oSDR1106	Cm Prp9	TTATCCACTTCCCACGTTACGTACGACTGGATCC
oSDR1107	Cm Prp11	TACTTCCAATCCCACGAGGAGAAATTAACTATGGACGCCTACGCA
oSDR1108	Cm Prp11	TTATCCACTTCCCACGCTAGTCTCTTCGCACATAC
oSDR1109	Cm Hsh155	TACTTCCAATCCCACGAGGAGAAATTAACTATGGAAGCGTTGGAAAG
oSDR1110	Cm Hsh155	TTATCCACTTCCCACGTTACAGAGTTGCGGC
oSDR1111	Cm Cus1	TACTTCCAATCCCACGAGGAGAAATTAACTATGGTGGAAGATTCAGAC
oSDR1112	Cm Cus1	TTA TCC ACT TCC CAC G TCA CGG GGG GTC TG
oSDR1113	Cm Rse1	TACTTCCAATCCCACGAGGAGAAATTAACTATGGCGGAGCCGCTG
oSDR1114	Cm Rse1	TTATCCACTTCCCACGTTACGTCGCATGCAACG
oSDR1115	Cm Rds3	TACTTCCAATCCCACGAGGAGAAATTAACTATGTCCCTGAAACATACG
oSDR1116	Cm Rds3	TTATCCACTTCCCACGTTACAAGTCACCAGCG
oSDR1117	Cm Prp5	TACTTCCAATCCCACGAGGAGAAATTAACTATGGAGCAACTGGGAAT
oSDR1118	Cm Prp5	TTATCCACTTCCCACGCTAGAACATCGTCTCTTCG
oSDR1119	Cm Prp43	TACTTCCAATCCCACGAGGAGAAATTAACTATGACGCATGTACCAAGC
oSDR1120	Cm Prp43	TTATCCACTTCCCACGTCACGGTGCCGCGGA
oSDR1082	Cm Hsh49	TACTTCCAATCCCACGAGGAGAAATTAACTATGAACGGCTCTGGGCTAGG
oSDR1083	Cm Hsh49	TTATCCACTTCCCACGTCACTTCTCTGAGAGCTGCTGCAC
oSDR1246	Cm Prp21	TACTTCCAATCCCACGAGGAGAAATTAACTATGGAGCCGCAGACG
oSDR1247	Cm Prp21	TTATCCACTTCCCACGCTATCGTGTGTGATGACCGTTAAAC
oSDR1084	pQLink	CCATTTGTCG AGAAATCATA AAAAATTTAT TTGCTTTGTG
oSDR1085	pQLink	CAACCG AGCGTTCTGA ACAAATCCAG

Ligation independent cloning was used to construct a co-expression vector containing all the U2 snRNP and SF3a genes by using the pQlinkH (Addgene plasmid 13667; Scheich *et al*. 2007) and the pQlinkN^{mod} vectors (Scheich *et al*. 2007; Figure 8). Plasmid names are listed in Table 4. All the vectors containing single genes of the U2 snRNP were combined into a single co-expression vector. In other words, individual genes were combined serially until they were all present in a single plasmid (Figure 9 and Figure 10). For this purpose, a pQlink vector with the first gene of interest was digested with the SwaI restriction enzyme, while the PacI restriction enzyme was used for the digestion of the other pQlink vector containing the second gene of interest. The pQlink consists of three restriction sites, two for PacI and one for SwaI. The PacI restriction sites are located at the 3′ of the

SwaI site and 5′ of the transcriptional promoter of the inserted gene. The SwaI restriction site is at the 3′ of the transcriptional terminator for the inserted gene.

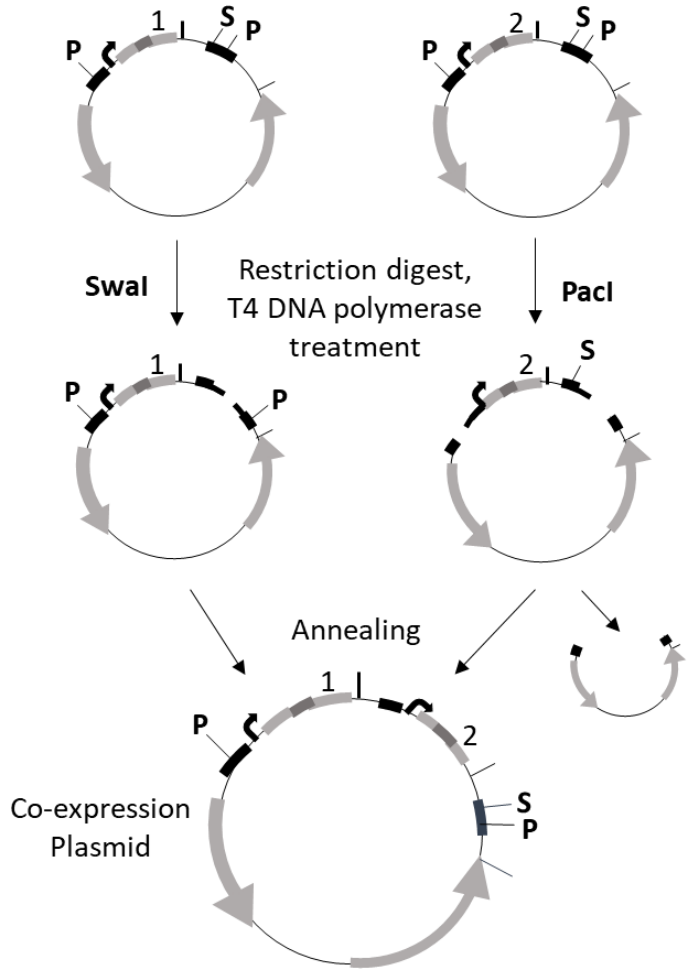


Figure 8. Construction of co-expression plasmid by using Ligation Independent Cloning technique. The map shows two plasmids with two different inserts, 1 and 2, S=SwaI and P=PacI (Scheich *et al*. 2007).

Table 4. U2snRNP and SF3a plasmids which were used to make the related co-expression vector.

Plasmid name	Backbone	Gene
pSR648	pQlinkN pSR617	Cm Rds3
pSR653	pQlinkN pSR617	Cm Prp11
pSR658	pQlinkN pSR617	Cm Prp9
pSR659	pQlinkN pSR617	Cm Rse1
pSR635	pQlinkN pSR617	Cm Hsh49
pSR646	pQlinkN pSR617	Cm Hsh155
pSR652	pQlinkN pSR617	Cm Cus1
pSR654	pQlinkN pSR617	Cm Prp5
pSR657	pQlinkN pSR617	Cm Prp43
pSR710	pQlinkN pSR617	Cm Prp9/Prp11
pSR730	pQlinkN pSR617	Cm Prp21
pSR711	pQlinkN pSR617	Cm Rse1/Hsh49
pSR713	pQlinkN pSR617	Cm Hsh155/Cus1
pSR757	pQlinkN pSR617	Cm Prp9/Prp11+ Prp21
pSR737	pQlinkN pSR617	Cm Hsh155/Cus1+ Rse1/Hsh49
pSR738	pQlinkN pSR617	Cm Prp43/Prp5
pSR760	pQlinkN pSR617	Cm Prp5/Prp43 + Prp9/Prp11/Prp21
pSR763	pQlinkN pSR617	Cm Hsh155/Cus1+ Rse1/Hsh49 + Cm Prp5/Prp43 + Prp9/Prp11/Prp21
pSR752	pQlinkH pSR627	Cm His-SmF/E3/G/D3/B/D1/D2
pSR754	pQlinkH pSR627	Cm His-SmF/E3/G/D3/B/D1/D2 + U2
pSR764	pQlinkH pSR627	Cm Hsh155/Cus1/ Rse1/Hsh49/Prp5/Prp43/Prp9/Prp11/Prp21+ Cm His-SmF/E3/G/D3/B/D1/D2/U2
pSR765	pQlinkH pSR627	Cm Hsh155/Cus1/ Rse1/Hsh49/Prp5/Prp43/Prp9/Prp11/Prp21+Rds3/ Cm His-SmF/E3/G/D3/B/D1/D2/U2
pSR793	pQlinkH pSR627	Cm Prp21
pSR795	pQlinkH pSR627	Cm Prp9/Prp11+Prp21

In summary, 500 ng of DNA plasmid was digested with 5 µL of SwaI or 5 µL of PacI restriction enzyme. BSA was used in each of SwaI and PacI digestion reaction. Incubation of PacI and SwaI digestion reaction was performed at 37 C and 25 C respectively for 3 hours. Afterward, heat inactivation of the enzyme was carried out at 75 C for 20 minutes. Next, the digested products were treated with 0.5 µL of T4 DNA polymerase in a 20 µL mixture reaction containing 50 mM Tris-HCl pH 8.0, 10 mM $MgCl_2$, 100 ng/µL BSA, 0.1 M DTT and 0.5 µL of each 50 mM dCTP and 50 mM dGTP for PacI and SwaI digests respectively. SwaI and PacI plasmids were mixed and were heated at 65 C for five minutes

then the reaction was allowed to cool at room temperature. Two µL of 25 mM EDTA was added and 5 µL of the reaction was transformed into 50 µL of $RbCl_2$ competent DH5α cells. Results of transformation were screened by performing colony PCR using gene specific forward and reverse primers to confirm that all of the U2 snRNP genes from the PacI and the SwaI-digested plasmid were present in the same pQlink expression vector. Restriction enzymes were used to ensure that the SF3a construct contained all three genes (Prp9, Prp11 and Prp21) in the pQlink expression vector. The ligation independent cloning was repeated to combine all individual genes into the same expression plasmid. This was done separately for the U2 snRNP and SF3a genes (Figure 9 and Figure 10).

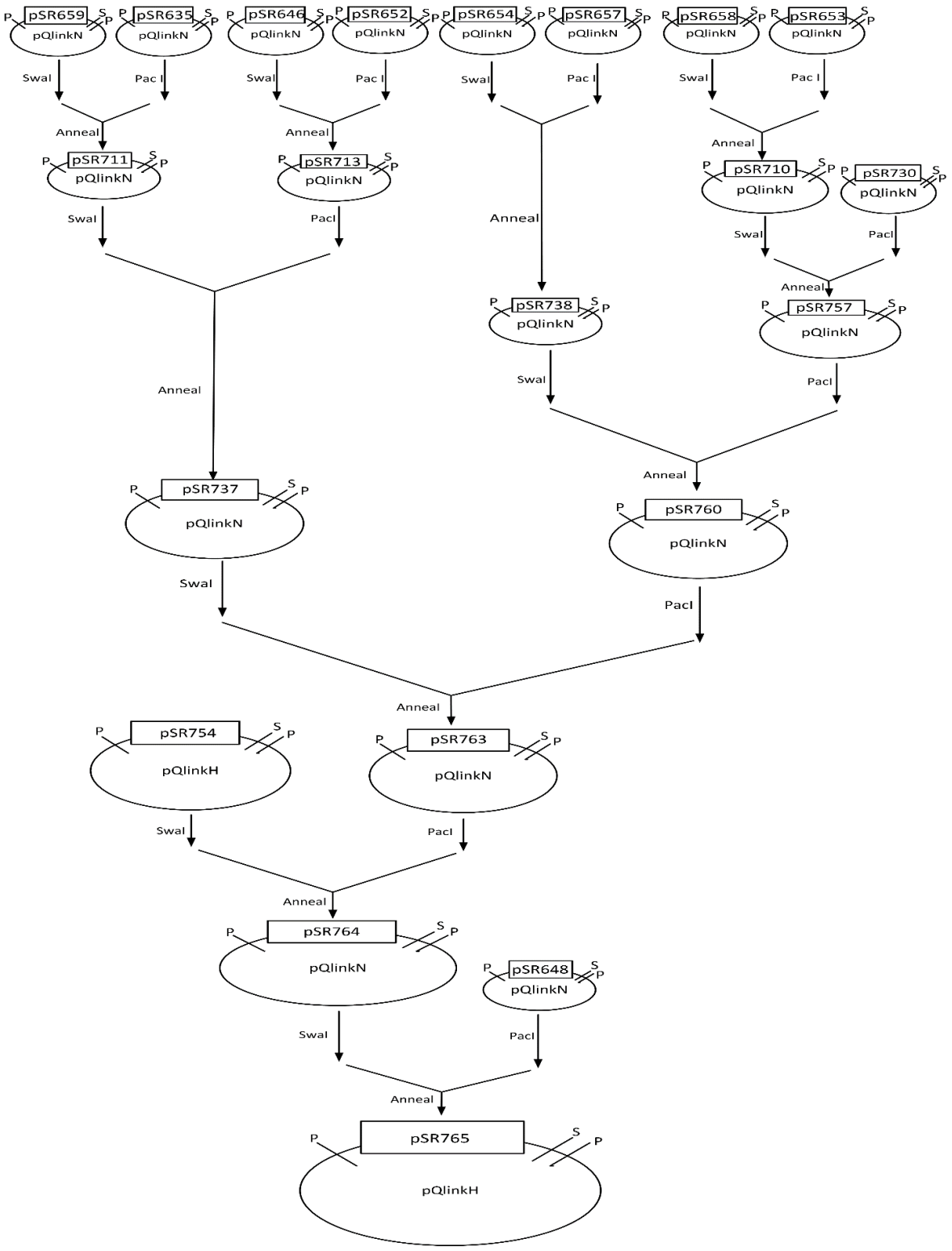


Figure 9. Construction of the U2 snRNP co-expression vectors. Each gene of the U2 snRNP was first introduced into a single plasmid and then all genes were combined serially into plasmids until they were present in one final plasmid. Plasmids were digested with SwaI (S) or PacI (P), followed by heat inactivation, T4 DNA polymerase treatment and annealing.

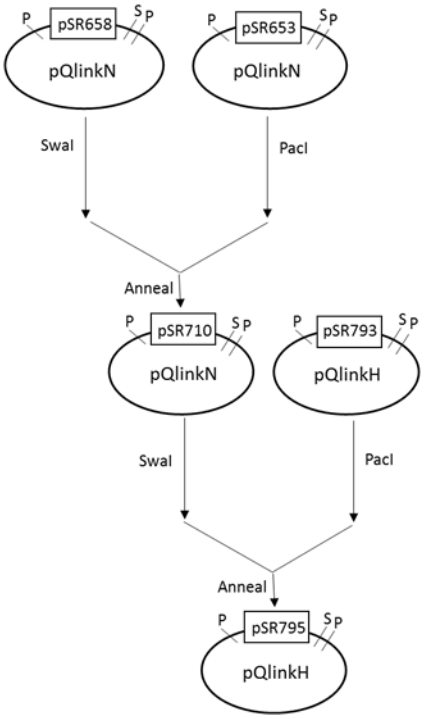


Figure 10. Construction of the SF3a co-expression vectors. Individual SF3a genes were first introduced into a single plasmid and then all genes were combined serially into plasmids until they were present in one final plasmid. Plasmids were digested with SwaI (S) or PacI (P), followed by heat inactivation, T4 DNA polymerase treatment and annealing.

2.1.3 Co-expression and Purification of the U2 snRNPs:

Rosetta pLysS cells were used for co-expression of both the U2 snRNP and SF3a complex by growing in Luria Bertani media. 25 mg/mL ampicillin and 25 mg/mL chloramphenicol were used to select for the pQlink vectors and for the Rosetta pLysS plasmids, respectively. The pQlink carries both of the genes and the Rosetta pLysS plasmid contains the genes for the synthesis of rare tRNA. I did small-scale co-expression of the U2 snRNP and SF3a complex. A 3 mL starter culture was grown at 37 C in a shaker rotating at a speed of 200 rpm to reach an $OD_{600\sim}$ 0.6 value. The starter culture was then used to inoculate a 22 mL culture and grown at 37 C, 200 rpm to reach the OD_{600} 0.6–0.8 range; at this point an induction was done by adding 1 mM isopropyl 1-thioβ-D-galactopyranoside (IPTG)

(Amresco; solon, Ohio) to the culture. The cells were left in the 16 C incubator for 12–18 hours afterward. Cells were harvested and stored at -80 C for further analysis.

A 25 mL cell pellet was resuspended in 500 mL of lysis buffer (20 mM HEPES-NaOH pH 7.5, 500 mM NaCl, 5 mM β-Mercaptoethanol and 20 mM Imidazole) to lyse the cells. The cell lysate was sonicated four times, in ten-second bursts, at five to seven Watts, and with ten second pauses on ice in between. To separate the soluble and insoluble fractions, the lysate was centrifuged at 13000 x g in the Beckman Coulter Allegra X-12R centrifuge for ten minutes at 4 C. The cells which appeared in supernatant are soluble and the cells which precipitated as a pellet are insoluble. The supernatant and pellet were run on a gel to check the solubility of the protein expression.

In order to do small scale protein purification, the clear lysate was passed through a Ni^{2+} resin which the histidine tagged protein will bind to. It is expected that other proteins of interest would co-purify via the connection between the Ni^{2+} resin and the tagged SmF for the U2 snRNP and Prp21 for the SF3a purification. Afterwards, a wash step was done to remove all bacterial proteins in the expression system from the bound proteins. Finally, the bound proteins were eluted.

For the purification of the U2 snRNA proteins and SF3a protein complex, a batch method (Thermo Scientific) were used. The supernatant was loaded on Pierce spin cups (Thermo Scientific #69700), which were equilibrated with two resin bed volumes of wash buffer (20 mM HEPES-NaOH pH 7.5, 500 mM NaCl, 5 mM β-Mercaptoethanol and 20 mM Imidazole) by resuspending the resin with wash buffer followed by a two minute spin at 700 x g in the Eppendorf 5417c centrifuge. Excess buffer was discarded and the lysate was then

mixed with resin and incubated for 30 minutes on an end-over-end rotator. The lysate was centrifuged as above and the lysate removed. The resin was washed twice with two resin-bed volumes of wash buffer and centrifuged for two minutes at 700 x g, and the excess buffer was removed. The U2 snRNA proteins and SF3a protein complex were eluted by using one resin bed volume of elution buffer (20 mM HEPES-NaOH pH 7.5, 500 mM NaCl, 5 mM β-Mercaptoethanol and 500 mM Imidazole) by repeating this step twice, and saving each supernatant fraction in a separate tube. The eluted proteins were analyzed by SDS-PAGE (12% high TEMED gel).

2.2 Results and Discussion

In order to develop *in vitro* reconstitution, all genes were cloned into the same pQlink expression vector. Results of transformation were screened by performing colony PCR using gene specific forward and reverse primers to confirm that all the U2 snRNP genes from the PacI and the SwaI-digested plasmid were present in the same pQlink expression vector (Table 4, Figure 11–Figure 14). Restriction enzyme digestion showed that all three genes of the SF3a construct were present in the pQlink expression vector as well. To see the difference between the size of digested and undigested plasmids, the 0.7% agarose ethidium bromide gel was run for different length of time (Figure 15).

In pQlink expression vectors, histidine tags were placed at the N-terminus of SmF and Prp21 genes located in the U2 snRNP and SF3a constructs, respectively. The six-histidine residues, were tagged with a protein of interest to facilitate the purification of the protein (Schieich *et al*. 2007).

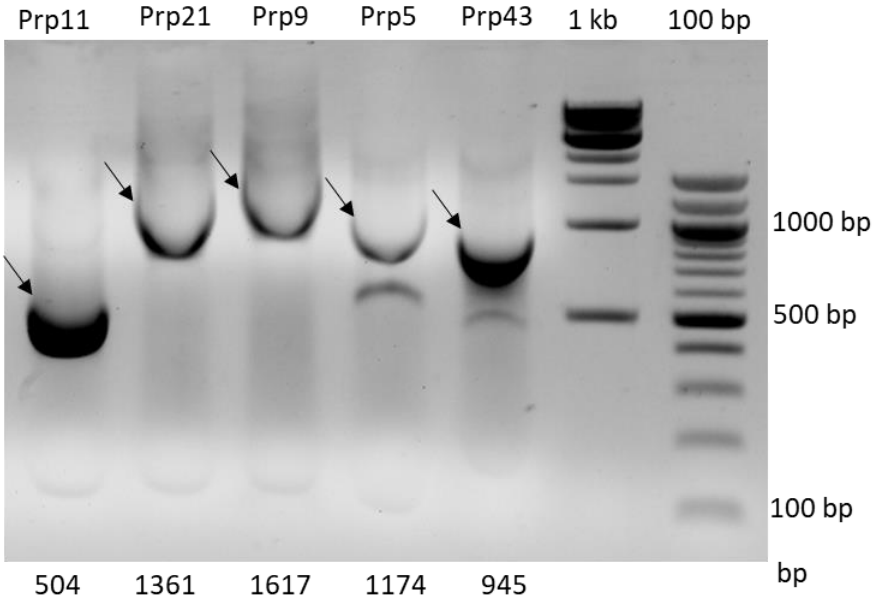


Figure 11. Determining the presence of Prp11, Prp21, Prp9, Prp5 and Prp43 genes in pQlink by PCR. PCR products were run on a 1% Agarose Ethidium Bromide gel. Gene names are listed above each lane with the expected size of the genes, which are listed below the lanes. Arrows indicate the presence of each gene in the pQlink vector. The 100 bp and 1 kb ladders are indicated on the right.

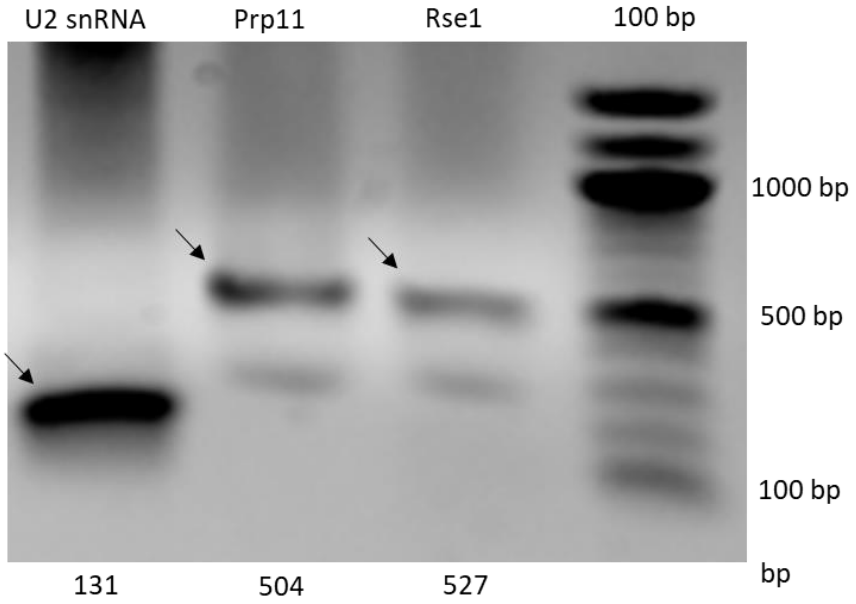


Figure 12. Determining the presence of U2 snRNA, Prp11 and Rse1 genes in pQlink by PCR. PCR products were run on a 1% Agarose Ethidium Bromide gel. Gene names are listed above each lane with the expected size of the genes, which are listed below the lanes. Arrows indicate the presence of each gene in the pQlink vector. The 100 bp ladder is indicated on the left.

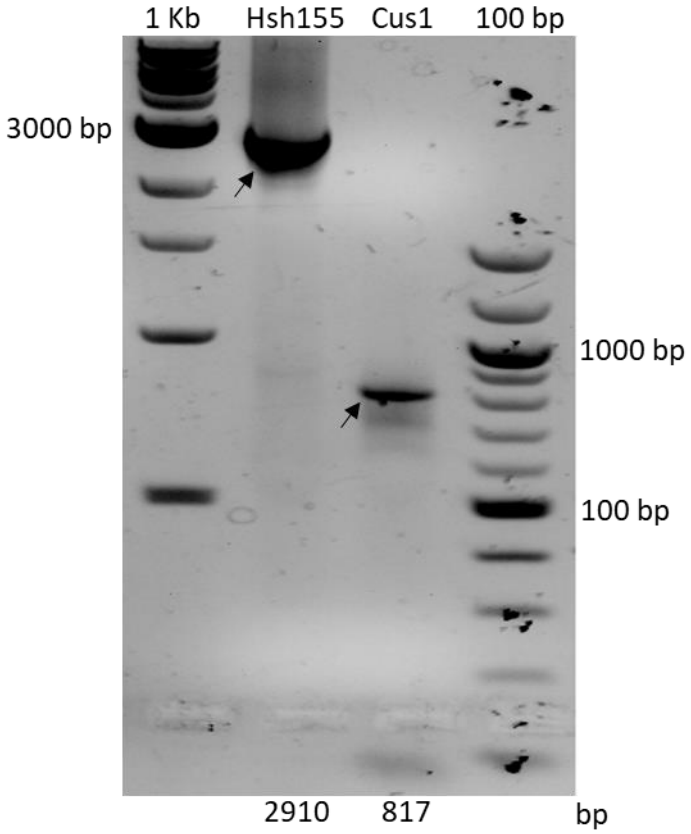


Figure 13. Determining the presence of Hsh155 and Cus1 genes in pQlink by PCR. PCR products were run on a 1% Agarose Ethidium Bromide gel. Gene names are listed above each lane with the expected size of genes, which are listed below the lanes. Arrows indicate the presence of each gene in the pQlink vector. The 1 kb and 100 bp ladders are indicated on the left and right of the gel respectively.

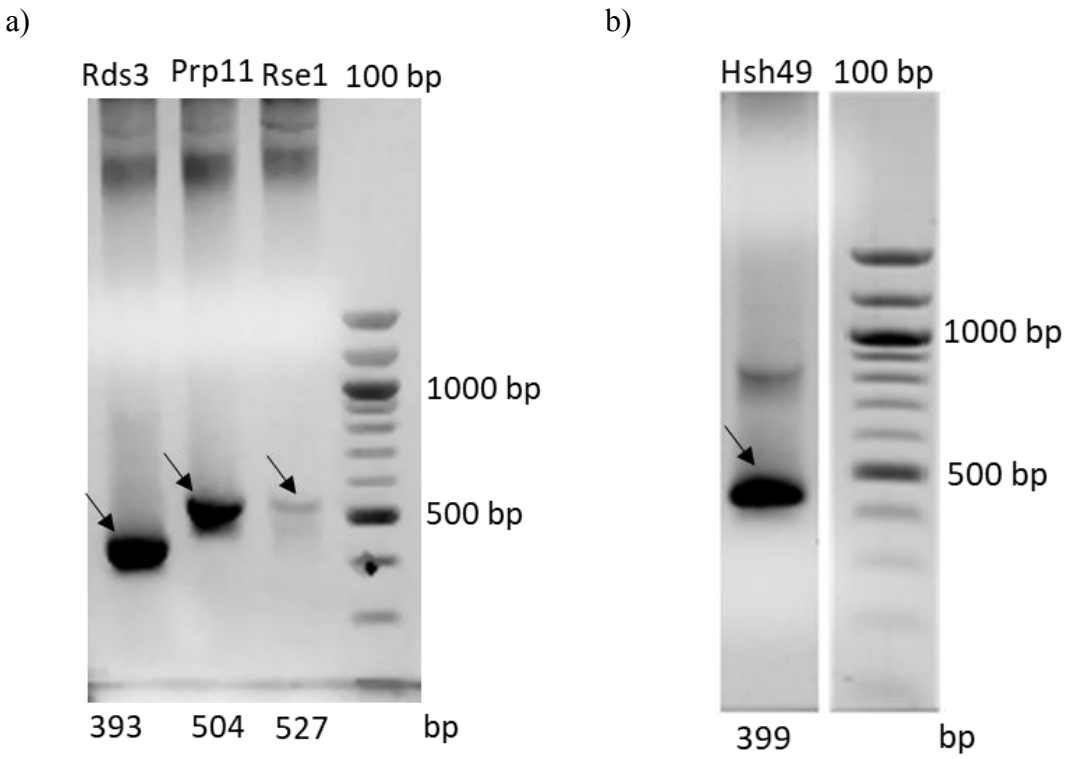


Figure 14. Determining the presence of Rds3, Prp11, Rse1 and Hsh49 genes in pQlink by PCR. Part (a) and Part (b) show the PCR products that were run on a 1% Agarose Ethidium Bromide gel. Gene names are listed above each lane with the expected size of genes, which are listed below the lanes. Arrows indicate the presence of each gene in the pQlink vector. The 100 bp ladder is indicated on the right of each gel respectively.

a) b)

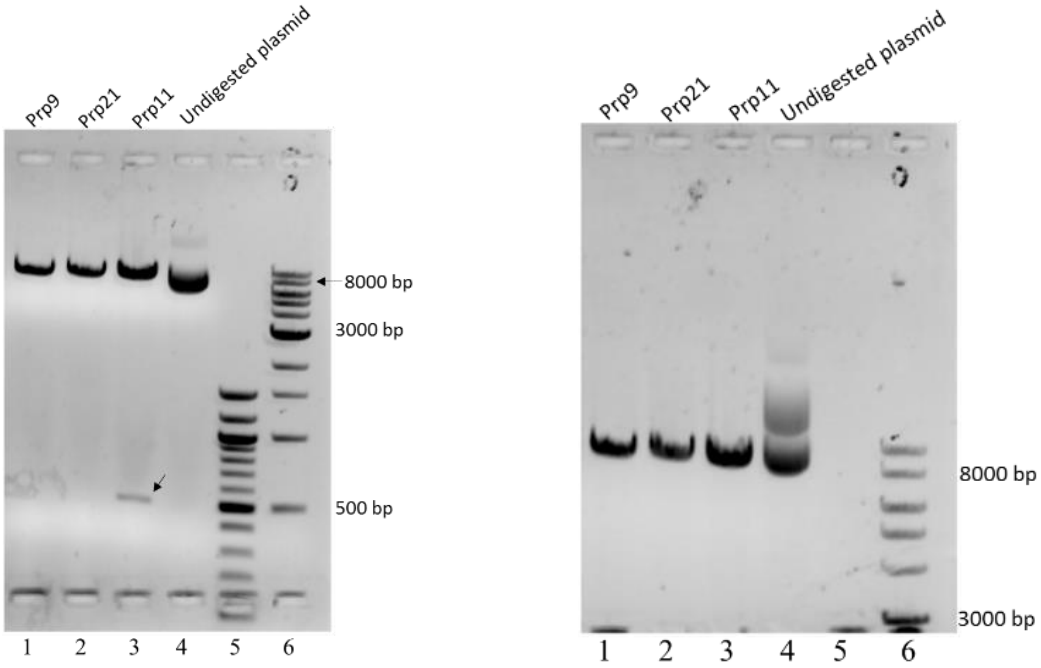


Figure 15. Determining the presence of each SF3a gene in pQlink. The same 0.7% agarose ethidium bromide gel indicating the presence of each gene in pQlink. The figure indicates the progress of the gel after 50 minutes (Part a) and two hours (Part b). In part a, the arrow shows the expected size of the Prp11. Prp9 and Prp21 were digested with BstBI and KpnI respectively (Lane 1 and lane 2). Also in Lane 3 of part a, NruI has two locations both in Prp9 and Prp11 resulting in two fragments that are 500 bp and ~10 Kb. Lane 4 shows the undigested SF3a plasmid and its size is 8938 bp.

