

**STRUCTURAL AND FUNCTIONAL STUDIES INTO BRANCHPOINT
SEQUENCE RECOGNITION AND THE U4/U6 DI-SNRNP IN
*CYANIDIOSCHYZON MEROLAE***

by

Corbin Black

B.Sc., University of Northern British Columbia, 2014

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN
BIOCHEMISTRY

UNIVERSITY OF NORTHERN BRITISH COLUMBIA

July 2018

© Corbin S. Black, 2018

Abstract

Pre-mRNA splicing ensures mRNA transcripts contain contiguous coding sequences prior to protein synthesis. 5' splice site recognition of pre-mRNA is unknown in *Cyanidioschyzon merolae*, because it lacks the U1 snRNP. Conservation among the core splicing proteins is low, and there is potential for novel adaptations. Similar to yeast, branchpoint sequence recognition appears driven by Msl5 and stabilized by Mud2. Electrophoretic mobility assays showed Msl5, with its canonical KH-QUA2 domain, specifically binds to the conserved branchpoint sequence. Mud2 is a novel *ScMud2/U2AF65* homolog. Conversely, U4/U6 di-snRNP assembly appears conserved. Fluorescence polarization assays with the U4 5' stem loop and the crystal structure of Snu13 revealed a conserved interaction. Snu13 and the Sm heptamer were able to bind base-paired U4/U6. Finally, circular dichroism spectroscopy and *in silico* mapping of intrinsic disorder found several proteins were partially unstructured; this would allow a minimal spliceosome to functionally fulfill the roles of otherwise necessary proteins.

Table of Contents

Abstract	II
Table of Contents	III
List of Tables	VI
List of Figures	IX
List of Equations	XIII
Acknowledgements	XIV
Table of Abbreviations	XV
References	i

Chapter 1: Introduction

1.1 Pre-mRNA Splicing and the Spliceosome.	1
1.2 <i>Cyanidioschyzon merolae</i> as a Model Organism.	3
1.3 Initiation of Pre-mRNA Splicing.	5
1.4 Bridging the Branchpoint Sequence to the 5' Splice Site.	7
1.5 The Branchpoint Bridging Protein Msl5 and U2 Auxiliary Factor Mud2.	8
1.6 U4/U6 Di-snRNP in Pre-mRNA Splicing.	11
1.7 Snu13 In di-snRNP Assembly.	15
1.8 Research Goals.	16

Chapter 2: Recognition of the pre-mRNA branch site by Msl5

2.1 Methods – Expression and Purification of <i>C. merolae</i> Msl5 and Mud2 Proteins.	18
2.2 Methods – Fluorescence Polarization of Msl5 Against Branchpoint RNA.	26
2.3 Methods – <i>In vitro</i> Transcription of Truncated <i>Cm</i> 072C RNA.	29

2.4 Methods – EMSAs of Msl5 with Various RNA.	33
2.5 Methods - Circular Dichroism Spectroscopy of Msl5 with <i>Cm</i> BPS RNA.	34
2.6 Results and Discussion.	36
Chapter 3: Functional and Structural Characterization of Snu13	
3.1 Methods - Expression and Purification of <i>C. merolae</i> Snu13 Protein.	74
3.2 Methods - Fluorescence Polarization of Purified Snu13.	77
3.3 Methods - Bioinformatics and Structural Modeling.	78
3.4 Methods - X-ray Crystallography of Snu13.	78
3.5 Results and Discussion.	79
Chapter 4: Assembly of the <i>Cm</i> U4/U6 di-snRNP	
4.1 Methods – Preparation of Proteins Used in U4/U6 di-snRNP Assembly.	97
4.2 Methods – <i>In vitro</i> Transcription of <i>Cm</i> U4 and U6 snRNA.	98
4.3 Methods – Assembly of the U4/U6 di-snRNP.	100
4.4 Methods – Recombinant Expression of <i>Cm</i> Proteins Prp3, Prp4, and Prp31.	102
4.5 Methods – Protein Disorder Predictions.	110
4.6 Results and Discussion.	110
Chapter 5: General Discussion and Conclusion	
5.1 Improving the yield of the Msl5/Mud2 dimer.	126
5.2 Role of the branchpoint bridging protein Msl5.	126
5.3 Bridge between Msl5 and the 5' splice site.	127
5.4 Role of the U2 Auxiliary Factor Mud2 in <i>C. merolae</i> .	128

5.5	Displacement of Msl5/Mud2 from the pre-mRNA.	129
5.6	Prp4 connects Snu13 and the 5' Stem Loop with Stem II.	129
5.7	Assembly of the U4/U6 di-snRNP.	130
5.8	Chaperones and solubility factors of thermophilic organisms.	130
5.9	Intrinsically disordered regions and splicing proteins.	131
5.10	Conclusion.	132

List of Tables

Table 1.1	Common names for the branchpoint bridging protein and U2 auxiliary factors in various model organisms.	9
Table 2.1	DNA primers used in the subcloning of Msl5 and Mud2 genes into pQLinkH plasmid vector.	19
Table 2.2	Composition of buffers used in the Ni-NTA purification of recombinant Msl5 and co-expressed Mud2/Msl5 from Rosetta pLysS (DE3) cells.	22
Table 2.3	Composition of buffers used in the cation exchange chromatography purification of recombinant Msl5.	24
Table 2.4	Composition of the buffer used in the size exclusion chromatography purification of recombinant Msl5 and co-expressed Mud2 and Msl5.	24
Table 2.5	Reagents used in the detection of Msl5 in purified Mud2/Msl5 dimer.	25
Table 2.6	Fluorescein-labeled RNA oligo probes used in FP Binding Assays.	27
Table 2.7	FP reagents used for the binding experiments of Msl5 with BP RNA Probe.	28
Table 2.8	Sequences of the <i>Cm</i> 072C transcripts used in binding experiments with Msl5.	29
Table 2.9	Reactions and thermocycler conditions used to generate the DNA template for <i>Cm</i> 072C pre-mRNA and mRNA.	30
Table 2.10	PCR primers used to generate <i>Cm</i> 072C DNA templates for <i>in vitro</i> transcription and their respective annealing temperatures.	30
Table 2.11	Composition of the IVT reactions that generated <i>Cm</i> 072C RNA transcripts.	31
Table 2.12	Composition of the 5' end-labeling reaction with [γ - 32 P]-labeled ATP.	32
Table 2.13	EMSA reagents used for the binding experiments of Msl5 with 072C RNA.	33

Table 2.14	Parameters used for far UV wavelength scans of Msl5.	34
Table 2.15	Parameters used for thermal denaturation of Msl5.	35
Table 2.16	Ratios for wavelengths 260:280 (nm) of protein and RNA.	38
Table 2.17	Binding affinities of Msl5 homologs for the branchpoint sequence of pre-mRNA.	49
Table 2.18.	Binding affinities of Msl5 for RNA oligonucleotides measured by FP.	52
Table 2.19	Binding affinities of Msl5 for RNA oligonucleotides measured by EMSA.	54
Table 2.20	Binding affinities of Msl5 for various <i>Cm</i> 072C Nascent transcripts.	62
Table 2.21	Secondary structure of Msl5 by Circular Dichroism and predictive software.	65
Table 2.22	Dissociation constants and Hill Coefficients for FP binding assays of the Msl5/Mud2 dimer against the <i>Cm</i> BPS.	71
Table 3.1	Primers used to clone Snu13 from the <i>Cm</i> genome into pMCSG7.	74
Table 3.2	Complete list of the buffers used in the purification of Snu13.	75
Table 3.3	Fluorescein-labeled RNA oligonucleotide probes used in FP Binding Assays.	77
Table 3.4.	FP reagents used for the binding experiments of Msl5 with 6 BP RNA probe.	77
Table 3.5	UniProt and RCSB Accession numbers for Snu13 homologs.	78
Table 3.6.	Data collection, phasing, and refinement statistics.	85
Table 3.7	Comparison of Protein-RNA Affinities.	94
Table 4.1	PCR primers used to generate <i>Cm</i> U4 and U6 DNA templates for <i>in vitro</i> transcription and their respective annealing temperatures.	98
Table 4.2	Composition of the IVT reactions that generated <i>Cm</i> U4 and U6 RNA transcripts.	99

Table 4.3	Sequences of the full-length U4 and U6 snRNA generated from IVT.	100
Table 4.4	Buffer components used in U4 and U6 snRNA annealing reactions.	100
Table 4.5	Primers used to modify pQLinkN into pQLinkC, as well as to subclone <i>Cm</i> Prp4 and Prp31 into pQLinkH and C plasmid vectors.	104
Table 4.6	Components of the MDG and ZYM-5052 media used in Auto-Induction.	109
Table 4.7	NCBI and UniProt accession numbers of U4/U6 di-snRNP proteins.	110

List of Figures

Figure 1.1	Pre-mRNA splicing, including the two transesterification reactions and snRNP conformational changes.	3
Figure 1.2	Pre-mRNA recognition and the Commitment Complex (Complex E).	6
Figure 1.3	Model for bridging the branchpoint sequence and 5' splice site via U1 and the Msl5 and Mud2 splicing factors.	7
Figure 1.4	Model for displacement of the Msl5 and Mud2 splicing factors, representing the transition to Pre-Spliceosome (Complex A).	8
Figure 1.5	The U4/U6 di-snRNP.	13
Figure 2.1	Expression and affinity purification of Msl5.	37
Figure 2.2	Cation exchange purification of Msl5.	39
Figure 2.3	Size exclusion purification of Msl5.	41
Figure 2.4	Alignment and predicted domain structures of the Msl5 against human (SF1) and yeast (BBP) homologs.	43
Figure 2.5	The crystal structure of the SF1-U2AF65 complex highlights the interaction between the ULM and UHM.	45
Figure 2.6	Predicted structure and domains of Msl5.	47
Figure 2.7	Simulated RNA folding of the U4 5' Stem Loop (nucleotides 22-50), corresponding to ro52 used in various binding experiments.	50
Figure 2.8	Saturation curve from fluorescence polarization measurements of Msl5 titrated against three RNA targets (15 nM).	51
Figure 2.9	EMSA measurements of Msl5 binding <i>Cm</i> BPS.	53
Figure 2.10	Saturation curve from EMSA binding assay of Msl5 titrated against two RNA targets (25 nM).	54
Figure 2.11	Simulated RNA folding and predicted secondary structure of WT pre-mRNA.	56
Figure 2.12	Simulated RNA folding and predicted secondary structure of mutant pre-mRNA.	57

Figure 2.13 Simulated RNA folding and predicted secondary structure of WT mRNA.	58
Figure 2.14 EMSA measurement of Msl5 binding the WT 072C pre-mRNA transcript.	59
Figure 2.15 EMSA measurements of Msl5 binding the mutant 072C pre-mRNA transcript.	60
Figure 2.16 EMSA measurements of Msl5 binding the 072C mRNA transcript.	60
Figure 2.17 Saturation curves of <i>Cm</i> Msl5 against various 072C RNA transcripts: WT pre-mRNA (black); mutant pre-mRNA (grey); mRNA (dotted grey).	61
Figure 2.18 CD spectra of Msl5 and the <i>Cm</i> consensus BPS.	63
Figure 2.19 Thermal stability analysis of Msl5.	64
Figure 2.20 Co-expression and Ni-NTA batch purification of His-Mud2 and Msl5 analyzed on a 12% SDS-PAGE.	65
Figure 2.21 Size exclusion chromatography purification of the Mud2/Msl5 heterodimer.	67
Figure 2.22 Western Blot of His- <i>Cm</i> Msl5 and His- <i>Cm</i> Mud2 affinity pull-downs.	68
Figure 2.23 The standardization of <i>Cm</i> Mud2/Msl5 by comparison against known amounts of <i>Cm</i> Msl5.	69
Figure 2.24 Fluorescence polarization of the <i>Cm</i> Mud2/Msl5 heterodimer titrated against the <i>Cm</i> BPS RNA.	70
Figure 2.25 Alignment and predicted secondary structure of the U2 Homology Motif (UHM) of Mud2 against human (SF1)) and yeast (BBP homologs).	73
Figure 3.1 The expression and affinity purification of Snu13.	80
Figure 3.2 Purification and Crystallography of Snu13.	81
Figure 3.3 X-ray diffraction of Snu13 crystals at the Stanford Synchrotron Light Source.	82

Figure 3.4	Sequence alignment with sequences from <i>G. sulphuraria</i> (Snu13g), <i>Homo sapiens</i> (15.5K), and <i>S. cerevisiae</i> (Snu13p).	84
Figure 3.5	Structural comparison of <i>C. merolae</i> Snu13 with orthologs.	87
Figure 3.6	Close-up of the RNA binding domain of human 15.5K (light blue) associated with the kink-turn of human U4 snRNA (pink) and Prp31 (human 61K, green).	89
Figure 3.7	Hydrophobic pocket of Snu13 is clearly seen from the molecular surface, while the rest of the protein is shown in ribbon.	90
Figure 3.8	Saturation curve of Snu13 against the U4 5' SL.	92
Figure 3.9	Functionally active fraction of Snu13 determined by RNA titration.	96
Figure 4.1	Overview of ligase-independent cloning to produce a single plasmid with two genes.	103
Figure 4.2	Overview of the reactions and incubations used in LIC to build the various <i>Cm</i> Prp3/4- and Prp31-containing constructs.	106
Figure 4.3	The <i>Cyanidioschyzon merolae</i> U4/U6 di-snRNP, based on human and yeast modeling.	111
Figure 4.4	SDS-PAGE analysis of purified <i>Cm</i> Snu13, LSm, and Sm proteins for use in the assembly assays.	112
Figure 4.5	U4 and U6 RNA annealing assay.	113
Figure 4.6	U4 and U6 <i>Cm</i> di-snRNP assembly assays done by electromagnetic mobility shift with 9% polyacrylamide gels.	115
Figure 4.7	U4 and U6 di-snRNP assembly assay with Snu13, LSm, and Sm proteins.	116
Figure 4.8	Cloning of Prp4 and Prp3 into pQLinkH. Amplification of Prp4gene by PCR.	118
Figure 4.9	SDS-PAGE analysis of <i>Cm</i> Prp3 and Prp4 co-expression through IPTG induction at 37°C.	119

Figure 4.10 SDS-PAGE analysis of <i>Cm</i> Prp3 and Prp4 co-expression through IPTG and auto-induction at various temperatures.	120
Figure 4.11 SDS-PAGE analysis of <i>Cm</i> Prp31 expression through IPTG induction at 37°C.	120
Figure 4.12 SDS-PAGE analysis of <i>Cm</i> Prp31 expression through IPTG induction at various temperatures.	121
Figure 4.13 Output for Predicted Intrinsic Disorder of the DisEMBL Intrinsic Protein Disorder Prediction 1.5 server.	122
Figure 4.14 Map of predicted intrinsically disordered regions within the U4/U6 di-snRNP proteins.	124

List of Equations

Equation 1	28
Equation 2	28
Equation 3	28
Equation 4	35
Equation 5	95
Equation 6	95
Equation 7	95
Equation 8	95

Acknowledgements

To my supervisor Dr. Stephen Rader, and my “hidden” supervisor Dr. Martha Stark, I give my utmost respect and gratitude. Together, Stephen and Martha provided world-class training both at the bench and in critical thinking. I pursued interesting research goals and gained independence as a junior scientist. Stephen was extremely generous with providing opportunities to attend workshops and conferences, from which I gained communication and networking skills. I am also deeply grateful to the brilliant Dr. Liz Dunn for wonderful conversations and support. I thank past and present members of the Rader Lab (Kirsten Reimer, Fatimat Shidi, Viktor Slat, Mona Aminorroayae, and the undergraduate students) for a highly enjoyable workplace. For insightful conversations and practical training, I thank Dr. Andrea Gorrell, Dr. Maggie Lee, and Dr. Chow Lee. To Dr. Daniel Erasmus, Dr. Jane Young, and committee member Dr. Keith Egger, I thank you for your recommendation and participation, for which I was able to become a graduate student. For their indispensable assistance during our collaboration, I thank Dr. Erin Garside and Dr. Andrew MacMillan. Finally, my family and friends – Shane, Colleen, Joel, Bode, Lorelie, Kurumi, Sebastian, and Victor – were more supportive and loving than they could ever know.

Table of Abbreviations

Abbreviation	Definition
15.5K	<i>H. sapiens</i> RNA kink-turn binding protein
3' SS	3' Splice Site
5' SS	5' Splice Site
6-FAM	6-Carboxyfluorescein (a fluorescent molecule)
BBP	<i>S. cerevisiae</i> branchpoint bridging protein
bp	base pair
BPS	branchpoint sequence
CD spec	Circular Dichroism Spectroscopy
<i>Cm</i>	<i>Cyanidioschyzon merolae</i>
di-snRNP	di-small nuclear ribonucleoprotein particle
dsRNA/dsDNA	double-stranded RNA/DNA
<i>E. coli</i>	<i>Escherichia coli</i> (Gram-negative bacteria)
EMSA	Electrophoretic Mobility Shift Assay
FA/FP	Fluorescence Anisotropy/Polarization
<i>Gs</i>	<i>Galdieria sulphuraria</i>
<i>Hs</i>	<i>Homo sapiens</i>
IVT	<i>In vitro</i> transcription
K _D	binding affinity/specificity (usually in nM)
L.B.	Luria Broth
LIC	ligase-independent cloning
ME	mean ellipticity
Msl5	<i>C. merolae</i> branchpoint bridging protein
Mud2	U2 Auxiliary Factor
PCR	Polymerase chain reaction
pre-mRNA	precursor messenger ribonucleic acid
RE	restriction enzyme
<i>Sc</i>	<i>Saccharomyces cerevisiae</i>
<i>Sc</i> Mud2	<i>S. cerevisiae</i> U2 Auxiliary Factor
SF1	<i>H. sapiens</i> branchpoint bridging protein
SL	Stem Loop/Kink-Turn motif (RNA fold)
snRNA	small nuclear ribonucleic acid
snRNP	small nuclear ribonucleoprotein particle
Snu13	<i>C. merolae</i> RNA kink-turn binding protein
Snu13p	<i>S. cerevisiae</i> RNA kink-turn binding protein
ssRNA/ssDNA	single stranded RNA/DNA
T _M	melting temperature/unfolding
tri-snRNP	tri-small nuclear ribonucleoprotein particle
U2AF65	<i>H. sapiens</i> U2 Auxiliary Factor (large subunit)

Chapter 1

Introduction

Life can be viewed as an information transmission process. Information is contained within DNA (deoxyribonucleic acid), transcribed to RNA (ribonucleic acid), and translated to proteins within a cell. After a precursor messenger RNA (pre-mRNA) molecule is created, processing of that pre-mRNA occurs prior to protein synthesis. These modifications include the addition of a methylated guanosine triphosphate to the end of the RNA in a 5' to 5' phosphodiester linkage to prevent degradation, polyadenylation to produce a 3' poly A-tail and to promote translation, and pre-mRNA splicing. Similar to protein synthesis with regards to the ribosome, pre-mRNA splicing requires an intricate protein-RNA complex to facilitate the necessary chemical reactions.

1.1 Pre-mRNA Splicing and the Spliceosome

Pre-mRNA splicing is critical to the survival of the Eukaryotic cell; previous research has shown that if it was affected by even a single mutation, the cell could exhibit temperature sensitivity or even lethality. Following transcription, pre-mRNA contains both non-protein coding (intronic) sequences and protein coding (exonic) sequences. Introns are non-coding pre-mRNA sequences that must be removed in order to have an uninterrupted series of coding sequence. Once the intronic sequences are removed and the exonic sequences are joined (or spliced) together, the pre-mRNA is now a mature mRNA biomolecule. This mature RNA (mRNA), containing nucleotides only from the start codon to the stop codon, is then transported into the cytosol for protein synthesis.

The machinery carrying out the process of pre-mRNA splicing, the spliceosome, is a multi-mega-Dalton ribonucleoprotein complex with over 200 distinct protein and RNA

components. The size and complexity of the spliceosome is likely attributed to a series of evolutionary adaptations that have allowed the spliceosome to differentiate between splice sites that change with certain genes through cellular differentiation (Graveley, 2001). The five snRNAs (U1, U2, U4, U5, and U6) each interact with certain spliceosome-associated proteins to form individual small nuclear ribonucleoproteins, or snRNPs.

Pre-mRNA splicing has been extensively researched, and the current model for yeast and humans follows a similar path (Fig. 1.1). The U1 snRNP first binds to the 5' splice site of an intron located on a pre-mRNA (Green, 1991). The U2 snRNA, facilitated by associated proteins, then base-pairs with the branchpoint sequence (BPS) and 3' splice site (Query *et al.*, 1996). At this point, the U4/U6.U5 tri-snRNP [notation used to indicate the U4/U6 di-snRNP in complex with U5 snRNP] enters and binds to the mRNA intron (Stark and Lührmann, 2006). The U1 snRNP, followed by the U4 snRNP each dissociate. The U2, U6, and U5 snRNP cluster completes two transesterification reactions and subsequently detach from the spliced mature mRNA (Green, 1991). In the first reaction, the 2' OH group of the conserved branchpoint sequence adenosine is a nucleophile for the attack on the conserved guanine of the 5' splice site. The second transesterification happens when the 3' OH of the released exon attacks the phosphodiester bond of the conserved guanine of the 3' splice site, thereby releasing the lariat intron. The snRNPs are then regenerated for future splicing reactions (Staley and Guthrie, 1998).

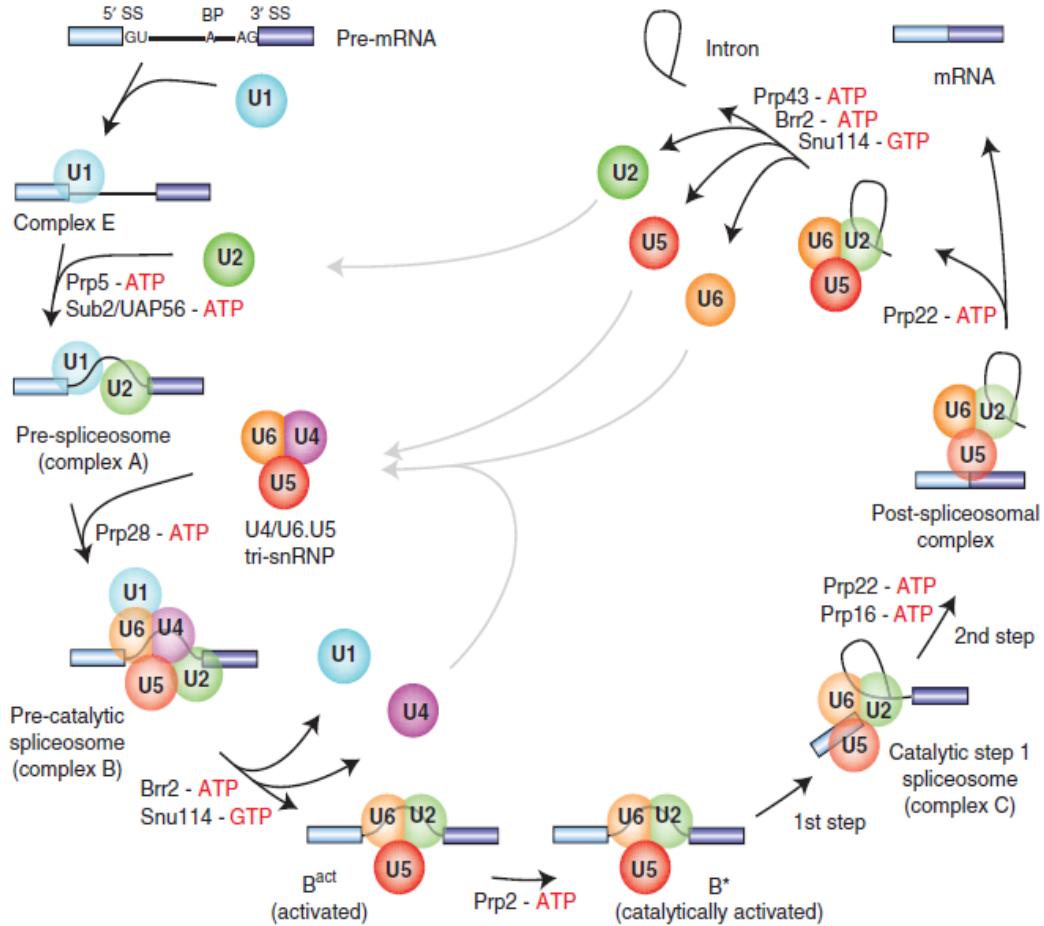


Figure 1.1. Pre-mRNA splicing, including the two transesterification reactions and snRNP conformational changes. This model for animals and budding yeast was adapted from Will and Lührmann, 2011.

1.2 *Cyanidioschyzon merolae* as a Model Organism

Given that the human spliceosome boasts 200+ proteins, 5 distinct snRNAs, and multiple conformations, it's not surprising that attention has been paid to other organisms that offer simplicity with regards to spliceosomes – namely yeast. *Saccharomyces cerevisiae* is a budding yeast species with 5 snRNAs and approximately half as many splicing factors. Identities of the individual *S. cerevisiae* splice factors have been determined, and a number of snRNA and proteins have been characterized (Stevens *et al.*, 2002; Fabrizio *et al.*, 2009). Similar to the human spliceosome, detailed structural

and mechanistic information is missing for the majority of the yeast spliceosome. In light of the tremendous amount of work necessary to understand the mechanism and structure of the spliceosome throughout the cycle, a novel organism has been selected to study pre-mRNA splicing: *Cyanidioschyzon merolae*. The human genome has over 18,000 introns. When the complete *C. merolae* genome was published, it was discovered that it had only 27 introns; the yeast genome, of similar size, has more than 250 (Spingola *et al.*, 1999). The reduced number of introns hinted at a more reduced splicing system, with possibly even fewer splicing factors than are expressed in *S. cerevisiae* (Stark *et al.*, 2015).

Using bioinformatics to look into the *C. merolae* genome, specifically at those genes pertaining to pre-mRNA splicing, has revealed only 4 snRNAs (U2, U4, U5, and U6) and 68 splicing factors. Of the 68 splicing factors, 49 core proteins have been identified as being directly associated with snRNP and spliceosomal assembly (Stark *et al.*, 2015). Based on current information from co-immunoprecipitation assays and extensive bioinformatics searches, there is no evidence to suggest the presence of either the U1 snRNA and its associated proteins, or the minor spliceosome (Stark *et al.*, 2015).

Tangential to the advantages of studying a reduced splicing system are the advantages gained by the inherent properties inherent to *C. merolae*. *C. merolae* is a unicellular red alga both thermophilic and acidophilic in nature, living at 45°C and at a pH of 1.5 (Matsuzaki *et al.*, 2004). Proteins of thermophilic organisms, i.e. *C. merolae*, would likely have more charged and/or polar residues to form additional ionic interactions and exhibit thermo-stability (Dalhus *et al.*, 2002). This is a desirable property when considering crystallization conditions, and another advantage over the human and yeast variants.

1.3 Initiation of Pre-mRNA Splicing

The first step of splicing is the identification of an intron-containing pre-mRNA transcript. Recognition occurs at both the 5' and 3' ends of the intron: specifically the 5' splice site (5' SS) and the branchpoint sequence (BPS) near the 3' SS. In humans and fission yeast (*Schizosaccharomyces pombe*), the 5' splice site of the pre-mRNA is base-paired with U1 snRNA (Fig. 1.2). At the 3' end, the BPS and downstream polypyrimidine tract are respectively recognized by the branchpoint bridging protein and the U2 auxiliary factor. Humans and fission yeast have two auxiliary factors U2AF65/U2AF35, while budding yeast (*S. cerevisiae*) has ScMud2. Shown below are the branchpoint bridging protein SF1/BBP/Msl5 and the large U2AF Mud2 (Fig. 1.2). It is likely that the auxiliary factors serve to guide the branchpoint binding protein because the BPS is degenerate in higher eukaryotes (Keller and Noon, 1984). While 5' and 3' splice site recognition are independent of each other, stability of the snRNA-5' SS complex is improved with the presence of the branchpoint bridging protein already bound to the BPS (Larson and Hoskins, 2017). Once the 5' and 3' splice sites are appropriately occupied, the pre-mRNA splicing reaction is at the Commitment Complex (CC) stage and will proceed to the pre-spliceosome (Rosbash and Sepharin, 1991; Colot *et al.*, 1996). The commitment complex and pre-spliceosome are equivalent to Complex E and Complex A, respectively (Fig. 1.1).

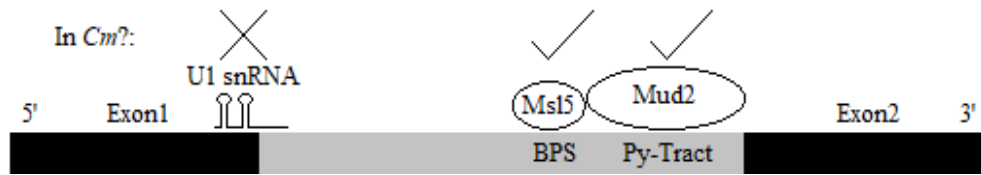


Figure 1.2. Pre-mRNA recognition and the Commitment Complex (Complex E). *Cyanidioschyzon merolae* contains Msl5 and Mud2 but lacks the U1 snRNA present in all other eukaryotes. This *C. merolae* Model was created using Paint.

Cyanidioschyzon merolae differs from previously studied models in both 5' and 3' splice site recognition due to the absence of several genes. The U1 snRNA and associated proteins are missing in *C. merolae*, a component otherwise believed to be necessary for 5' splice site (5' SS) recognition. Without the U1 snRNP, 5' SS recognition is unknown in *C. merolae*. This is the first organism ever studied that lacks U1, and there are currently several hypotheses for 5' SS definition that include 2 possible snRNA candidates (U5 or U6) or proteins bypassing the function of U1 (Stark *et al.*, 2015). Similar to budding yeast (*S. cerevisiae*), the polypyrimidine tract and one of the U2 auxiliary factors is also missing in *C. merolae* (Romfo and Wise, 1997; Stark *et al.*, 2015). This leaves only the branchpoint bridging protein (Msl5) and the larger U2 auxiliary factor (Mud2) responsible for initially recognizing the BPS/3' SS.

Without a polypyrimidine tract, it is unclear how Mud2 functions in recognizing the BPS. It is possible that Msl5 and Mud2 bind to the branchpoint region in a stepwise manner, with one protein guiding or stabilizing the other. It may also be that Msl5 and Mud2 form a heterodimer prior to binding the BPS. Msl5 and Mud2 have been previously shown to interact without RNA, but that was in fission yeast and the resulting complex was a heterotrimer due to the presence of the smaller U2 auxiliary factor (Huang *et al.*, 2002).

1.4 Bridging the Branchpoint Sequence to the 5' Splice Site

The current budding yeast model for the Commitment Complex (or Complex E in animals) involves both binding and base-pairing interactions between proteins and RNA (Fig. 1.2; Abovich *et al.*, 1994; Becerra *et al.*, 2015). Through these interactions, the BPS is connected to the 5' SS through a “bridge” of protein and RNA (Fig. 1.3). Integral to the CC is the branchpoint bridging protein Msl5 that directly interacts with Mud2, the BPS, and the U1 splicing protein Prp40 (Kao and Siliciano, 1996; Abovich and Rosbash, 1997). Prp40 serves as the final link to the bridge as it is bound to U1 snRNA which is base-paired to the 5' SS (Abovich and Rosbash, 1997; Rutz and Sepharin, 1999). Of course, Prp40 along with every other U1 related gene has not been found in *Cyanidioschyzon merolae* and no obvious candidate exists (Stark *et al.*, 2015).

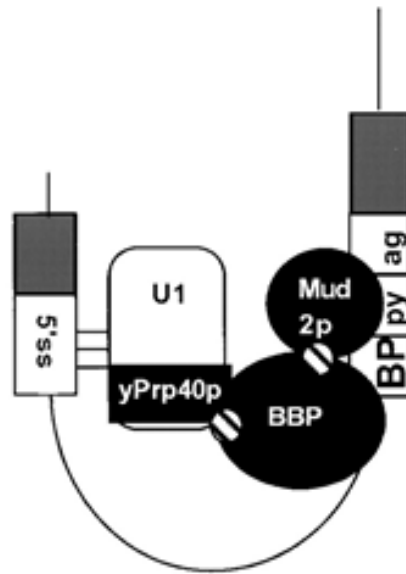


Figure 1.3. Model for bridging the branchpoint sequence and 5' splice site via U1 and the Msl5 and Mud2 splicing factors. This budding yeast model was adapted from Abovich and Rosbash (1997).

In order for U2 snRNA to base-pair with the BPS and 3' SS and form the “Pre-Spliceosome” (or Complex A in animals), Msl5 and Mud2 are replaced by U2 snRNP associated proteins (Fig. 1.4). While the detailed mechanism of displacement remains unknown, genetic and biochemical work revealed two DExD/H box proteins Sub2 and Prp5 are involved (Perriman and Ares, 2000; Kistler and Guthrie, 2001; Liang and Chen, 2015). Following incorporation of the U2 snRNP, the U4/U6.U5 tri-snRNP is added and the rest of the splicing cycle continues as outlined above.

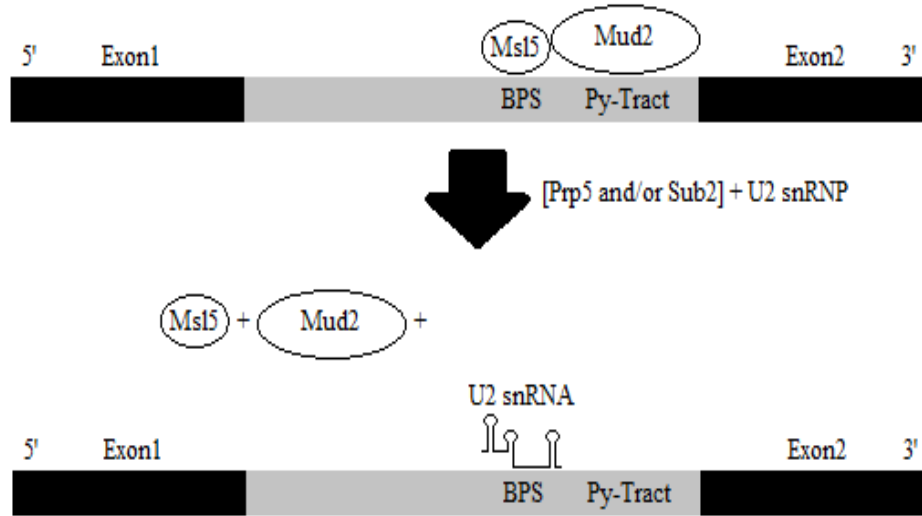


Figure 1.4. Model for displacement of the Msl5 and Mud2 splicing factors, representing the transition to Pre-Spliceosome (Complex A). This *C. merolae* model was created using Paint.

1.5 The Branchpoint Bridging Protein Msl5 and U2 Auxiliary Factor Mud2

The majority of splicing research has been done on human and yeast (both budding and fission). Not surprisingly, orthologous gene products have amassed multiple names. The branchpoint binding protein and U2 auxiliary factors are SF1 and U2AF65/35 (respectively) in humans and fission yeast. In budding yeast, these proteins are BBP and ScMud2, while in *C. merolae* they are Msl5 and Mud2 (Table 1.1). With a degenerate

BPS and conserved polypyrimidine tract (Py-tract), human U2AF65 drives branchpoint sequence (BPS) selection through interactions with both SF1 and the polypyrimidine tract (Ruskin and Green, 1985; Zamore *et al.*, 1992). Indeed, the binding affinity of human U2AF65 for the Py-tract is 50-300 X greater than SF1 for the BPS; similarly, there is a 20x increase in binding affinity for the BPS when SF1 is in complex with U2AF65 (Berglund *et al.*, 1998a; Guth and Valcárcel, 2000). Cross-linking assays have revealed that there is only 1 free nucleotide between the BPS and Py-tract when SF1 and U2AF65 are bound to human pre-mRNA (Berglund *et al.*, 1998b). Footprinting assays have found that SF1/BBP will protect the 7 nucleotide BPS UACUAAC plus two nucleotides on either side, and U2AF65 will only protect the Py-tract (Berglund *et al.*, 1997).