12% high TEMED SDS-PAGE was used to analyze the expression and purification of the recombinant proteins. Based on Figure 16, which relates to the Sms and U2 snRNP construct, one band was observed between 6.5 and 14.4 kDa (Lane 3). This band could indicate SmF which was tagged with histidine and eluted after purification. Since the sizes of SmE and SmG are close to the size of SmF, these proteins can co-migrate in the gel and the band can designate these proteins as well (Table 2). Also, several bands were seen in the eluted fraction (Lane 7) after purification of the U2 snRNP proteins, which might relate to

Cus1, and Prp11. In order to confirm that these proteins were purified, I decided to conduct large-scale expression and purification of these proteins.

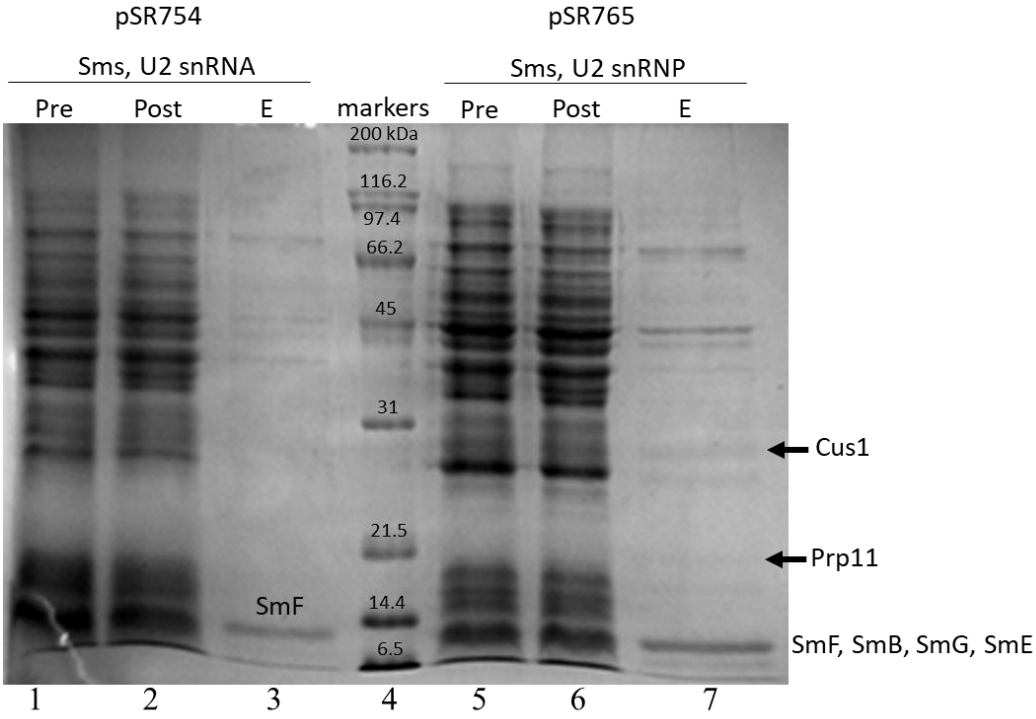


Figure 16. Small scale co-expression and purification of Sms-U2 snRNA and Sms-U2 snRNP constructs. 12% high TEMED SDS gel, showing (Pre) pre-induction, (Post) post-induction and (E) elution of Ni-NTA resin purification for each sample. Protein sizes are as follows: SmF (10 kDa), SmB (8.85 kDa), SmG (10.68 kDa), SmE (11.93 kDa), Hsh155 (104.04 kDa), Cus1 (29.09 kDa) and Prp11 (19.09 kDa).

In the large-scale expression of the U2 snRNP as shown in Figure 17, one band was seen between 6.5 and 14.4 kDa in all eluted fractions: Lane 3 (Sms construct), Lane 6 (Sms, U2 snRNA construct), and Lane 9 (Sms-U2 snRNA and U2 snRNA proteins). These bands could indicate SmF (His-tagged protein), SmG and SmE. However, it seems that other U2 snRNP proteins did not express in Rosetta pLysS since no differences in intensity between the post-induced cell extract lane and pre-induced cell extract lane were observed. Two other possibilities are either the induction was weak or the presence of too many bands of other

proteins make it hard to see the changes in both the post-induced and pre-induced cell extract lanes.

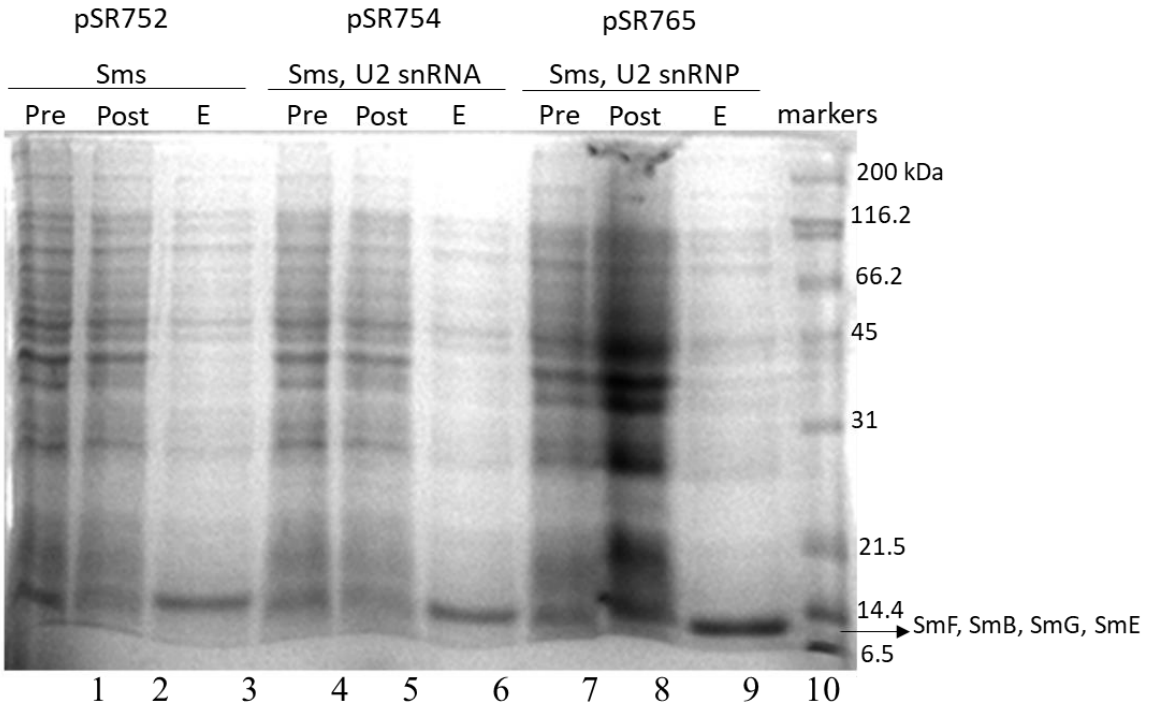


Figure 17. Large-scale co-expression and purification of Sms, Sms-U2 snRNA and Sms-U2 snRNP constructs. 12% high TEMED SDS gel, showing (Pre) pre-induction, (Post) post-induction and (E) elution of Ni-NTA resin purification for each sample. Protein sizes are as follows: SmF (10 kDa), SmB (8.85 kDa), SmG (10.68 kDa), SmE (11.93 kDa), Hsh155 (104.04 kDa), Cus1 (29.09 kDa) and Prp11 (19.09 kDa).

The co-expression of the SF3a complex was inspected in 12% high TEMED SDS-PAGE gel (Figure 18). By comparing pre-induced and post-induced cell extract lanes it seems that Prp9 and Prp21 proteins were expressed. There is a band around 21.5 kDa that could be Prp11, contaminant or an endogenous protein that was running the same size as the protein of interest. Separation of soluble and insoluble fractions by centrifuge demonstrates that the candidate bands for Prp9, Prp21 and Prp11 are all in the insoluble fraction.

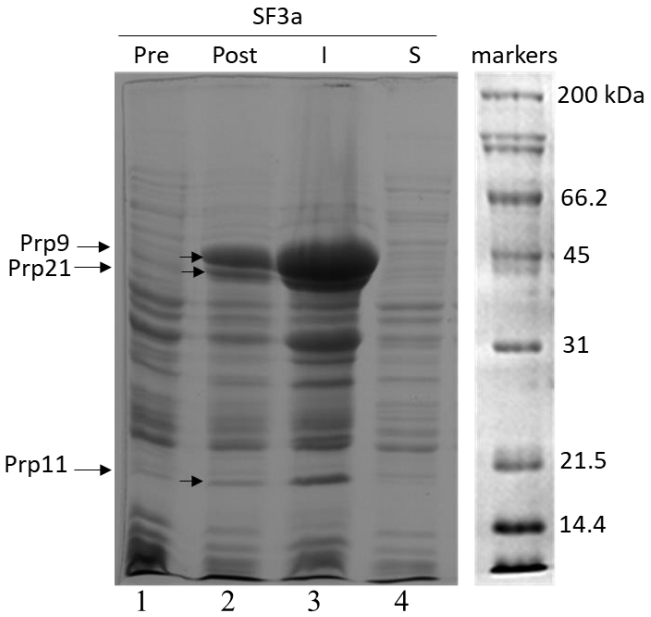


Figure 18. Co-expression of the SF3a from pQlink. 12% high TEMED SDS gel, showing (Pre) Pre-induction, (Post) Post-induction (I) insolubility, and (S) solubility of SF3a. The protein sizes are as follows: Protein sizes are as follows: Prp9 (60.28 kDa), Prp21 (50.36 kDa) and Prp11 (19.09 kDa). Prp21 is His-tagged (2.8 kDa). Arrows indicate the candidate bands for Prp9, Prp21 and Prp11 in the post induced lane and also demonstrates that SF3a appeared in the insoluble fraction.

In this study, I started to co-express the U2 snRNP proteins and the SF3a protein complex with the hope that through co-expression and purification of these proteins, I can get a better understanding of the function and the structure of these proteins. Unfortunately, as shown previously from co-expression of the U2 snRNP proteins, I could not express and purify all proteins. In conclusion, the expression strategy for the U2 snRNP and SF3a proteins did not work. A possibility for the unexpected results is that perhaps proteins were not stable in the expression vector. This issue could be due to the lack of other U2 snRNA proteins or the lack of an efficient amount of snRNA that led to instability of the U2 snRNP construct. Another possibility is that since *C. merolae* grows at up to 45 C (Matsuzaki *et al*, 2004) and my induction was 37 C there could have been a loss of some proteins.

A co-expression technique aids the protein partners to fold better and increase the solubility of the recombinant expression protein (Romier *et al*. 2006). However, co-expression of the SF3a protein complex resulted in an insoluble particle. According to a study by Lin and Xu, *Saccharomyces cerevisiae* SF3a complex co-expressed in *E.coli* (BL21 strain) successfully. GST tag was used as a fusion tag for the expression of Prp9, while Prp11 expressed as a poly-His-tagged fusion protein (Lin *et al*. 2012). The GST tag helps protein to fold properly and increase the solubility of the recombinant proteins. It is possible that using His-tag as a fusion tag with Prp21 could result in the insolubility of the SF3a complex in my project. As co-expression and purification of the U2 snRNP and the SF3a complex proteins were unsuccessful, I decided to express and purify the U2 snRNP proteins individually for the biochemical and functional analyses.

Chapter 3
Attempted individual expression and purification of Prp9, Prp11, Prp21 and Rds3 proteins

As a result of the unsuccessful co-expression and purification of the U2 snRNP proteins and the SF3a protein complex, as discussed in the previous chapter, I decided to express the SF3a protein and Rds3 protein individually. Rds3 is a subunit of the SF3b protein complex that interacts with the other four subunits of SF3b in yeast: Cus1p, Hsh49p, Hsh155p, Rds3p, and Ist3p/Snu17p. Rds3p is an essential factor for the stability of the spliceosome. The absence of Rds3p prevents spliceosome formation and blocks U2 snRNP addition (Wang and Rymond. 2003).

I have generated separate plasmids for each gene of the SF3a proteins and Rds3 protein. A polymerase chain reaction was performed to obtain the genes of these proteins from *C. merolae* genomic DNA. Using the ligation independent cloning method, those genes were introduced in two different expression vectors: pQlink and PMCSG23 plasmids. PMCSG23 was derived from the PMCSG21 backbone and this vector contains histidine tags (His_6-tag), a maltose-binding protein tag (MBP-tag), and sequences that encode the tobacco etch virus (TEV) protease cleavage site. The MBP tag was used for further purification of protein and it increases the solubility of the protein. The TEV cleavage site facilitates protein purification. Since co-expression of the SF3a protein complex resulted in an insoluble particle, having the MBP fusion protein may aid to obtain soluble protein expression. I tried different conditions in order to express these proteins individually.

3.1 Material and Methods

3.1.1 Preparation of *C. merolae* genomic DNA

I used *C. merolae* 10D strain (NIES-1332) provided by the Microbial Culture Collection at the National Institute for Environmental Studies in Tsukuba, Japan (mcc.nies.go.jp/). *C. merolae* genomic DNA was prepared by Martha Stark as previously described (Stark *et al*. 2015).

3.1.2 Construction of expression Vectors containing Prp9, Prp11, Prp21, Rds3 and Hsh49 genes

Amplification of the SF3a, Rds3 and Hsh49 genes were performed by PCR from one microgram of *C. merolae* genomic DNA. Gene specific primers (prepared by Eurofins Genomics) were employed for amplification of these genes (Table 3).

In order to construct an expression vector, each gene of the SF3a subunits and Rds3 was cloned separately into two different expression vectors: pQlink and PMCSG23 (Table 6). The PMCSG23 expression vector containing the Hsh49 gene was constructed using the same procedure that was employed for Rds3.

This was accomplished using the ligation independent cloning into the PMCSG23 expression vector which will add a MBP and a six-histidine tag at the N-terminus of the SF3a subunits, Rds3 and Hsh49 (Chapter 2). PMCSG23 was derived from the PMCSG21 backbone and this vector contains six histidine tags, maltose-binding protein (MBP), and sequences that encode the tobacco etch virus (TEV) protease cleavage site. The MBP tag was

used for further protein purification, which increases the solubility of the protein, and the TEV cleavage site facilitates protein purification.

Before cloning the genes, 10 µg of the PMCSG23 vector was digested with 5µL SspI (100 U) in 100 µL volume by incubation at 37 C for six hours. Five µL of annealing reaction was transformed into 50 µL of $RbCl_2$ competent DH5α cells. Results of the transformation were inspected via sequencing. Sequencing of each gene was done by the UNBC genetic facility and the results of sequencing were analyzed by 4Peaks software (http://nucleobytes.com/index.php/4peaks). Gene specific primers along with oSDR1283 and oSDR1284 primers were used for the sequencing of the Prp9, Prp11, Prp21, Rds3 and Hsh49 in the PMCSG23 vector (Table 5). Gene specific primers and oSDR1084 and oSDR1085 primers were also recruited for the sequencing of these genes in the pQlink vector (Table 4).

Table 5. DNA oligonucleotides which were used to amplify and sequence the desired gene. The first DNA oligonucleotide for each gene that are listed, is the forward and the second DNA oligonucleotide is the reverse primer.

oSDR#	U2 snRNA Genes	Sequence
oSDR1319	Cm Prp9	TACTTCCAATCCAATGCAAACGCGCTCACCGC
oSDR1320	Cm Prp9	TTATCCACTTCCAATGTTACGTACGACTGGATCCAG
oSDR1129	*Cm Prp9	CAGAGTGTTTAGATGCC
oSDR1321	Cm Prp11	TACTTCCAATCCAATGCAGACGCCTACGCACTCAAAAAC
oSDR1322	Cm Prp11	TTATCCACTTCCAATGCTAGTCTCTTCGCACATACAG
oSDR1323	Cm Prp21	TACTTCCAATCCAATGCAGAGCCGCAGACGGTTC
oSDR1324	Cm Prp21	TTATCCACTTCCAATGCTATCGTGTGTGATGACCG
oSDR1223	Cm Rds3	TACTTCCAATCCAATGCATCCCTGAAACATACCG
oSDR1224	Cm Rds3	TTATCCACTTCCAATGTTACAAGTCACCAGCG
oSDR1294	Cm Hsh49	TACTTCCAATCCAATGCAAACGGCTCTGGGCTG
oSDR1295	Cm Hsh49	TTATCCACTTCCAATGTCACTTCTCTGAGAGCTGC
oSDR1283	pMCSG23	TGATCAACGCCGCCAGC
oSDR1284	pMCSG23	GCAGCGGTTTCTTTACC

*is the sequencing primer to read the Prp9 gene.

Table 6. SF3a, Rds3 and Hsh49 plasmids that were used to make the related expression vector.

Plasmid name	Backbone	Gene
pSR828	pMCSG23 pSR614	Cm Prp21
pSR827	pMCSG23 pSR614	Cm Prp11
pSR826	pMCSG23 pSR614	Cm Prp9
pSR766	pMCSG23 pSR614	Cm Rds3
pSR800	pMCSG23 pSR614	Cm Hsh49

3.1.3 Expression and purification of Prp9, Prp11, Prp21 and Rds3 proteins

The Rosetta pLysS cell was used for expression of the Prp9, Prp11, Prp21, and Rds3 by growing in two different media: Luria Bertani and 2XYT media. 50 mg/mL ampicillin and 25 mg/mL chloramphenicol were used to select for the pQlink vectors and for the Rosetta pLysS plasmids, respectively. 25 mg/mL spectinomycin was used to select for the PMCSG23 vector as well. Rosetta pLysS cell was used for the expression of the Hsh49 by

growing into 2XYT media. 25 mg/mL chloramphenicol was used to select for the Rosetta pLysS plasmids, and 25 mg/mL spectinomycin was used to select for the PMCSG23 vector.

The entire process of the expression of these proteins was performed as explained in the materials and methods section of Chapter 2. However, different strategies were applied for better expression of these proteins; for example, two different concentrations of isopropyl 1-thioβ-D-galactopyranoside (IPTG), 0.5 mM and 1 mM were used for induction, and different temperatures (16 C, 25 C, 37 C) were applied after induction. Once any of the proteins expressed well, I inoculated the starter culture to one liter and it was grown at 37 C, at 200 rpm to reach the OD_{600} 0.6–0.8 range. At this point an induction was done by adding 1 mM or 0.5 mM IPTG (VWR) to the culture. The cells were left at different temperatures (16 C, 25 C, 37 C) for 12–18 hours. Aside from using IPTG, auto-induction was performed for the expression of the SF3a subunits in the PMCSG23 expression vector. After transformation to the expression vector, a single colony was selected and grown in 5 mL MDG non-inducing media (Table 7). 25 mg/mL spectinomycin and 25 mg/mL chloramphenicol were added to the culture and it was incubated at 37 C for 24 hours. An overnight saturated culture containing OD_{600} ~5–10 was spun down and resuspended in 5 mL ZYM-5052 Auto-inducing media (Table 8).

Table 7. MDG non-inducing media.

Ingredients	**Final Concentration**
Na_2HPO_4	25 mM
KH_2PO_4	25 mM
NH_4Cl	50 mM
Na_2SO_4	5 mM
$MgSO_4$	2 mM
Glucose	50%
TRACE METALS (see below)	0.2x
Aspartic acid	0.25%
TRACE METALS	1000x
$FeCl_3$	50 mM
$CaCl_2$	20 mM
$MnCl_2$	10 mM
$ZnSO_4$	10 mM
$CoCl_2$	2 mM
$CuCl_2$	2 mM
$NiCl_2$	2 mM
Na_2MoO_4	2 mM
$NaSeO_3$	2 mM
H_3BO_3	2 mM
HCl	60 mM

Table 8. ZYM-5052 Auto-inducing media.

Ingredients	**Final Concentration**
N-Z-amine	1%
yeast extract	0.5%
$MgSO_4$	2 mM
TRACE METALS	1000x
M stock solution (see below)	50x
5052 stock solution (see below)	50x
M stock solution	50x
Na_2HPO4	1.25 M
KH_2PO4	1.25 M
Na_2SO4	0.25 M
NH_4Cl	2.5 M
5052 stock solution	50x
Glycerol	25%
Glucose	2.5%
alpha-lactose monohydrate	10%

The saturated culture was grown at 37 C for 24 hours. Cells were harvested and stored at -80 C for further analysis. A one liter cell pellet was resuspended in 10 mL of lysis buffer (20 mM HEPES-NaOH pH 7.5, 500 mM NaCl, 5 mM β-Mercaptoethanol and 20 mM Imidazole) containing half of a tablet of Roche Complete, EDTA-free protease inhibitor cocktail. The cell lysate was sonicated four times, in ten-second bursts, at five to seven W, and with ten second pauses on ice in between. The lysate was centrifuged at 25000 x g in a JA25.50 rotor (Beckman Coulter Avanti HP-20 XPI) centrifuge for 30 minutes at 4 C. One percent Streptomycin Sulfate was added to the lysate in order to precipitate some of the genomic DNA. The lysate was centrifuged as above to get clear lysate. The supernatant was filtered by passing through a 0.45 µm nylon syringe filter and was then passed over a 1 mL

His-Trap column. AKTA FPLC system with Unicorn software version 5.01 was used for the protein purification. Rds3 bound to the His-Trap column was washed with ten column volumes (10 mL) of buffer A (wash buffer: 20 mM HEPES-NaOH pH 7.5, 500 mM NaCl, and 20 mM Imidazole). Then it was eluted with 5 column volumes (25 mL) of buffer B (elution buffer: 20 mM HEPES-NaOH pH 7.5, 500 mM NaCl, and 500 mM Imidazole). The flow rate for both the washed and the elution steps was 0.5 mL/min over the Ni-NTA His-trap column. The fraction that was collected was 5 mL and it was quantified with a Nanodrop ND-1000 spectrophotometer. The eluted fraction was then passed over a 5 mL MBP Trap column and washed with 3 column volumes (15 mL) of buffer A (wash buffer: 20 mM HEPES-NaOH pH 7.5, and 500 mM NaCl). Then the fraction was eluted with 5 column volumes (25 mL) of buffer B (elution buffer: 20 mM HEPES-NaOH pH 7.5, 500 mM NaCl and 10 mM Maltose). This second purification step was performed for better purification of the Rds3 and SF3a proteins. The flow rate for purification of Rds3 over the MBP-Trap column was 1 mL/min. The purification steps were performed at 4 C in AKTA FPLC system. The collected fraction was 2.5 mL and it was incubated with one milligram of TEV protease per 50 mg of purified protein at 4 C for 18 hours. The TEV cleaved protein was dialyzed with 20 mM HEPES-NaOH pH 7.5, 250 mM NaCl. After this step, the removal of cleaved His tags, His_6-TEV Protease, or proteins of interest retaining His-tags was required. To do so, the dialyzed protein was passed through Ni^{2+} resin in a batch bind purification. This time the cleaved protein did not bind to the Ni^{2+} resin, but the His_6 tag and TEV protease did bind. An amylose-resin batch bind purification was carried out on the flow-through to get more purified protein. The samples were analyzed by SDS-PAGE.

The Hsh49 protein was purified using the same protocol explained as the Rds3 protein, described in the material and methods of this chapter. It is necessary to mention that both the expression and the purification were conducted at a large scale to be able to complete the purification using the FPLC system. I used a 4 L culture for purification of Hsh49 protein.

After performing FPLC using the AKTA FPLC system, the eluted fraction was quantified with a Nanodrop ND-1000 spectrophotometer. The eluted fraction was passed over a 5 mL MBP-trap column for better purification and was then incubated with one milligram of TEV protease per 50 mg of purified protein at room temperature for 18 hours. The TEV cleaved protein was dialyzed with 20 mM HEPES-NaOH pH 7.5, 300 mM NaCl. After this step, the dialyzed protein was passed through a His-trap column purification to remove cleaved His and MBP tags, His_6-TEV protease, or any proteins that were still with His or MBP tags. The Flow through was dialyzed with 20 mM HEPES-NaOH pH 7.5, 300 mM NaCl. The dialyzed Hsh49 sample was collected and then concentrated in a YM-3 centrifuge filter unit (Microcon Ultracel) to get the concentration of 10 mg/mL according to the manufacture's protocol. The concentrated flow through was quantified with the Nanodrop ND-1000 spectrophotometer. The samples were analyzed by SDS-PAGE.

3.2 Results and Discussion

In this Chapter, I focused on expressing and purifying the SF3a protein subunits and the Rds3 protein separately. Each of the Prp9, Prp11, Prp21 and Rds3 genes were cloned in the PMCSG23 expression vector, and histidine and MBP tags were placed at the N-terminus

of these genes. The Prp9, Prp11 and Prp21 genes were also cloned in the pQlinkN^{mod} expression vector. Restriction enzyme digestion confirmed the presence of each gene in the PMCSG23 vector (Figure 19).

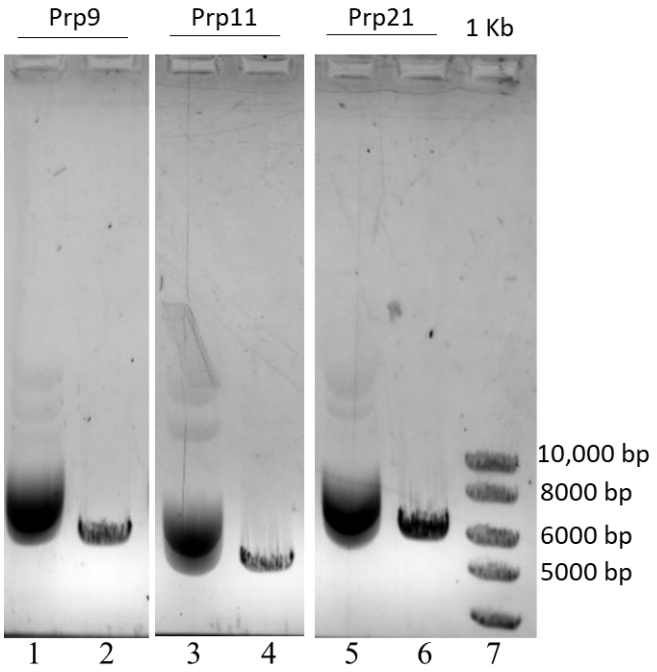


Figure 19. Determining the presence of each SF3a gene in PMCSG23. The 0.7% agarose ethidium bromide gel indicating the presence of each gene in the PMCSG23 vector. The figure indicates the progress of the gel after 1 hour and 30 minutes. Prp9, Prp11 and Prp21 were digested with unique restriction enzymes: SphI, NruI and SspI respectively (Lane 2, Lane 4 and Lane 6). Lanes 1, 3 and 5 contain undigested plasmids for each SF3 a genes. The size of undigested plasmids for Prp9, Prp11 and Prp21 are 6592 bp, 5479 bp and 6337 bp respectively (Lane1, Lane 3 and Lane 5).

SDS-PAGE was used to analyze the expression and purification of the Prp9, Prp11, Prp21 and Rds3 proteins in both the pQlink and the PMCSG23 expression vectors. Based on Figure 20, which shows the expression of the Prp9, Prp11, and Prp21 in the pQlinkN^{mod} expression vector, one band was observed in a post-induced and auto-induced cell extract

(Lane 5 and Lane 6). This band might indicate the expression of the Prp9 protein (60.28 kDa) which was lower than 66.2 kDa. A very faint band was observed in the auto-induced cell extract lane (Lane 9) in the range of 50 kDa which could correspond to the expression of the Prp21 protein with the molecular mass of 50.36 kDa. There was no clear band for Prp11 in either the post-induced or the auto-induced cell extract lane (lane 2 and lane 3). The expected size of the Prp11 protein is 19.09 kDa.

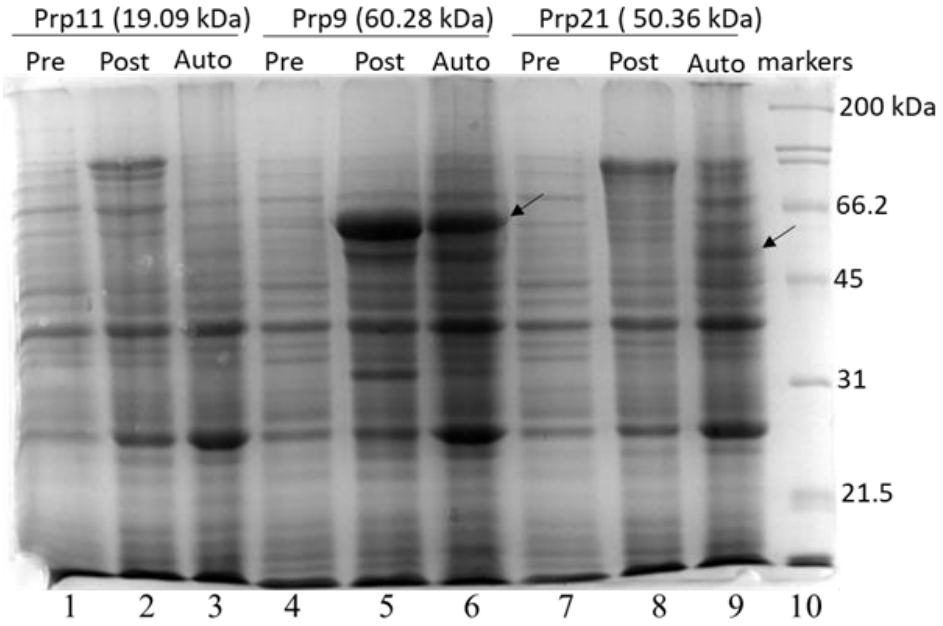


Figure 20. SF3a protein expression from pQlink. 12% high TEMED SDS gel, showing the (Pre) pre-induction, (Post) post-induction and (Auto) auto-induction of SF3a protein expression. The arrow points to the band that corresponds to the expression of the desired proteins.

In Figure 21, the expression of Prp9, Prp11 and Prp21 from the PMCSG23 vector were investigated to analyze if there is any difference in expression of each SF3a subunit protein between the pQlink and PMCSG23 expression vectors. By comparing the pre-induced cell extract with either the post-induced or the auto-induced cell extract lane, no expression of the Prp9, Prp11 and Prp21 proteins was observed. For analyzing the expression

of each subunit of the SF3a protein complex in the PMCSG23 expression vector, as well as the molecular mass of the protein, the molecular mass of the MBP tag (42.5 kDa) and the 6x-His (2.8 kDa) should be considered.

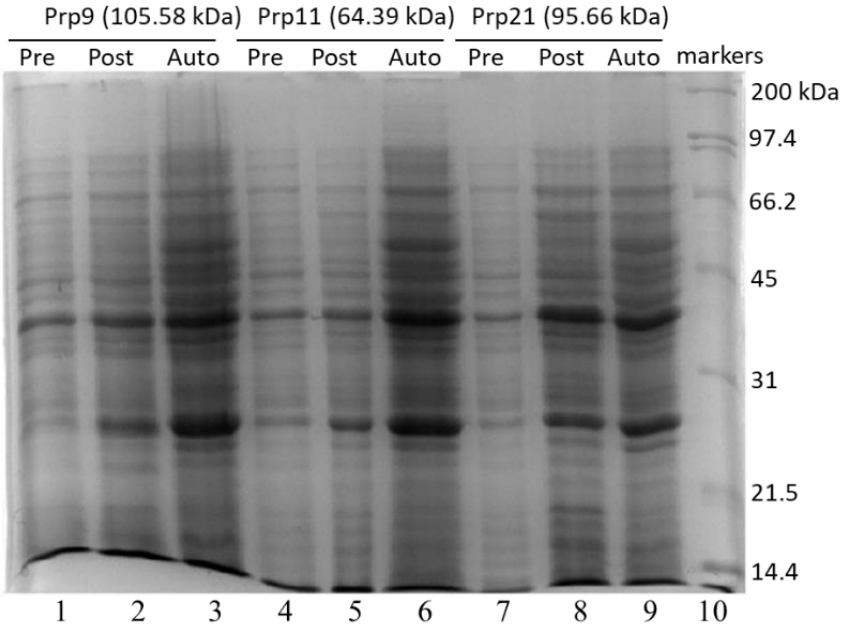


Figure 21. SF3a protein expression from PMCSG23 expression vector. 12% high TEMED SDS gel, showing the (Pre) pre-induction, (Post) post-induction and (Auto) auto-induction of the SF3a protein complex expression.

According to Figure 22, by comparing pre-induced and post-induced cell extract lanes (Lane 1 and Lane 2) the Rds3, which was cloned in the pQlink expression vector, was not expressed (14.04 kDa). After cloning the Rds3 gene in the PMCSG23 expression vector, the target protein was transformed into the host cells and cells were grown in two different media: Luria Bertani (37 C) and 2XYT (16 C for before induction and 25 C for post-induction). Rds3 protein was expressed well in these two media (Figure 23). The Rds3 protein appeared as a soluble protein and partially as an insoluble protein (Figure 24). As was described in the materials and methods of Chapter 2, to separate the soluble and insoluble

fractions, the lysate was centrifuged at 13000 x g in the Beckman Coulter Allegra X-12R centrifuge for ten minutes at 4 C. The proteins that appeared in the supernatant were soluble and the proteins that precipitated as a pellet are insoluble. The supernatant and pellet were run on a gel to check the solubility of the protein expression.

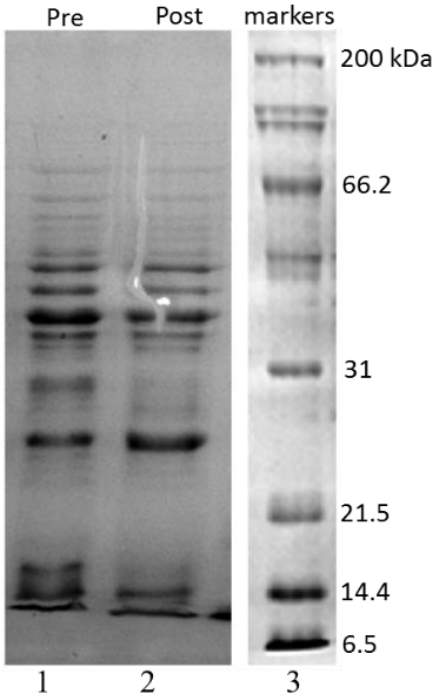


Figure 22. Expression of Rds3 from pQlink. 12% high TEMED SDS gel showing the (Pre) pre-induction, (Post) post-induction of the Rds3 expression.

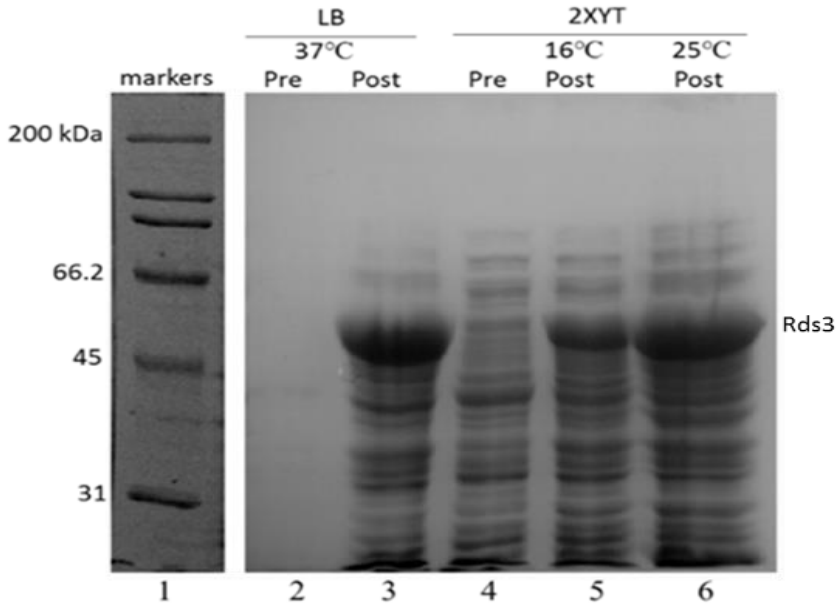


Figure 23. Expression of Rds3 from PMCSG23 expression vector. 10% polyacrylamide gel, showing the expression of Rds3 in two different media: LB and 2XYT. Lanes 2 and 3 represent (Pre) pre-induction and (Post) post-induction of Rds3 at 37 C. Lanes 4, 5 and 6 represent (Pre) pre-induction at 37 C and (Post) post-induction lanes at 16 C and 25 C respectively. The size of the Rds3 protein is 14 kDa, which with His-MBP tags will be around 59.3 kDa (The His-MBP tag size is 45.3 kDa).

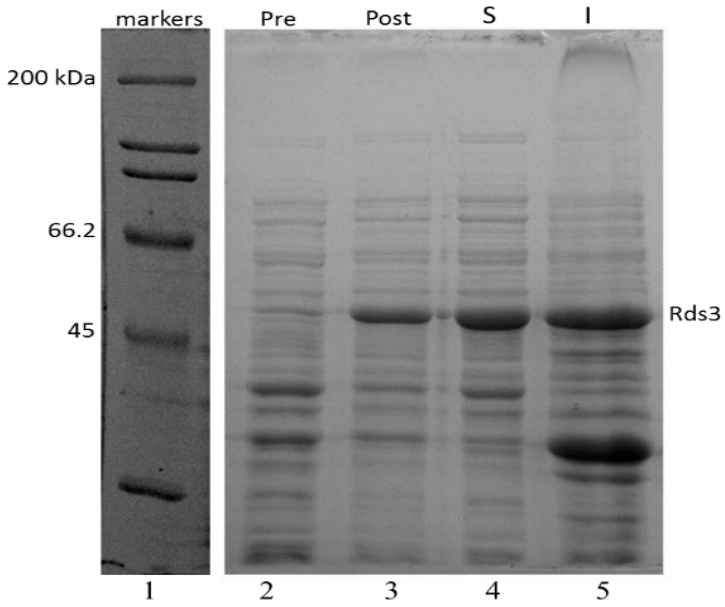
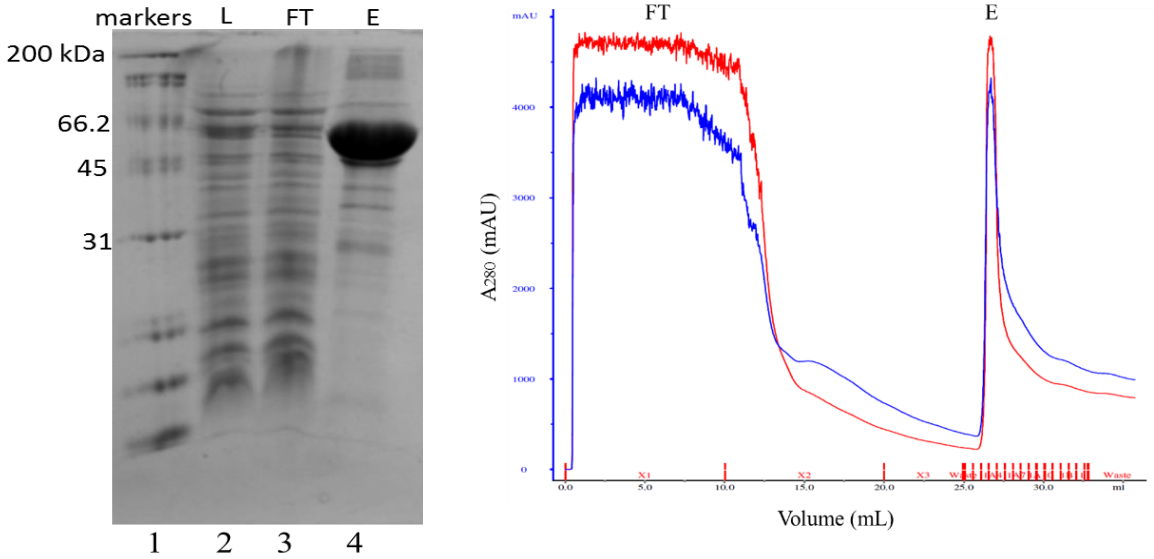


Figure 24. Solubility test of the Rds3. 10% polyacrylamide gel, showing the expression and solubility of Rds3 in 2XYT media. Lanes 2 and 3 represent (Pre) pre-induction and (Post) post-induction of Rds3 at 37 C. Lanes 4 and 5 represent (S) solubility and (I) insolubility of the Rds3 respectively.

Following the successful expression of the Rds3 protein, purification of this protein was performed by FPLC. Since the Rds3 protein contains histidine and MBP tags, it was first purified through the His-trap column (Figure 25a) and then the MBP trap column (Figure 25b).

To remove the His-MBP tags from the Rds3 protein, a TEV protease cleavage was performed at 4 C since TEV protease cleavage at room temperature resulted in the Rds3 protein precipitation (Figure 26a). It was then dialyzed to reduce the concentration of salt and imidazole which were present in the elution fraction following the second purification. A small amount of the Rds3 protein was precipitated after dialysis as shown in Lane 5 in Figure 26b.

a)



b)

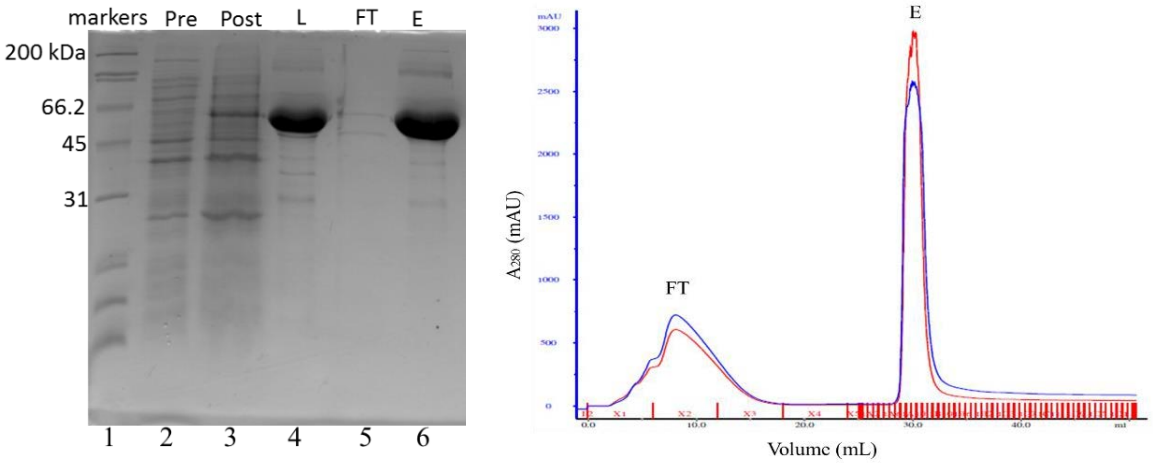


Figure 25. Purification of Rds3. Part (a) right showing the first FPLC chromatogram. Part (a) left indicates first FPLC purification through His-trap column: (L) is the lysate that was gained from the Rosetta pLysS cell, (FT) is the first FPLC flow through, (E) first FPLC elution. Part (b) right showing the second FPLC chromatogram. Part (b) left represents second FPLC purification through MBP trap column: Pre-induced (Pre), Post-induced (Post), (L) is the elution sample from His-Trap column purification, (FT) is the FPLC flow through and (E) is the FPLC elution. The blue and red lines indicate UV absorption at 280 and 260 nm respectively.

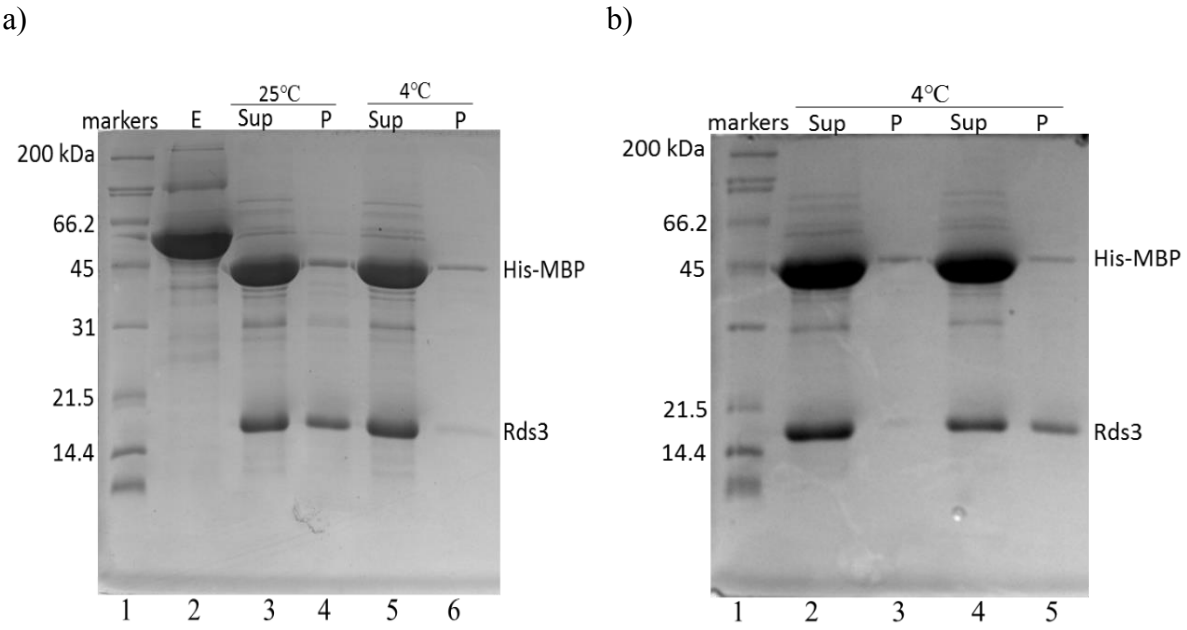


Figure 26. TEV protease cleavage and dialysis of the Rds3. Part (a) represents TEV protease cleavage of the Rds3 at 4 C: Lane 2, (E) is the elution sample after MBP-Trap column purification via FPLC; Lanes 3–4 are the (S) supernatant and (P) pellet of TEV protease cleavage of Rds3 at 25 C; Lane 5–6 are the (S) supernatant and pellet (P) of TEV protease cleavage of Rds3 at 4 C. Part (b) indicates dialysis of Rds3 at 4 C after TEV protease cleavage: Lane 2–3 are the (S) supernatant and (P) pellet of TEV protease cleavage of Rds3 which ran as a control; Lane 4–5 are the (S) supernatant and (P) pellet of dialysis Rds3 at 4 C which was done after TEV protease cleavage. These analyses was performed on 15% high TEMED SDS gels.

Another purification was performed by doing batch binding Ni^{2+}-NTA resin purification to obtain pure Rds3 protein. This purification helps to eliminate His-MBP tags, His-TEV protease or any Rds3 protein that retained tags. However, after this step, the flow through sample still contained tags and most of the Rds3 protein appeared in the elution fraction (Lane 3 and Lane 4, Figure 27a). However, it is expected that Rds3 should appear in the flow through fraction since it is assumed that His-MBP tags have been already removed from the Rds3 protein and there is no tag to bind to the Ni^{2+} resin. Therefore, I ran another purification through the amylose-resin batch binding on the flow through and the elution

fraction that was obtained from the Ni^{2+}-NTA resin purification. This last purification was done to remove tags from these two fractions and produce a band that was almost entirely Rds3 protein. Purification on the flow through fraction resulted in the loss of Rds3 protein (Figure 27b, Lane 3), and purification on the elution fraction showed that the Rds3 retained tags because no differences were observed when comparing Lane 5 (before purification) with Lane 6 (after purification). However, I was able to remove Rds3 protein from the eluted fraction after doing amylose-resin purification (Lane 7, Figure 27b).

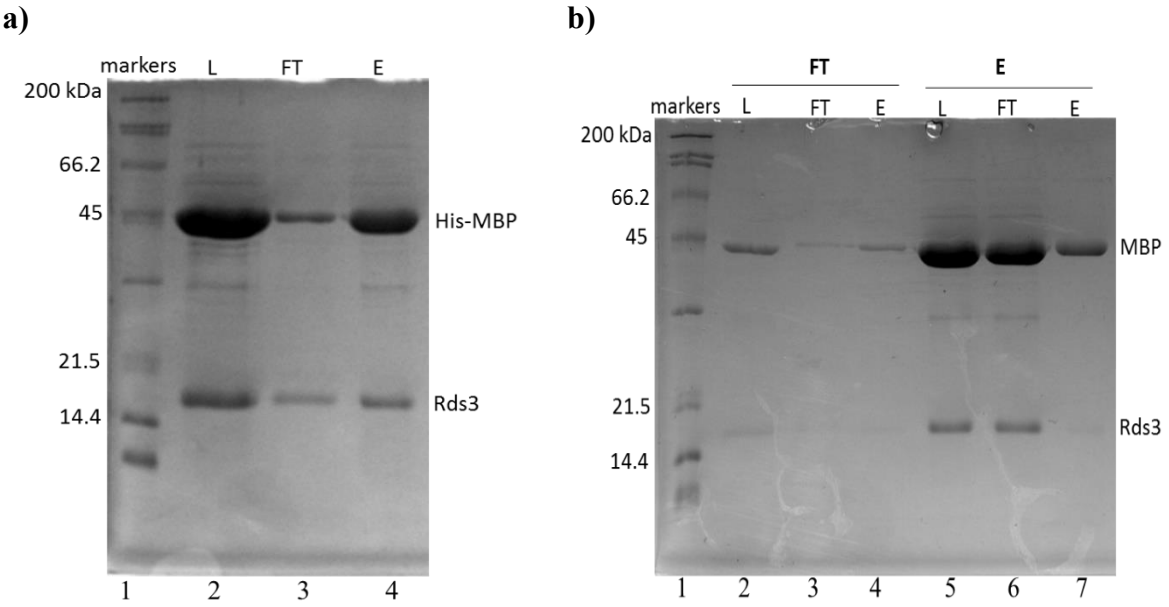


Figure 27. Purification of Rds3 after TEV protease cleavage and dialysis. Part (a) indicates batch binding Ni^{2+}-NTA resin purification on the (FT) flow through in Lanes 2–4. (L) is the load of Rds3 sample after TEV cleavage and dialysis, (FT) is the flow through of the Ni^{2+}-NTA resin purification, (E) is the elution of the Ni^{2+}-NTA resin purification. Part (b) represents amylose-resin batch binding purification that was done on the (FT) flow through in Lanes 2–4 and the (E) elution fraction Lanes (5–7) that was obtained from Ni^{2+}-NTA resin purification: Lane 2, (L) is the sample that was gained from flow through of the batch binding Ni^{2+}-NTA resin purification; Lane 3, (FT) is the flow through of the amylose-resin purification; Lane 4, (E) is the elution of the amylose-resin purification; Lane 5, (L) is the sample that was obtained from the elution that was derived from the Ni^{2+}-NTA resin purification; Lane 6, (FT) is the flow through of the amylose-resin purification; Lane 7, (E) is the elution of the amylose-resin purification. The quality of the Rds3 purification was analyzed on the 15% high TEMED SDS gel.

The Hsh49 gene was cloned in the PMCSG23 expression vector, and histidine and MBP tags were placed at the N-terminus. To confirm the presence of this gene in the PMCSG23 vector, the gene was sequenced by the UNBC genetic facility.

After cloning the Hsh49 gene in the PMCSG23 expression vector, the target protein was transformed into host cells and cells were grown in 2XYT media (37 C for pre- and post-induction). SDS-PAGE was used to analyze the expression and purification of the Hsh49 in the PMCSG23 expression vector. Based on Figure 32a, Hsh49 protein was expressed well in the 2XYT media. The Hsh49 protein appeared as a partially soluble protein and partially as an insoluble protein based on Figure 32b. As was described in the materials and methods of Chapter 2, to separate the soluble and insoluble fractions, the lysate was centrifuged at 13000 x g for ten minutes at 4 C. The proteins that appeared in the supernatant were soluble and the proteins that precipitated as a pellet were insoluble. The supernatant and pellet were run on a gel to check the solubility of the protein expression.

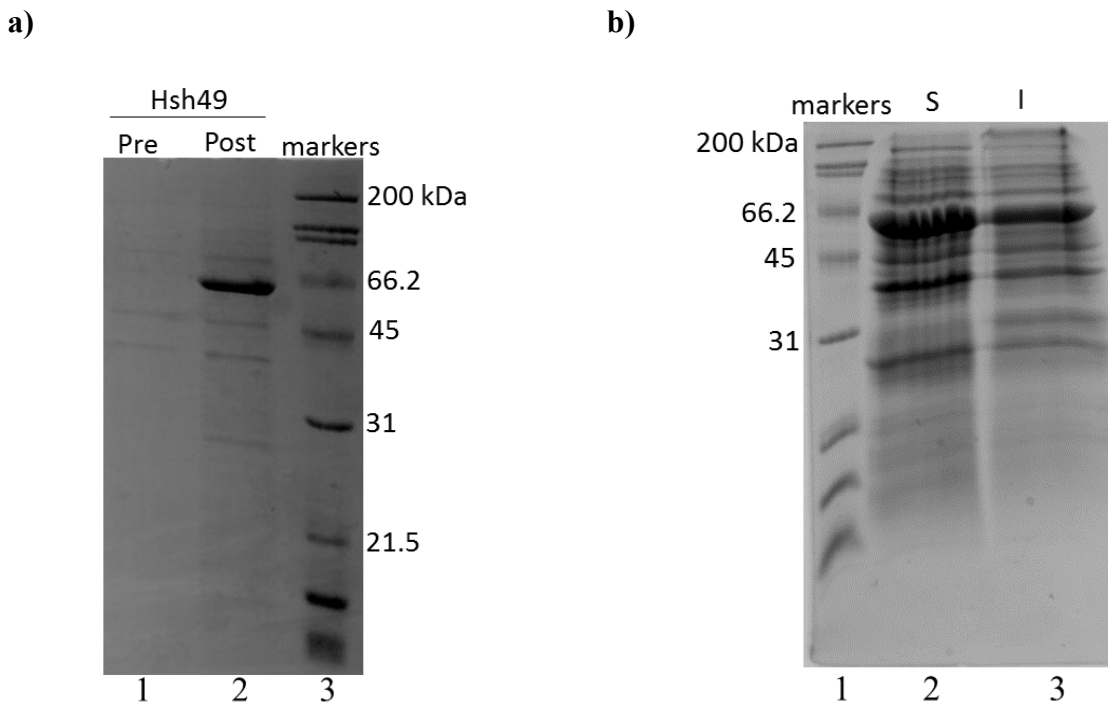


Figure 28. Expression of the Hsh49 protein from PMCSG23 expression vector. The expression of the Hsh49 in Rosetta pLysS (Part a) also, (S) solubility and (I) insolubility of the Hsh49 (Part b) were analyzed on 15% high TEMED SDS gels.

Following the successful expression of the Hsh49 protein, purification of this protein was performed by FPLC. Since the Hsh49 protein contains histidine and MBP tags, it was first purified through the His-trap column and then the MBP-trap column. Part of the Hsh49 protein comes off the column during the wash step of the His-trap purification (Lane 3, Figure 33a). It is possible that the column was overloaded and not all the proteins were able to bind to the nickel-resin.

Since a pure protein was not given off by purification of the Hsh49 through the MBP-trap column at a smaller scale, I decided to remove this step and do another His-trap column purification on the Hsh49 to remove tags after TEV removal. Before the second His-trap column purification, TEV removal was performed to remove the His-MBP tags from the

Hsh49 protein. TEV protease cleavage was conducted at room temperature since TEV protease is more active at room temperature (Figure 33b). It was then dialyzed to reduce the concentration of salt and imidazole which were present in the elution fraction. As I mentioned above, the second purification through the His-trap column was done to obtain pure Hsh49 protein. This purification helps to eliminate His-MBP tags, His-TEV protease or any Hsh49 protein that retained tags (Figure 33c). The flow through was dialyzed and then concentrated. Finally, the concentrated Hsh49 was run on a 15% high TEMED SDS gel which showed that the majority of tags were removed from this protein (Figure 33d).