Table 1.1. Common names for the branchpoint bridging protein and U2 auxiliary factors in various model organisms.

Protein	Species	Specific Name	Size (kDa)	UniProt ID
Branchpoint Bridging Protein	<i>H. sapiens</i>	SF1	68.3	ZNF162
Branchpoint Bridging Protein	<i>S. cerevisiae</i>	BBP	53.0	YLR116W
Branchpoint Bridging Protein	<i>C. merolae</i>	Msl5	36.2	CM1292C
U2 Auxiliary Factor (large)	<i>H. sapiens</i>	U2AF65	53.5	U2AF65
U2 Auxiliary Factor	<i>S. cerevisiae</i>	ScMud2	60.5	YKL074C
U2 Auxiliary Factor	<i>C. merolae</i>	Mud2	142.5	CMS438C

In budding yeast, there is no polypyrimidine tract but there is enrichment in uridine content between the BPS and 3' splice site, named the poly-U region (Patterson and Guthrie, 1991; Stark *et al.*, 2015). Introns are more efficiently spliced when transcripts contain a poly-U region, which is thought to cause more contacts and stronger binding with the spliceosome (Ma and Xia, 2011). Through binding experiments with several

modified α -tropomyosin pre-mRNA transcripts, it was discovered that 11 continuous uridines placed anywhere between the BPS and 3' SS were optimal for 3' SS selection (Ma and Xia, 2011). Any reduction of the poly-U region reduced splicing efficiency, and required that the reduced poly-U be immediately adjacent to the 3' SS (Ma and Xia, 2011). While the poly-U region varies wildly across the *S. cerevisiae* intronome, there is generally uridine enrichment 5-18 nucleotides upstream of the 3' SS (Patterson and Guthrie, 1991). Furthermore, uridine enrichment 9 nucleotides upstream of the 3' SS in budding yeast functions in the second transesterification reaction of splicing (Patterson and Guthrie, 1991).

Just as the Py-tract has been lost in *S. cerevisiae*, so too is ScMud2 divergent from U2AF65 of human and fission yeast *S. pombe* (Abovich *et al.*, 1994). Since ScMud2 is so different from its human/yeast homologs and is not required for survival, there may have been a weaker selection constraint on the evolution of the Py-tract in *S. cerevisiae* (Ma and Xia, 2011). Indeed, the *S. cerevisiae* genome is AT rich, and it stands to reason that cytosines may have been continually replaced by thymines, which would ultimately enrich uridine content (Ma and Xia, 2011). Since the poly-U region does enhance 3' SS selection and splicing efficiency, perhaps the evolution of the poly-U region owes itself to both random mutation and selection (Ma and Xia, 2011). In *C. merolae*, there is no Py-tract or poly-U region. While Mud2 presumably recognizes a sequence downstream of the BPS that is not enriched with uridine or cytosine, no other information is yet known which brings to mind several interesting questions. Whether Mud2 and Msl5 recognize the BPS as a dimer, or whether one of the proteins bind RNA first is unknown. The sequence or binding determinant region recognized by Mud2 remains unidentified. It

would be interesting to investigate whether or not there is still a preference for uridine despite the lack of a Py-tract or poly-U region? The 145 kDa Mud2 is also unique in size, as it is much larger than the 54 kDa human or 60 kDa yeast homologs (NCBI; Stark *et al.*, 2015). Various structures of yeast and human Msl5 and Mud2 homologs have been determined through X-ray Crystallography, Small Angle X-ray Scattering (SAXS), and Nuclear Magnetic Resonance (NMR) (Liu *et al.*, 2001; Selenko *et al.*, 2003; Zhang *et al.*, 2012; Wang *et al.*, 2013; Jacewicz *et al.*, 2015).

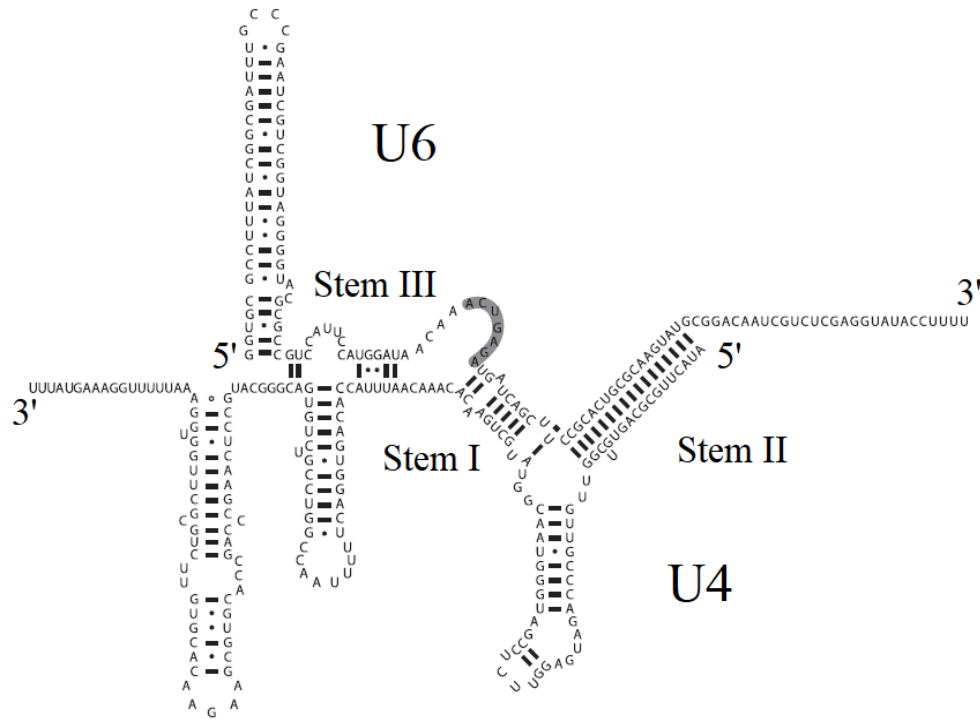
1.6 U4/U6 Di-snRNP in Pre-mRNA Splicing

One of the essential components of the spliceosome, the U4 snRNA, acts as a chaperone and regulator of the U6 snRNA. And while U1, U2, U5, and U6 are all found to exist as individual ribonucleoprotein particles, almost all U4 is found conjugated with U6 (Bringmann *et al.*, 1984; Hashimoto and Steitz, 1984; Jandrositz and Guthrie, 1995). At any given time, there would be a very small portion of the U4 snRNP existing as an individual entity after it is dislodged from the active spliceosome and before it is reunited with the U6 snRNP. The catalytic activity of the spliceosome could be seen as the result of a series of conformational changes associated with the U6 snRNP. Thus U4 likely also has an inhibitory role on splicing, preventing the event by base pairing to U6. Prior to its engagement with the intron, the U6 snRNP exists in di-snRNP form with U4 through base pairing (Nottrott *et al.*, 2002). The di-snRNP then assembles into a tri-snRNP with U5 (the U4/U6.U5 tri-snRNP). After the tri-snRNP binds the 5' intronic sequence and forms a complex with the U1 and U2 snRNPs, U1 dissociates from the 5' splice site. The U4/U6 di-snRNP then undergoes a conformational shift so as to disrupt the base pairs

causing U4 to dissociate and a new series of interactions between U6, U2, and the pre-mRNA nucleotides on the 5' end to occur (Fig. 1.1; Nottrott *et al.*, 2002)

Structurally, U4 snRNA is dynamic in that it can be found in three different states: the U4 snRNP, the U4/U6 di-snRNP, and the U4/U6.U5 tri-snRNP. It contains several conserved stem loops and features a kink-turn motif common to protein-RNA interactions. Both the 5' stem loop and the core domain have been solved for human U4 snRNA (Vidovic *et al.*, 2000; Leung *et al.*, 2011). The core domain is a single-stranded uridine-rich RNA sequence that is targeted by the Sm core proteins prior to re-entry into the nucleus (Fischer *et al.*, 1993). The *C. merolae* U4 snRNA secondary structure appears to be well conserved when compared with the human homolog. In Stem Loop III of the base-paired U4/U6 particle, however, there is an additional 18 nucleotides (Fig. 1.5 A; Stark *et al.*, 2015).

A



B

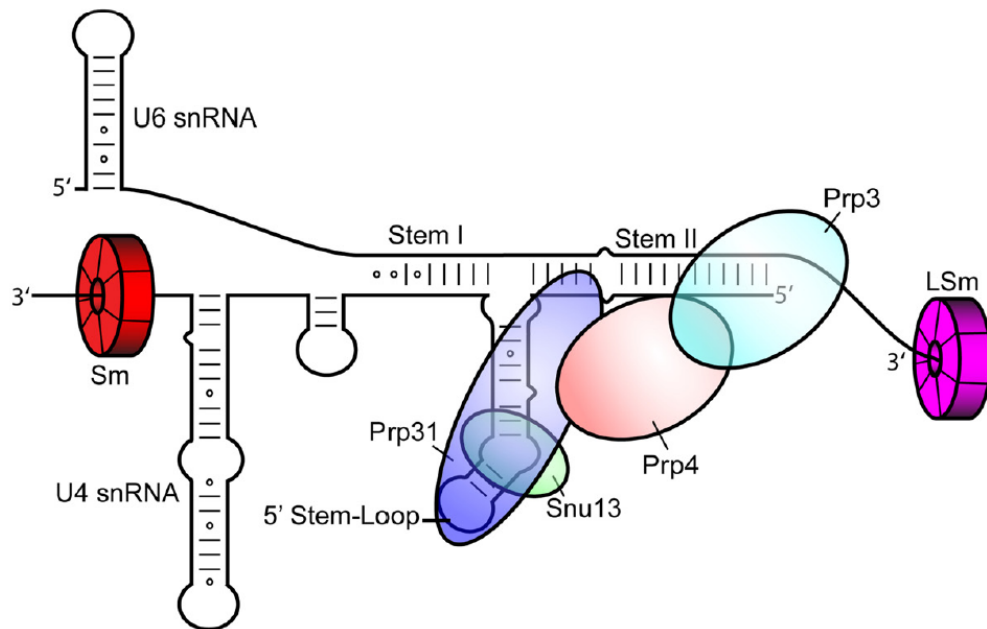


Figure 1.5. The U4/U6 di-snRNP. (A) Predicted secondary structure of the U4/U6-disnRNA in *C. merolae*. Adapted from Stark *et al.* (2015). (B) Predicted secondary structure of the human U4/U6 di-snRNP with all associated proteins.

Adapted from Hardin *et al.* (2015).

Assembly of the U4/U6 di-snRNP has been studied in humans and yeast (Nottrott *et al.*, 2002; Hardin *et al.*, 2015). The di-snRNP consists of U4 and U6 snRNAs, along with associated proteins Sms, LSms, Snu13, Prp3, Prp4, and Prp31 (Horowitz *et al.*, 1997; Nottrott *et al.*, 1999; Sander *et al.*, 2006). U4 and U6 snRNAs are transcribed in the nucleus, followed by nuclear-localized LSm proteins binding the 3' uridine-rich region of U6 snRNA (Mayes *et al.*, 1999; Achsel *et al.*, 1999). U4 is exported to the cytoplasm where the Sm proteins similarly assemble onto the 3' uridine-rich tail; U4-Sm is then imported into the nucleus (Karaduman *et al.*, 2006). With U4-Sm and U6-LSms together in the nucleus, di-snRNP assembly continues with Watson Crick base-pairing between the two snRNAs (Fig. 1.5 A; Bringmann *et al.*, 1984; Hashimoto and Steitz, 1984). The small, globular Snu13 then binds the U4 5' stem loop and acts as an anchor for the subsequent assembly of Prp3, Prp4, and Prp31 (Nottrott *et al.*, 2002). Prp3 and Prp4 exist as a heterodimer prior to di-snRNP assembly and require Snu13 already present on the U4 5' stem loop, as does Prp31 (Nottrott *et al.*, 2002). Functional studies with human and yeast models suggest that the Prp3/4 dimer and Prp31 bind independently of each other (Hardin *et al.*, 2015). Recent cryogenic electron microscopy (cryo-EM) structures of human and yeast spliceosomes have revealed several details pertaining to the architecture of the di-snRNP (Fig. 1.5 B; Nguyen *et al.*, 2015; Agafonov *et al.*, 2016). Both the Prp3/4 dimer and Prp31 contact Snu13, the U4 5' stem loop, and Stem II of the di-snRNP (Nguyen *et al.*, 2015). A high-resolution structure under 2 Angstroms for U4 snRNA has yet to be solved. *C. merolae* U4 snRNA is a good candidate for crystallization due to the thermo-stability and reduction of splicing factors. Fewer components would also behoove functional studies to study the interaction between complete U4 and U6 snRNPs. The

simplicity of the *C. merolae* splicing system is reflected in the U4 snRNP: no homolog has been found for the human CypH U4/U6 associated splicing factor (Stark *et al.*, 2015).

1.7 Snu13 In di-snRNP Assembly

One of the proteins associated with the U4 snRNA is Snu13, a globular protein with an α - β - α fold reminiscent of the L30 ribosomal protein (Vidovic *et al.*, 2000). While the size similarity between human (15.5 kDa) and algal (15.8 kDa) homologs may suggest similarity in structure, no experimental data is available for validation. The classic RNA motif to which Snu13 binds is the helix-bulge-helix, or kink-turn motif (Marmier-Gourrier *et al.*, 2003). In general, the kink-turn motif is characterized as having an asymmetrical internal loop that causes the flanking stems to fold towards each other, creating a “kink” (Klein *et al.*, 2001). The U4 5' stem loop is highly conserved across all organisms, with the internal loop consisting of 5 nucleotides on one side of the flanking stems, and 2 on the other (Nottrott *et al.*, 1999). Interestingly, the Archaeal homolog is unique in that it does not discriminate between the kink-turn and similar motifs (Oruganti *et al.*, 2005). Although the fidelity of *C. merolae* Snu13 is as of yet uncharacterized, specificity and structure is expected to be comparable to previously studied Eukaryotic homologs and not the Archaeal variant (Oruganti *et al.*, 2005). The structure of the human and yeast variants have been solved by X-ray crystallography. There has not been functional or structural characterization on any component of the *C. merolae* spliceosome. Structures for Snu13 homologs both free and bound to U4 snRNA have been previously characterized, making the algal variant an attractive starting point for structural and functional characterization of U4 snRNP splicing factors (Vidovic *et al.*, 2000; Oruganti *et al.*, 2005).

1.8 Research Goals

The first research objective was to study recognition of an intron-containing pre-mRNA transcript in *Cyanidioschyzon merolae* using fluorescence anisotropy and electrophoretic mobility shift assays. Recombinant Msl5 and Mud2 proteins were to be titrated against labeled synthetic pre-mRNA to measure binding. These experiments were meant to determine whether branchpoint sequence recognition was a stepwise process (cooperative) or if Msl5 and Mud2 were a heterodimer prior to binding. If time permitted, recombinant Prp5 and Sub2 would be titrated against the Msl5-Mud2-RNA complex to determine which DExD/H box protein was necessary for Msl5 and Mud2 displacement.

The second research goal was to study structural properties of Msl5 both free and bound to RNA. Circular dichroism spectroscopy would measure conformational changes in Msl5 structure upon binding the branchpoint sequence of pre-mRNA. X-ray crystallography was to be used to elucidate high-resolution structures of the protein both free and bound to RNA, which would have revealed details of branchpoint sequence recognition.

The third objective was to study the interaction between Snu13 and the 5' stem loop of U4 snRNA. The binding affinity of Snu13 for U4 was to be measured by fluorescence polarization. This objective would be reached by measuring the anisotropy of a fluorescently labeled U4 snRNA 5' stem loop oligonucleotide titrated with recombinant Snu13. The data would then be analyzed to determine affinity and cooperativity.

The fourth research objective was to characterize the order of assembly of the *Cm* U4/U6 di-snRNP. Using gel shift assays, *in vitro* transcribed U4 and U6 snRNAs would be mixed with recombinant di-snRNP associated proteins (Snu13, LSms, Sms, Prp3,

Prp4, Prp31). EMSAs allow for a visual representation of step-wise assembly because bands would appear sequentially higher as more components interact with each other.

The fifth research objective was to crystallize Snu13 both bound and free of U4 snRNA. *C. merolae* offers the very small, globular, thermo-stable protein Snu13. Upon successfully acquiring crystals, X-ray crystallography would be done to obtain diffraction patterns and elucidate a structure for each crystal. Structure determination of *C. merolae* Snu13 would then complete the proof of principle for X-ray crystallography of the algal spliceosome. Upon achieving the primary objectives, we will have a better understanding of the structure and assembly of the U4 snRNP.

Chapter 2

Recognition of the pre-mRNA branch site by Msl5

This chapter contains the experiments performed in order to investigate the interaction of the Msl5 and Mud2 proteins with the branchpoint sequence inherent to *C. merolae* introns. To that end, several techniques were employed including fluorescent polarization, radioactive electrophoretic mobility shift assay, and circular dichroism spectroscopy.