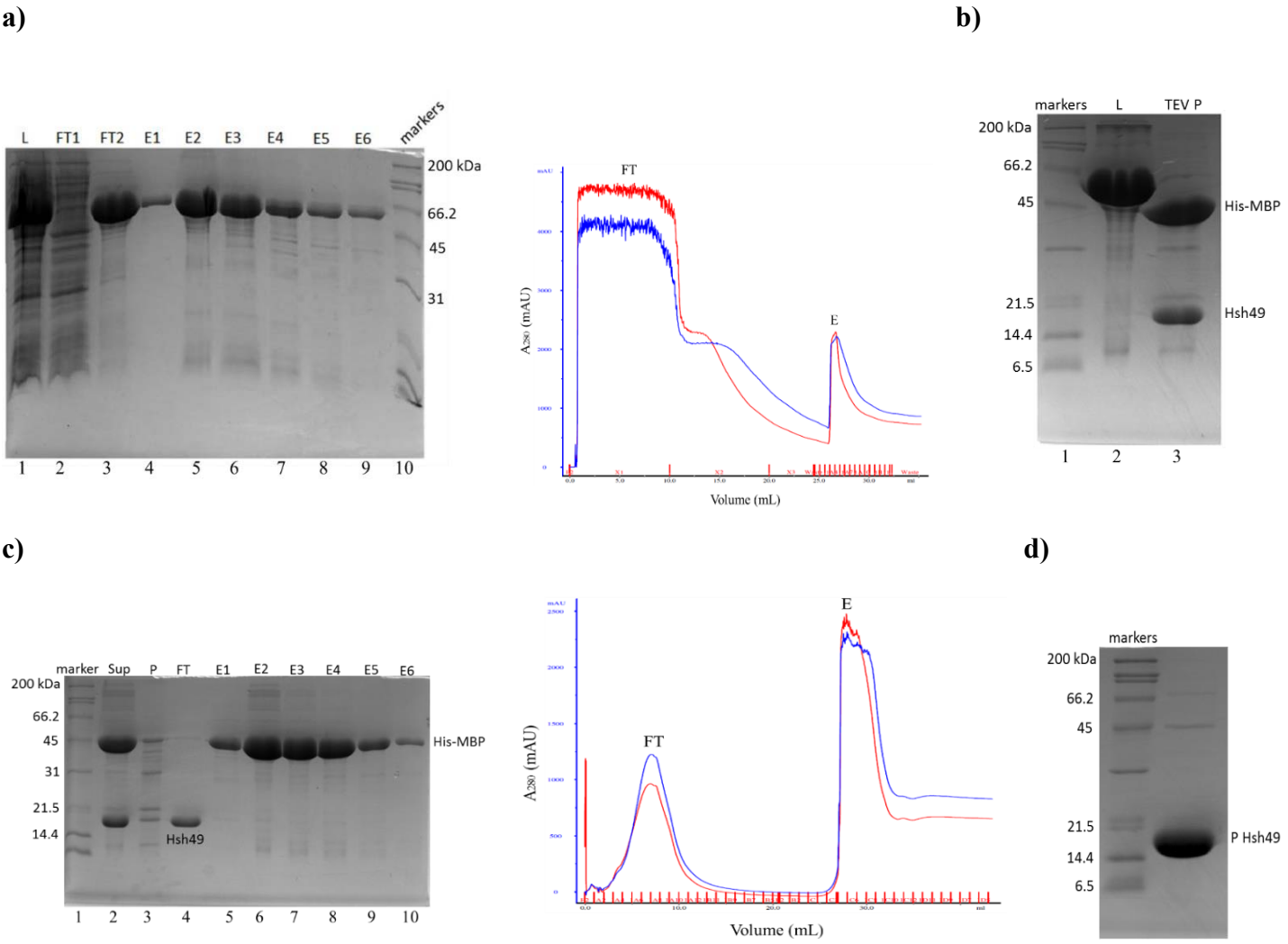


Figure 29. Purification of the Hsh49 was analyzed on 15% high TEMED SDS gel. Part (a) right showing the first FPLC chromatogram. Part (a) left represents first FPLC purification of Hsh49: (L) is the lysate that was obtained from the Rosetta pLysS cell, (FT1) is the first FPLC flow through, (FT2) second FPLC flow through, (E1–E6) is the first to sixth FPLC elution. Part (b) indicates the TEV protease cleavage of Hsh49: (L) is the elution pool that was collected after the FPLC purification, (TEV P) His-MBP tags was cleaved from Hsh49 using TEV protease. Part (c) right shows the second FPLC chromatogram. Part (c) left represents the second FPLC purification of Hsh49 after TEV cleavage: (Sup) supernatant and (P) pellet from dialysis sample, (FT) first FPLC flow through, (E1–E6) is the first to sixth FPLC elution. Part (d) shows the purified Hsh49 protein (P Hsh49) after the majority of His-MBP tag was removed and concentrated. The blue and red lines indicate UV absorption at 280 and 260 nm respectively.

Based on the results, it was surprising that expression of proteins after cloning in pQlinkN^{mod} and PMCSG23 expression vectors was not successful. The reason that the expression failed is hard to address but it is possible that the protein could be toxic for the cell when expressed in a higher yield. Another possibility is that the protein might be unstable due to the absence of some proteins with which it forms a complex (Howe. 2007). In this project, the SF3a protein complex was co-expressed well but an individual expression of each subunit of the SF3a failed. This might be due to the lack of one or two of the SF3a subunits and could have led to protein degradation.

The Rds3 protein was expressed very well and purified through the His and MBP trap column in several steps to isolate this protein and attain a pure result. Unfortunately, I was not successful in removing the tags from the Rds3 protein. Potentially, it is possible that the protein may have folded in a way that made the tags inaccessible to the Ni^{2+}-NTA resin or amylose resin columns. Removal of tags is necessary for the functional assays since tags can affect the function of proteins. Moreover, tags may affect growing crystals for the structural determination of proteins. Also during this time, expression and purification of the Hsh49 protein was performed and the tags were removed from Hsh49 in order to carry out the functional assays. In order to maintain the scope and timeline of the project, I did not proceed with any further analysis of Rds3.

Chapter 4
Functional Analysis of *C. merolae* Hsh49 protein

Given the difficulty I had with the expression of the SF3a protein complex and purification of the Rds3 protein (Chapter 3), I started to express and purify another U2 snRNA protein called Hsh49, to characterize the function and the interaction of this protein with U2 snRNA. Hsh49 is one of the five subunits of the SF3b protein complex in *C. merolae*. The SF3b complex is required for recognition of the intron's branch point, and precise excision of introns from pre-mRNA. In fact, SF3b plays an essential role during pre-spliceosome assembly (Golas *et al*. 2003).

Hsh49 consists of two RNA recognition motifs (RRMs): N-terminal RRM (RRM1) and C-terminal RRM (RRM2). In yeast, RRM1 interacts with the Cus1 protein (another subunit of the SF3b complex) and it seems that association of the Cus1 to the Hsh49 is necessary for the interaction of the U2 snRNA to the branch site of the pre-mRNA (Igel *et al*. 1998; Gozani *et al*. 1996; Champion-Arnaud *et al*. 1994). In yeast, Hsh49 has been shown to have RNA binding activity in which RRM2 associates with the pre-mRNA, and for this association Cus1 protein is not required (Igel *et al*. 1998). As shown in Figure 28, human Hsh49 interacts with several sites of the U2 snRNA, notably the 5′ end of the U2 snRNA and nucleotides located in stem loop I and IIb (Dybkov *et al*. 2006). Moreover, human Hsh49 creates a stable connection with part of the 3′ end of the pre-mRNA (Staknis *et al*. 1994). In this chapter, I describe the examination of the Hsh49 with U2 snRNA to determine whether

they interact in *C. merolae*. If this interaction occurred, I would be able to determine which part of the U2 snRNA interacts with Hsh49.

Circular Dichroism (CD) spectroscopy was used to determine the secondary structure and folding property of the Hsh49 protein. To investigate the RNA-protein base pairing interaction between the U2 snRNA and the Hsh49 protein in *C. merolae*, fluorescence anisotropy (FA), and electrophoretic mobility shift assays (EMSA) were performed as functional analyses. Ultimately, isolation of the Hsh49 protein from the *C. merolae* extract was examined by using co-immunoprecipitation analysis. Following co-immunoprecipitation, a western blot was carried out to identify if the Hsh49 protein was isolated from the *C. merolae* whole cell extract, the immunoprecipitated RNA was then analyzed by Northern Blot assay to detect any Hsh49-U2 snRNA interaction.

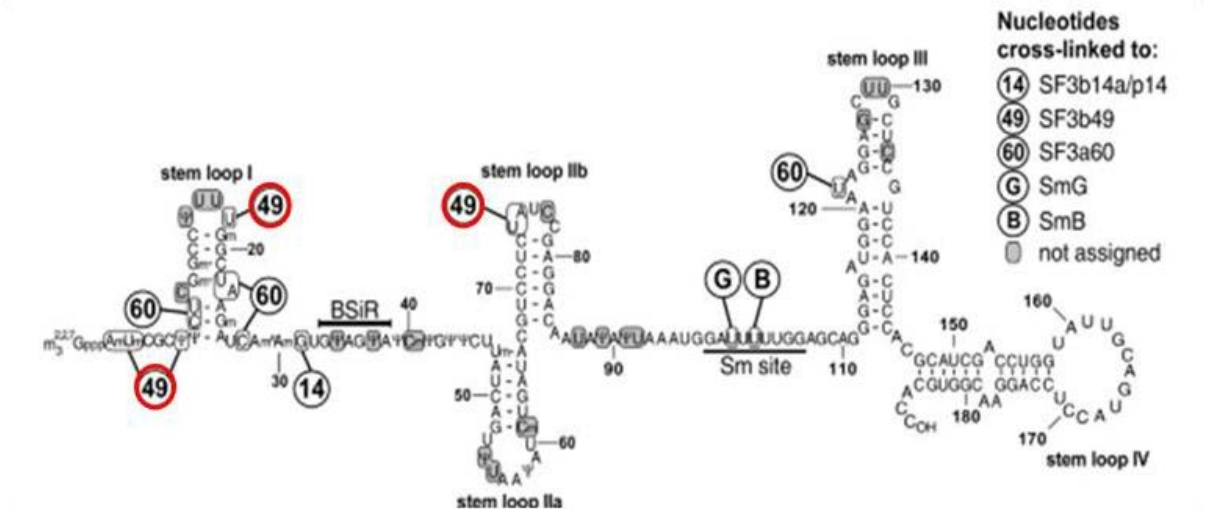


Figure 30. HumanU2 snRNA-protein crosslinks (Dybkov *et al.* 2006).

4.1.1 Fluorescence Anisotropy Assay

Fluorescence anisotropy (FA) or fluorescence polarization (FP) is a powerful, rapid and sensitive technique that is used in molecular interaction subjects such as protein-protein, nucleic acid-protein, and carbohydrate-protein (Lea and Simeonov. 2011; Kakehi *et al*. 2001). The principle of FP, first introduced by Perrin (1926), is that a small fluorescent molecule is excited by the polarized light. The fluorescent molecule will emit polarized light which will then be depolarized (Perrin. 1926). Since the speed of rotation inversely depends on the size of the molecule, the small molecules rotate quickly during the excited state and have low polarization values. Large molecules or complexes, in this case formed from the binding of a second molecule, rotate slowly and therefore have high polarization values (Held and Jolley. 2008). The principle of the fluorescence polarization is shown in Figure 31.

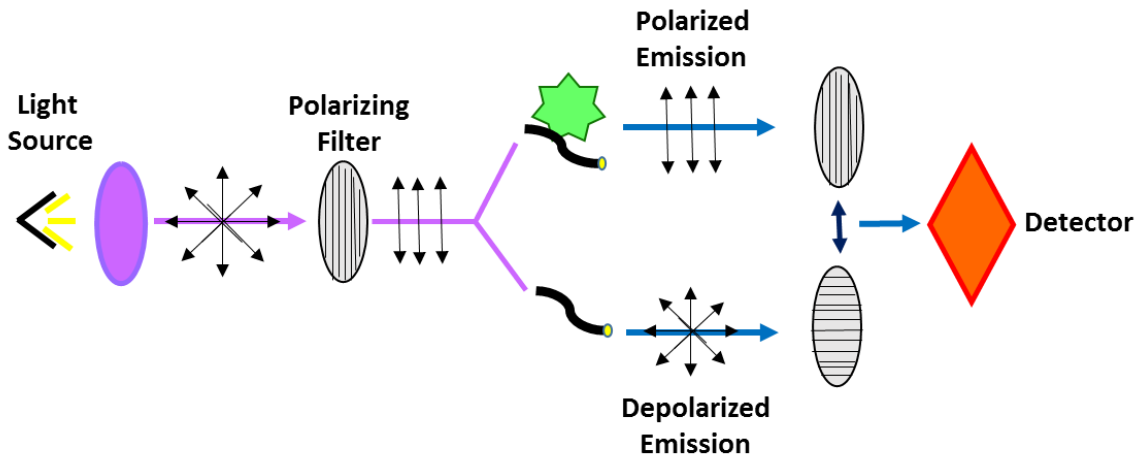


Figure 31. The basic principle of fluorescence polarization (Pagano *et al.* 2011).

4.2 Materials and Methods

4.2.1 Circular Dichroism Spectropolarimetry

A. Estimation of the secondary structure of Hsh49

CD spectra were performed to determine the secondary structure and stability of recombinant expressed, purified Hsh49. Stability and conformational changes of Hsh49 were assigned under thermal stress by using Jasco J-815 CD spectropolarimeter (Jasco Inc). The Hsh49 protein first was dialyzed and diluted in NaF (250 mM NaF and 20mM HEPES-NaOH pH 7.5) because chloride ions from NaCl have a high absorbance in the far UV spectral region (260–190 nm). Hsh49 protein samples were scanned in the far-UV region using a 0.1 cm path length quartz cuvette. CD spectra were assessed at a constant temperature of 25C which was controlled by a Peltier controller (Table 9). This wavelength was used to estimate the fraction of the alpha helices and beta sheets in the Hsh49 protein. The K2D3 circular dichroism web server (Louis-Jeune *et al*. 2011) was used to estimate the secondary structure of Hsh49 from its CD spectrum.

Table 9. Measurement parameters for the CD spectra of Hsh49.

Photometric mode:	CD,HT	**Scanning speed:**	50 nm/minute
Measure range:	260–190 nm	**Baseline correction:**	Baseline
Data pitch:	0.1 nm	**Shutter control:**	Manual
Sensitivity:	Standard	**CD detector:**	PMT
D.I.T.:	2 Seconds	**PMT voltage:**	Auto
Bandwidth:	1.00 nm	**Accumulations:**	3
Start mode:	Immediately		

B. Thermal Stability of Hsh49

The melting temperature was assessed by increasing the temperature to unfold this protein. The concentration of Hsh49 was 1.6 mg/mL (~ 40μM) for CD scanning. Thermal stability of the Hsh49 was measured in the range of 28.79–84.44 C by using a thermal ramp (Peltier temperature controller) based on the α-helix content at 222 nm as measured by CD spectrum (Jasco J-815).

4.2.2 RNA-Protein Binding Assays

This part of Chapter 4 covers the experiments performed to determine the base pairing interaction between the U2 snRNA and the Hsh49 protein. First, a fluorescence polarization (FP) assay was done by using 3′ fluorescein labelled U2 snRNA. The in vitro electrophoretic mobility shift assay (EMSA) experiment was then performed by recruiting radiolabelled full-length U2 snRNA and recombinant Hsh49.

4.2.2.1 Fluorescence Anisotropy Assay

A. Generation of fluorescence labelled U2 snRNA

The U2 snRNA oligonucleotide (35 nucleotides) was synthesized with fluorescein (6-FAM) at its 3′ end by Integrated DNA Technologies (IDT).

5′-CUCAUGGUGUAUCGAGAGCUUCAAGCUUCGAUAUU-3′-F

B. Analysis of the U2 snRNA-Hsh49 Protein Interaction using Fluorescence Anisotropy Assay

In order to determine the U2 snRNA-Hsh49 interaction by fluorescence anisotropy, different concentrations of the Hsh49 protein were incubated with 10 nM of fluorescein labelled U2 snRNA oligo and different types of buffer to find out the optimal conditions for this experiment (Table 10–Table 16). Incubation was done in a 1.5 mL microcentrifuge tube at room temperature (20 C –25 C) for 10 minutes. The trials were run in triplicates using the excitation and emission filter settings of 485 and 521 respectively by Synergy 2 (BioTek Instruments, Inc). Synergy 2 (Fluorescence Anisotropy) provided the values of free and bound RNA and the data were analyzed on the basis of fluorescence anisotropy values. The data were plotted using Prism 7 software.

Table 10. The fluorescence anisotropy reagent (Buffer 1) used for the detection of the U2 snRNA-Hsh49 interaction. (Hsh49 with benzonase and without benzonase).

FA binding reaction reagents (300 uL)	**Volume**	**Final Concentration**
Binding Buffer (see below)	90 µL	-
Hsh49	-	*0 to 200000 nM 0 to 140000 nM 0 to 400 nM
Fluorescent labelled U2 RNA oligo (50 nM)	60 µL	10 nM
H_2O	To 300 µL	
Binding Buffer 1 Reagents (per 1ml)		
1 M HEPES-NaOH pH 7.5	20 µL	20 mM
3 M NaCl	83.3 µL	250 mM
H_2O	896.7 mL	

*Concentrations of Hsh49 used for fluorescence anisotropy experiments (Figure 44 and Figure 45d).

Table 11. The fluorescence anisotropy reagent (Buffer 2) used for the detection of the U2 snRNA-Hsh49 interaction.

FP binding reaction reagents (300 uL)	**Volume**	**Final Concentration**
Binding Buffer (see below)	90 µL	-
Hsh49	-	0 to 40000 nM
Fluorescent labelled U2 RNA oligo (50 nM)	60 µL	10 nM
H_2O	To 300 µL	
Binding Buffer 2 Reagents (1 mL)		
1 M HEPES-NaOH pH 7.5	20 µL	20 mM
1 M $MgCl_2$	2.5 µL	2.5 mM
H_2O	977.5 µL	

Table 12. The fluorescence anisotropy reagent (Buffer 3) used for the detection of the U2 snRNA-Hsh49 interaction.

FA binding reaction reagents (300 uL)	**Volume**	**Final Concentration**
Binding Buffer (see below)	90 µL	-
Hsh49	-	0 to 40000 nM
Fluorescent labelled U2 RNA oligo (50 nM)	60 µL	10 nM
H_2O	To 300 µL	
Binding Buffer 3 Reagents (1 mL)		
1 M HEPES-NaOH pH 7.5	20 µL	20 mM
1 M $MgCl_2$	2.5 µL	2.5 mM
10% Tritonx-100	10 µL	0.1%
H_2O	967.5 µL	

Table 13. The fluorescence anisotropy reagent (Buffer 4) used for the detection of the U2 snRNA-Hsh49 interaction.

FA binding reaction reagents (300 uL)	**Volume**	**Final Concentration**
Binding Buffer (see below)	90 µL	-
Hsh49	-	0 to 40000 nM
Fluorescent labelled U2 RNA oligo (50 nM)	60 µL	10 nM
H_2O	To 300 µL	
Binding Buffer 4 Reagents (1 mL)		
1 M HEPES-NaOH pH 7.5	20 µL	20 mM
3 M NaCl	83.3 µL	250 mM
10% Tritonx-100	10 µL	0.1%
H_2O	886.7 µL	

Table 14. The fluorescence anisotropy reagent (Buffer A) used for detection of conformational changes of the U2 snRNA.

FA binding reaction reagents (300 uL)	**Volume**	**Final Concentration**
Binding Buffer A (see below)	150 µL	-
Fluorescent labelled U2 RNA oligo (50 nM)	60 µL	10 nM
H_2O	To 300 µL	
Binding Buffer A Reagents (10ml)		
100 mM $MgCl_2$	0.3 mL	3 mM
3 M NaCl	0.33 mL	100 mM
50% glycerol	4 mL	20%
H_2O	5.37 mL	

Table 15. The fluorescence anisotropy reagent (Buffer B) used for detection of conformational changes of the U2 snRNA.

FA binding reaction reagents (300 uL)	**Volume**	**Final Concentration**
Binding Buffer B (see below)	150 µL	-
Fluorescent labelled U2 RNA oligo (50 nM)	60 µL	10 nM
H_2O	To 300 µL	
Binding Buffer B Reagents 2 (10ml) (Igel *et al*. 1998)		
40 mM HEPES, pH 7.9	0.4 mL	1.6 mM
250 mM KCl	2.5 mL	62.5 mM
2 mM DTT	0.02 mL	0.004 mM
10% glycerol	2 mL	2%
H_2O	5.08 mL	

Table 16. The fluorescence anisotropy reagent (Buffer C) used for detection of conformational changes of the U2 snRNA.

FA binding reaction reagents (300 uL)	**Volume**	**Final Concentration**
Binding Buffer C (see below)	150 µL	-
Fluorescent labelled U2 RNA oligo (50 nM)	60 µL	10 nM
H_2O	To 300 µL	
Binding Buffer C Reagents (1ml)		
1 M HEPES-NaOH pH 7.5	20 µL	20 mM
3 M NaCl	83.3 µL	250 mM
H_2O	896.7 mL	

4.2.2.2 Fluorescence-based electrophoretic mobility shift assay (F-EMSA)

The binding reaction was prepared for the F-EMSA assay which contained the fluorescein labelled U2 snRNA oligo (ro65) and Hsh49 in varying concentrations. F-EMSA binding reaction reagents are listed in Table 17.

Table 17. The F-EMSA reagent used for the detection of the U2 snRNA-Hsh49 interaction.

F-EMSA binding reaction reagents (20 uL)	**Volume**	**Final Concentration**
Binding Buffer (see below)	10 µL	-
10 mg/mL Ecoli-tRNA	1 µL	0.5 nM
Hsh49	-	0 to 10000 nM
Fluorescent labelled U2 RNA oligo (100 nM)	2 µL	10 nM
H_2O	To 20 µL	
Binding buffer reagents 1 (10ml)		
100 mM $MgCl_2$	0.3 mL	3 mM
3 M NaCl	0.33 mL	100 mM
50% glycerol	4 mL	20%
10% Tritonx-100	0.2 mL	0.2%
H_2O	5.17 mL	

The reagents of the binding reaction were added to a 1.5 mL microcentrifuge tube and mixed well. 3 µL of 6x native loading dye (50% glycerol, 1.5% bromophenol blue, 1.5% xylene cyanol) was added to each sample. 20 µL of samples were resolved by loading on to a prechilled, 6% polyacrylamide (29:1 Acrylamide/Bis, Bio-rad) non-denaturing gel in 1x TBE running buffer. The gel was run at 200 V for 45 minutes in the cold room (4 C) and it was exposed to a cyclone storage imager system (Packard Instrument Company, Inc.) for 18 hours at -80 C. Visualization and analysis of the gel image were performed by using software version 0.4 of the Packard Instrument Company, Inc.

4.2.2.3 Electrophoretic mobility shift assay (EMSA)

4.2.2.4 Generation of [32p]-labelled U2 snRNA by IVT

In vitro transcription of the full length *C. merolae* U2 snRNA (131 nt) was a gift from Martha Stark. Following this step, a phenol-chloroform extraction and an ethanol precipitation were performed. For the U2 probe, the oligonucleotide oSDRro65 was 5′ end labelled with γ-^{32}p ATP (Table 18).

Table 18. The labelling reaction

16.5 µL	H_2O
2.5 µL	PNK Buffer
2.5 µL	10 uM oligo
2.5 µL	γ-^{32}P
1 µL	PNK (NEB)
25 µL	Total Reaction Volume

The first three components were prepared at the bench, while the last two components were added in the hot room. The heat block was pre-heated to 37 C for ten minutes and volume of the reaction was brought up to 50 µL by adding 25 µL of H_2O. One µL of the reaction was then counted by using a liquid scintillation counter. The rest of the reaction was spun through a G-25 sized exclusion column using the refrigerated centrifuge. The resin was re-suspended in the column by vortexing. The cap was loosened and the bottom closure twisted off, and the column was then placed in the collection tube and spun for one minute at 735 x g. After this step, the column was placed into a fresh DNase-free 1.5 mL microcentrifuge tube and an appropriate amount of the sample was applied to the top-centre of the resin. The sample was spun for two minutes at 735 x g. The purified sample was collected in the bottom of the 1.5 mL microcentrifuge tube and quantified using a liquid scintillation counter and stored at -20 C.

A. Analysis of the U2 snRNA-Hsh49 protein interaction using EMSA

The binding reaction which was prepared for the EMSA assay contained the [^{32}p]-labelled IVT U2 snRNA probe (ro65) (7000 cpm) and the Hsh49 in varying concentrations. Two types of binding buffers were also used for EMSA binding reactions (Table 19 and Table 20).

Table 19. EMSA reagents (Binding Buffer Reagent 1) used for the detection of the U2 snRNA-Hsh49 interaction.

EMSA binding reaction reagents (20 uL)	**Volume**	**Final Concentration**
Binding Buffer (see below)	10 µL	-
10 mg/mL Ecoli-tRNA	1 µL	0.5 nM
Hsh49	-	0 to 5000 nM
[^{32}p]-labelled U2 RNA probe	2 µL	7000 cpm/reaction
H_2O	To 20 µL	
Binding Buffer Reagents 1 (10 mL)		
100 mM $MgCl_2$	0.3 mL	3 mM
3 M NaCl	0.33 mL	100 mM
50% glycerol	4 mL	20%
10% Tritonx-100	0.2 mL	0.2%
H_2O	5.17 mL	

Table 20. EMSA reagents (Binding Buffer Reagent 2) used for the detection of the U2 snRNA-Hsh49 interaction.

EMSA binding reaction reagents (20 μL)	**Volume**	**Final Concentration**
Binding Buffer (see below)	10 μL	-
10 mg/mL Ecoli-tRNA	1 μL	0.5 nM
Hsh49	-	100 to 5000 nM
[^{32}p]-labelled U2 RNA probe	4 μL	7000 cpm/reaction
H_2O	To 20 μL	
Binding Buffer Reagents 2 (10 mL) (Igel *et al*. 1998)		
40 mM HEPES, pH 7.9	0.4 mL	1.6 mM
250 mM KCl	2.5 mL	62.5 mM
2 mM DTT	0.02 mL	0.004 mM
10% glycerol	2 mL	2%
0.2% Tritonx-100	0.2 mL	0.004%
2 mM EDTA	0.04 mL	0.008 mM

The reagents of the binding reaction were added to a 1.5 mL microcentrifuge tube and mixed well. In order to determine the optimal incubation temperature, a range of temperatures were tested for EMSA binding reactions. The samples were directly run onto a prechilled, 6% polyacrylamide (29:1 Acrylamide/Bis, Bio-rad) non-denaturing gel in 1x TBE running buffer.

The gel was run at 200 V for 90 minutes in the cold room and exposed to a cyclone storage imager system (Packard Instrument Company, Inc.) for 18 hours at -80 C. Visualization and analysis of the gel image were performed by using software version 0.4 of the Packard Instrument Company, Inc.

4.2.3 Co-immunoprecipitation of Hsh49

4.2.3.1 Dot blot assay

Strips of nitrocellulose membrane were cut and several small circles were drawn on these strips. Purified Hsh49 protein was serially diluted in 20 mM HEPES-NaOH pH 7.5 and 250 mM NaCl. In the center of each circle, 2 µL of each protein dilution was loaded. After the membrane was dried, it was incubated in a blocking solution (5% W/V stain milk powder in 0.1% PBS-Tween) for 1 hour at room temperature. The membrane was washed three times briefly with 0.1% PBS-Tween before being incubated with rat anti-Hsh49 antibody (1:500, 1:1000 and 1:1500) for 1 hour at room temperature. The membrane was washed as before and incubated with goat anti-rat-HRP antibody (diluted 1:1000) for 1 hour at room temperature. The wash step was processed before using the SantaCruz ImmunoCruz Western Blotting Luminol Reagent (SC-2048) for visualization.

4.2.3.2 Western blot assay

Different concentrations of purified Hsh49 protein, including 10 ng, 100 ng and 1 µg, was run on the 15% high TEMED SDS polyacrylamide gel. The gel was then transferred to the nitrocellulose membrane and incubated with anti-Hsh49 (primary antibody, diluted 1:500) in PBST, then incubated with the goat anti-rat-HRP antibody (secondary antibody, diluted 1:1500). To visualize the blot, Western Blotting Luminol Reagent (Santa Cruz Biotechnology) was used according to the manufacturer's protocol. The procedure for Western blot analysis was explained previously (Reimer *et al*. 2017).