2.1 Methods – Expression and Purification of *C. merolae* Msl5 and Mud2 Proteins

2.1.1. Transformation of Rosetta pLysS (DE3) cells with pQLinkH-Msl5

Full length Msl5 was amplified from the *Cyanidioschyon merolae* genome by PCR and subcloned into the pQLinkH plasmid vector (Addgene plasmid 13667) using primers found in Table 2.1. Rader lab member Fatimat Shidi made modifications to pQLinkH prior to its use by inserting *Swa*I and *Pac*I restriction enzyme cut sites to make the plasmid appropriate for ligase independent cloning (LIC). LIC is illustrated in section 4.4.1. The gene is flanked by an N-terminus 6x His-tag followed by TEV protease cleavage site ENLYFQS, whereby cleavage occurs between the Q and S. pQLinkH also contains an ampicillin resistance gene for selection and a T7 promoter for expression. Once Msl5 was subcloned into pQLinkH, pQLinkH-Msl5 was then transformed into DH5 α cells for preparation of a glycerol stock. The plasmid was later isolated and made available to use for expression purposes.

pQLinkH-Msl5 was used to transform Rosetta (DE3) pLysS competent *E. coli* cells (Novagen) which contain a plasmid that features genes for rare codon tRNA (optimized to humans) and chloramphenicol antibiotic resistance expressed through a T7

promoter. 50 uL of Rosetta pLysS (DE3) cells were thawed on ice in a 1.5 mL Eppendorf tube (Fisher Scientific) for 15 minutes before adding 150 ng of pQLinkH-Msl5. The tube sat on ice for 30 minutes followed by 45 seconds at 42°C. 250 uL of Luria Broth (L.B., Fischer Scientific) was added then the cell suspension was incubated at 37°C for 30 minutes. L.B. consists of tryptone (10 g/L), yeast extract (5 g/L), and sodium chloride (10 g/L) dissolved in 1 L of distilled water, which was autoclaved prior to use.

Cells were spun down (Eppendorf, model Centrifuge 5417 C) for 10 seconds then resuspended in 100 uL fresh L.B. broth and plated on L.B. agar supplemented with 20 ug/mL ampicillin (IBI Scientific) and chloramphenicol (Sigma-Aldrich) antibiotics. The plate was incubated at 37°C for 18 hours and subsequently resulted in (~50) colonies containing the pQLinkH-Msl5 plasmid.

Table 2.1. DNA primers used in the subcloning of Msl5 and Mud2 genes into pQLinkH plasmid vector.

Gene	Sequence
Msl5 (fwd)	TAC TTC CAA TCC CAC GCA CCG CGG AAC AGT GAG
Msl5 (rev)	TTA TCC ACT TCC CAC G TTA AGA ATC GCC CTG GAC C
Mud2 (fwd)	TAC TTC CAA TCC CAC GCA CCG AGA ACA GGC TCG G
Mud2 (rev)	TTA TCC ACT TCC CAC G TTA CCG CAG GTA GGG C

2.1.2. Transformation of Rosetta pLysS (DE3) cells with pQLinkH-Mud2/Msl5

A modified pQLinkH plasmid vector (Addgene) was previously assembled with full length tagged Mud2 (N-terminus 6x His-tag and TEV protease site), as well as full length untagged Msl5. Primers used to subclone the genes are detailed in Table 2.1. The transformation was performed as previously described using 200 ng of plasmid DNA. The plasmid pQLinkH-Mud2/Msl5 was used to transform 50 uL Rosetta pLysS cells. After 19 hours of incubation, the ampicillin- and chloramphenicol-supplemented L.B. agar plate yielded about 20 colonies.

2.1.3. Induction of pQLinkH-Msl5 transformed Rosetta pLysS (DE3) cells

One transformant colony containing the pQLinkH-Msl5 plasmid was used to inoculate 50 mL L.B. broth supplemented with 20 ug/mL each of ampicillin and chloramphenicol in a 250 mL Erlenmeyer flask. The culture was incubated in a shaker set at 300 RPM (New Brunswick Scientific, model Innova 40) at 37°C for 18 hours. 10 mL of the overnight broth was added to 4X 3L Fernbach flasks each containing 1L L.B. broth (20 ug/mL ampicillin and chloramphenicol) for large-scale expression. The 1 L cultures were put in the same shaker for 3.5 hours at 300 RPM and 37°C until they reached an OD₆₀₀ of 0.55 (Beckman Coulter, model DU 800). The cells were induced with 1 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG, VWR). Induction was carried out for 4 hours under the same conditions before each litre was centrifuged for 15 min at 3200 RPM and 4°C (Beckman Coulter, model Avanti HP-20 XPI) and resuspended in 10 mL Msl5 Ni Wash buffer (Table 2.2). The 4 cell suspensions were combined and split between two 50 mL conical tubes (VWR) then centrifuged for 15 min at 3200 RPM and 4°C (Beckman Coulter, model Allegra X-12R). The supernatants were discarded and the two pellets were flash-frozen in liquid nitrogen before storage in -80°C.

Both before and after induction, 1 mL of each culture and resuspended in SDS loading buffer to a final concentration of 1 OD unit per 100 uL (A600) before being stored at -20°C. The samples were later run on a 12% polyacrylamide (37.5:1 acrylamide/Bis, Bio-Rad) gel via sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). For the rest of this section, unless otherwise stated, all SDS-PAGE was performed under the following conditions: samples boiled at 95°C for 5 minutes then centrifuged at 20000 Xg for 3 minutes. 10 uL of each sample was loaded

into a lane of the gel and ran at 22 mA for 50 minutes. For all standard SDS-PAGE, the Broad Range Molecular Weight Marker (Bio-Rad) was used.

2.1.4 Induction of pQLinkH-Mud2/Msl5 transformed Rosetta pLysS cells

Mud2 and Msl5 were co-expressed using 2 L of L.B. media inoculated with pQLinkH-Mud2/Msl5-containing cells. 1 mM IPTG (VWR) was used for induction when an OD₆₀₀ of 0.67 was reached. Induction was carried out for 16 hours at 22°C before the cells were harvested at an OD₆₀₀ of 3.54. Both the L.B. media and the Wash Buffer were supplemented with 200 uM of L-arginine and L-aspartate (Sigma-Aldrich) at a pH of 7.5. The Wash Buffer was also supplemented with 0.5 M urea (Fisher Scientific).

2.1.5 Multi-step Chromatography of Recombinant Msl5 and Mud2/Msl5

Recombinant *Cm* Msl5 was purified using 3 chromatographic techniques that included affinity, cation exchange, and size exclusion chromatography. Co-expressed recombinant *Cm* Mud2 and Msl5 were purified using affinity and size exclusion chromatography.

2.1.5.1. Step 1: Ni-NTA Purification

Recombinant Msl5 and the Mud2/Msl5 dimer were each isolated from Rosetta pLysS (DE3) cells using the following buffer and purification scheme covered by Table 2.2 and below. The only difference in the treatment of Mud2/Msl5 was the inclusion of amino acids (L-Arg and L-Asp) and urea in the wash and elution buffers.

Table 2.2. Composition of buffers used in the Ni-NTA purification of recombinant Msl5 and co-expressed Mud2/Msl5 from Rosetta pLysS (DE3) cells.

Components	Msl5 Ni Wash Buffers	Mud2Msl5 Ni Wash Buffer	Msl5 Ni Elution Buffer	Mud2Msl5 Ni Elution Buffer	Msl5 Dialysis Buffer
HEPES-NaOH pH 7.5	30 mM	30 mM	30 mM	30 mM	20 mM
NaCl	300 mM	300 mM	300 mM	300 mM	90 mM
Imidazole	30 mM	30 mM	500 mM	500 mM	15 mM
β -mercaptoethanol	5 mM	5 mM	5 mM	5 mM	5 mM
L-Arginine	-	200 μ M	-	200 μ M	-
L-Arspartate	-	200 μ M	-	200 μ M	-
Urea	-	0.5 M	-	0.5 M	-

Two frozen Rosetta pLysS cell pellets were each treated with half of a tablet of protease inhibitor (Sigma-Aldrich) then resuspended in 20 mL Msl5 Ni Wash Buffer. After re-suspension, the samples were sonicated (Fisher Scientific, model 100) in 15 second bursts with 1 minute rests. Max power was used and 12 bursts were executed. Streptomycin sulfate (Amresco) was added to the cells at 0.1 g/mL and the conical tubes were swapped for centrifuge tubes. The lysate was cleared by centrifugation (Beckman Coulter, model Avanti HP-20 XPI) at 25000 Xg for 30 minutes at 4°C. The supernatant was mixed with 1 mL of Msl5 Ni Wash buffer-equilibrated nickel nitriloacetic acid (Ni NTA) agarose beads (Thermo Fisher) in a 50 mL conical (VWR) and incubated at 4°C fastened to a tube rotator (Mandel Scientific) for 2 hours. The pellet was collected via centrifugation at 800 X g and the lysate was removed by aspiration. The resin was washed 6 times with 7 mL Msl5 Ni Wash Buffer. Msl5 was eluted in 20 X 1 mL fractions using Msl5 Ni Elution Buffer. Based on A280 values from a spectrophotometer (Nanodrop Spectrophotometer, model ND-1000), only the first 15 elution samples were kept and combined. The Ni-NTA-purified protein was mixed with 1 mg of recombinant TEV protease and put in 10 cm of dialysis tubing (Spectrum Lab, 6-8000 MWCO). TEV

protease is a cysteine protease from the Tobacco Etch Virus, which was recombinantly expressed and purified from pRK793-transformed Rosetta pLysS *E. coli* and according to the recommended conditions (Addgene). The sample was dialyzed in Msl5 Dialysis Buffer for 8 hours at 4°C. The buffer was exchanged three times over 24 hours. The TEV-protease/Ni-NTA elution mixture was then incubated with 1 mL equilibrated (dialysis buffer) Ni NTA in a 15 mL conical (VWR) for 2 hours with rotation at 4°C. The sample was spun and the supernatant (containing cleaved Msl5) was collected and quantified on a 12% polyacrylamide gel.

The cell pellet (from 2L) with expressed Mud2/Msl5 was treated similarly to above but was not dialyzed nor treated with TEV protease. Ni-NTA wash and elution buffers were supplemented with L-Arg, L-Asp, and urea for improved solubility. The protein-bound resin was washed 6 times with 8 mL Mud2Msl5 Ni Wash Buffer and eluted in 1 mL fractions (Table 2.2). The Nanodrop spectrophotometer was used to analyze the fractions, of which 12 were pooled and saved for further purification by size exclusion chromatography. The sample was concentrated from 12 mL to 3 mL using a 15 mL 10 k spin concentrator (Millipore) and centrifuging at 3200 RPM and 4°C (Beckman Coulter, model Allegra X-12R).

2.1.5.2. Step 2: Cation Exchange Column Chromatography Purification

The second purification step for Msl5 was performed with an 8 mL strong cation exchange column (MonoS 10/100 GL) attached to a fast protein liquid chromatography system (ÄKTA Purifier 10, GE Healthcare Life Sciences). The purified Msl5 (no His-tag present) was concentrated from 15 mL to 1.2 mL using the 15 mL 10 k spin concentrator at 3200 RPM and 4°C. Concentrated Msl5 was then quickly mixed with 3.9 mL of a “pH

emersion buffer” (pHE Buffer) and syringed into the FPLC sample injection loop before a MonoS purification program was ran. The program ran from 0 to 100% in terms of Buffer A to Buffer B (Table 2.3).

Table 2.3. Composition of buffers used in the cation exchange chromatography purification of recombinant Msl5. All buffers were filter sterilized and de-gassed.

Components	pHE Buffer	Buffer A	Buffer B
2-(N-morpholino)ethanesulfonic acid (MES)	100 mM	30 mM	30 mM
NaCl	90 mM	90 mM	1 M
β -mercaptoethanol	5 mM	5 mM	5 mM

From the chromatogram and 12% polyacrylamide electrophoresis of the MonoS elution fractions, only fractions B7-B4 were kept (8 mL). The 8 mL was concentrated to 4 mL using a 15 mL 10 k spin concentrator as previously described.

2.1.5.3. Step 3: Size Exclusion Column Chromatography Purification

The final purification step of both recombinant Msl5 and the Mud2/Msl5 dimer was done using a 24 mL size exclusion column (Superdex200 10/300 GL, GE Healthcare Life Sciences) attached to the ÄKTA Purifier 10. In 0.5 mL portions, protein was syringed into the sample injection loop and loaded onto the size exclusion column (SEC). In total, 8 runs were made for Msl5 and 6 were made for the Mud2/Msl5 dimer. The method involved either Msl5 or Mud2Msl5 SEC Buffer with a constant flow of 0.25 mL/min and collection of 0.5 mL fraction sizes (Table 2.4).

Table 2.4. Composition of the buffer used in the size exclusion chromatography purification of recombinant Msl5 and co-expressed Mud2 and Msl5.

The buffers were filter sterilized and de-gassed.

Components	Msl5 SEC Buffer	Mud2Msl5 SEC Buffers
HEPES-NaOH pH 7.5	25 mM	25 mM
NaCl	100 mM	100 mM
β -mercaptoethanol	5 mM	5 mM
L-Arginine	-	200 uM
L-Aspartate	-	200 uM
Urea	-	0.5 M

From the chromatogram and 12% polyacrylamide electrophoresis of the SEC elution fractions, fractions that looked the cleanest by SDS-PAGE were kept and pooled together. Recombinant Msl5 was over 95% pure as determined by band intensity quantification. Msl5 was ready for experimentation.

2.1.6. Western blot of the Msl5/Mud2 Heterodimer

In order to verify the presence of Msl5 in the presumed Mud2/Msl5 dimer after Ni-NTA and SEC purification, a Western blot was performed using the reagents listed in Table 2.5.

Table 2.5. Reagents used in the detection of Msl5 in purified Mud2/Msl5 dimer.

Reagent	Components
10 x Towbin buffer	25 mM Tris pH 8.3, 192 mM glycine, 20% methanol (v/v)
10 x Phosphate Buffered Saline (PBS)	137 mM NaCl, 2.7 mM KCl, 4.3 mM Na ₂ HPO ₄ ·7H ₂ O, 1.4 mM KH ₂ PO ₄
PBS – Tween 20 (PBS-T) solution	0.1% Tween 20 in 1 x PBS solution (v/v)
5% Blocking solution	5% w/v skim milk powder in PBS-T
Wash solution	1% w/v skim milk powder in PBS-T
1° antibody solution (polyclonal rat α Msl5, Dr. John Burke Lab UVIC)	1:1500 in 1% milk/PBS-T (w/v)
2° antibody solution (monoclonal goat α rat, Santa Cruz Biotechnology sc-2065)	1:1800 in 1% milk/PBS-T (w/v)

A 10% polyacrylamide (37.5:1 acrylamide/Bis) gel was loaded with ~14 μ g purified Msl5, Dual Color™ pre-stained protein marker (Bio-Rad), and 2 μ L purified Mud2/Msl5. Electrophoresis was done for 45 min at 22 mA before the gel was taken off and trimmed for transfer. Nitrocellulose and Whatman filter papers were cut to gel size (8.2 Cm X 5.6 Cm) and 3 pieces of Whatman were stacked in the Owl™ semidry electroblotting system (Thermo Fisher). 1X Towbin buffer was poured onto each layer as the sandwich was stacked, and pooled buffer was spotted before the transfer began. The transfer was ran at 70 mA for 1 hour, after which the nitrocellulose membrane was placed

in on a rocker at 4°C with 5% Blocking solution for 16 hours. The membrane was treated with 1° antibody solution (rat α Msl5, gift of Dr. Burke, UVIC) on a rocker at room temperature for 4 hours then washed 3X with Wash solution for 10 minutes each. 1° antibody was generated from rat by Dr. Robert Burke's Laboratory at the University of Victoria. The 2° antibody solution (goat α rat, sc-2065, Santa Cruz Biotechnology) was applied for 12 hours on a shaker at 4°C. The membrane was then washed 3X with Wash solution before treatment with Luminol Reagent (Santa Cruz Biotechnology) and exposed using the Fluor Chem Q imager (Protein Simple).

After the presence of Msl5 was confirmed by Western, the Mud2/Msl5 dimer needed to be quantified. This was not easily quantified by common methods (i.e. spectrophotometry or BCA assay) because of the presence of two proteins of unknown stoichiometry. A concentration gradient of previously purified Msl5 (16667 nM to 1.7 nM) was run next to 2 uL of concentrated Mud2/Msl5 dimer on a 10% polyacrylamide (37.5:1 acrylamide/Bis) gel. All bands were quantified using a program built-in the Fluor Chem software. The band intensities of Msl5 in the dimer were compared with the band intensities of previously purified Msl5 to determine the concentration of dimer, which was assumed to have a stoichiometry of 1:1.

2.2 Methods - Fluorescence Polarization of Msl5 Against Branchpoint RNA

2.2.1. Generation of RNA probes for the *Cm* BPS and Controls

The interaction between Msl5 and the conserved *Cm* branchpoint sequence was measured using fluorescence polarization (FP). Fluorescein-labeled RNA probes for the branchpoint sequence and two negative controls were designed and later synthesized by Integrated DNA Technologies (Table 2.6). Two negative controls were used for FP in

order to see if Msl5 would discriminate against single or double-stranded RNA. An additional unlabeled probe (ro62) was also designed and ordered to act as a negative control in a subsequent EMSA experiment.

Table 2.6. RNA oligo probes used in FP Binding Assays.