4.2.3.3 Preparation of *C. merolae* whole-cell extract

C. merolae whole-cell extract was prepared from the 10D strain from (NIES-1332) provided by the Microbial Culture Collection at the National Institute for Environmental Studies in Tsukuba, Japan (mcc.nies.go.jp/) (Stark *et al*. 2015). The cells were harvested and squeezed in liquid nitrogen. Cell pellets were grounded slowly in a mortar and pestle by adding liquid nitrogen. Cells were thawed by the addition of cold lysis buffer. The lysate was centrifuged to remove cell debris. Glycerol was added to the cell extract and stored at -80°C. Detailed procedures were performed as previously described (Reimer *et al*. 2017).

4.2.3.4 Generating anti-CmHsh49 antibody

Polyclonal anti-serum against recombinant purified Hsh49 (see methods Chapter 3) was generated from a rat following the standard protocol (Harlow and Lane. 1988).

4.2.3.5 Hsh49 co-immunoprecipitation

Immunoprecipitation procedures were performed as described previously (Reimer *et al*. 2017). Briefly, in order to immunoprecipitate the Hsh49 from *C. merolae* whole cell extract, anti-Hsh49 serum, and non-immune serum were coupled to protein G Sepharose 4 Fast Flow (GE Healthcare) or magnetic protein G Dynabeads according to the manufacturer's protocol. The Dynadeads-Ab-Ag complex was then incubated with *C. merolae* extract. To eliminate nonspecific proteins, several wash steps were performed. Proteins were eluted with laemmli buffer followed by Western blot analysis.

4.2.3.6 Northern blot analysis

To determine if Hsh49 binds to U2 snRNA, the immunoprecipitated sample was analyzed via Northern blot. Proteinase K digestion and Phenol: Chloroform extraction were

performed to elute RNA from the beads, followed by EtOH-precipitation. The RNA was resuspended and electrophoresed on a 6%, 7 M urea polyacrylamide gel. The RNA was transferred onto a Hybond N^+ nylon membrane in 1X TBE buffer on a semi-dry blotter (Panther Semidry Electroblotter HEP-3 Owl). The membrane was hybridized to a ^{32}P-labelled U2 oligonucleotide complementary to U2 snRNA. The blot was washed several times and visualized on a phosphorimager with a Cyclone Phosphor Imager using OptiQuant software (Packard Instruments) (Reimer *et al*. 2017).

4.2.4 Alignment of Hsh49 amino acid and U2 snRNA sequences

An amino acid sequence of Hsh49 and U2 snRNA sequence from *Homo sapiens* and *Saccharomyces cerevisiae* were obtained through BLAST searches at the National Centre for Biotechnology Information at the National Library of Medicine and at the Universal Protein knowledgebase (http://www.ncbi.nlm.nih.gov/gene/). U2 snRNA and Hsh49 amino acid sequences of *C. merolae* were obtained from the *C. merolae* genome database (*http://*merolae.biol.s.u-tokyo.ac.jp/*)*. Sequences of Hsh49 and U2 snRNA were aligned using EMBL (European Bioinformatics Institute) (http://www.ebi.ac.uk/Tools/psa/emboss_needle/). The *C. merolae* Hsh49 aligned with *Homo sapiens* RRM1 AND RRM2 using PROMALS3D (Pei *et al*. 2008).

4.3 Results

4.3.1 Expression and Purification of *C. merolae* Hsh49

The Hsh49 gene was cloned in the PMCSG23 expression vector, and histidine and MBP tags were placed at the N-terminus. To confirm the presence of this gene in the PMCSG23 vector, the gene was sequenced by the UNBC genetic facility.

After cloning the Hsh49 gene in the PMCSG23 expression vector, the target protein was transformed into host cells and cells were grown in 2XYT media (37 C for pre- and post-induction). SDS-PAGE was used to analyze the expression and purification of the Hsh49 in the PMCSG23 expression vector. Based on Figure 32a, Hsh49 protein was expressed well in the 2XYT media. The Hsh49 protein appeared as a partially soluble protein and partially as an insoluble protein based on Figure 32b. As was described in the materials and methods of Chapter 2, to separate the soluble and insoluble fractions, the lysate was centrifuged at 13000 x g for ten minutes at 4 C. The proteins that appeared in the supernatant were soluble and the proteins that precipitated as a pellet were insoluble. The supernatant and pellet were run on a gel to check the solubility of the protein expression.

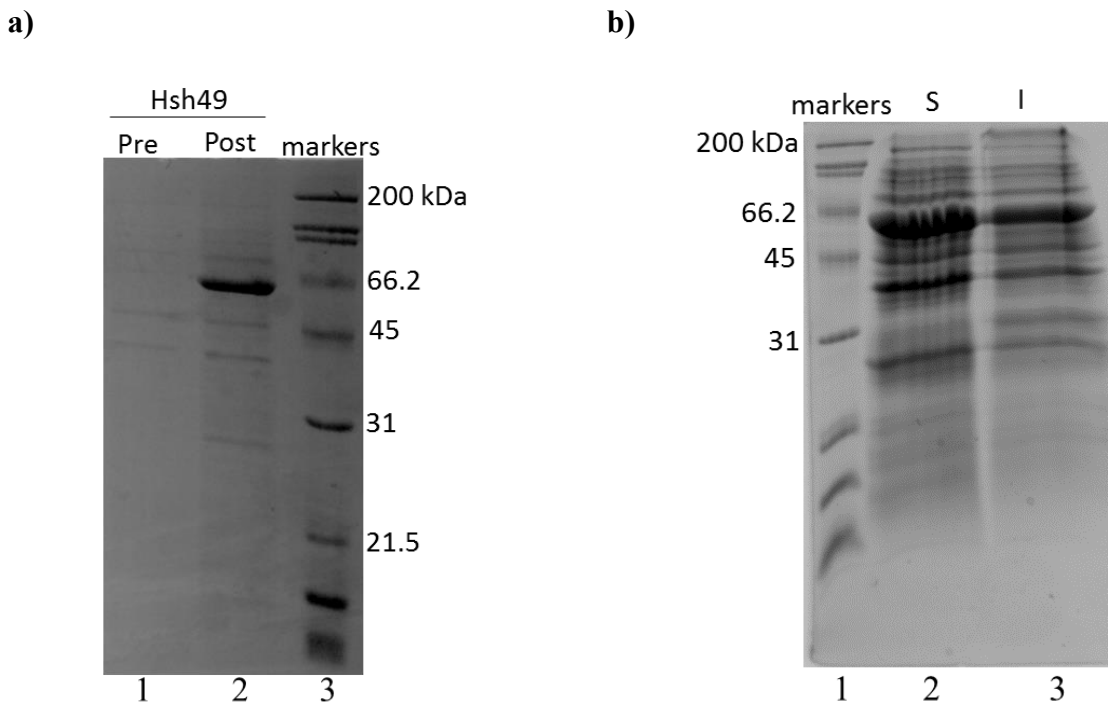


Figure 32. Expression of the Hsh49 protein from PMCSG23 expression vector. The expression of the Hsh49 in Rosetta pLysS (Part a) also, (S) solubility and (I) insolubility of the Hsh49 (Part b) were analyzed on 15% high TEMED SDS gels.

Following the successful expression of the Hsh49 protein, purification of this protein was performed by FPLC. Since the Hsh49 protein contains histidine and MBP tags, it was first purified through the His-trap column and then the MBP-trap column. Part of the Hsh49 protein comes off the column during the wash step of the His-trap purification (Lane 3, Figure 33a). It is possible that the column was overloaded and not all the proteins were able to bind to the nickel-resin.

Since a pure protein was not given off by purification of the Hsh49 through the MBP-trap column at a smaller scale, I decided to remove this step and do another His-trap column purification on the Hsh49 to remove tags after TEV removal. Before the second His-trap column purification, TEV removal was performed to remove the His-MBP tags from the

Hsh49 protein. TEV protease cleavage was conducted at room temperature since TEV protease is more active at room temperature (Figure 33b). It was then dialyzed to reduce the concentration of salt and imidazole which were present in the elution fraction. As I mentioned above, the second purification through the His-trap column was done to obtain pure Hsh49 protein. This purification helps to eliminate His-MBP tags, His-TEV protease or any Hsh49 protein that retained tags (Figure 33c). The flow through was dialyzed and then concentrated. Finally, the concentrated Hsh49 was run on a 15% high TEMED SDS gel which showed that the majority of tags were removed from this protein (Figure 33d).

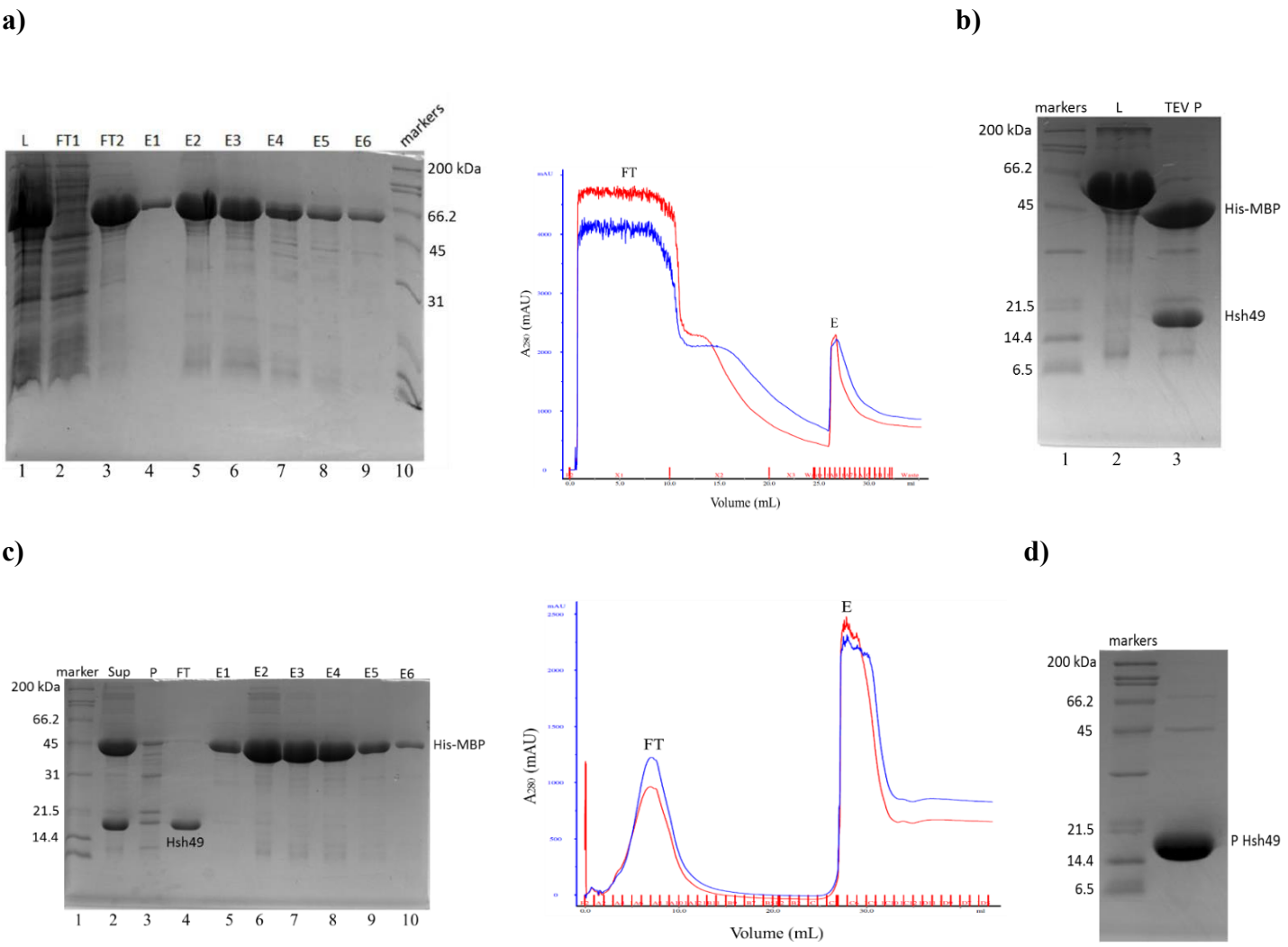


Figure 33. Purification of the Hsh49 was analyzed on 15% high TEMED SDS gel. Part (a) right showing the first FPLC chromatogram. Part (a) left represents first FPLC purification of Hsh49: (L) is the lysate that was obtained from the Rosetta pLysS cell, (FT1) is the first FPLC flow through, (FT2) second FPLC flow through, (E1–E6) is the first to sixth FPLC elution. Part (b) indicates the TEV protease cleavage of Hsh49: (L) is the elution pool that was collected after the FPLC purification, (TEV P) His-MBP tags was cleaved from Hsh49 using TEV protease. Part (c) right shows the second FPLC chromatogram. Part (c) left represents the second FPLC purification of Hsh49 after TEV cleavage: (Sup) supernatant and (P) pellet from dialysis sample, (FT) first FPLC flow through, (E1–E6) is the first to sixth FPLC elution. Part (d) shows the purified Hsh49 protein (P Hsh49) after the majority of His-MBP tag was removed and concentrated. The blue and red lines indicate UV absorption at 280 and 260 nm respectively.

4.3.2 Sequence alignment of *C. merolae* U2 snRNA and Hsh49 with *S. cerevisiae* and *Homo sapiens* U2 snRNA and Hsh49

Sequence alignment of *C. merolae* U2 snRNA with the U2 snRNA in *S. cerevisiae* showed that the similarity of the U2 snRNA in these two organisms is 8% (Figure 34). However, the similarity of the U2 snRNA in *C. merola*e and *Homo sapiens* is 23% (Figure 35). Hsh49 sequence alignment in *C. merolae* indicating 12% identical and 19% similarity to the yeast Hsh49 (Figure 36). Moreover, Hsh49 in *C. merolae* is 9% identical and 13% similar to the *Homo sapien* Hsh49 (Figure 37). Since there are some gaps in the *C. merolae* Hsh49 sequence alignment of Hsh49 from *S. cerevisiae* and *Homo sapiens*, those sequences were trimmed to improve the alignment. The results showed that the similarity of the Hsh49 in *C. merolae* and *S. cerevisiae* is 52%. However, the similarity of the Hsh49 in *C. merola*e and *Homo sapiens* is 50%. In addition, the Hsh49 sequence alignment in *C. merolae* is 34% identical to the yeast Hsh49 and 35% identical to the *Homo sapien* Hsh49.

In order to know if the *C. merolae* Hsh49 contains RRM domains, multiple sequence alignments of Hsh49 coding regions were performed against *S. cerevisiae* and *Homo sapiens* (Figure 38). I also manually aligned the *C. merolae* Hsh49 against each RRM domain individually (Figure 39 and Figure 40). The results showed that the first sequences of *C. merolae* Hsh49 contain RRM1 and the second half matches contain RRM2. In these three sequence alignments, the predicted secondary structure follows the canonical RRM fold: βαββαβ (Figure 38–Figure 40). Moreover, half of the RRMs of *C. merolae* Hsh49 were aligned against both RRMs of *Homo sapiens* Hsh49 that showed 48% similarity and 33% identity. Each *Homo sapiens* Hsh49 RRM1 and RRM2 were separatly aligned against *C. merolae* Hsh49 to compare the RRM similarity. The sequence alignment results demonstrate

that *Homo sapiens* Hsh49 RRM1 is more similar to *C. merolae* Hsh49 RRM1 than RRM2 with 50% similarity. However, the similarity of Hsh49 RRM2 between *C. merolae* and *Homo sapiens* is 37%.

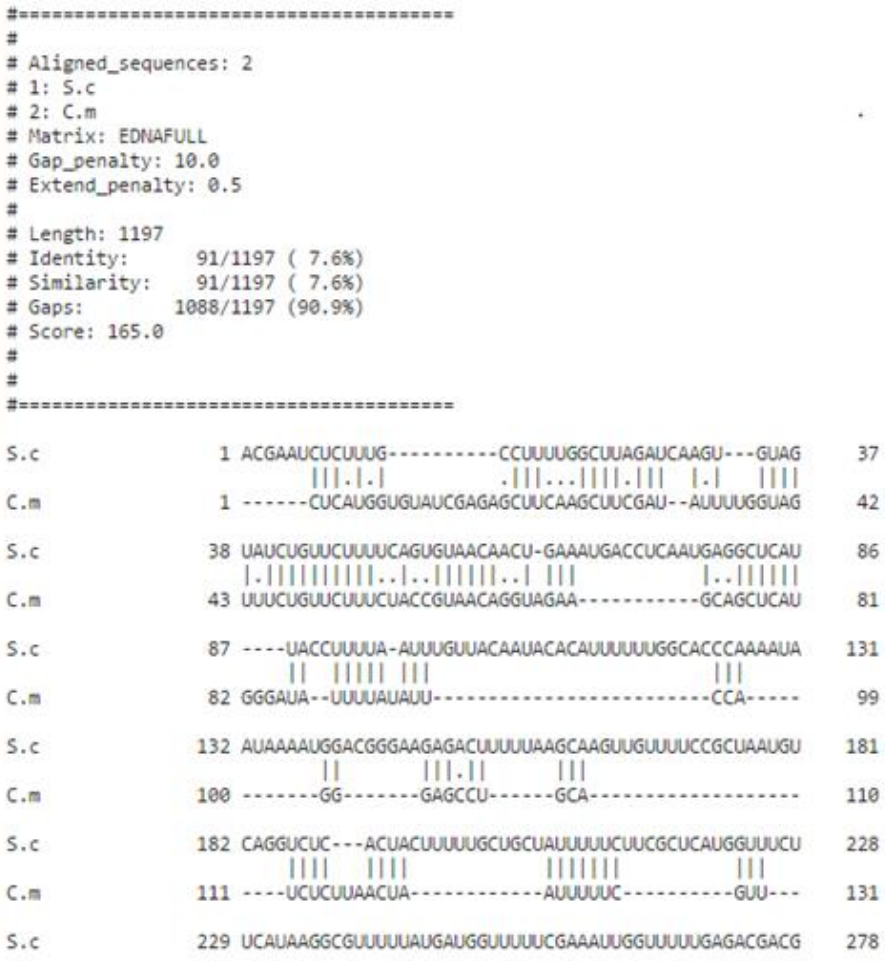


Figure 34. Sequence alignment of *C. merolae* (C.m) U2 snRNA with *S. cerevisiae* (S.c) U2 snRNA.

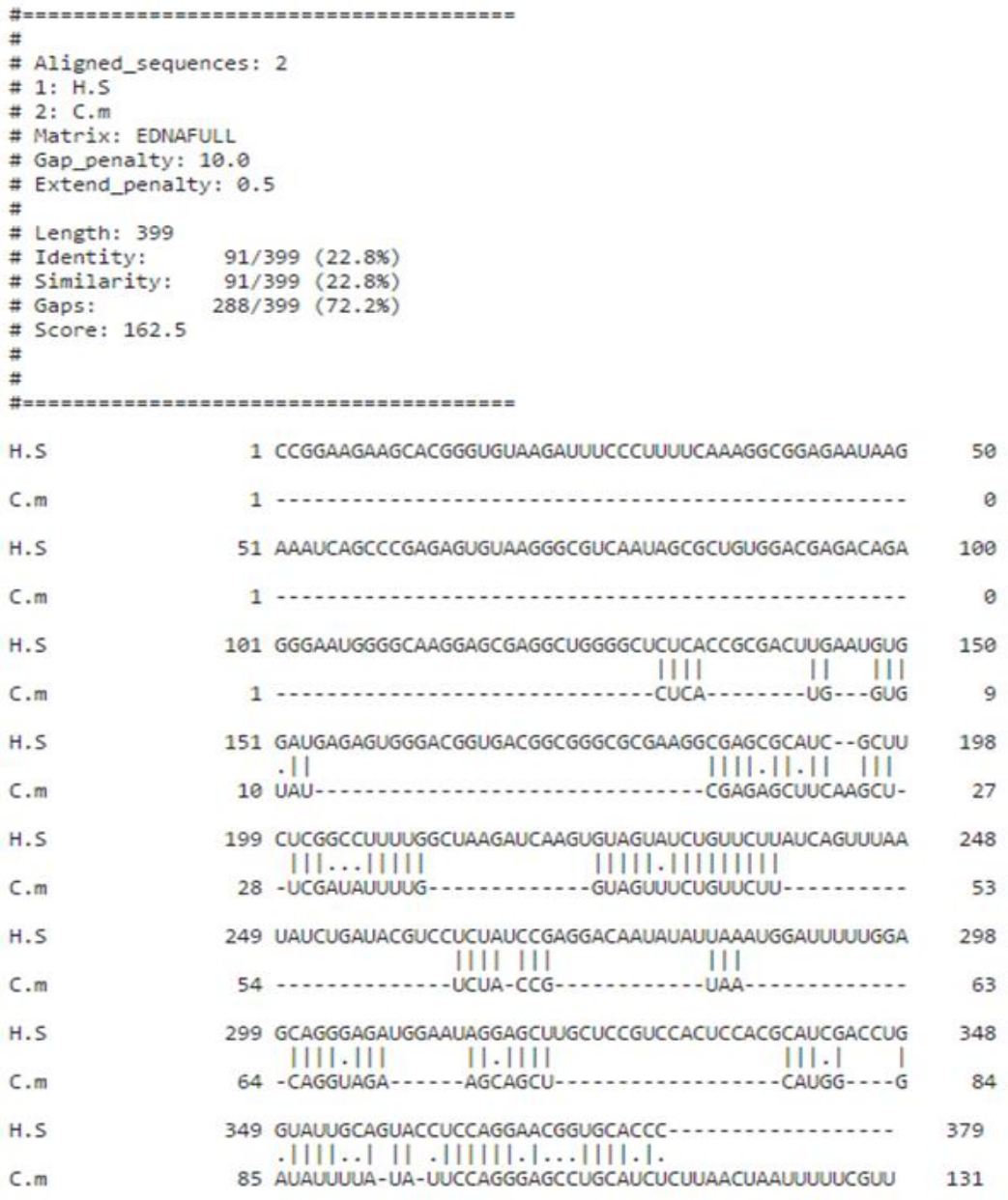


Figure 35. Sequence alignment of *C. merolae* (C.m) U2 snRNA with *Homo sapiens* (H.S) U2 snRNA.

```
#=======================================
#
# Aligned_sequences: 2
# 1: S.c
# 2: C.m.
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 260
# Identity:       32/260 (12.3%)
# Similarity:     49/260 (18.8%)
# Gaps:          175/260 (67.3%)
# Score: 114.0
#
#
#=======================================

S.c                1 ------------------------------------------MNYSADSG      8
                                                                .:.|.:||
C.m.               1 MNGSGLRRAPTVPEADGIAAYSLATRSLNNPDQIVFVSGFVAADESLNSG     50

S.c                9 NTVYVGNIDPRITKEQLYELFIQINPVLRIKYPKDKVLQAYQG-YAFIEF     57
                     ....|.:...|:    |.|||.|..||.|:..|:|::.....| |||:||
C.m.              51 EGAAVTDSPERL----LAELFTQFAPVRRVSIPRDRITGKILGNYAFVEF     96

S.c               58 YNQGDAQYAIKIMN----NTVRLYDRLIKVRQVTNSTGTTNLPSNISKDM    103
                     ::..||.||.|:::    |..||..|...|:|::..
C.m.              97 FSTEDASYAAKVVDGVRFNGRRLLARGPSVQQLSEK--------------    132

S.c              104 ILPIAKLFIKNLADSIDSDQLVKIFNKFGKLIREPEIFYLSNGKLKCAYV    153

C.m.             133 ------------------------------------------------    132

S.c              154 YFEDFEKADLAIKSLNNQLVANNRITVDYAFKENGKGNAKYGDDVDRLLN    203

C.m.             133 ------------------------------------------------    132

S.c              204 KEALKHNMLK    213

C.m.             133 ----------    132
```

Figure 36. Sequence alignment of *C. merolae* (C.m) Hsh49 protein with *S. cerevisiae* (S.C) Hsh49 protein.

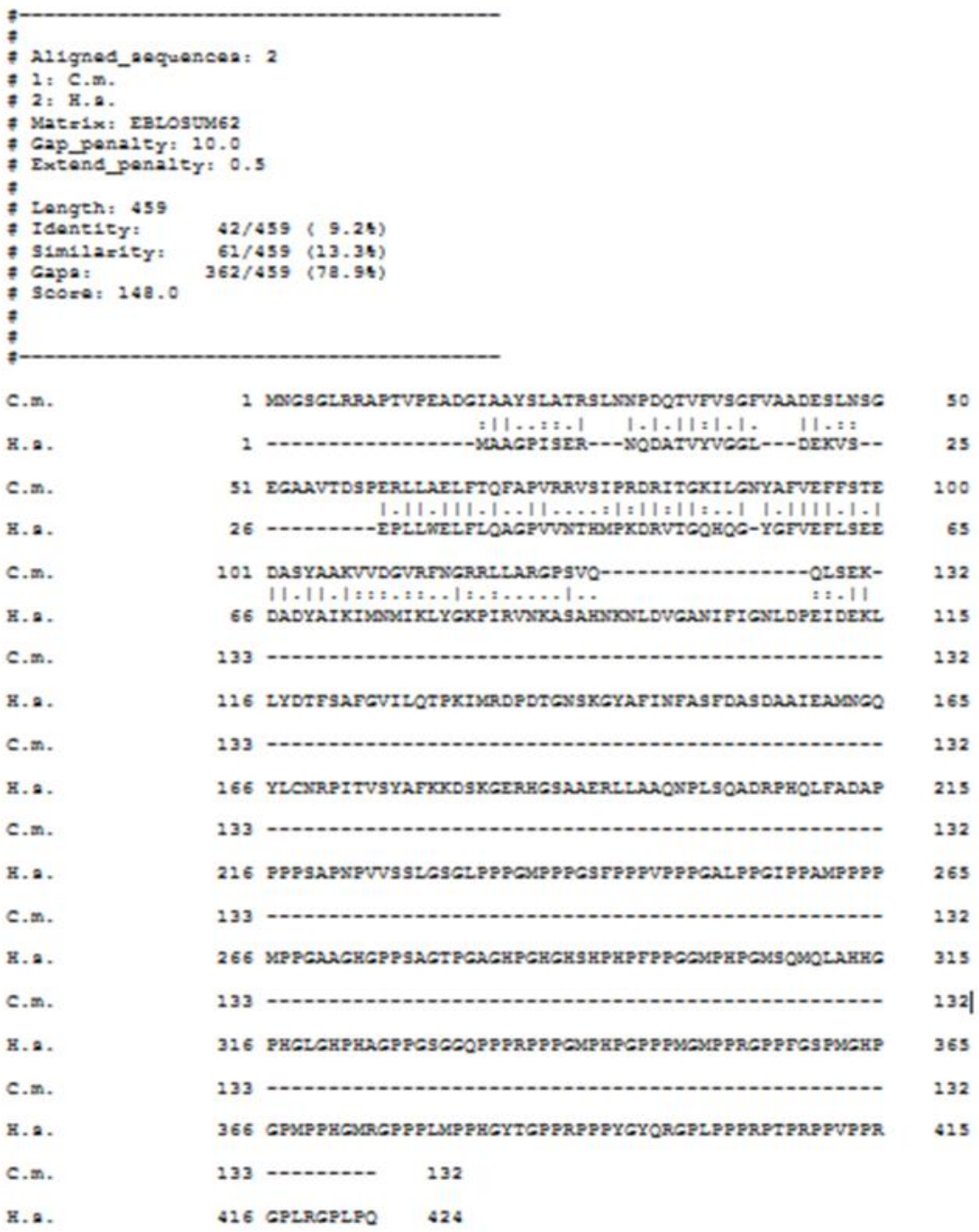


Figure 37. Sequence alignment of *C. merolae* (C.m) Hsh49 protein with *Homo sapiens* (H.S) Hsh49 protein.

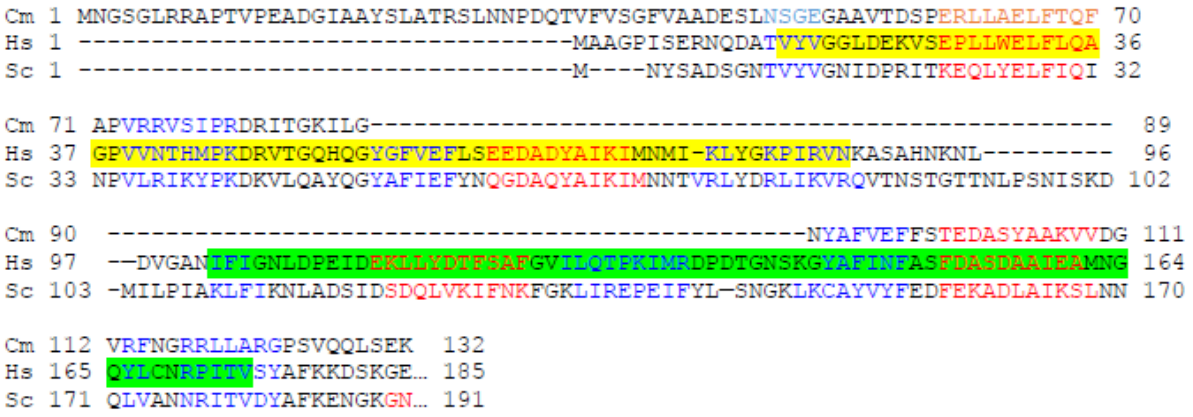


Figure 38. A Multiple sequence alignment of *C. merolae* (C.m) Hsh49 with *S. cerevisiae* (S.C) and *Homo sapiens* Hsh49 (H.S) protein. Yellow and green highlighted sequences relate to RRM1 and RRM2 respectively.