Identity	Sequence	Fluorescein
ro52: U4 5' stem	5'-UUGCCCAGAUGAGG UUCUCCGAUGGGUAA-3'	5'-(6-FAM)
ro62: U6 3' end	5'-AGGUAUACCUUUUU-3'	none
ro64: Conserved Branchpoint Sequence	5'-AAACUAACCAU-3'	3'-(6-FAM)
ro66: U4 3' end	AAAUUUUUGGAAAGUAUUU-3'	5'-(6-FAM)

2.2.2. Set-up of the Fluorescence Polarization Binding Assay

Msl5 or Mud2/Msl5 dimer was mixed with RNA probe in 90 uL reactions using 1.5 mL Eppendorf tubes according to Table 2.7 and pipetted them into 384-well black microplates (VWR). The FP Reaction Buffer was 25 mM HEPES-NaOH pH 7.5, 100 mM NaCl, and 5 mM β -mercaptoethanol. Plates were incubated at 42°C then on ice for 10 minutes each before FP was measured with the Synergy 2 Multi-Mode Microplate Reader (Biotek Instruments). Excitation and emission of the fluorescein-labeled samples was done with 485 nm and 528 nm, respectively, which was well within the peak ranges for these parameters (Sjöback *et al.*, 1995). The Msl5 FP assay was ran in triplicate 8X and the dimer in triplicate once. A negative control was always included corresponding to either a U6 snRNA single stranded or U4 5' stem loop double stranded probe.

Table 2.7. FP reagents used for the binding experiments of Msl5 with BP RNA probe. Reactions were each set-up in 1.5 mL microtubes with total volumes of 90 uL.

FP Reaction Reagents	Volume	Final Concentration
Msl5	40 uL	0 to 10000 nM
Mud2/Msl5	40 uL	0 to 222 nM
FP Reaction Buffer	45 uL	1 X
Fluorescein-labelled probe (ro64, ro66, ro52)	5 uL	15 nM

2.2.3. Analysis of FP data

Fluorescence polarization measurements (parallel and perpendicular) were converted into anisotropy values with Gen5 Data Analysis Software (Biotek Instruments) and subsequently into percent of RNA probe bound by protein. RNA % bound values were plotted with their respective protein concentration using Kaleidagraph Software against a Hill Equation (see below) to determine binding affinity (K_D). Equation 1 converted anisotropy to percent of RNA probe bound, while Equations 2 and 3 fit percentages of bound fractions to floating/variable and fixed Hill Equations, respectively:

$$\text{Fraction}_{\text{BOUND of } n} = (\text{Anisotropy}_n - \text{Anisotropy}_{\text{FREE}}) / (\text{Anisotropy}_{\text{BOUND}} - \text{Anisotropy}_{\text{FREE}}) \quad \text{Equation 1}$$

Hill equations used to determine binding affinity:

Floating (non-constrained):

$$m1 + ((m2 - m1) / (1 + ((m3 / m0)^{m4}))); m1=0.001; m2=100; m3=200; m4=\# \quad \text{Equation 2}$$

Fixed (constrained, forced Hill Coefficient):

$$m1 + ((m2 - m1) / (1 + ((m3 / m0)^{\#}))); m1=0.001; m2=1; m3=100 \quad \text{Equation 3}$$

Where $m1$ = offset, $m2$ = max bound (%), $m3$ = K_D , $m4$ = Hill Coefficient.

2.3 Methods – *In vitro* Transcription of Truncated *Cm* 072C RNA

2.3.1. Preparation of DNA Templates For *In vitro* Transcription

Three versions of *Cm* 072C pre-mRNA DNA templates were used for *in vitro* transcription (IVT): wild-type (WT) pre-mRNA, mutant pre-mRNA, and mRNA. All transcripts retain full intron length but specific truncations were made to the 5' and 3' ends of the flanking exons. Both pre-mRNA transcripts kept the full 85 nucleotide-long intron but only 40 nucleotides of the 5' exon and 50 of the 3' exon. WT and mutant pre-mRNA transcripts were 175 nucleotides, while mRNA was 90 (Table 2.8).

Table 2.8. Sequences of the *Cm* 072C transcripts used in binding experiments with Msl5.

Transcript	Sequence (BPS underlined, 5' and 3' intron boundaries bold)
WT 072C pre-mRNA	5'-UUGCGGCAUUCUAGGUACCACCUACACCGGACAAUUUG AAG CAAGU UGACGAAACUUCUGGGGCUUUUCCUCAGAAA UAACCGAACAUGAA <u>ACUAACCA</u> UUGCGAAACAAUCGAUUC GAGUCUAGGACAUUGCCACUUUGGACGCUCUCGUGACGCA GAUGAACGAGCAGAAGGG-3'
Mutant 072C pre-mRNA	5'-UUGCGGCAUUCUAGGUACCACCUACACCGGACAAUUUG AAG GCCGU UGACGAAACUUCUGGGGCUUUUCCUCAGAAA UAACCGAACAUGAA <u>CGGGCCA</u> UUGCGAAACAAUCGAUUC GAGUCUAGGACAUUGCCACUUUGGACGCUCUCGUGACGCA GAUGAACGAGCAGAAGGG-3'
072C mRNA	5'-UUGCGGCAUUCUAGGUACCACCUACACCGGACAAUUUG AAGACAUUGCCACUUUGGACGCUCUCGUGACGCAGAUGAA CGAGCAGAAGGG-3'

WT pre-mRNA cDNA template was previously generated and subcloned into a plasmid by Dr. Martha Stark. Using the WT cDNA template, mutations were made to the 5' splice site and the branchpoint sequence to create the mutant pre-mRNA cDNA template. Primers were then designed to make an exon-truncated cDNA template of WT mRNA from previously extracted *Cm* total RNA. The T7 promoter sequence was incorporated upstream of the 5' end of the forward primer. The reactions and

thermocycler conditions that generated the mRNA cDNA template is outlined below (Tables 2.9 and 2.10).

Table 2.9. Reactions and thermocycler conditions used to generate the DNA template for *Cm* 072C pre-mRNA and mRNA.

<u>Components of Reaction</u>		<u>Thermocycler Program</u>		
<i>Cm</i> total RNA (250 ng/uL)	2 uL	65°C	5 minutes	1 cycle
Reverse primer (20 uM)	1.5 uL	22°C	10 minutes	
Water	up to 12.5 uL			
To the previous reaction was added:				
Superscript III RT (200 units/uL)	0.7 uL			1 cycle
1 st Strand Buffer (5x)	5 uL			
DTT (100 mM)	2.5 uL			
SUPERaseIn (20 units/uL)	1 uL	42°C	60 minutes	
dNTPs (10 mM)	2.5 uL	70°C	15 minutes	
Total volume	24.2 uL			
Previous Reaction (24.2 uL)	7 uL			35 cycles
Taq Polymerase (5000 units/mL)	0.4 uL	95°C	2 minutes	
Taq Reaction Buffer (10x)	5 uL	95°C	20 seconds	
Forward primer (10 uM)	2.5 uL	56°C	20 seconds	
Reverse primer (10 uM)	2.5 uL	68°C	45 seconds	
dNTPs (10 mM)	1 uL	68°C	5 minutes	
Water	up to 50 uL			

Table 2.10. PCR primers used to generate *Cm* 072C DNA templates for *in vitro* transcription and their respective annealing temperatures.

Identity	Sequence
072C fwd	GCAGATGAATTCTTGCGGCATTCTAGGTACCACC
072C rev	GCAGATGGATCCCCCTTCTGCTCGTTCATCTGC
072C mutate 5' SS	CTG AGG AAA AGC CCC AGA AGT TTC GTC AAC GGC CTT CAA ATT GTC C
072C mutate BPS	CT GGG GCT TTT CCT CAG AAA TAA CCG AAC ATG AAA CGG GCC ATT GCG AAA C
072C mRNA fwd	GCAATGAATTC TAATACGACTCACTATAGGG AGC TTGCGGCATTCTAGGTACCACC
072C mRNA rev	GCAGATGGATCCCCCTTCTGCTCGTTCATCTGC

Cm 072C mRNA cDNA template was isolated by electrophoresis with a 1.2% agarose gel ran at 120 V for 80 minutes. The mRNA-sized gel band was excised and purified it as per the gel extraction protocol of the Nucleospin Gel and PCR cleanup kit (Macherey-Nagel).

2.3.2. *In vitro* Transcription of *Cm* 072C pre-mRNA and mRNA

In vitro transcription (IVT) of truncated *Cm* 072C WT and mutant pre-mRNA, as well as mRNA was done cold as per the HiScribe T7 High Yield RNA Synthesis kit (New England Biolabs). Reactions were carried-out in 0.2mL clear PCR tubes (Corning). and 5' end labeled using [γ - 32 P]-labeled ATP. IVT reactions and the subsequent end labeling are outlined below (Table 2.11).

Table 2.11. Composition of the IVT reactions that generated *Cm* 072C RNA transcripts. The reactions were carried out in 20 uL for 12 hours at 37°C.

IVT Reaction Reagents	WT pre-mRNA Reaction	Mutant pre-mRNA Reaction	mRNA Reaction
Template DNA	4 uL linear plasmid (1 ug)	10 uL PCR product (500 ng)	5 uL PCR product (500 ng)
T7 RNA Polymerase Mix	1.5 uL	1.5 uL	1.5 uL
T7 Reaction Buffer (10x)	1.5 uL	1.5 uL	1.5 uL
ATP (100 mM)	1.5 uL	1.5 uL	1.5 uL
UTP (100 mM)	1.5 uL	1.5 uL	1.5 uL
GTP (100 mM)	1.5 uL	1.5 uL	1.5 uL
CTP (100 mM)	1.5 uL	1.5 uL	1.5 uL
Water	7 uL	1 uL	6 uL

Each IVT product was mixed with equal volume formamide, heated at 65°C for 3 minutes then purified on a 6% polyacrylamide (19:1) 7M urea gel which ran in 1 x TBE at 400 V for 60 minutes. RNA gel bands were visualized by UV shadowing, excised into a 1.5 mL Eppendorf tube and crushed with a micro pestle (DWK Life Sciences). The crushed gel was mixed with 400 uL water and heated for 10 minutes at 70°C. The gel

slurry was subsequently loaded onto a Performa Spin Column (Edge Bio) and spun for 3 minutes at 850 xg. RNA precipitation was done by adding 1.5 uL glycogen (20 mg/mL), 1/10 volume 3M sodium acetate, and 1 mL cold ethanol (100%) and subsequently being stored for 18 hours at -20°C. The precipitated RNA was pelleted by centrifugation at 17000 xg for 30 minutes at 4°C. Each RNA pellet was washed twice with 70% cold ethanol before being resuspended in 50 uL water and stored at -20°C before being end-labeled.

2.3.3. 5' End Labeling of RNA Probes And *Cm* 072C RNA Transcripts

Five RNA sequences were end-labeled: 3 072C RNA IVT transcripts and 2 probes ro64 and ro62 (Table 2.6). All were prepared to 10 uM and reactions were set up according to Table 2.12 and subsequently incubated at 37°C for 10 minutes.

Table 2.12. Composition of the 5' end-labeling reaction with [γ -³²P]-labeled ATP.

Reagents for 5' end-labeling	Volumes added (total 25 uL)
RNA (10 uM)	2.5 uL
T4 Polynucleotide Kinase (PNK, 10000 units/mL, NEB)	1 uL
PNK Buffer (10 x)	2.5 uL
[γ - ³² P]-labeled ATP	3 uL
Water	16 uL

When the reactions were completed, they were briefly centrifuged then diluted to 50 uL with water. Each diluted sample was centrifuged through a Micro G-25 Spin Column (Santa Cruz Biotechnology) and quantified (1 uL into 1 mL scintillation fluid) before and after the column using a Hidex scintillation counter. Each labeled RNA was stored at -20°C.

2.4. Methods - EMSAs of Msl5 With Various RNA

2.4.1. Set-up of the Msl5 binding reactions using EMSA

Electrophoretic mobility shift assays (EMSAs) were set up for Msl5 with each of the 3 072C RNA transcripts (WT pre-mRNA, mutant pre-mRNA, WT mRNA) according to Table 2.13. Within each reaction was a final concentration of Msl5 that ranged from 0 to 10000 nM, and 1 uL of RNA had between 24000 to 28000 cpm depending on the RNA. Each reaction was 20 uL. The FP Reaction Buffer (1 X) was 25 mM HEPES-NaOH pH 7.5, 100 mM NaCl, and 5 mM β -mercaptoethanol.

Table 2.13. EMSA reagents used for the binding experiments of Msl5 with 072C RNA.

EMSA Reaction Components	Volume	Final Concentration
FP Reaction Buffer	10 uL	1 X
<i>E. coli</i> tRNA (Sigma-Aldrich)	0.5 uL	5 ug/mL
BSA (Sigma-Aldrich)	0.1 uL	2.5 ug/mL
Supersin (Thermo Fisher)	0.25 uL	0.2 units/uL
Water	4 uL	-
Msl5 protein	4 uL	0 to 10000 nM
[γ - ³² P]-labeled RNA	1 uL	25 nM (24000-28000 cpm)

Each EMSA reaction was incubated at 42°C for 30 minutes prior to running on a 6% polyacrylamide (acrylamide/Bis, Bio-Rad 29:1) native gel in 1 x TBE for 72 minutes at 200 V. The gels were then exposed for 16 hours at -80°C.

2.4.2. Analysis of EMSAs

After sufficient exposure at -80°C, gels were visualized with a phosphor imager (Packard Instrument Company). Bands were quantified with Optiquant Software. Band intensities were converted to percentages of free and bound RNA. Those percentages were then fit to the Hill Equation (Equations 2 and 3, section 2.2.3) against their respective protein concentration. Kaleidagraph (Synergy Software) was used to determine binding affinity (k_D) and cooperativity.

2.5. Methods - Circular Dichroism Spectroscopy of Msl5 with *Cm* BPS RNA

2.5.1. Wavescan of Msl5 free and bound to ro64

Circular dichroism spectroscopy (CD) was used to study the structure and thermal stability of both free and RNA-bound Msl5 protein with the Jasco J-815 CD Spectropolarimeter. Msl5 was diluted to 2.75 μM in CD Buffer (15 mM NaH_2PO_4 pH 7.5, 40 mM NaF) and either scanned alone or with 10 μM RNA corresponding to the BPS (ro64, Table 2.6). Each sample was incubated at 42°C then put on ice for 15 minutes before injection into a quartz cuvette (1 mm path length). For an assessment of total α -helical and β -strand content of Msl5, wavelength scans were performed in the far UV range (260 nm to 185 nm) at 25°C (Peltier temperature controller). Blank measurements were taken for buffer subtraction using the identical make-up excluding presence of protein (when RNA was mixed with protein, it was also present in the blank sample). Program parameters for wavelength scans are included below in Table 2.14.

Table 2.14. Parameters used for far UV wavelength scans of Msl5.

Parameter	Value	Parameter	Value
Accumulations	6 x	Photometric modes	CD and HT
Bandwidth	1.00 nm	Scanning mode	Continuous
Baseline correction	CD Buffer (baseline)	Scanning speed	10 nm/minute
D.I.T.	2 seconds	Sensitivity	Standard (100 mdeg)
Data pitch	0.5 nm	Wavelength range	260 to 185 nm

2.5.2. Secondary Structure Estimations of Free And RNA-Bound Msl5

CD Spectra data points were data-dumped and converted from CD units (mdeg) into mean ellipticity (ME; $\text{deg} \times \text{Cm}^2 \text{dmol}^{-1}$) using SpectraManager software (Jasco). Importantly, the data-dump was also done with CD spectra points from the buffer and RNA together, and was subtracted from the protein + RNA data, so that the effect of RNA was removed from the analysis of protein conformation. CD spectra graphs were

also generated from SpectraManager. The data points were then transferred to Microsoft Excel and converted into mean residue molar ellipticity via the Equation 4 below.

$$\text{ME}(\text{deg X Cm}^2\text{dmol}^{-1}) = [\text{Ellipticity(mdeg X } 10^6)] / [\text{Pathlength(mm)X Protein(uM) X \#peptides}] \quad \text{Equation 4}$$

Mean residue molar ellipticity values for Msl5 and BP RNA-bound Msl5 were each uploaded to the K2D3 server for estimations of total α -helical and β -strand content. From this analysis, anything not attributed to either structural motif is assumed to be an intrinsically disordered region (i.e. flexible linker or random coil). Due to limited (45) input values allowed by the K2D3 server, data points collected from 190 to 245 nm at 1 nm intervals were submitted for analysis. Upon submission, K2D3 utilized a database of theoretical CD spectra provided by DichroCalc to estimate secondary structure (Louis-Jeune *et al.*, 2011).

2.5.3. Thermal Stability Of Free Msl5

The Jasco J-815 spectropolarimeter and Peltier temperature controller were used for a thermal denaturation experiment to measure the melting temperature (or stability) of Msl5 by observing the CD signal at a wavelength of 222 nm over the temperature range 25°C to 90°C. Msl5 was diluted to 18 uM in CD Buffer (15 mM NaH₂PO₄ pH 7.5, 40 mM NaF) and immediately transferred to the 1 mm quartz cuvette. Program parameters are as previously described except where outlined in Table 2.15.

Table 2.15. Parameters used for thermal denaturation of Msl5.

Parameter	Value
D.I.T.	4 seconds
Data pitch	0.5°C
Ramp rate	2°C/minute
Wavelength	222 nm
Temperature Range (°C)	30°C - 80°C

The melting temperature (T_M) is the temperature at which the sample has lost 50% ellipticity, or the protein is 50% unfolded (Consalvi *et al.*, 2000). Once a thermal spectrum was acquired, T_M of Msl5 was calculated from a built-in T_M algorithm in the SpectraManager software suite (Jasco).