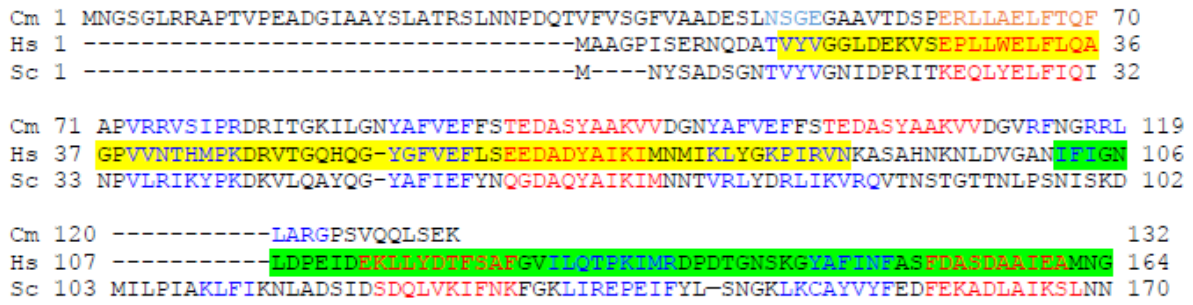


Figure 39. A Multiple sequence alignment of *C. merolae* (C.m) Hsh49 with *S. cerevisiae* (S.C) and *Homo sapiens* Hsh49 (H.S) protein. *C. merolae* Hsh49 manually aligned with the first RRM1 of *Homo sapiens* and *S. cerevisiae* (excluding non-*C. merolae* regions).

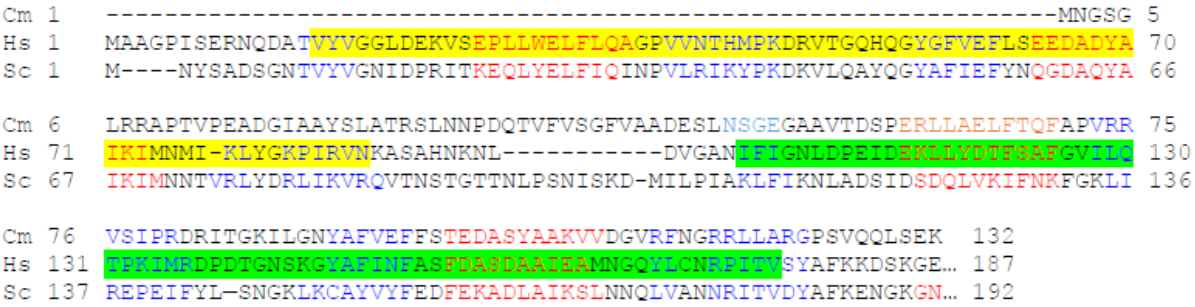


Figure 40. A Multiple sequence alignment of *C. merolae* (C.m) Hsh49 with *S. cerevisiae* (S.C) and *Homo sapiens* Hsh49 (H.S) protein. *C. merolae* Hsh49 manually aligned with the RRM2 of *Homo sapiens* and *S. cerevisiae* (excluding non-*C. merolae* regions).

4.3.3 Circular Dichroism Spectropolarimetry

CD spectroscopy is a powerful technique to determine the secondary structure of the protein. Far-UV spectral region (190–260) of CD data showed that the secondary structure of Hsh49 contains 91% α-helices, 0% β-sheets and 9% random coils (Figure 41).

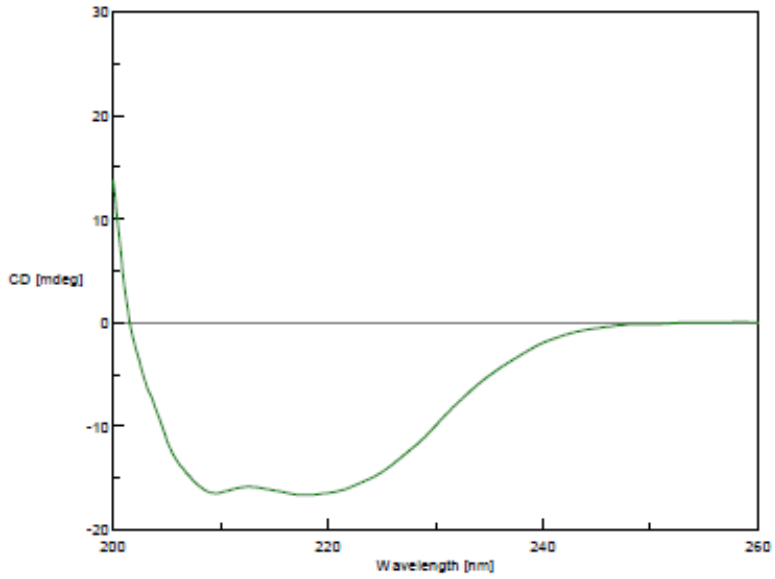


Figure 41. CD spectra of the Hsh49 secondary structure. Circular dichroism spectra of the recombinant Hsh49 protein. The protein was scanned in 250 mM NaF and 20 mM HEPES-NaOH (pH 7.5). The CD spectrum was generated from 3 accumulations using a Jasco J-815.

In order to determine if the Hsh49 was folded, CD spectra was done by increasing the temperature to unfold this protein. The melting temperature of the Hsh49 that was obtained from CD was 76.48 C and it showed a very good transition between the folded and unfolded state of the protein. The unfolding of Hsh49 was reversible and it showed conformational stability in this range of temperatures (Figure 42).

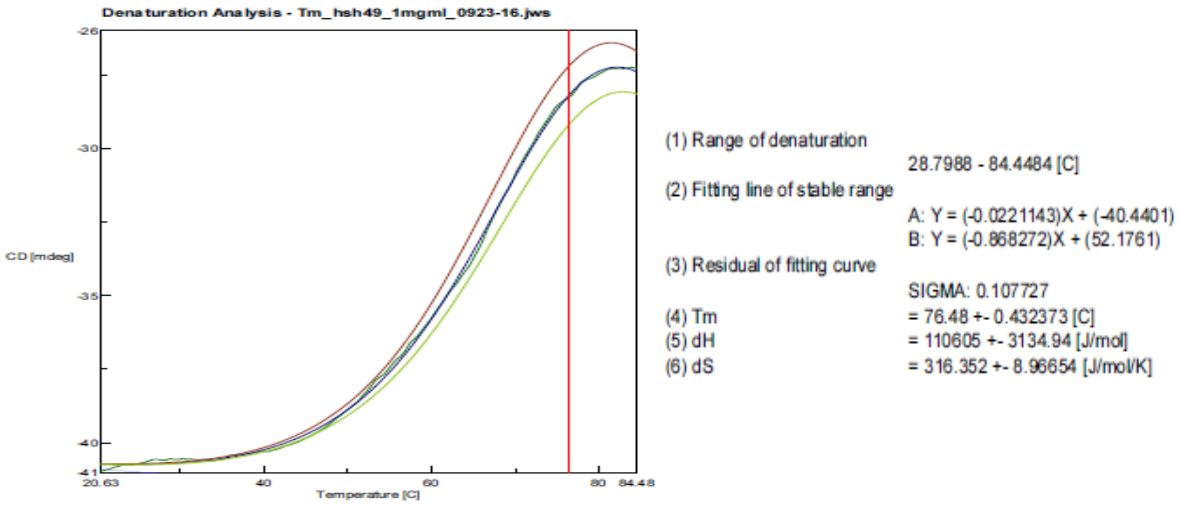


Figure 42. Melting temperature (Tm) of the recombinant Hsh49 protein. Tm calculation and related computed data were obtained from the thermal melting curve of Hsh49.

4.3.4 Fluorescence Anisotropy (FA) Assay

Fluorescence anisotropy was performed on the Snu13 protein with the fluorescein labelled U4 snRNA oligo as a positive control and confirmed the interaction between the U4 snRNA and the Snu13 protein (Figure 43). The U4 oligo is a 5′ kink-turn of *C. merolae* U4 snRNA that was previously demonstrated to interact with the Snu13 protein (Black *et al*. 2016). The fluorescein U4 snRNA oligo is 29 nucleotides in length and fluorescein is attached on the 5′ of this oligo. The concentration of Snu13 was between 0 to 400 nM. Snu13 was serially diluted into 30 mM HEPES-NaOH (pH 7.5) and 100 mM NaCl. The binding buffers that were used for this experiment included HEPES-NaOH (pH 7.5) and NaCl (Table 10).

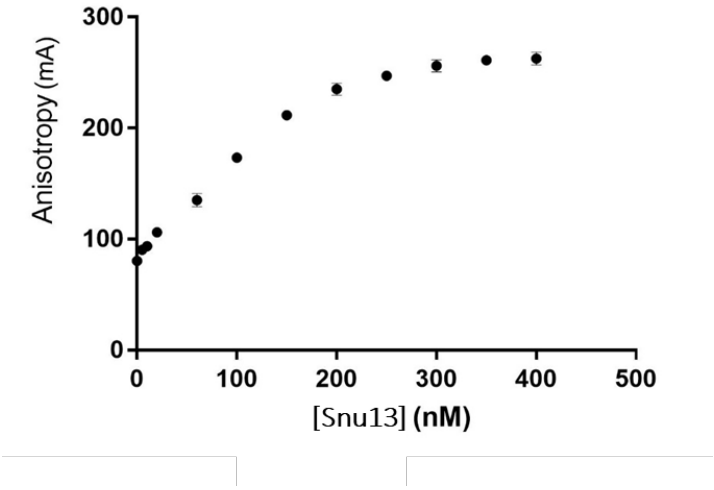


Figure 43. The result of the fluorescence anisotropy experiment of the U4 snRNA-Snu13 interaction. The concentration of the Snu13 was between 0 to 400 nM.

The 35 nucleotide length, fluorescein U2 snRNA oligonucleotide was synthesized to study the interaction of Hsh49 to this part of the U2 snRNA using a fluorescence anisotropy assay. Different types of buffer were used for these FA assays to optimize the binding conditions. Hsh49 was serially diluted into 20 mM HEPES-NaOH (pH 7.5) and 250 mM NaCl for each FA assay.

The fluorescence anisotropy (FA) assay (Figure 44a) was done with Hsh49 protein (0 to 200,000 nM) which was treated with benzonase and reagents that are listed in Table 10. Benzonase was added during the purification of the Hsh49 protein to remove all forms of single stranded or double stranded DNA and RNA. Therefore, it prevents any non-specific interactions that may occur between the Hsh49 and the bacterial RNA. No interaction was observed between Hsh49 and U2 snRNA (Figure 44a). There is a trend in this anisotropy assay, but we do not expect that the interaction is occurring as the trend is different from the U4 snRNA-Snu13 interaction (Figure 43), specifically in that the polarization decreases, the opposite of what we expect for a binding interaction.

There was potential that the benzonase with the Hsh49 may degrade the U2 snRNA, so in this case, there would not be any RNA that could interact with the protein. So another FA assay was performed on the Hsh49 protein (0 to 140,000 nM) samples treated without benzonase to determine if there is any interaction between Hsh49-U2 snRNA. As shown in Figure 44b, there was no interaction between Hsh49-U2 snRNA. The reagents were the same as used in the first FA assay (Table 10).

a) **b)**

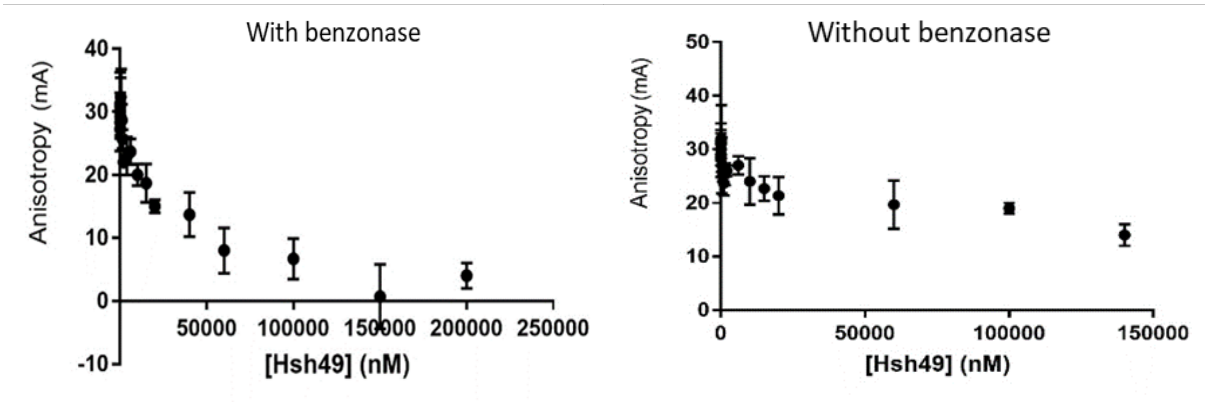


Figure 44. The results of the fluorescence anisotropy of the recombinant Hsh49 and fluorescein labelled of the U2 snRNA. In which Hsh49 was treated with benzonase during purification and the concentration of the Hsh49 for this experiment was between 0 to 200,000 nM (a), also Hsh49 was treated without benzonase and Hsh49 concentration for this experiment was between 0 to 140,000 nM (b).

I performed another FA assay with Hsh49 which was treated with no benzonase and its concentration varied from 0 to 400 nM (Figure 45a–d). To determine how $MgCl_2$ and detergent will affect the Hsh49-U2 snRNA interaction, I ran FA assays with different types of binding buffers.

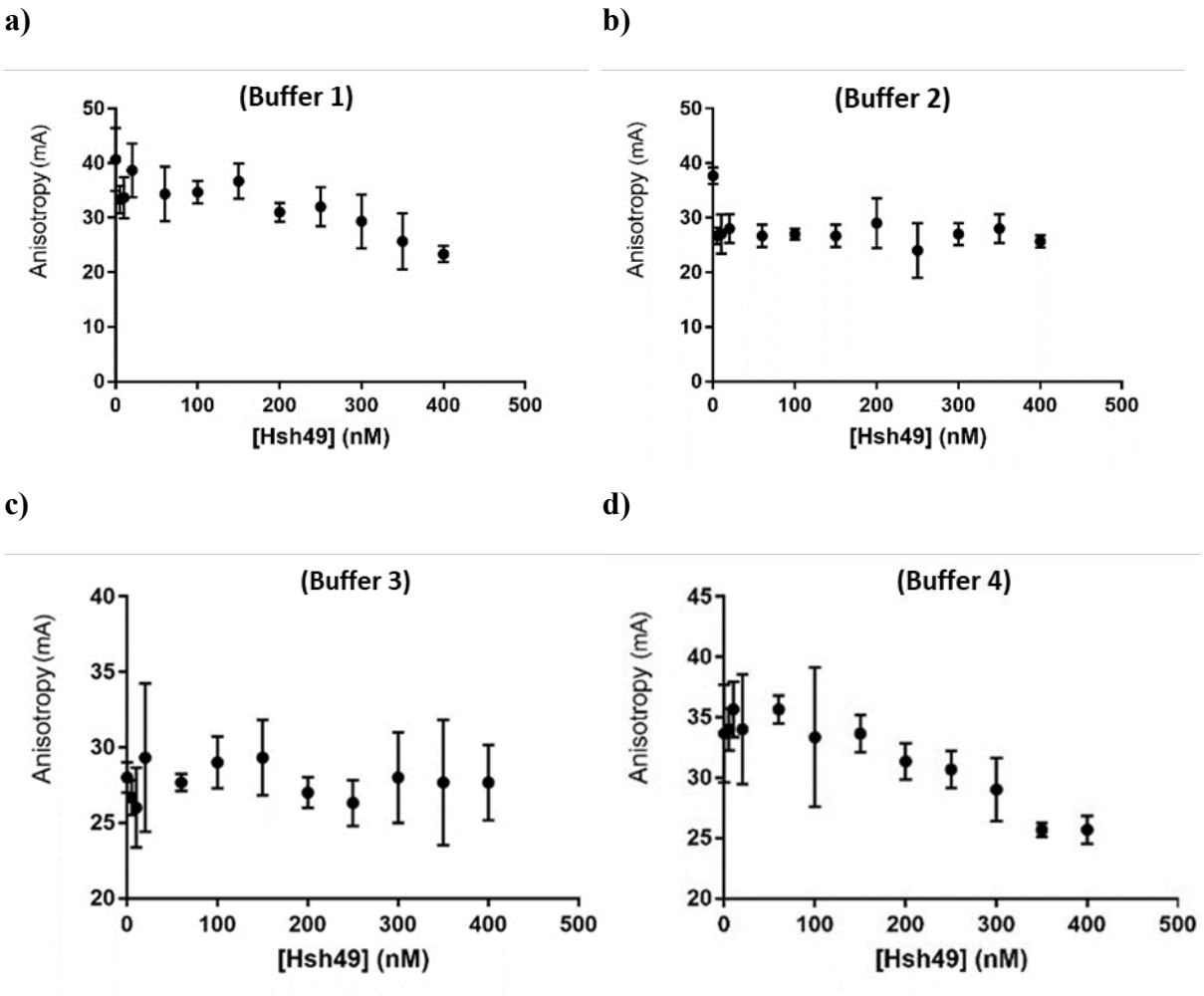


Figure 45. The results of the fluorescence anisotropy experiments of the U2 snRNA-Hsh49 interaction. Using four different types of binding buffers (1–4) (Table 10–Table 13).

The first FA assay was performed (Figure 45a) using $MgCl_2$ and HEPES-NaOH (pH 7.5) as a binding buffer (Table 11), because $MgCl_2$ increases the specificity of DNA and protein interaction, I assumed it might increase the specificity of the RNA and protein interaction. $MgCl_2$ also increases the stability of the RNA (Moll *et al*. 2002; Misra and Draper. 1998). For the second FA assays (Figure 45b), HEPES-NaOH (pH 7.5), $MgCl_2$ and Triton X-100 was used (Table 12). Triton X-100 is a detergent and inhibits nonspecific binding. HEPES-NaOH (pH 7.5), NaCl and Triton X-100 (Table 13) were employed for the

third FA assay (Figure 45c). Finally, the fourth FA assay (Figure 45d) was done by utilizing NaCl and HEPES-NaOH (pH 7.5) (Table 10). Upon comparing the results of these four FA assays, no interactions were observed that indicated the Hsh49-U2 snRNA formed a complex.

Due to buffer composition, the conformation of the RNA molecule can be varied which might affect protein- nucleic acid interaction. The conformation of the RNA indicates transition between open and folded configurations. For instance, the presence of salts such as KCl and $MgCl_2$ increase the RNA stability and $MgCl_2$ affects proper RNA folding (Draper. 2004)

In order to compare how the conformation of the fluorescein U2 snRNA oligonucleotide will change and how the U2 snRNA conformation will affect RNA-protein interaction, three different types of binding buffers (A–C) (Table 14–Table 16) were used in this study. Moreover, incubating of the fluorescein labelled oligonucleotide with binding buffers was compared at two different temperatures, 65 C and 25 C. Therefore, there were two mixtures for each binding buffer. The first mixture was incubated at room temperature for two minutes and the second mixture was heated at 65 C for three minutes and slowly cooled to room temperature for two minutes. According to the bar graph (Figure 46), the labelled U2 snRNA oligo with buffer A at room temperature had the highest fluorescence anisotropy value. The t-test results showed that the fluorescence anisotropy value of the U2 snRNA oligo in Buffer A was higher than the U2 snRNA oligo in Buffer B which was significantly different with a p value of 6.3E-14. The higher anisotropy value of the U2 snRNA oligo in Buffer A is because the U2 snRNA tumbles slowly and the oligo is probably in a more extended conformation in this buffer. Moreover, at 25 C the anisotropy value of the

U2 snRNA oligo in Buffer A was significantly different from the U2 snRNA oligo in buffer C with a p value 7.3E-17.

Based on all fluorescence anisotropy results, no interaction was observed between the fluorescein labelled 35 nucleotide U2 snRNA oligo and Hsh49 protein in *C. merolae*.

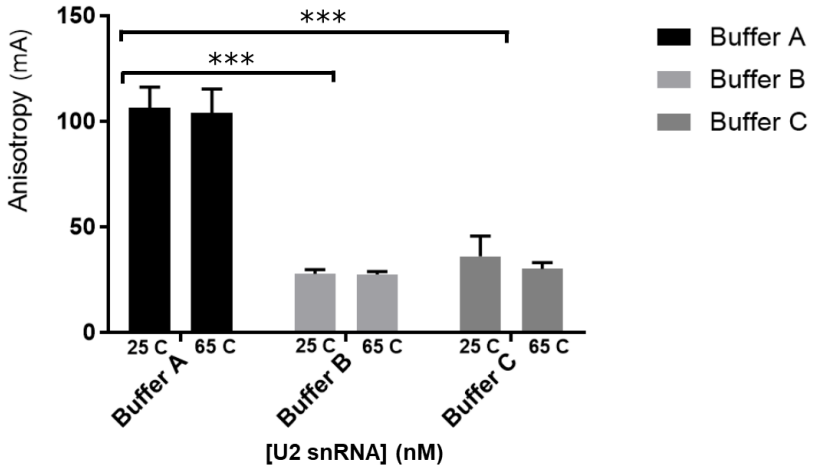


Figure 46. A comparison of the conformational changes of the fluorescein U2 snRNA oligonucleotide in three different types of binding buffers by using the fluorescence anisotropy assay. *** = p< 0.001.

4.3.5 Fluorescence-based electrophoretic mobility shift (F-EMSA) Assay

Since the FA results did not show any Hsh49-U2 snRNA interaction, it was possible that the fluorescent U2 oligo has been degraded. In order to assess this possibility, F-EMSA was performed. In addition, F-EMSA was conducted to detect Hsh49-U2 snRNA interactions and Snu13-U4 snRNA was run on the gel as a control.

In Figure 47, no binding interaction between U2 oligo and Hsh49 was observed as demonstrated in Lanes 3–7. However, there was an interaction between the U4 oligo and Snu13 as it is shown in Lane14.

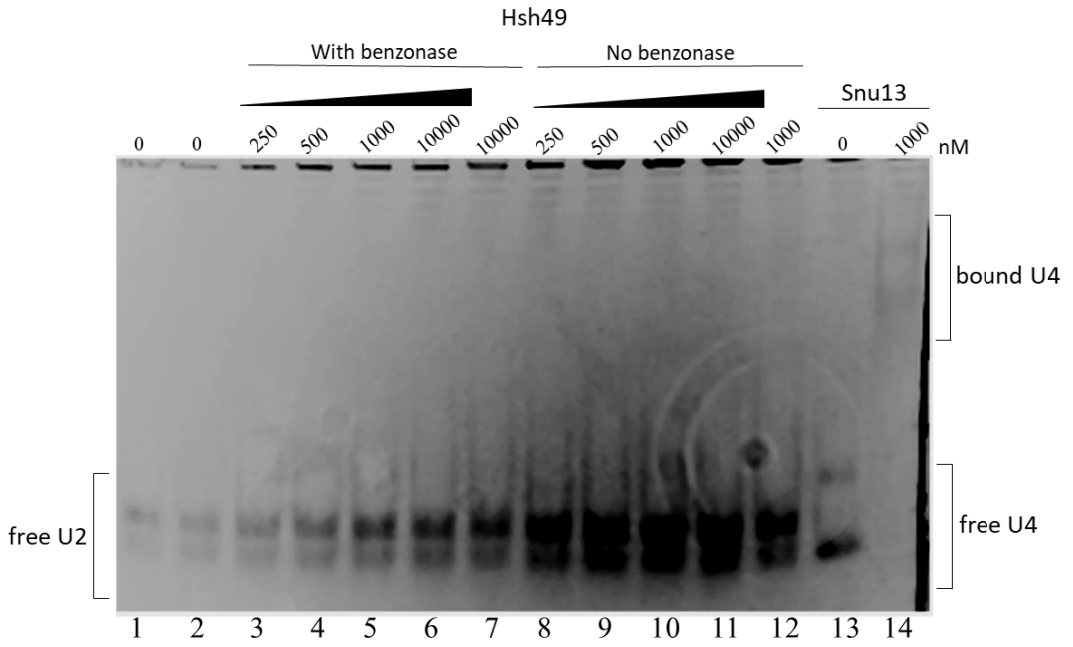


Figure 47. Fluorescence-based electrophoretic mobility shift assay (F-EMSA) with recombinant Hsh49 and fluorescent labelled U2 snRNA. Hsh49 and Snu13 concentrations are indicated on top. Lane 13 contains free U4 snRNA. Lane 1 contains free U2 snRNA with no incubation and Lane 2 contains free U2 snRNA incubated at 25 C for 30 minutes. Lanes 3–7 contain the Hsh49 with benzonase and Lanes 8–12 contain Hsh49 with no benzonase incubated with U2 snRNA at 25 C for 30 minutes. Lanes 13 and 14 contain snu13 incubated with U4 snRNA at 25 C for 30 minutes. The U2 and U4 snRNA oligo concentrations were 10 nM. The 6% native gel was run for 45 minutes at voltage 200. F-EMSA was performed by using binding buffer reagents which are listed in Table 17.

4.3.6 EMSA for detecting the Hsh49 protein-U2 snRNA interaction

The Hsh49 protein interaction with the full-length U2 snRNA was studied using the Electrophoretic mobility shift assay. If the Hsh49 protein forms a complex with the U2 snRNA, it appears as a band which has not traveled as far on the gel as the free U2 snRNA. The full length radioactive [^{32}P]-labelled U2 snRNA was titrated with the Hsh49.

The first EMSA was completed by using binding buffer reagents 1 (Table 19) and the EMSA binding reactions mixture was incubated at room temperature for 30 minutes. Comparing the resulting free U2 snRNA lane (Lane 1) to the lanes containing Hsh49 and U2

snRNA (Lanes 2–5), no bands were observed that indicated the Hsh49-U2 snRNA formed a complex (Figure 48).

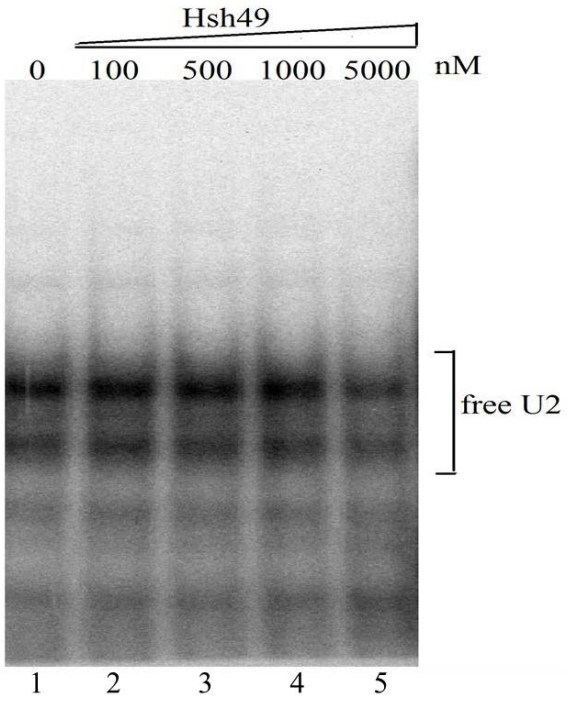


Figure 48. Electrophoretic mobility shift assay with recombinant Hsh49 and radioactive [32p]-labelled U2 snRNA. Hsh49 concentrations are indicated on top and free U2 snRNA is shown in Lane 1. The U2 RNA oligo concentration was 10 nM and incubation was done at 25 C, for 30 minutes. The gel was run for 90 minutes at Voltage 200. The binding buffer reagents for this EMSA are listed in Table 19.

I repeated the EMSA with a different binding buffer, Binding Buffer Reagents 2 (Table 20), and incubation temperature, 25 C, for 20 minutes (Igel *et al*. 1998). The binding buffer reagents 2 was used for the remaining EMSA procedures. Even by changing these conditions, there was no shift indicating the Hsh49-U2 snRNA interaction (Figure 49).