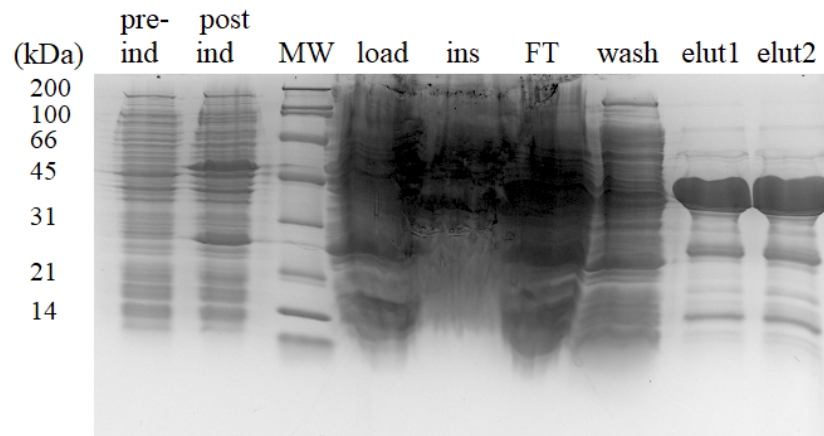
2.6 Results and Discussion

2.6.1. Recombinant Expression and Purification of Msl5

His-CmMsl5 was affinity purified by incubation with Ni-NTA resin, washed, and then eluted by centrifugation. The expression and elution fractions were subsequently analyzed on a 12% polyacrylamide SDS-PAGE (25 mA, 50 min) (Fig 2.1 A). The His-Msl5 construct is 38.64 kDa; on 10-12% polyacrylamide SDS-PAGE, it migrated very near the 45 kDa band of the Broad Range Molecular Weight Marker.

After batch purification, His-Msl5 was mixed with His-TEV protease (28.62 kDa) and dialyzed. Cleaved Msl5 was isolated from non-cleaved Msl5, cleaved His-tag, and TEV protease (which contains a His-tag), by mixing with the Ni-NTA resin followed by centrifugation. The flow-through should contain cleaved Msl5, while everything with a His-tag should remain bound to the Ni-NTA resin. Cleaved Msl5 is 36.39 kDa in size and ran slightly below the non-cleaved product that is 38.64 kDa (lanes 2-3, Fig. 2.1 B). Also visible is a band near 31 kDa marker that corresponds to the TEV protease (lane 3, Fig. 2.1 B). The flow-through looked to contain cleaved Msl5 and some contaminants, but not non-cleaved product or TEV protease (lanes 4-5, Fig. 2.1 B).

A



B

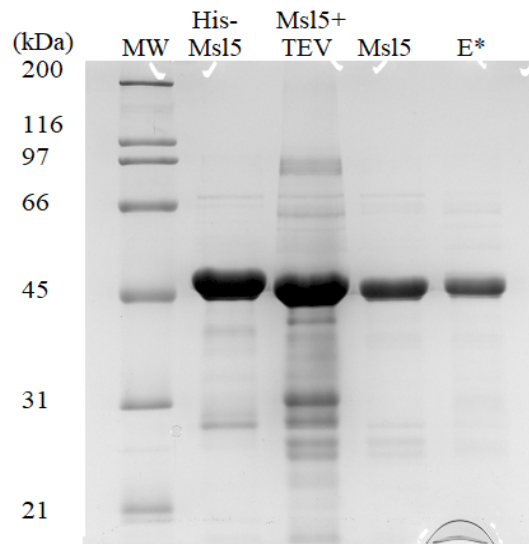


Figure 2.1. Expression and affinity purification of Msl5. (A) The expression and Ni-NTA batch purification of His-*CmMsl5* analyzed on a 12% SDS-PAGE. Pre- and post-induced lysates (lanes 1-2) revealed the expressed protein was near the 45 kDa marker. The load, insoluble, flow-through, and wash are not readable (lanes 4-7). Both elution lanes show high amounts of His-Msl5. (B) SDS-PAGE analysis of Msl5 after cleavage by TEV-protease and subsequent purification through Ni-NTA resin. The cleaved protein (lanes 3-4) migrates slightly faster than His-Msl5 (lanes 2 and 4). E* contains uncleaved His-Msl5 and cleaved His-tag separated from cleaved Msl5 (lane 5).

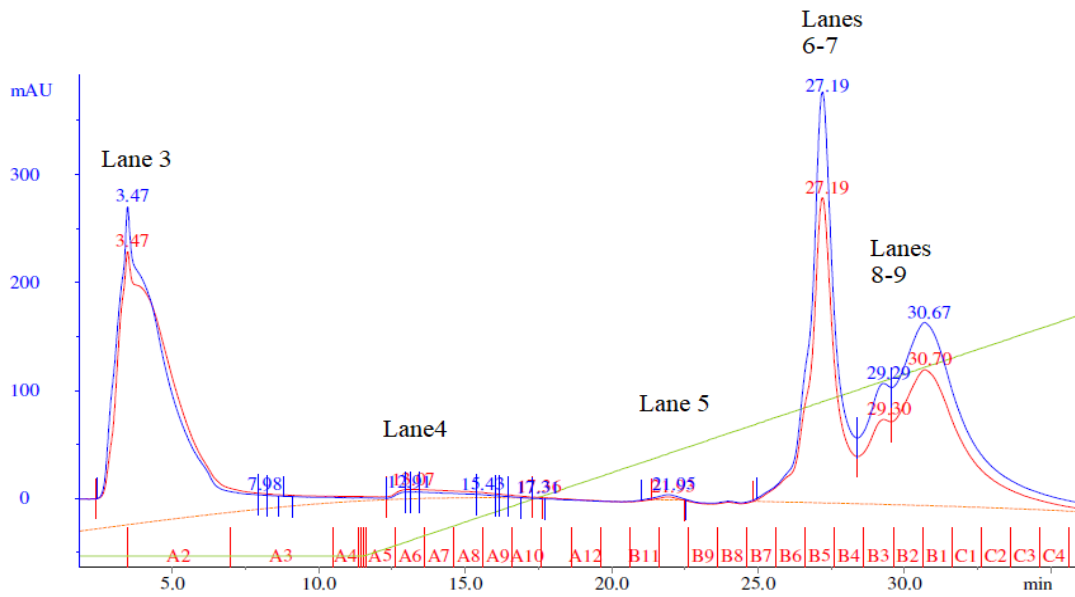
Affinity purified Msl5 was injected onto the 8 mL MonoS 10/100 GL cation exchange column attached to the ÄKTA Purifier10 FPLC system (Fig. 2.2 A). In the chromatogram, only 3 colored lines are shown: red (260 nm wavelength), blue (280 nm),

and green (gradient of 100 mM to 1 M salt in the MES buffer). The 260 and 280 nm wavelengths are excellent in analytical biochemistry; amino acids tryptophan and tyrosine optimally absorb light at approximately 280 nm, while nucleic acids absorb best at 260 nm. Thus the ratio of 260:280 nm is widely used as a gauge for protein purity and nucleic acid contamination (Glase1, 1995). Approximate 260:280 ratios for various amounts of protein and nucleic acid are outlined in Table 2.16 below.

Table 2.16. Ratios for wavelengths 260:280 (nm) of protein and RNA (Held 2001).

Amount Protein (%)	Amount Nucleic Acid (%)	260 nm : 280 nm
100	0	0.60
90	10	1.2
70	30	1.6
50	50	1.75
0	100	1.95

A



B

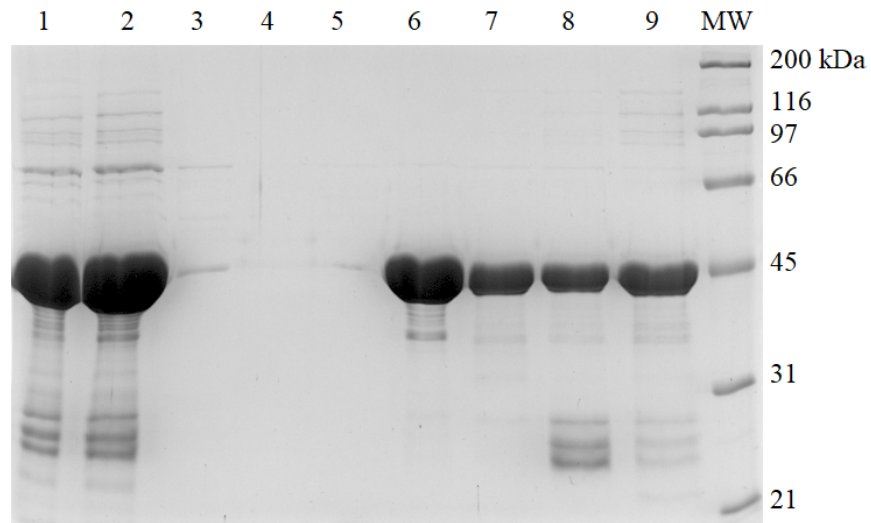


Figure 2.2. Cation exchange purification of Msl5. (A) Chromatogram of cation exchange (MonoS) purification using the ÄKTA Purifier10 for *C. merolae* Msl5. (B) SDS-PAGE analysis of elution fractions. Lanes 1 and 2 are the protein sample (2.5 ug) before and after buffer exchange, and are not represented in the chromatogram. Lane 3 is the flow-through. Lanes 4-5 are the wash fractions. Lanes 6-7 are each elution fractions from the first elution peak (27.19 min). Lanes 8 and 9 are the elution fractions at 29.29 and 30.67 min, respectively. MW refers to the protein marker.

In the chromatogram are 3 major peaks; fractions from each of the peaks were analyzed by SDS-PAGE (Fig. 2.2 B). From the profile of each peak (height, width, and 260:280), the middle peak looked to be the cleanest (Fig. 2.2 A). Indeed, from inspecting

the load (lanes 1-2), peak 1 (lane 3), peak 2 (lanes 6-7), and peak 3 (lanes 8-9), it's clear that cation exchange was useful in removing contaminants (Fig. 2.2 B). Interestingly, several light bands slightly below the Msl5 band (around the 45 kDa marker) followed Msl5 throughout both affinity and cation exchange chromatography (lanes 6-7, Fig. 2.2 B). This lead me to suspect that they are C-terminally truncated versions of Msl5 made during recombinant expression. Reasons for truncated products include cryptic initiation sites in frame with the coding region, premature stop codons, and limited proteolysis; of these, the latter two are most likely due to the role of the N-terminal His-tag in affinity purification (Kramer and Farabaugh, 2007; Jennings *et al.*, 2016; Baggett *et al.*, 2017). Even with a shortened tail, truncated Msl5 would have the His-tag and have a very similar amino acid constitution, resulting in it co-purifying with full-length protein (Jennings *et al.*, 2016). There are genetic and biochemical ways to ensure only full-length protein is purified: silent mutations and a C-terminal tag. Silent mutations made to the gene would lower the frequency of premature termination and an additional tag (i.e. FLAG) on the C-terminal end of the protein would exclude C-terminal truncations (Jennings *et al.*, 2016).

In an effort to remove the truncated products (and other contaminants) from Msl5, size exclusion chromatography was used. According to chromatograms from Gel Filtration Standards (Bio-Rad) being run on the same system, the peak at 14.5 mL translates to 45 +/- 10 kDa (Fig. 2.3 A; La Verde *et al.*, 2017). While the standards and work by La Verde suggested the Superdex200 column would not be suitable, I performed the purification due to the availability of size exclusion columns and because Msl5 had not yet been attempted on the system (2017). Upon analysis by SDS-PAGE, it's not clear

that fractions from the SEC peak (lanes 3-6) are more pure than the load (Fig. 2.3 B). In subsequent biological replicates, *CmMsl5* was purified by batch affinity chromatography followed by cation exchange chromatography.

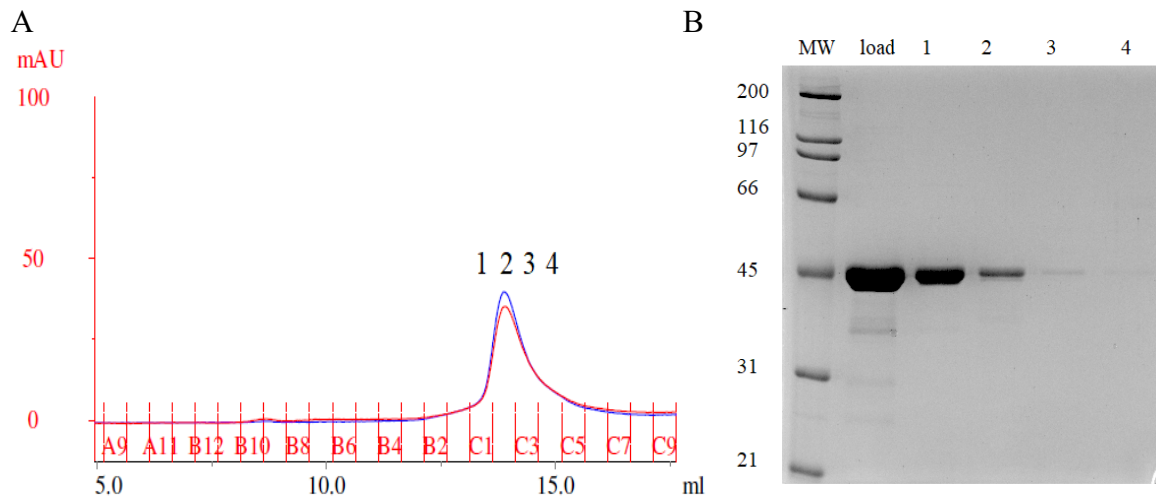


Figure 2.3. Size exclusion purification of Msl5. (A) Chromatogram of *Cm Msl5* after size exclusion (Superdex200) purification using the ÄKTA Purifier10. Peaks are labeled according to the fractions analyzed. (B) SDS-PAGE analysis of peak fractions from the chromatogram in A. Lanes 1-4 in the gel refer to fractions C1-C4 in the chromatogram.

2.6.2. Bioinformatics of Msl5

As discussed in the introductory chapter, there is no Py-tract or poly-U region in *C. merolae*. While Mud2 presumably recognizes a sequence downstream of the BPS that is not enriched with uridine or cytosine, no other information is yet known. This suggests Msl5 and Mud2 are likely behaving similarly to budding yeast, where BBP (the yeast homolog) defines the BPS rather than a Py-tract-driven interaction (Berglund *et al.*, 1997; Garrey *et al.*, 2006). To address this question, I first wanted to see how similar the Msl5 sequence was to its yeast and human counterparts (Selenko *et al.*, 2003; Chang *et al.*, 2012; Jacewicz *et al.*, 2015).

An initial BLAST (Basic Local Alignment Search Tool) search was not fruitful, as protein sequences nearest *C. merolae* were often other extremophiles such as *Galdieria sulphuraria*. As such, I performed a multiple sequence alignment against homologs to *C. merolae* Msl5 that included human (SF1) and budding yeast (BBP). The alignment was for the purpose of observing similarity and mapping predicted secondary structures (Fig. 2.4). I've included the solution structure of a domain of human SF1 (*C. merolae* Msl5 homolog) in complex with RNA, as well the crystal structure of human SF1-U2AF65 interacting domains (Fig. 2.5-2.7; Liu *et al.*, 2001; Wang *et al.*, 2013). The alignment was done using STRAP and CLUSTALW software (Fig. 2.4; Gille *et al.*, 2014).

Msl5 is 36.6% similar to human and 25.2% identical (Fig. 2.4). Msl5 is 32.3% similar to yeast and 21.5% identical. Yeast and human are 45.7% similar and 36.9% identical. Thus Msl5 is more similar to human than yeast, but yeast and human proteins are the most similar based on sequence (Fig. 2.4). Protein structure prediction done using the Phyre2 (Protein Homolgy/AnalogY Recognition Engine) server revealed the highest confidence and coverage (100.0% confident, 43% Coverage) was the PDB entry 4WAL, which is the crystal structure of yeast BBP (Kelley *et al.*, 2015; Jacewicz *et al.*, 2015). Furthermore, the Phyre2 homology modeling of Msl5 with yeast BBP produced the same alignment as done with STRAP. Together these results give confidence in the secondary structure predictions and domain assignment for Msl5. To that end, using alignments and modeling, it appears that Msl5 is missing canonical sequences for the ULM, SPSP/PPxY, and zinc knuckle motifs (Fig. 2.4).

```

                                ULM
MSL5 ..... 0
SF1 .....MATGANATPLDFPSKKRKRSRWNQDTMEQKTVIPGM. 36
BBP MSFRRINSRYFENRKGSMEEEKKAKVPPNVNLSLWRKNTVESDVHRFNSL 50
                                * * *

                                ULM                                SPSP motif
MSL5 .....MPRNSERLQQIRERVEQIGRLLRGEERINFRVLEPVEKDSII 43
SF1 PTVIPPGLTREQERAYIVQLQIEDLTRKLR...TGDLGIPPNPEDRSPSP 83
BBP PSKISGALTREQIYSYQVMFRIQEITIKLR...TNDF.VPPSRKNRSPSP 96
    * * **!***** ** * ** * ** * ** * ** *

PPxY motif
MSL5 APVYDRAGRRVNTAFQRACAILAERDDLLFEAFALDPKFRLLPPGCPPP. 92
SF1 EPIYNSEGKRLNTREFRTRKKLEERHNLITEMVALNPDKPPADYKPPA 133
BBP PPVYDAQGKRTNTREQRYRKKLEDERIKLVEIALKTIPYFVPPDDYKRP. 145
    !*!* **! **! **! **! **! **! **! **! **! **! **!

                                KH-QUA2
MSL5 .RAESKVYFPVDKYPHMNFAGLVLGPRGVTQKRIEERFRCLLIRGRGAR 142
SF1 TRVSDKVMIPQDEYPEINFVGLLIGPRGNTLKNIEKECNAKIMIRGKGSV 183
BBP TKFQDKYYIPVDQYPDVNFVGLLLGPRGRTLRLKLQEDSNCKIAIRGRGSV 195
    ** *!****!*! **! **! **! **! **! **! **! **! **!

                                KH-QUA2
MSL5 SRDAG.....DDALHARIEAVG.ADARQ RVAACAKYLKEEILVP 180
SF1 KEGKVGRKDGQMLPGEDEPLHALVTANTMENVKKAVEQIRNILKQGIETP 232
BBP KEGKNASDLPPGAMNFEDPLHCLIIADSEDKIQKGIKVCQNIVIKAVTSP 244
    **** **! **! **! **! **! **! **! **! **!

                                KH-QUA2                                BBP Zn1/SF1 Zn
MSL5 RRDEENALKIAQLRELAAMNGTLREDARIARLAAERAVRAAVAHRHETEP 229
SF1 ..EDQNDLRKMQLRELARLNGTLREDDNRILRPWQSSETR SITNTTVCTK 281
BBP ..EGQNDLKRGQLRELAELNGTLREDNR.....PCPI 275
    * *! **! **! **! **! **! **! **! **!

                                Zinc Knuckle                                Zinc Knuckle 2
MSL5 QVDSLEQEMESFLQEVGVSV EADATHRSGDASAHDPSGVPALNFDQLGKT 279
SF1 CGGAGHIASDCKFQRP G D P Q S A Q D K A R M D K E Y L S L M A E L G E A P V P A S V G S 331
BBP CGLKDHKRYDCPNRKIPNIQGIVCKICGQTGHFSRDCNS..... 314
    ** * ** * * * * * *

MSL5 SPTSNISTEKRILESEAEASPQANSNQSVMRDARKRPRSASVQGDS 325
SF1 TSGPATTPLASAPRPAAPANNPPPPSLMSTTQSRPPWMNSGPSES RPYHG 381
BBP .....SSQRM SRFDRNATVNNSAPIQSN DVH. 340
    * * * * * * * *

```

Figure 2.4 Alignment and predicted domain structures of the Msl5 against human (SF1) and yeast (BBP) homologs. Similarity was shown with “*” and identity with “!”. The UniProt identifiers for the sequences in (B) for *Cm* Msl5, *Hs* SF1, and *Sc* BBP are CM129C, Q15637, and Q12186, respectively.