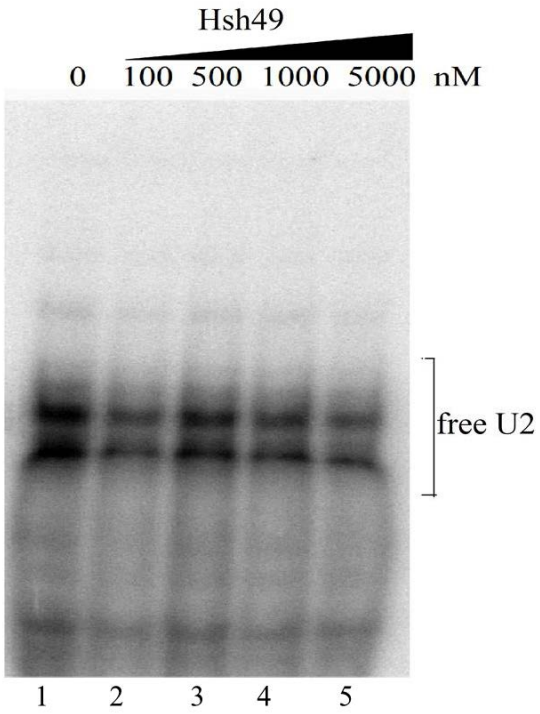


Figure 49. Electrophoretic mobility shift assay with recombinant Hsh49 and radioactive [32p]-labelled U2 snRNA. Hsh49 concentrations are indicated on top and free U2 snRNA is shown in Lane 1. The U2 RNA oligo concentration was 10 nM and Incubation was done at 25 C, for 20 minutes. The gel was run for 90 minutes at voltage 200. This EMSA was done by using the binding buffer reagents in Table 20.

Prior to mixing the reagents, an RNA probe was heated at 65 C for three minutes and allowed to cool on ice for two minutes to ensure proper folding of the snRNA. An incubation temperature of 25 C was used for Hsh49 with no benzonase and 40 C and 25 C were used for the Hsh49 which was with benzonase. I chose a 40 C incubation temperature to be close to the temperature that the *C. merolae* grows in their natural settings (45 C) in the hope that this temperature would help the RNA to unfold and facilitate the protein-RNA interaction (Matsuzaki *et al*, 2004). No band representing the U2 snRNA-Hsh49 complex formation was detected when comparing free U2 snRNA with the other lanes (Figure 50).

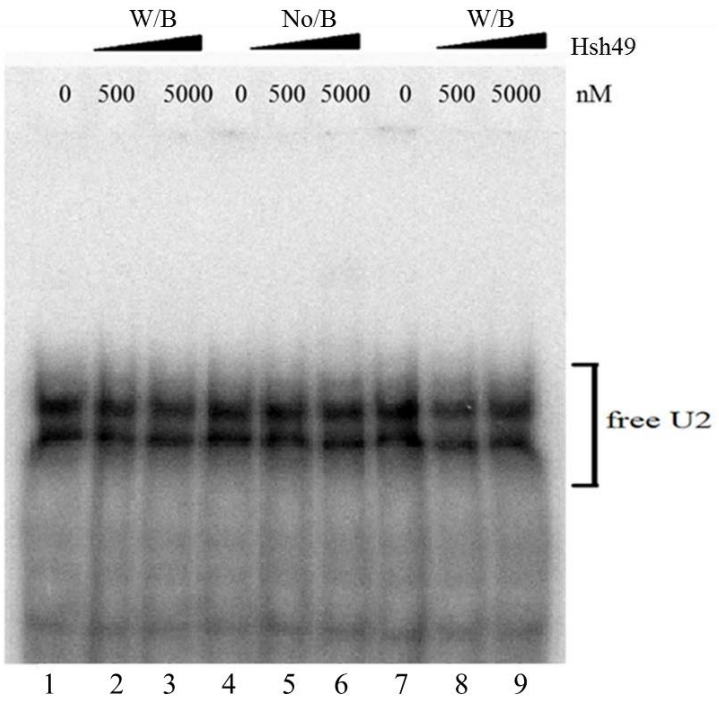


Figure 50. Electrophoretic mobility shift assay with recombinant Hsh49 and radioactive [32p]-labelled U2 snRNA. Hsh49 concentrations are indicated on top and free U2 snRNA is shown in lane 1, Lane 4 and Lane 7. Lanes 2 and 3 contain the Hsh49 with benzonase (W/B) and incubated with U2 snRNA at 25 C for 20 minutes. Lanes 5 and 6 contain the Hsh49 with no benzonase (No/B) and incubated with U2 snRNA at 25 C for 20 minutes. Lanes 7 and 8 contain the Hsh49 with benzonase and incubated with U2 snRNA at 40 C for 20 minutes. The concentration of the U2 oligo was 20 nM and the oligo was heated at 65 C for 3 minutes and cooled on ice for 2 minutes before incubation. The gel was run for 90 minutes at voltage 200.

In order to determine whether cleaving His and MBP tags from Hsh49 caused this protein to misfold, I performed EMSA to compare the results with the cleaved and uncleaved Hsh49 (Figure 51). When comparing uncleaved (Lanes 2–6) and cleaved (Lanes 8–12) Hsh49 in Figure 51, a slower-migerating band was formed in uncleaved Hsh49 lanes. This band corresponds to a protein-RNA complex formation and the intensity of the band increased with the protein concentration.

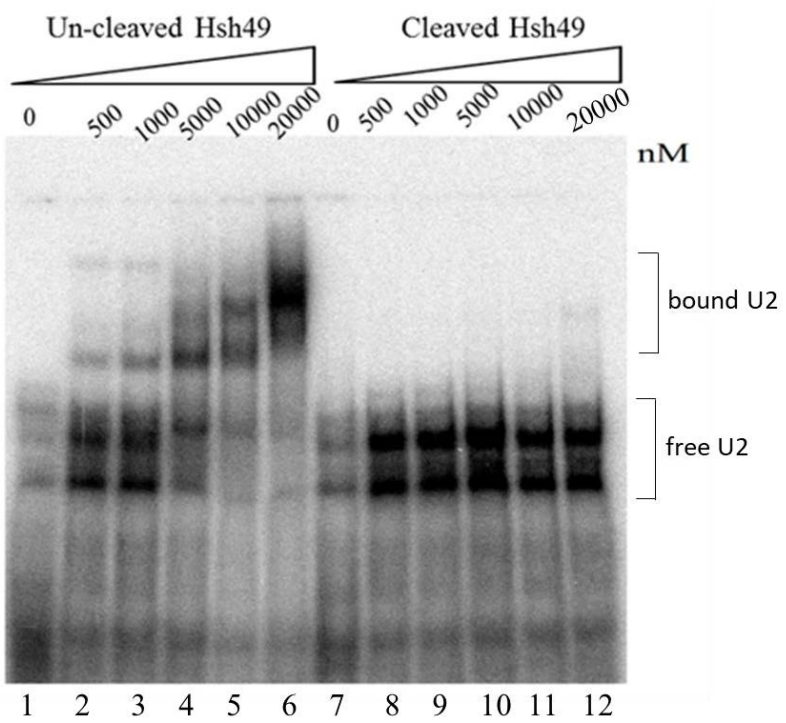


Figure 51. Electrophoretic mobility shift assay with cleaved and uncleaved recombinant Hsh49 and radioactive [32p]-labelled U2 snRNA. Cleaved and uncleaved Hsh49 concentrations are indicated on top and free U2 snRNA is shown in Lane 1 and Lane 7. The concentration of the U2 snRNA oligo was 20 nM. The U2 snRNA oligo was heated at 65 C for 3 minutes and cooled on ice for 2 minutes before incubation (incubation was at 25 C for 30 minutes). The gel ran for 90 minutes in voltage 200.

To insure if the U2 snRNA interacts with the uncleaved Hsh49 protein, I ran another EMSA to test the interaction of the U2 snRNA with the uncleaved Hsh49 and used the His-MBP tags as a negative control. Surprisingly, bands appeared in His-MBP tags lanes (negative control, Lanes 8–12) which were similar to the bands that were observed in the uncleaved Hsh49 Lanes 2–6 of Figure 52. This observation suggests that U2 snRNA does not associate with the uncleaved Hsh49, but it forms a complex with the His-MBP tags.

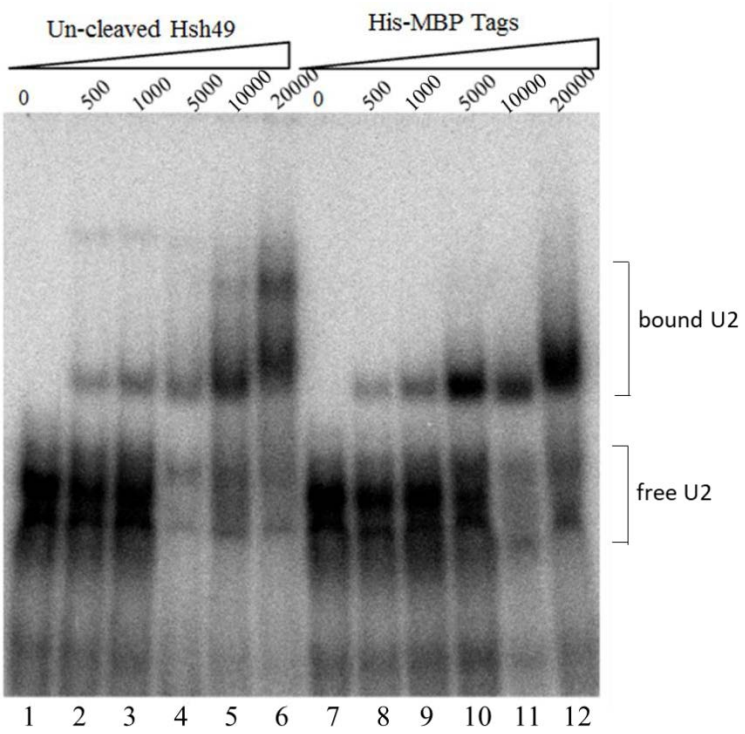


Figure 52. Electrophoretic mobility shift assay of uncleaved recombinant Hsh49 and His-MBP tags (negative control) with radioactive [32p]-labelled U2 snRNA. Uncleaved recombinant Hsh49 and His-MBP tags, concentrations are indicated on top and free U2 snRNA is shown in Lane 1 and Lane 7. The concentration of the U2 snRNA oligo was 20 nM. The U2 snRNA oligo was heated at 65 C for 3 minutes and cooled on ice for 2 minutes before incubation (incubation was at 25 C for 30 minutes). The gel ran for 90 minutes at 200 volts.

4.3.7 Co-Immunoprecipitation

4.3.7.1 Hsh49 Dot Blot Assay

Before doing co-immunoprecipitation, a dot blot assay was performed to determine the sensitivity and effectiveness of the polyclonal Hsh49 antibody which was raised against purified, recombinant Hsh49 protein (Figure 53). Based on the dot blot result, the Hsh49 antibody could detect the purified Hsh49 by comparing anti-Hsh49 antiserum with non-immune serum (pre-bleed).

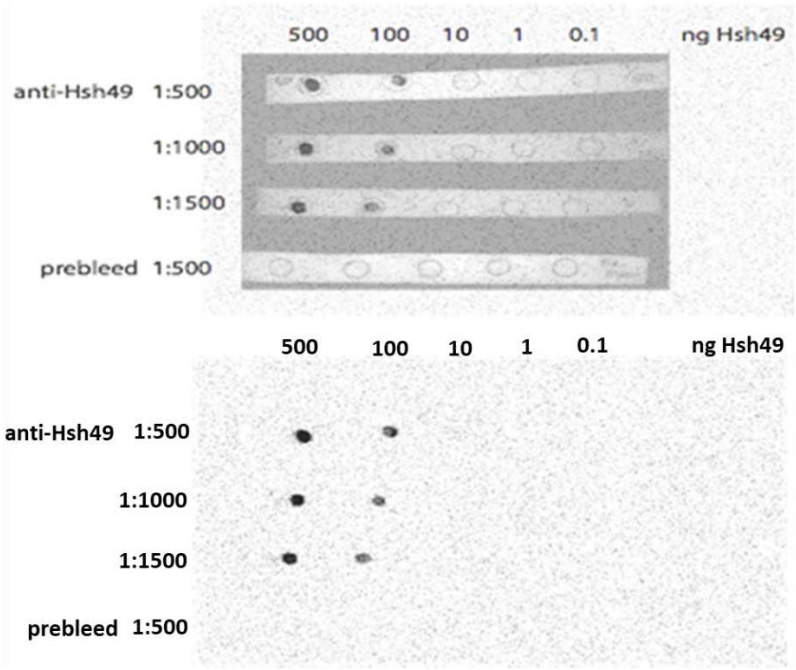


Figure 53. Dot blot test. Determining the sensitivity of the anti-Hsh49 antiserum generated against recombinant *C. merolae* Hsh49 protein.

4.3.7.2 Western Blot Assay

Also, a Hsh49 Western blot test was performed (Figure 54) to study the specificity of the anti-Hsh49 antiserum. Different concentrations of recombinant Hsh49 protein were probed with the anti-Hsh49 antiserum (primary antibody) and goat anti-rat-HRP antibody (secondary antibody). Western blot analysis showed that recombinant Hsh49 protein with the concentrations of 1µg and 100 ng could be detected by the anti-Hsh49 antiserum. A band of the expected size for Hsh49 (~14.2 kDa) was observed as shown in Figure 54. Also, some nonspecific bands appeared in the Western blot analysis which could be due to the high amount of total protein loaded on the gel and the lower band is related to the protein degradation.

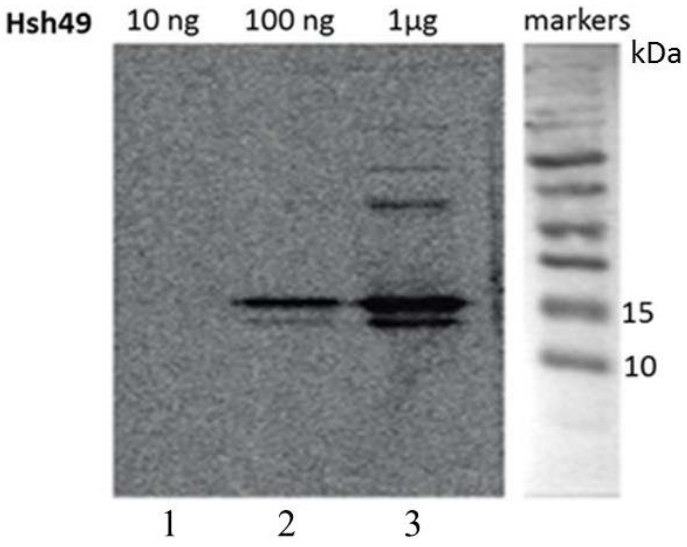


Figure 54. Specificity of the anti-Hsh49 antiserum via Western blot analysis using different concentrations of the recombinant *C. merolae* Hsh49.

4.3.7.3 Hsh49 Co-Immunoprecipitation

I needed to determine if the anti-Hsh49 antiserum—generated against the recombinant Hsh49 protein—could detect and immunoprecipitate Hsh49 protein from *C. merolae* extract. Immunoprecipitation of the Hsh49 allows for the removal and identification of other particles from the *C. merolae* extract that are associated with Hsh49. Based on Western blot analysis in Figure 55, the Hsh49 protein was immunoprecipitated with anti-Hsh49 antiserum as shown in lane 2, but not with non- immune serum (lane 3) as was expected. As the band in Lane 2 co-migrated with the recombinant Hsh49 (rHsh49), both bands indicating a size of ~14.2 kDa, I conclude that the anti-Hsh49 antiserum could immunoprecipitate the Hsh49 protein from *C. merolae* extract.

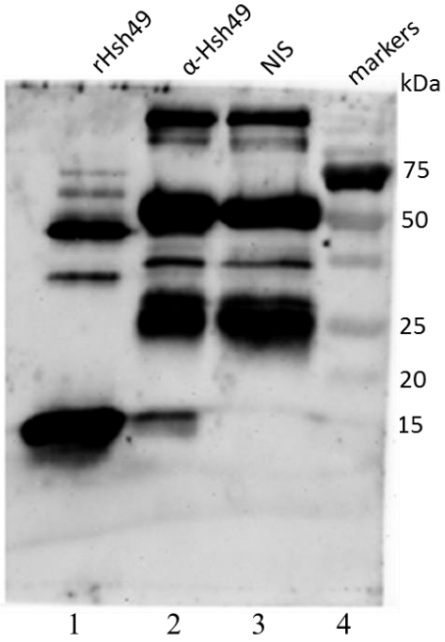


Figure 55. Western blot of Hsh49 *C. merolae* whole cell extract. Lane1, 25ng of the recombinant and purified Cm Hsh49 protein (rHsh49); Lane 2, immunoprecipitate from anti-Hsh49 antiserum (α-Hsh49) and Lane 3, immunoprecipitate from non-immune serum (NIS).

4.3.7.4 Northern Blot Analysis

Northern analysis of the immunoprecipitated RNA showed that no band was observed which might correspond to the U2 snRNA. By comparing the supernatant and immunoprecipitated pellets, no U2 snRNA was precipitated by the anti-Hsh49 antiserum from *C. merolae* extract (Figure 56).

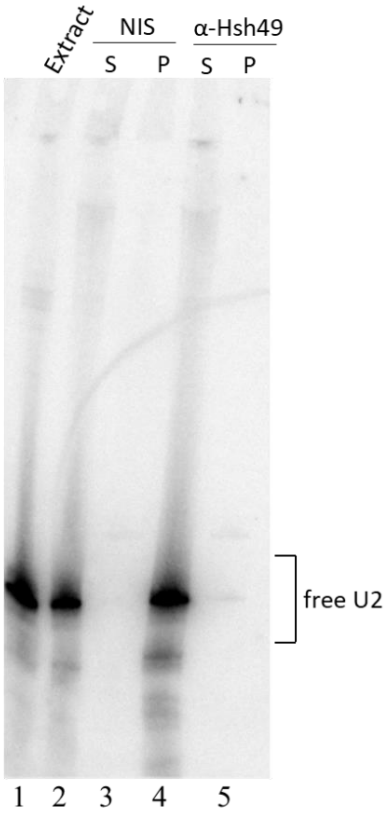


Figure 56. Northern blot analysis of the immunoprecipitated RNA probed for U2 snRNA. Lane1, total RNA from *C. merolae* whole cell extract; Lane 2–3, supernatant (S) and pellet (P) from non-immune serum which is the control and Lane 4–5, supernatant (S) and pellet (P) from the anti-Hsh49 antiserum.

4.4 Discussion

Successful expression and purification of the Hsh49 protein made it possible to pursue further investigations and research on the Hsh49 protein. To study the Hsh49 and U2 snRNA interaction, first, circular dichroism (CD) spectropolarimetry was performed. The CD spectra helped to determine if the Hsh49 is folded, and the secondary structure of the recombinant Hsh49 protein was also investigated. It is worth considering the possibility of

misfolding Hsh49 during the purification process. If Hsh49 misfolds or loses its secondary structure, this can affect the formation of a complex between the Hsh49-U2 snRNA.

Based on the CD spectra result, recombinant Hsh49 contains 91% α-helices and 0% β-sheets. Also, CD spectra showed the thermal stability of this protein over a temperature range of 28.79 to 84.44 C.

I compared the α-helix and β-sheet content of *C. merolae* Hsh49 with that of *Homo sapien* and *S. cerevisiae* Hsh49. The secondary structure of human Hsh49 consists of 19% α-helices and 23% β-sheets (Figure 57– Figure 60) while *S. cerevisiae* Hsh49 contains 23% α-helices and 26% β-sheets (Figure 61 and Figure 62). The content of α-helices and β-sheets in the secondary structure of Hsh49 in yeast and humans are close to each other.

Based on the CD spectra results, the content of α-helices and β-sheets in *C. merolae* is very different from yeast and humans. The sequence alignment of the Hsh49 protein in *C. merolae* demonstrated that there is less similarity to the *S. cerevisiae* and *Homo sapiens.* This result suggests that the Hsh49 does not have the same function in *C. merolae* as it has in *S. cerevisiae* and *Homo sapiens*. Therefore, I decided to perform the functional analyses.

As shown previously, the secondary structure of Hsh49 in *C. merolae* consists of 91% α-helices and 0% β-sheets (Figure 41). Hsh49 sequence alignment in *C. merolae* indicates it is 34% identical and 52% similar to the *S. cerevisiae* Hsh49. On the other hand, Hsh49 in *C. merolae* is 35% identical and 50% similar to the human Hsh49. Sequence alignment results indicate that the first sequences of *C. merolae* Hsh49 contain RRM1 and the second half matches contain RRM2. In addition, the results show that *Homo sapiens* Hsh49 RRM1 is more similar to *C. merolae* Hsh49 RRM1 than RRM2.

Therefore, based on the stability and folding of the recombinant Hsh49 protein that was obtained from CD spectra result, FA, EMSA and immunoprecipitation analysis were performed to characterize the Hsh49-U2 snRNA interaction.

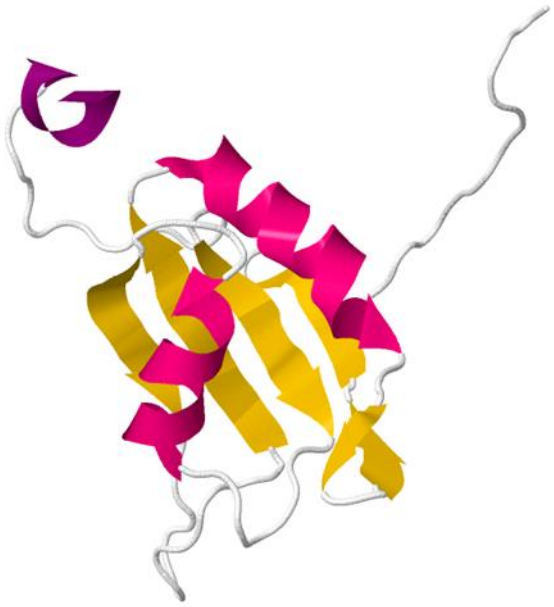


Figure 57. The crystal structure of the first RRM of *Homo sapiens* Hsh49 (PDB file 5GVQ; Kuwasako *et al*. 2017).

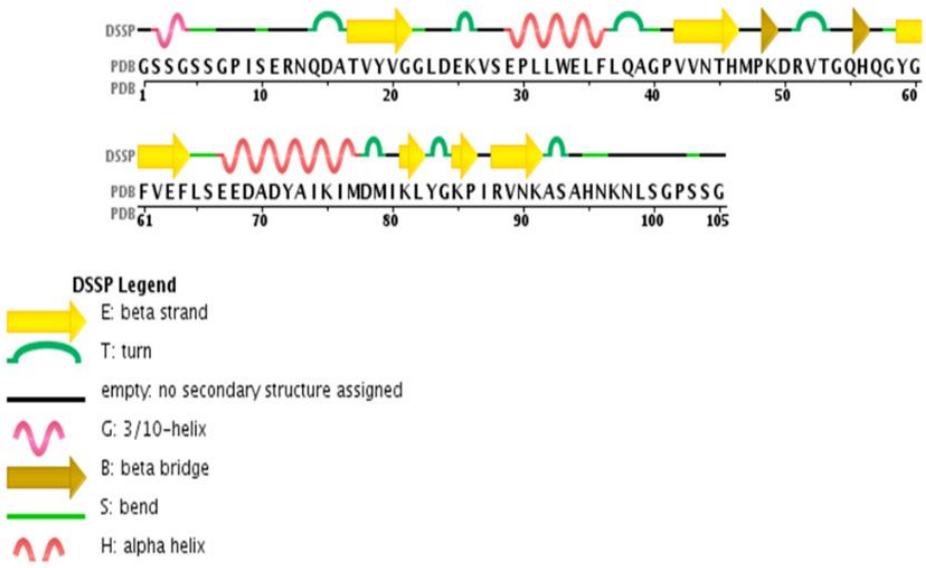


Figure 58. The secondary structure of the first RRM of *Homo sapiens* Hsh49 (PDB file 5GVQ; Kuwasako *et al*. 2017).

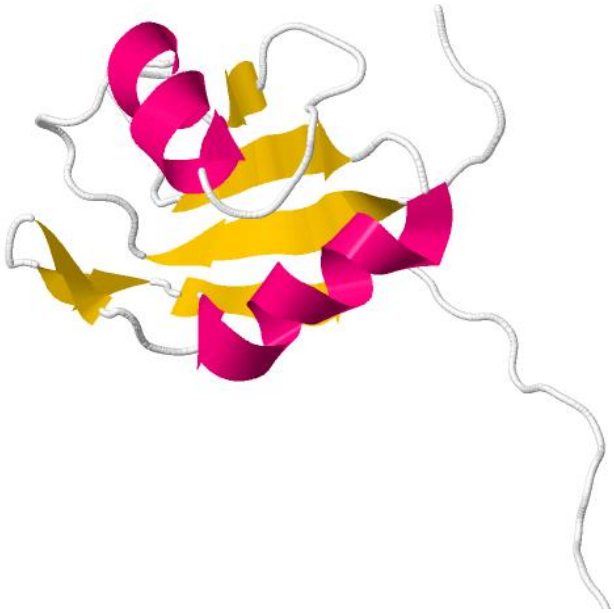


Figure 59. The crystal structure of the second RRM of *Homo sapiens* Hsh49 (PDB file 1X5T; Sato *et al*., to be published).

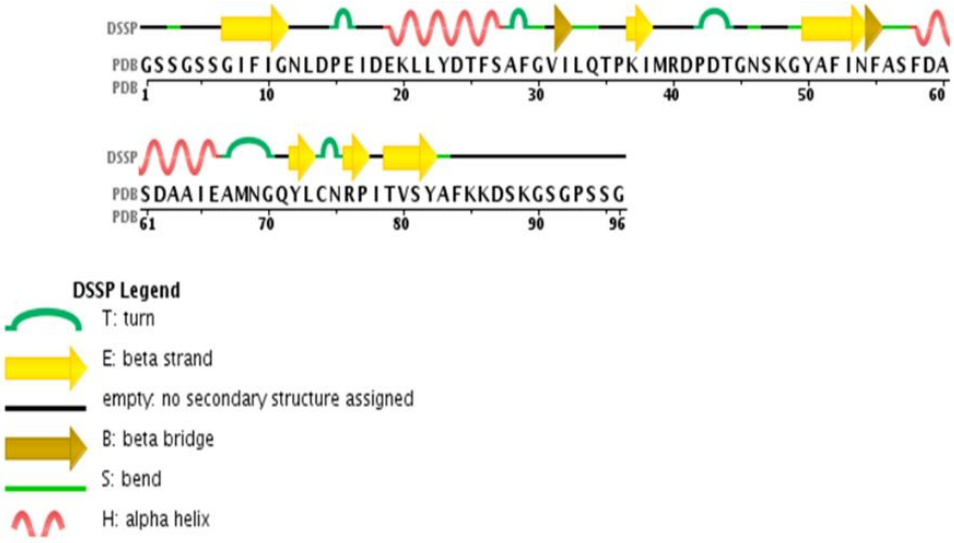


Figure 60. The secondary structure of the second RRM of *Homo sapiens* Hsh49 (PDB file 1X5T; Sato *et al*., to be published).

Figure 61. The Crystal structure of the *S. cerevisiae* Hsh49 protein (PDB file 5LSB; vanRoon *et al*. 2017).

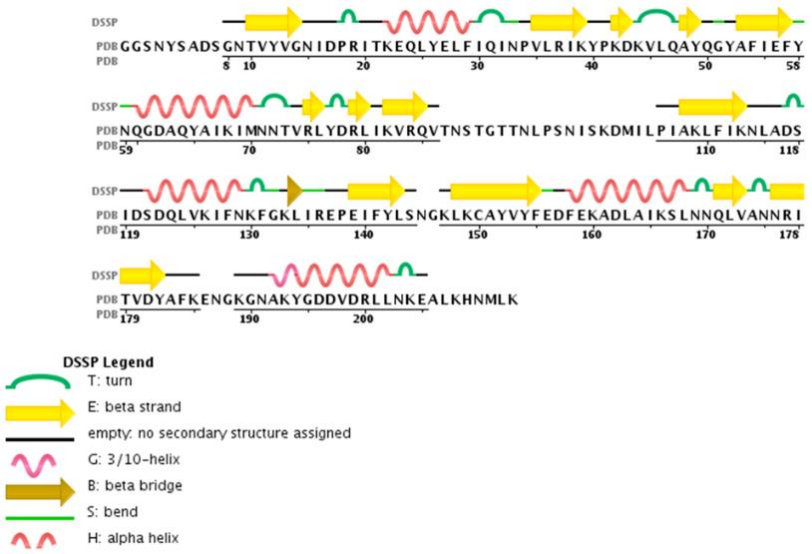


Figure 62. The secondary structure of the *S. cerevisiae* Hsh49 (PDB file 5LSB; vanRoon *et al*. 2017).

F-EMSA indicated that the fluorescent U2 oligo was not degraded; therefore, fluorescence anisotropy was performed and the results did not show any interaction between the Hsh49 and the fluorescein labelled 35 nucleotide U2 snRNA oligo in *C. merolae*. It is possible that the Hsh49 interacts with different sites of the U2 snRNA which does not include

the 35 oligonucleotides that were used in the FA assays. Therefore, I decided to study the interaction of the full-length U2 snRNA with the recombinant Hsh49 protein by EMSA to determine if any interaction occurs.