Budding yeast and some other invertebrates have two zinc knuckle motifs; human SF1 and vertebrate homologs have a single conserved zinc knuckle, while Msl5 has none (Berglund *et al.*, 1997). The single zinc knuckle of SF1 is labeled “SF1 Zn/Zinc Knuckle,” while the second zinc motif of yeast BBP is labeled “Zinc Knuckle 2” (Fig. 2.4). The conserved zinc knuckle is implicated in protein-protein interactions and possibly negatively regulates RNA binding activity of BBP/SF1, as a truncated variant that lacks the conserved zinc knuckle has an 8-fold improved affinity for RNA (Garrey *et al.*, 2006). It is also possible that the conserved zinc knuckle is involved in the displacement of BBP/SF1 from the BPS when U2 base pairs with the BPS/3' SS (Garrey *et al.*, 2006; Wang *et al.*, 2008; Chang *et al.*, 2012). Interestingly, the additional non-conserved knuckle in yeast is involved in RNA binding and removal of that motif results in a complete loss of RNA binding ability (Garrey *et al.*, 2006). Furthermore, the presence of hydrophobic residues linking the two yeast zinc knuckles suggests an additional alpha helix is present in yeast BBP (Fig. 2.4; Garrey *et al.*, 2006). Those elements are unique for a branchpoint bridging protein and may be related to the preference of yeast BBP for a BPS with an upstream stem loop (Berkowitz *et al.*, 1995; Berglund *et al.*, 1997; Garrey *et al.*, 2006). No structures of Msl5 homologs currently exist for any domains other than the KH-QUA2 and ULM.

The ULM (U2AF Ligand Motif) interacts with the reciprocal UHM (U2AF Homology Motif) of Mud2/U2AF65 (Fig. 2.5; Selenko *et al.*, 2003; Chang *et al.*, 2012). The ULM is characterized by several positively charged amino acid residues (usually Lys and Arg) followed by a tryptophan that is known to interact with non-canonical RRM, such as the UHM of *ScMud2/U2AF65* (Fig. 2.5; Kielkopf *et al.*, 2001; Zhang *et al.*,

2012; Wang *et al.*, 2013). The UHM-interacting tryptophan is position 22 in human and 35 in yeast (Fig. 2.4). Msl5 does not contain a single tryptophan residue, or a stretch of contiguous positively charged residues. This brings to question the legitimacy of an Msl5-Mud2 interaction in *C. merolae*, which is addressed later in this chapter.

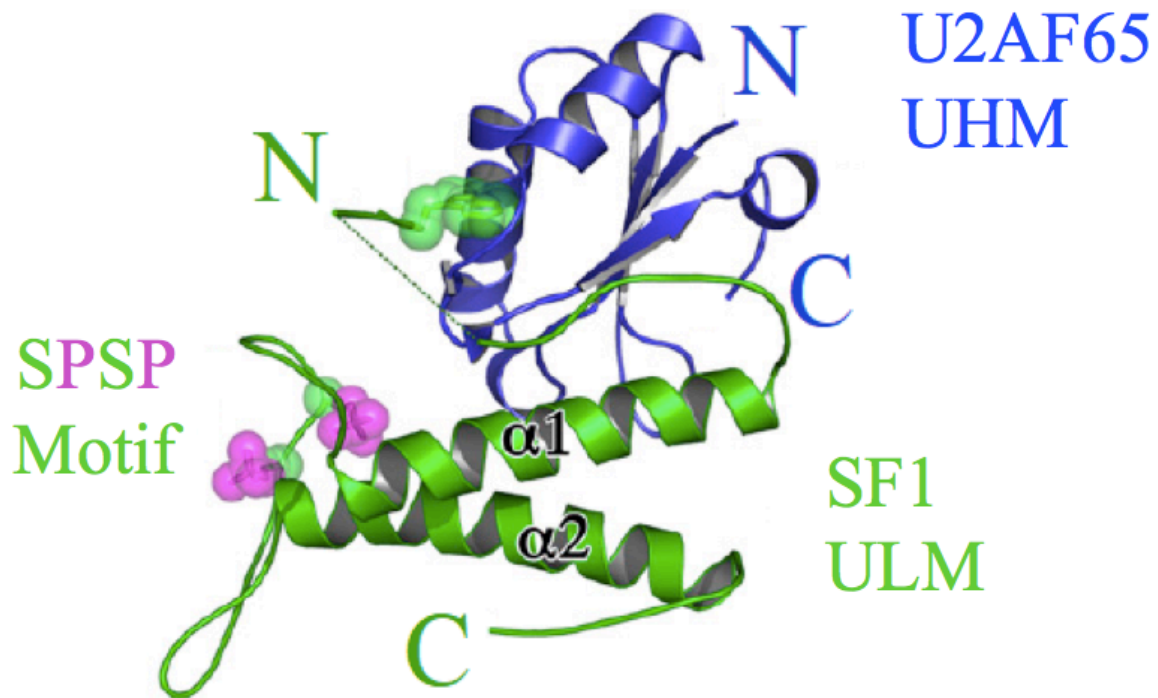


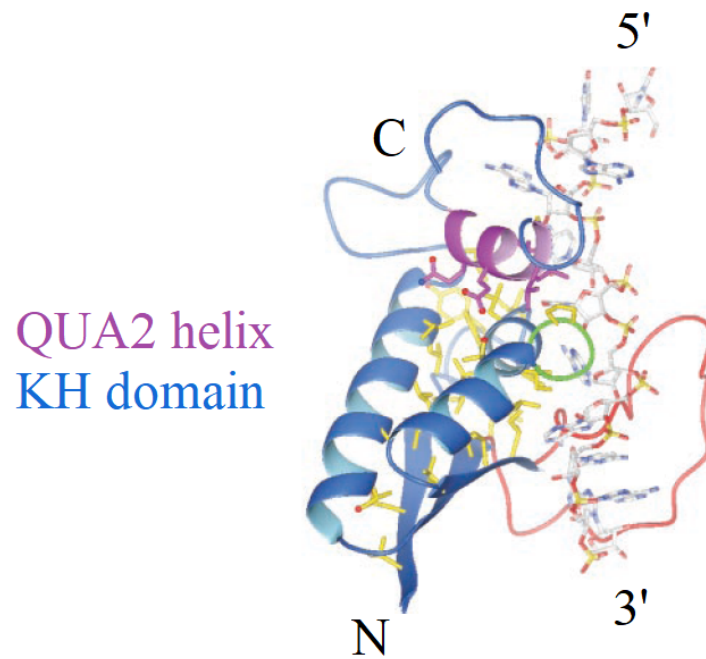
Figure 2.5. The crystal structure of the SF1-U2AF65 complex highlights the interaction between the ULM and UHM. This human model was adapted from Wang *et al.*, 2013.

Msl5 is also missing the canonical SPSP and PPxY motifs, which have been shown to interact with Prp40 and other U1 snRNP proteins (Fig. 2.5; Chang *et al.*, 2012). These motifs are within a coiled-coil domain, and have been found to regulate the splicing of a subset of human pre-mRNA transcripts (Lipp *et al.*, 2015). In addition, phosphorylation of the SPSP motif induces an “arginine claw” fold around the phosphorylated serines. This event changes the structure of the ULM and makes the coiled-coil interface more accessible to the UHM of *ScMud2*/U2AF65 (Manceau *et al.*,

2006; Zhang *et al.*, 2013). Phosphorylation of the SPSP also increases the specificity of SF1 for UHM targets, but decreases the affinity for RNA binding (Chatrikhi *et al.*, 2016).

Surprisingly, only the KH-QUA2 (hnRNP K Homology-Quaking 2 Homology) domain is conserved in *Cm* Msl5 (Fig. 2.6 A). The structure of the KH-QUA2 domain in complex with RNA has been solved with both solution NMR (Nuclear Magnetic Resonance) and X-ray crystallography (Fig. 2.6; Liu *et al.*, 2001; Jacewicz *et al.*, 2015). KH-QUA2 folds into the canonical $\beta 1-\alpha 1-\alpha 2-\beta 2-\beta 3-\alpha 3-\alpha 4$ topology adopted by the STAR (Signal Transduction and Activation of RNA) family of proteins (Vernet and Artzt, 1997; Liu *et al.*, 2001). For clarification, the KH domain itself refers to $\beta 1-\alpha 1-\alpha 2-\beta 2-\beta 3-\alpha 3$ and QUA2 makes up $\alpha 4$ (Fig. 2.4; Liu *et al.*, 2001; Valverde *et al.*, 2008). A hydrophobic groove brings the N- and C-terminal regions of the KH-QUA2 domain together, which interact with the 3' and 5' ends of the BPS, respectively (Fig. 2.6 B; Liu *et al.*, 2001). The conserved QUA2 motif, the GPRG loop, and a loop within the KH domain all interact to create the hydrophobic cleft rich in arginine and serine that binds the single-stranded negatively charged RNA (Fig. 2.6 B; Liu *et al.*, 2001). Binding experiments with various truncations have shown that this domain is necessary and sufficient for binding RNA for SF1 (Berglund *et al.*, 1998a; Liu *et al.*, 2001).

A



B

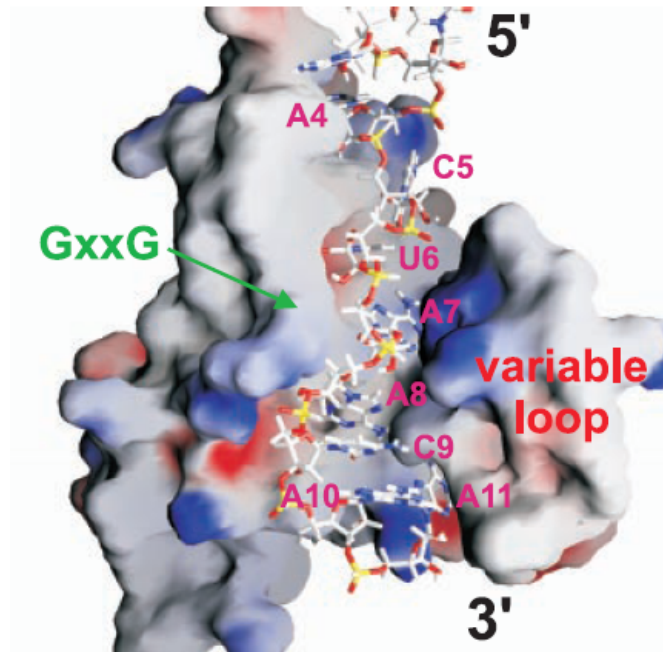


Figure 2.6. Predicted structure and domains of Msl5. (A) Ribbon representation of the SF1 KH-QUA2 domain in complex with 11 nucleotides of the human BPS. (B) Surface representation of the hydrophobic cleft that interacts with the BPS. The GPRG (GxxG) loop and KH domain (variable loop) are highlighted. These human models were adapted from Liu *et al.*, 2001.

2.6.3. Functional Interactions of Msl5 with the Branchpoint Sequence

The human BPS is quite degenerate; accordingly, SF1 is able to recognize YXCURAY, where Y is a pyrimidine, R is a purine and X is any nucleotide (Keller and Noon, 1984; Green, 1986). Yeast BBP binds the highly conserved UACUAAC sequence, which human SF1 prefers as well (Moore *et al.*, 1993; Berglund *et al.*, 1997). The BPS in *C. merolae* is xACUAAC and is conserved across all 27 introns of the *Cm* intronome (Matsuzaki *et al.*, 2004; Stark *et al.*, 2015). To test whether or not Msl5 had similar affinity for the BPS than its yeast and human homologs (Table 2.17), I performed a series of RNA binding experiments and employed both fluorescence polarization and electrophoretic mobility shift assays (EMSAs).

Table 2.17. Binding affinities of Msl5 homologs for the branchpoint sequence of pre-mRNA.

Protein	RNA target	K _D (nM)	Method	Literature
SF1	22nt rp51A pre-RNA	30000	EMSA	Berglund <i>et al.</i> , 1997
BBP	22nt rp51A pre-RNA	500	EMSA	Berglund <i>et al.</i> , 1997
SF1	34nt AdML pre-mRNA	6000	EMSA	Berglund <i>et al.</i> , 1998b
U2AF65	34nt AdML pre-mRNA	1000	EMSA	Berglund <i>et al.</i> , 1998b
SF1 + U2AF65	34nt AdML pre-mRNA	300	EMSA	Berglund <i>et al.</i> , 1998b
SF1	24nt rp51A pre-RNA (upstream SL)	340	EMSA	Garrey <i>et al.</i> , 2006
SF1	22nt rp51A pre-RNA	310	EMSA	Garrey <i>et al.</i> , 2006
BBP	22nt rp51A pre-RNA	25	EMSA	Garrey <i>et al.</i> , 2006
BBP	24nt rp51A pre-RNA (upstream SL)	4	EMSA	Garrey <i>et al.</i> , 2006
U2AF65	57nt B3P3 pre-mRNA (Mullen <i>et al.</i> , 1991)	509 ± 4	ITC	Chatrikhi <i>et al.</i> , 2016
SF1 (phosphorylated)	57nt B3P3 pre-mRNA	78 ± 9	ITC	Chatrikhi <i>et al.</i> , 2016
SF1	57nt B3P3 pre-mRNA	40 ± 1	ITC	Chatrikhi <i>et al.</i> , 2016
SF1 + U2AF65	57nt B3P3 pre-mRNA	2 ± 1	ITC	Chatrikhi <i>et al.</i> , 2016

I used fluorescence polarization as a binding assay to test the specificity of Msl5 for BPS RNA. RNA oligonucleotides ro64, ro66, and ro52 corresponded to the *Cm* BPS (3'-fluorescein), 3' U4 snRNA (5'-fluorescein), and U4 5' stem loop (5'-fluorescein). Using the mfold server, secondary structures were predicted for each of the RNA oligonucleotides used in FP and EMSA (Zuker and Jacobson, 1998; Waugh *et al.*, 2002; Zuker, 2003). Of the 4 RNA sequences, only ro52 (the U4 5' Stem Loop) generated a secondary structure (Fig. 2.7).

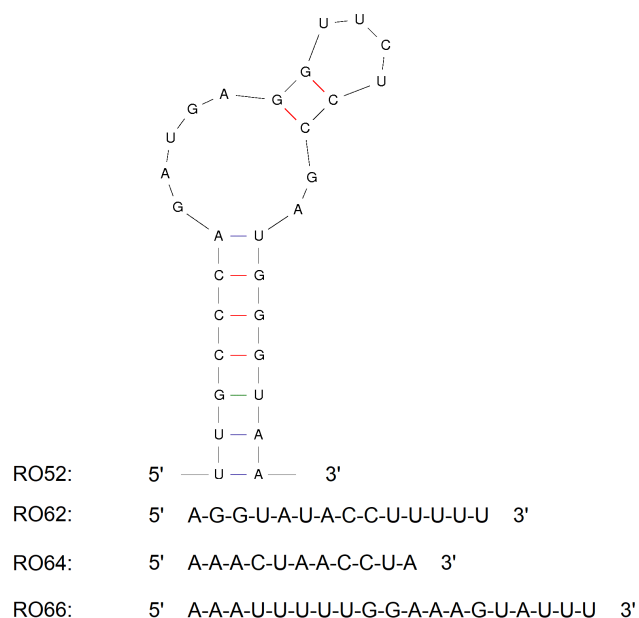


Figure 2.7. Simulated RNA folding of the U4 5' Stem Loop (nucleotides 22-50), corresponding to ro52 used in various binding experiments. ro52 is the *C. merolae* U4. The remaining RNA oligonucleotides did produce reasonable structures and are mostly single-stranded. The predicted secondary structure was generated with the mfold server, and had a predicted free energy of Delta G = -10.10 (initially -9.60) kcal/mol.

Fluorescein was chosen as a fluorophore because it has strong excitation and emission peaks, and because it can handle several exposures with minimal photo bleaching (Jöback *et al.*, 1995). The saturation curve below shows the interaction between the *C. merolae* branchpoint bridging protein Msl5 against ro64, ro66, or ro52 (Fig. 2.8). Notably, the solid black line (ro64, *Cm* BPS, ssRNA) is very steep and quickly becomes a plateau, while spaced (ro66, U6 3' end, ssRNA) and dotted (ro52, U4 5' Stem Loop, dsRNA) never plateau (Fig. 2.8). The “plateau” on a curve indicates the labeled ligand population has become fully bound.