Based on previous studies that were done on Hsh49 originating from *Homo sapiens* and *S. cerevisiae* cells, it was shown that Hsh49 has an RNA binding activity and associates to the U2 snRNA. Igel *et al*. (1998) found that the *S. cerevisiae* GST-Hsh49 protein interacts with a 309-nt *in vitro* U2 snRNA transcript. This finding came from EMSA. Furthermore, *Homo sapiens* Hsh49 interacts with several regions of the full-length U2 snRNA which consists of the 5′ end of the U2 snRNA, and stem loops I and IIb (Dybkov *et al*. 2006). This is contrasted with my EMSA findings which showed that there was no interaction between the Hsh49-U2 snRNA in *C. merolae*. The FA and EMSA results confirm the Northern analysis which indicated that no interaction occurs between Hsh49-U2 snRNA. In all F-EMSA and EMSA results, free U2 snRNA oligo showed two bands in all lanes and these bands could be related to the conformational changes of the U2 snRNA.

Western blot analysis showed that the anti-Hsh49 antiserum which was generated against the recombinant Hsh49 could recognize and immunoprecipitate the Hsh49 from *C. merolae* extract. This result supports the successful recombination of the Hsh49 protein.

It was demonstrated that Hsh49 in yeast interacts with the Cus1 but the presence of the Cus1 is not essential for the U2 snRNA-Hsh49 interaction in yeast (Igel *et al*. 1998). It is possible that the pre-mRNA splicing in *C. merolae* occurs differently between yeast and human cells. Perhaps, the presence of Cus1 or some other factors are necessary for this interaction in *C. merolae*.

Chapter 5
General Discussion

C. merolae has a small genome, containing only 26 introns and a small number of core splicing proteins. In *C. merolae*, 43 splicing proteins are used compared to almost 90 in budding yeast and around 140 in *Homo sapiens*. Furthermore, *C. merolae* only contains four snRNAs including U2, U4, U5 and U6 snRNA, since no candidate for the U1- associated proteins or U1 snRNA were found (Stark *et al*. 2015). The reduced number of splicing proteins and introns enables *C. merolae* to be considered a simpler organism to study splicing. The U2 snRNP, with ten specific proteins and seven Sm proteins, is one of the most complex particle in *C. merolae.*

In order to better understand the molecular function and structural role of the U2 snRNP and SF3a complex, the first step is to generate an in vitro reconstitution system. The prerequisite step towards this approach was the expression and purification of these small nuclear ribonucleoproteins. In Chapter 2, my major attempt was to co-express and purify seventeen recombinant proteins. Current studies suggest that the co-expression of proteins results in stability of each protein and prevents protein degradation. In addition, protein co-expression increases the solubility and yield of protein (Romier *et al*. 2006; Stefan *et al*. 2015). Previous studies showed that both Lsm and Sm complexes co-expressed and copurified successfully (Dunn, 2014, unpublished). However, co-expression of the U2 snRNP and SF3a protein complex was not successful. Not all of the U2 snRNA proteins could be expressed and SF3a appeared as an insoluble particle.

As a result of the unsuccessful co-expression of the U2 snRNPs and SF3a protein complex, I decided to express and purify some U2 snRNA proteins individually. Chapter 3 is concerned with the expression of the Rds3 and SF3a protein to study the function and structure of these proteins. Of the SF3a and Rds3 genes which were cloned individually into the pQlink vector, only Prp9 and Prp21 were expressed. None of the SF3a genes which were cloned in the PMCSG23 vector were expressed. The Rds3 protein was expressed and purified well after its gene was cloned in the PMCSG23 vector. However, I could not remove the His and MBP tags completely from this protein, and the tags could later interfere with the protein interactions or with obtaining the crystal structure of the desired protein.

Chapter 4 describes the individual expression and purification of one of the U2 snRNA protein subunits, Hsh49. Since this protein could express and purify successfully, I performed some functional assays to determine the interaction between the Hsh49 protein and U2 snRNA. Before performing this step, I pursued CD spectra for two reasons: First to determine if the recombinant Hsh49 is folded, which would help me to do functional assays, and second, to assess the content of α-helices and β-sheets of the secondary structure of the Hsh49. The CD spectrum results indicated that the recombinant Hsh49 is folded and the secondary structure of the recombinant Hsh49 contains 91% α-helices, 0% β-sheets and 9% random coils.

Unexpectedly, no interaction was observed between the portion of the U2 snRNA oligonucleotide (35 nucleotide) and the recombinant Hsh49 via fluorescence anisotropy assay. Electrophoretic mobility shift assay (EMSA) did not show any interaction between the recombinant Hsh49 and full length IVT U2 snRNA oligonucleotide. The EMSA results were

confirmed by the northern blot analysis of the immunoprecipitated RNA, for which there was no band observed indicating Hsh49-U2 snRNA interaction.

The question that arises: is it possible that the Hsh49 was misidentified and it is not really the right protein? Sequence alignments of Hsh49 in *C. merolae* showed amino acids sequence identity with their *S.cerevisiae* and *Homo sapiens* homologs, with 12% and 9% respectively. Comprehensive bioinformatics survey of the *C. merolae* splicing machinery recognized four snRNAs (U2, U4, U5 and U6) and 43 core splicing proteins. In addition, BLAST searches using Reciprocal Best Hit methodology could identify all 10 U2 snRNA associated proteins including Hsh49 with E-values below the cutoff of 1E-13 (Stark *et al*. 2015). Importantly, 2′O-methyl- Mass spectrometry data confirmed the association of Hsh49 protein with U2 snRNA and some U2 snRNA proteins. This was performed using 2′O-methyl antisense oligonucleotide pull-downs against U2 snRNA of *C. merolae* extract. This experiment was followed by Northern analysis and mass spectrometry which showed that Hsh49 copurified with U2 snRNA (Reimer *et al*. 2017). These results provides confirmatory evidence that the Hsh49 is the right protein.

The data gathered in this study showed that the recombinant Hsh49 did not interact with the U2 snRNA. It may be possible that the interaction between the Hsh49 and the U2 snRNA does not occur directly. More specifically, the presence of one or a group of the U2 snRNA proteins are necessary to mediate this interaction. Previous studies showed that the Cus1 protein directly interacts with the Hsh49 protein in yeast through the first RRM and this interaction is essential for the tethering of the U2 snRNA to the branch point of the pre-mRNA (Igel *et al*. 1998; Champion-Arnaud *et al*. 1994).

Moreover, another subunit of the SF3b containing Hsh155 and Rse1 proteins interact with one another through the protein-protein interaction. These two proteins also associate with the Hsh49 and Cus1 protein complex. On the other hand, biochemical studies indicate that the U2 snRNA interacts with the Hsh155, Hsh49 and Cus1 in yeast (Champion-Arnaud *et al*. 1994).

5.1 Future Directions

In light of the fact that the interaction of Hsh49 to U2 snRNA might need other proteins of the SF3b complex, assembling U2 snRNP could help to detect the interaction of the Hsh49 protein with the U2 snRNA. One approach would be checking the interaction of the U2 snRNA-Hsh49 in the presence of the Cus1 protein. It seems that a deeper understanding of the U2 snRNA and Hsh49 protein is needed; therefore, other subunits of the SF3b complex should be studied.

Literature Cited

Abovich, N., Lia, X. C., and Rosbash, M. (1994). The yeast MUD2 protein: an interaction with PRP11 defines a bridge between commitment complexes and U2 snRNP addition. Genes and Development. 8: 843-854.

Abovich, N., and Rosbash, M. (1997). Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. Cell. 89: 403-412.

Arenas, J. E., and Abelson, J. (1997). Prpp43: An RNA helicase-like factor involved in spliceosome disassembly. Proc Natl Acad Sci USA. 94 (22): 11798–11802.

Ares, M., Jr. (1986). U2 RNA from yeast is unexpectedly large and contains homology to vertebrate U4, U5, and U6 small nuclear RNAs. Cell. 47: 49-59.

Black, C. S., Garside, E., L, MacMillan, A., M, Rader, S., D. (2016). Conserved structure of Snu13 from the highly reduced spliceosome of *Cyanidioschyzon merolae*. Protein Science, 25: 911-916.

Brosi, R., Groning, K., Behrens, S. E., Luhrmann, R., Kramer, A. (1993). Interaction of mammalian splicing factor SF3a with U2 snRNP and relation of its 60-kD subunit to yeast PRP9. Science, 262 (5130): 102-105.

Brow, D. A. (2002). Allosteric cascade of spliceosome activation, Annu. Rev. Genet, 36: 333-360.

Caspary, F., and Seraphin, B. (1998). The yeast U2A/U2B complex is required for pre-spliceosome formation. EMBO Journal, 17(21): 6348-6358.

Chabot, B., Black, D. L., LeMaster, D. M., and Steitz, J. A. (1985). The 3′ splice site of pre-messenger RNA is recognized by a small nuclear ribonucleoprotein. Science, 230 (4732): 1344-1349.

Champion-Arnaud, P., and Reed, R. (1994). The prespliceosome components SAP 49 and SAP 145 interact in a complex implicated in tethering U2 snRNP to the branch site. Genes and Development. 8: 1974-1983.

Chang, T. H., Tung, L., Yeh, F. L., Chen, J. H., Chang, S. L. (2013). Functions of the DExD/H-box proteins in nuclear pre-mRNA splicing. Elsevier, 1829(8): 764-774.

Chang, J., Schwer, B., and Stewart Shuman, S. (2012). Structure function analysis and genetic interactions of the yeast branchpoint binding protein Msl5. Nucleic Acids Research, 40(10): 4539-4552.

Chiou, N. T., and Lynch, K. W. (2014). Mechanism of spliceosomal assembly. In Hartel, K. J (Eds), Spliceosomal pre-mRNA splicing (pp. 35-43). Heidelberg New York: Springer.

Combs, D. J., Nagel, R. J., Ares, Jr. M., Stevens, W. S. (2006). Prp43p Is a DEAH-Box Spliceosome Disassembly Factor Essential for Ribosome Biogenesis. Mol Cell Biol. 26(2): 523-534.

Cretu, C., Schmitzová, J., Ponce- Salvetierra, A., Dybkov, O., De Laurentiis, E. I., Sharma, K., Will, C. L., Urlaub, H., Lührmann, R.,Vladimir Pena, V. (2016) Molecular Architecture of SF3b and structural consequences of its cancer-related mutations. Molecular and Cell Biology, 64(2): 307-319.

Cvitkovic and Jurica, Nucleic Acids Research [Internet]. c2013. [cited 2015 June 10]. Available from: Spliceosomedb.ucsc.edu/proteins.

Draper, D. E. (2004). A guide to ions and RNA structure. RNA Society, 10(3):335–343.

Donmez, G., Hartmuth, K., Lührmann, R. (2004). Modified nucleotides at the 5' end of human U2 snRNA are required for spliceosomal E-complex formation. RNA Society, 10(12):1925-1933.

Dunn EA. 2014. “Investigation of free U6 small nuclear ribonucleoprotein structure and function.” Ph.D. thesis, University of British Columbia, Vancouver.

Dybkov, O., Will, C. L., Deckert, J. Behzadnia, N., Hartmuth, K., and Reinhard Luhrmann. (2006). U2 snRNA-protein contacts in purified human 17S U2 snRNPs and in spliceosomal A and B complexes. Molecular and Cellular Biology. 26(7): 2803-2816.

Dyson, M. R. (2007). Expression strategy. In Dyson, M. R., and Durocher, Y (Eds.), Expression systems (pp. 1-11). Bioxham, Oxforshire: Scion.

http://www.ebi.ac.uk/Tools/psa/emboss_needle/

Fabrizio, P. Dannenberg, J., Dube, P., Kastner, B., Stark, H., Urlaub, H., Lührmann, R. (2009). The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. Molecular and Cellular Biology, 36(4): 593-608.

Faustino, N. A., and Cooper, T. A. (2003). Pre-mRNA splicing and human disease. Cold Spring Harbor Laboratory Press, Gene and Development, 17(4): 419-437.

Golas, M. M., Sander, B., Will, C. L., Lührmann, R., Stark, H. (2003). Molecular architecture of the multiprotein splicing factor SF3b. Science. 300(5621): 980-984.

Gozani, O., Feld, R., Reed, R. (1996). 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 and Development, 10: 233-243.

Harlow, Ed., Lane, D. (1988). Antibodies: A laboratory manual. New York: Cold spring harbor laboratory, Chapter 5.

Held, P., Jolley, M. E. (2008). Fluorescence polarization detection of Brucella abortus and Mycobacterium bovis specific antibodies using kits from Diachemix, BioTek Instruments, Inc.

Hoskins, A. A., Rodgers, M. L., Friedman, L. J., Gelles, J., Moore, M. J. (2016). Single molecule analysis reveals reversible and irreversible steps during spliceosome activation. eLIFE Research article, 5: 1-22.

Howe, C. (2007). Gene cloning and manipulation. New York: Cambridge university press. Chapter 8.

Igel, H., Wells, S., Perriman, R., and Ares, M. (1998). Conservation of structure and subunit interactions in yeast homologues of splicing factor 3b (SF3b) subunit. RNA society, 4(1): 1-10.

Igel, A. H. and M. Ares, Jr. (1988). Internal sequences that distinguish yeast from metazoan U2 snRNA are unnecessary for pre- mRNA splicing. Nature 334: 450-453.

Juneau, K., Curtis Palm, C., Miranda, M., and Davis, R. W. (2007). High-density yeast-tiling array reveals previously undiscovered introns and extensive regulation of meiotic splicing. National Academy of Sciences, 104(5): 1522-1527.

Jurica, M. S., and Moore, M. J. (2003). Pre-mRNA splicing: awash in a sea of proteins. Molecular and Cellular Biology 12(1): 5-14.

Kakehi, K., Oda, Y., and Kinoshita, M. (2001). Fluorescence Polarization: analysis of carbohydrate-protein interaction. Analytical Biochemistry, 297(2): 111-116.

Kholod, N., and Mustelin, T. (2001). Novel vectors for co-expression of two proteins in E. coli. Biotechniques, 31(2): 322-323, 326-328.

Kim, R. (2007). Protein expression in Escherichia coli. In Dyson, M. R., and Durocher, Y (Eds.), Expression systems (pp. 13-28). Bioxham, Oxforshire: Scion.

Kistler, Al., Guthrie, C. (2001). Deletion of Mud2, the yeast homolog of U2AF65, can bypass the requirement for Sub2, an essential spliceosomal ATPase. Genes and Development.15: 42-49.

Kramer, A. (1988). Presplicing complex formation requires two proteins and U2 snRNP. Genes and Development, 2(9): 1155-1167.

Kramer, A., Mulhauser, F., Wersig, C., Groning, K., Bilbe, G. (1995). Mammalian splicing factor SF3a120 represents a new member of the SURP family of proteins and is homologous to the essential splicing factor PRP21p of *Saccharomyces cerevisiae*. RNA, 1(3): 260-272.

Kuwasako, K. Nameki, N. Tsuda, K. M., Takahashi, M. Sato, A. Tochio, N. Inoue, M. Terada, T. Kigawa, T. Kobayashi, N. Shirouzu, M. Ito, T. Sakamoto, T. Wakamatsu, K. Güntert, P. Takahashi, S.Yokoyama, S., and Muto. Y. (2017). Solution structure of the first RNA recognition motif domain of human spliceosomal protein SF3b49 and its mode of interaction with a SF3b145 fragment. Protein science, 26(2): 280-291.

Lardelli, R. M., Thompson, J. X., Yates, J. R., and Stevens, S. W., (2010). Release of SF3 from the intron branchpoint activates the first step of pre-mRNA splicing. RNA biology, 16(3):516-528.

Lea, W. A., and Simeonov, A. (2011). Fluorescence polarization assays in small molecule screening. Expert Opin Drug Disco. 6(1): 17-32.

Legrain, P., and Chapon, C. (1993). Interaction between PRP11 and SPP91 yeast splicing factors and characterization of a PRP9-PRP11-SPP91 complex. Science, 262(5130): 108-110.

Liang, W-W., and Cheng, S-C. (2015). A novel mechanism for Prp5 function in prespliceosome formation and proofreading the branch site sequence. Gene and Development, 29(1): 81-93.

Lin, P. C., and Xu, R. M. (2012). Structure and assembly of the SF3a splicing factor complex of U2 snRNP. EMBO J, 31(6): 1579-1590.

Lopez-Bigas, N., Audita, B., Ouzounisa, C., Parrab, G., Guigo, R. (2005). Are splicing mutations the most frequent cause of hereditary disease?. Federation of European Biochemical Societies, 579(9): 1900-1903.

Louis-Jeune, C., Andrade-Navarro, M., Perez-Iratxeta, C. (2011). Prediction of protein secondary structure from circular dichroism using theoretically derived spectra. Proteins, Structure, Function, and Bioinformatics. 80(2): 347-381.

Martin, A., Schneider, S., and Schwer, B. (2002). Prp43 Is an Essential RNA-dependent ATPase Required for Release of Lariat-Intron from the Spliceosome, J Biol Chem, 277(20): 17743-17750.

Matsuzaki, M., Misumi, O., Shin-i, T., Maruyama, S., Takahara, M., Miyagishima, S. M., Mori, T., Keiji Nishida, K., Yagisawa, F., Nishida, K., Yoshida, Y., Nishimura, Y., Nakao, S., Kobayashi, T., Momoyama, Y., Higashiyama, T., Minoda, A., Sano, M., Nomoto, S., H., Oishi, K., Hayashi, H., Ohta, F., Nishizaka,S., Haga,S., Miura, S., Morishita, T., Kabeya, Y., Terasawa, K., Suzuki, Y., Ishii, Y., Asakawa, S., Takano, H., Ohta, N., Kuroiwa,H., Tanaka, K., Shimizu, N., Sugano, S., Sato, N., Nozaki, H., Ogasawara, N., Kohara, Y., and Kuroiwa, T. (2004). Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D, Nature, 428(6983): 653-657.

Massenet, S., Mougin, A., Branlant, C. (1998). Post-transcriptional modifications in the U small nuclear RNAs. In Modifications and editing of RNA (eds. H. Grosjean and R. Benne), (pp. 201-228). ASM Press, Washington, DC.

Misra, V., K. Draper, D., E. (1998). On the role of magnesium ions in RNA stability. Nucleic Acid Sciences, 48(2): 113-135.

Misumi, O., Matsuzaki, M., Nozaki, H., Miyagishima, S. Y., Mori, T., Nishida, K., Yagisawa, F.,Yoshida, Y., Kuroiwa, H., and Kuroiwa, T. (2005). *Cyanidioschyzon merolae* Genome. A tool for facilitating comparable studies on organelle biogenesis in photosynthetic eukaryotes. Plant Physiology, 137(2): pp. 567-585.

Moll, J., R. Acharya, A. Gal, J. Mir, A., A. and Vinson, C. (2002). Magnesium is required for specific DNA binding of the CREB B-ZIP domain. Nucleic Acids Research. 30(5): 1240-1246.

NCBI Gene [Internet]. [date unknown]. National Center for Biotechnology Information [NCBI]; [cited 2015 June 10]. Available from: http://www.ncbi.nlm.nih.gov/gene/

Nilsen, T.W. (1998). RNA-RNA interactions in nuclear pre-mRNA splicing. In RNA Structure and Function, M. Grundber-Manago and R.W. Simons (Eds.), Cold Spring Harbor Laboratory Press, 35: 279-307.

Noble, J. S. C., Prives, C., and Manley, J. L. (1988). Alternative splicing of SV40 early pre-mRNA is determined by branch site selection. Genes and Development, 2(11): 1460-1475.

Nozaki, H., Takano, H., Misumi, O., Terasawa, K., Matsuzaki, M., Maruyama, S., Nishida, K., Yagisawa, F., Yoshida, Y., Fujiwara, T., Takio, S., Tamura, K., Sung Jin Chung, S. J., Nakamura, S., Kuroiwa, H., Tanaka, K., Sato, N., and Kuroiwa, T. (2007). A 100%-complete sequence reveals unusually simple genomic features in the hot-spring red alga *Cyanidioschyzon merolae*. 5(28): 1-8.

http://nucleobytes.com/index.php/4peaks

O'Day, C. L., Dalbadie, M. G., Abelson, J. (1996). The saccharomyces cerevisiae Prp5 protein has RNA-dependent ATPasae activity with specificity for U2 small nuclear RNA. Biol. Chem. 271: 33261-33267.

Parker, R., Siliciano, P. G., and Guthrie, C. (1987). Recognition of the TACTAAC box during mRNA splicing in yeast involves base pairing to the U2-like snRNA. Cell, 49(2): 229-239.

Pagano, J., M. Clingman, C., C. and Ryder, S., P. (2011). Quantitative approaches to monitor protein–nucleic acid interactions using fluorescent probes. RNA society. 17(1): 14-20.

Patel, A. A., and Joan A. Steitz, J. A. (2003). Splicing double: insights from the second spliceosome. Mol. Cell, 4: 960-970.

Pauling, M. H., McPheeters, D, S., Ares, Jr, M. (2000). Functional Cus1p is found with Hsh155p in a multiprotein splicing factor associated with U2 snRNA. Molecular and Cellular Biology, 20(6): 2176-2185.

Pei, J., Kim, B. H., and Grishin, N. V. (2008). PROMALS3D: a tool for multiple sequence and structure alignment. Nucleic Acids Res. 36(7), 2295-2300.

Perrin, F. (1926) Polarization de la Lumiere de Fluorescence. Vie Moyenne de Molecules dans L'etat Excite. J. Phys Radium. 7(12): 390-401.

Rymond, B. C. and Rosbash, M. (1985). Cleavage of 5' splice site and lariat formation are independent of 3' splice site in yeast mRNA splicing. Nature 317: 735-737.

Rymond, B. C., Torrey, D. D., Rosbash. M. (1987). A novel role for the 3' region of introns in pre-mRNA splicing of *Saccharomyces cerevisiae*. Genes Dev. 1: 238-246.

Reimer, K. A., Stark, M. R., Aguilar, L-C., Stark, S. R., Burke, R. D., Moore, J., Fahlman, R. P., YIP, C. K., Kuroiwa, H., Oefinger, M., and Rader, S., D. (2017). The sole LSm complex in *Cyanidioschyzon merolae* associates with pre-mRNA splicing and mRNA degradation factors. RNA society, 23:1-16.

Romier, C., Ben Jelloul, M., Albeck, S., Buchwald, G., Busso, D., Celie, P. H., Christodoulou, E., De Marco, V., van Gerwen, S., Knipscheer, P., Lebbink, J. H., Notenboom, V., Poterszman, A., Rochel, N., Cohen, S. X., Unger. T., Sussman, J. L., Moras, D., Sixma, T. K., Perrakis, A. (2006), Co-expression of protein complexes in prokaryotic and eukaryotic hosts: experimental procedures, database tracking and case studies. Biological crystallography, 62: 1232-1242.

Ruskin, B., Zamore, P. D., and Green, M.R. (1988). A factor, U2AF is required for U2 snRNP binding and splicing complex assembly. Cell, (52): 207-219.

Rutz, B., and Seraphint, B. (1999).Transient interaction of BBP/ScSF1 and Mud2 with the splicing machinery affects the kinetics of spliceosome assembly. RNA society. 5: 918-831.

Saccharomyces Genome Database [Internet]. [date unknown]. Stanford (CA): Stanford University; [cited 2015 June 10]. Available from: http://www.yeastgenome.org

Sashital, D. G., Cornilescu, G., McManus, C. J., Brow, D. A., Butcher, S. E. (2004). U2–U6 RNA folding reveals a group II intron-like domain and a four-helix junction. Nat. Struct. Mol. Biol. 11: 1237–1242.

Sashital, D. G., Venditti, V., Angers, C. G., Cornilescu, G., and Butcher, S. E. (2007). Structure and thermodynamics of a conserved U2 snRNA domain from yeast and human. Cold Spring Harbor Laboratory Press, 13(3): 328-338.

Sato, A. et al. (no date) 'Solution structure of the second RRM domain in splicing factor = 3B', to be Published. doi: 10.2210/PDB1X5T/PDB.

Scheich, C., Kummel, D., Soumailakakis, D., Heinemann, U., and Bussow, K. (2007). Nucleic Acids Research, 35(6): 1-7.

Schmid-Burgk, J. L., Xie, Z., and Benenson, Y. (2014). Hierarchical ligation-independent assembly of PCR fragments. Springer science+business media, 1116: 49-58.

Staknis, D., and Reed, R. (1994). Direct Interactions between Pre-mRNA and Six U2 Small Nuclear Ribonucleoproteins during Spliceosome Assembly. Molecular and Cellular Biology. 14(5): 2994-3005.

Stark, H., Dube, P., Lührmann, R., Kastner, B. (2001). Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle. Nature, 409 (6819): 539-42.

Stark, M. R., Dunn, E. A.,Dunn, W. S. C., Grisdalec, C. J., Danielea, A. R., Halsteada, M. R. G., Fast, N. M., Rader, S. D. (2015). Dramatically reduced spliceosome in *Cyanidioschyzon merolae*. PNAS, 112(11): 1-10.

Stefan, A., Ceccarelli, A., Conte, E., Monton, S. A., Hochkoeppler, A. (2015). The multifaceted benefits of protein co-expression in *Escherichia coli*. Journal of Visualized Experiment. 5(96): 1-9.

Sun, J. S. and Manley, J. L. (1995). A novel U2-U6 snRNA structure is necessary for mammalian mRNA splicing. Genes & Dev. 9: 843–854.

Suzuki, K., Ohta, N. and Kuroiwa, T. (1992). lsolation of the cell-nuclear, mitochondrial, and chloroplast DNA from the ultrasmall eukaryote *Cyanidioschyzon merolae*. Protoplasma, 171(1-2): 80-4.

Tan, S., Kern, R. C., and Selleck,W. (2005). The pST44 polycistronic expression system for producing protein complexes in Escherichia coli. Protein Expr. Purif. 40(2): 385-395.

Tauchert, M. J., Fourmann, J-B., Christian, H., Luhrmann, R., Ficner, R. (2016) structural and functional analysis of the RNA helicase Prp43 from the thermophilic eukaryote chaetomium thermophilum. Structural Biology Communications. 72: 112-120.

https://thebiogrid.org/

www.uniprot.org

van Roon, A. M., Oubridge, C., Obayashi, E., Sposito, B., Newman, A. J., Seraphin, B. and Nagai, K. (2017) 'Crystal structure of U2 snRNP SF3b components: Hsh49p in complex with Cus1p binding domain', RNA, p. rna.059378.116. doi: 10.1261/rna.059378.116.

Wang, Q., He, J., Lynn, B., and Rymond, B. C. (2005). Interactions of the yeast SF3b splicing factor. Molecular and Cellular Biology, 25(24): 10745-10754.

Wang, Q., and Rymond, B. C. (2003). Rds3p is required for stable U2 snRNP recruitment to the splicing apparatus. Molecular and Cellular Biology. 23(20): 7339-7349.

Wahl, M. C.,Will, C. L., and Luhrmann, R. (2009). The Spliceosome: Design Principles of a Dynamic RNP Machine, Cell, 136, 701-718.

Wiest, D. K., O'Day, C. L., and Abelson, J. (1996).Vitro studies of the Prp9. Prp11. Prp21 complex indicate a pathway for U2 small nuclear ribonucleoprotein activation. Biological Chemistry, 271(52): 33268-3327.

Will, C. L., Urlaub, H., Achsel, T., Gentzel, M., Wilm, M., Luhrmann, R. (2002). Characterization of novel SF3b and 17S U2 snRNP proteins, including a human Prp5p homologue and an SF3b DEAD box protein. EMBO J, 21: 4978-4988.

Will, C. L., and Lührmann, R. (2011). Spliceosome Structure and Function. Cold Spring Harbor Laboratory Press, 3(7): 1-23.

Wu, S., C. Romfo, T. Nilsen, and M. Green. (1999). Functional recognition of the 3′ splice site AG by the splicing factor U2AF35. Nature, 402: 832-835

Wu, J., and Manley, J. L. (1989). Mammalian pre-mRNA branch site selection by U2 snRNP involves base pairing. Genes and Development. Cold Spring Harbor Laboratory Press, 3, 1553-1561.

Wu, T., and Fu, X. D. (2015). Genomic functions of U2AF in constitutive and regulated splicing. RNA Biology, 12(5):479-785.

Yan, D., Perriman, R., Igel, H., Howe, K. J., Nevile, M., and JR. M. A. (1998). CUS2, a yeast homolog of human Tat-SF1, rescues function of misfolded U2 through an unusual RNA recognition motif. Molecular and Cellular Biology, 18(9): 5000-5009.

http://www.yeastgenome.org/locus/S000002201/interaction

Zhuang, Y., and Weiner, A. M. (1989). A compensatory base change in human U2 snRNA can suppress a branch site mutation. Genes and Development. Cold Spring Harbor Laboratory Press, 3: 1545-1552.