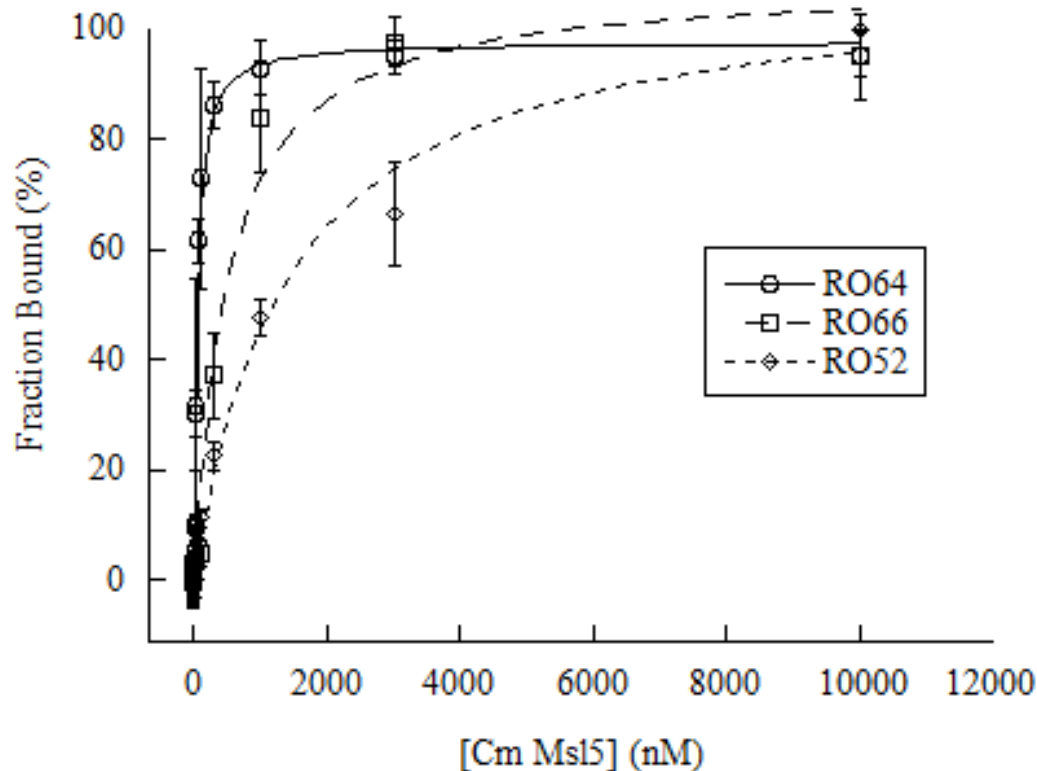


Figure 2.8. Saturation curve from fluorescence polarization measurements of Msl5 titrated against three RNA targets (15 nM). Each curve represents the binding of *Cm*Msl5 to the *Cm* BPS (ro64, solid), U4 3' end (ro66, slashed), and U4 5' stem loop RNA (ro52, dotted). Curve fitting for all 3 data sets was done using Equation 2.

The data were fitted to the Hill Equation to quantify cooperativity and binding affinity (section 2.2.3), both with the Hill Coefficient floating (Equation 2) and fixed to 1 (Equation 3). While minimizing constraints is often the best way to analyze an interaction, I wanted to see how the model behaved if the Hill Coefficient was set to 1. The Hill Coefficient describes the cooperativity of an interaction; a value of 1 indicates 1 molecule of the ligand is interacting with one other molecule. Given that the interaction between homologs of Msl5 and the BPS have been characterized and shown to be 1:1, I expected Msl5 to behave similarly (Liu *et al.*, 2001; Garrey *et al.*, 2006; Chatrikhi *et al.*, 2016). As such, there should be little difference in binding affinity between a variable

and fixed coefficient if Msl5 truly interacts with the BPS RNA in a non-cooperative (1:1) manner. Indeed, the K_D of Msl5 for the BPS RNA (ro64) was 44 ± 4 nM when the Hill Coefficient was fixed to 1, and was 42 ± 2 nM when the Hill Coefficient was allowed to float (Table 2.18). Each data point on the curve represents the average from 12 replicates, and the error bars were generated from the standard deviations of those replicates (Fig. 2.8). From these binding experiments, Msl5 has very similar affinity for the conserved *Cm* BPS relative to published work with its homologs (Tables 2.17 and 2.18). The affinity of Msl5 against 3' end of U4 (ro66) was 10 X worse (400-500 nM), the error and cooperativity were very high, further indicative of a nonspecific interaction (Table 2.18).

Table 2.18. Binding affinities of Msl5 for RNA oligonucleotides measured by FP.

RNA target	K_D (nM)	Hill Coefficient	R^2 (goodness of fit)	Fit to Equation
ro64 (<i>Cm</i> BPS)	42 ± 2	1.3 ± 0.07	0.99878	2
ro64 (<i>Cm</i> BPS)	44 ± 4	1	0.99716	3
ro66 (ssRNA)	440 ± 3000	27 ± 500	0.61990	2
ro66 (ssRNA)	490 ± 120	1	0.98849	3
ro52 (dsRNA)	3400 ± 1000	0.7 ± 0.05	0.99841	2
ro52 (dsRNA)	1400 ± 180	1	0.99498	3

I next wanted to test how well Msl5 binds a *Cm* pre-mRNA transcript, and how well it discriminates against a mutated BPS. To that end I used the electrophoretic mobility shift assay (EMSA) and titrated increasing concentrations of Msl5 against 25 nM RNA. In order to ensure the technique was analogous to FP, ro64 and ro62 were 5' end-labeled with [32 P] (Fig. 2.9). A single gel shift of each was performed as a proof of principle for comparing Msl5 binding via FP and EMSA (Fig. 2.9). Msl5 begins to noticeably shift the [32 P]-labeled ro64 at 50 nM, although not quite by 50% as seen in FP (Fig. 2.9 A). Surprisingly, even at 5000 nM Msl5 does not shift ro62 (Fig. 2.9 B). These results highlight the specificity Msl5 has for the BPS.

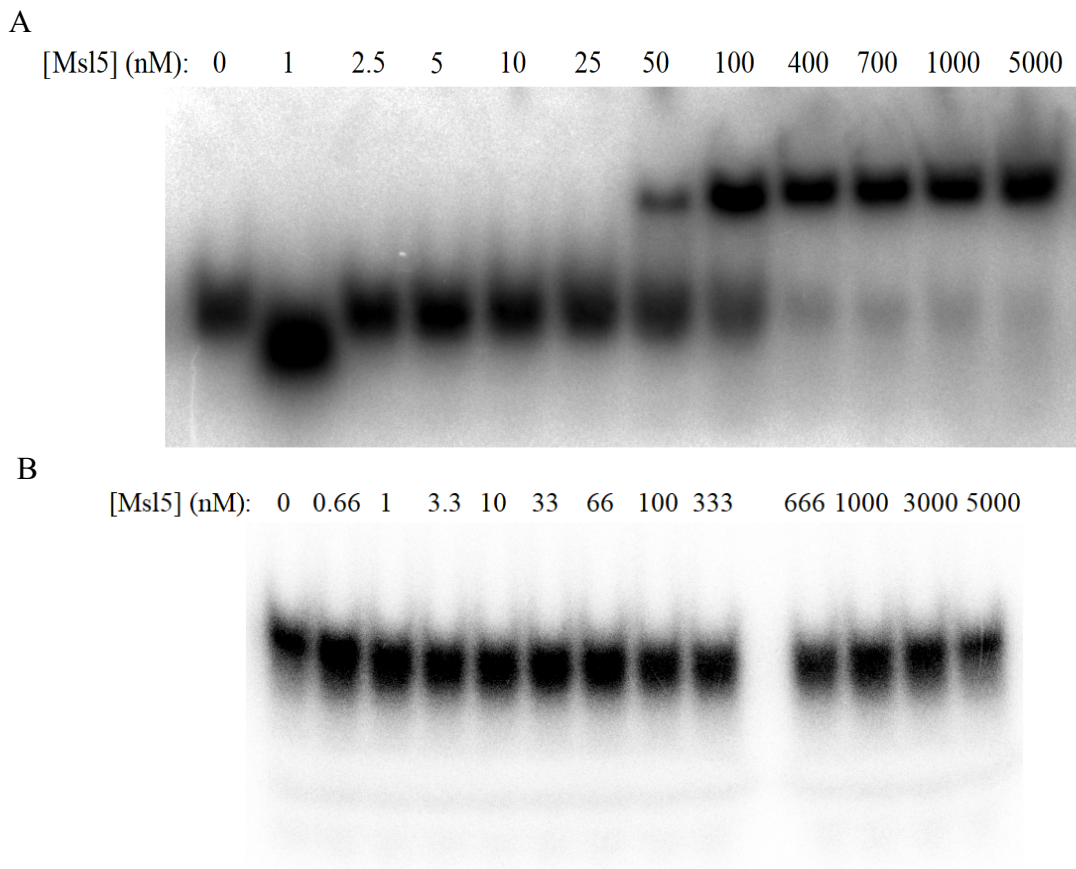


Figure 2.9. EMSA measurements of Msl5 binding *Cm* BPS. (A) Msl5 titrated against the *Cm* BPS (ro64). (B) Msl5 titrated against the 3' end of U6 snRNA (ro62). Both RNA probes were [32 P]-labeled and 25 nM (24000-28000 cpm) was mixed in each reaction.

The bands were quantified as intensities, which were converted into percentage of RNA free and bound. Percentages of RNA bound were plotted against [Msl5] (nM) and fitted against the Hill Equation to generate saturation curves and binding affinities (Fig. 2.10; Table 2.19).

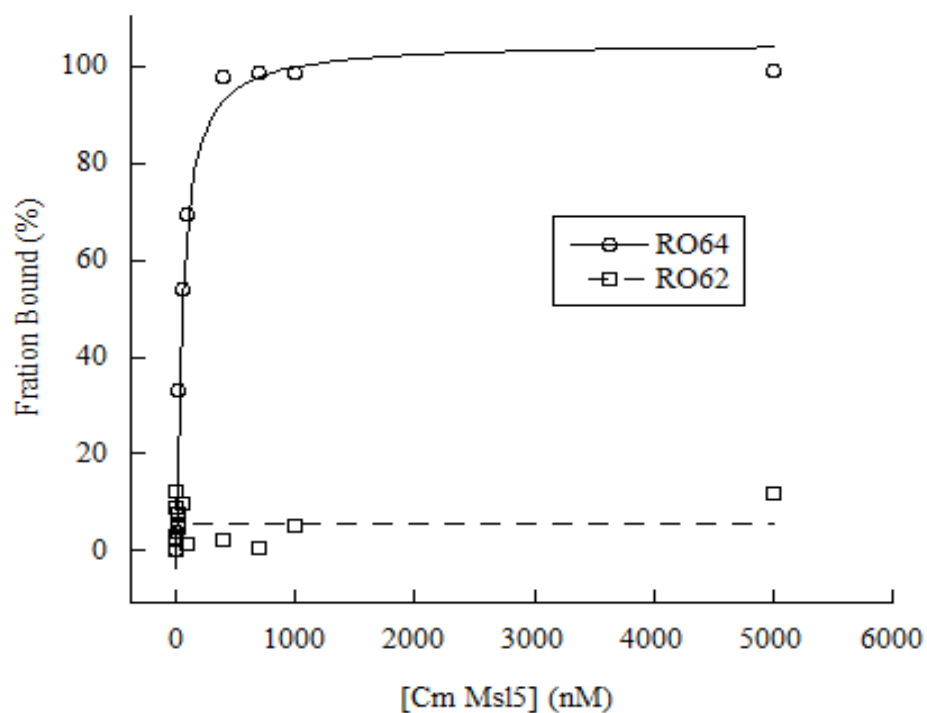


Figure 2.10. Saturation curve from EMSA binding assay of Msl5 titrated against two RNA targets (25 nM). Each curve represents the binding of *Cm*Msl5 to the *Cm* BPS (ro64, solid), U6 3' end (ro62, slashed). Curve fitting for both data sets was done using Equation 2.

Table 2.19. Binding affinities of Msl5 for RNA oligonucleotides measured by EMSA.

RNA target	K_D (nM)	Hill Coefficient	R^2 (goodness of fit)	Fit to Equation
ro64 (<i>Cm</i> BPS)	64 ± 4	1.4 ± 0.15	0.99792	2
ro64 (<i>Cm</i> BPS)	66 ± 8	1	0.99494	3
ro62 (ssRNA)	> 5000	n/a	n/a	2
ro62 (ssRNA)	> 5000	n/a	n/a	3

The RNA oligonucleotide ro64 was used in both FP and EMSA binding experiments to confirm that the results obtained from either experiment was similar. Msl5 bound ro64 with similar affinity using either FP or EMSA, and these values (K_D of 40-70 nM) were in-line with published work (Tables 2.16-2.19). Gel shift assays were next performed using 3 different variants of nascent 072C *Cm* RNA: wild type (WT) pre-mRNA, mutant (MUT) pre-mRNA, and mRNA (Fig. 2.11-2.13; section 2.3.1). The

purpose of the experiment was to determine if Msl5 was able to discriminate against mutant pre-mRNA or mRNA transcripts. Given that *C. merolae* seems to lack many of the splicing factors necessary for improved fidelity, it would be advantageous for Msl5 to avoid binding anything in the nucleus that is not a normal pre-mRNA (Stark *et al.*, 2015). The pre-mRNA transcripts included the full intron and 40 nucleotides upstream and 50 downstream of each flanking exon. The mRNA transcript had a contiguous stretch of exonic sequence (section 2.3.1). The secondary structures of each of these RNA were predicted with the mfold server and associated software, as previously described (Zuker and Jacobson, 1998; Waugh *et al.*, 2002; Zuker, 2003). The 5' splice site (5' SS), 3' splice site (3' SS) and BPS are highlighted in the pre-mRNA structures of wild type (WT) and mutant (MUT) transcripts (Fig. 2.11-2.12). A black dot marks the branchpoint adenosine (A95) for the WT structure (Fig. 2.11) as well as the A95G mutation in the mutant transcript (Fig. 2.12). The exon-exon junction of the splicing event is marked for the mRNA transcript (Fig. 2.13). I wanted to visualize the most energetically favourable secondary structures for each transcript to check if the 5' SS and BPS mutations dramatically altered the structure of the pre-mRNA. If the structures were similar, and the BPS was single stranded/accessible for both WT and MUT, I could be reasonably sure that the mutations would only affect the Msl5-BPS interaction through an altered BPS.

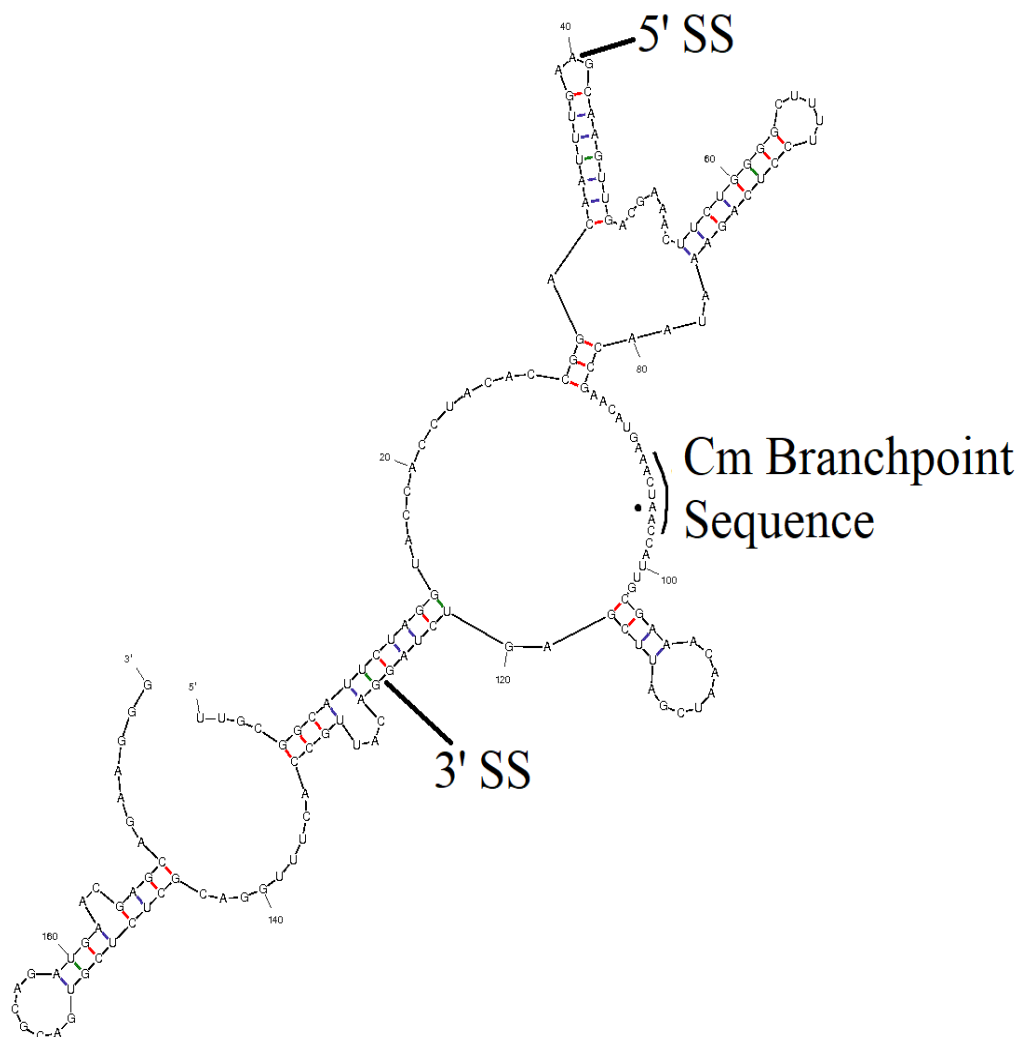


Figure 2.11. Simulated RNA folding and predicted secondary structure of WT pre-mRNA. The predicted secondary structure of *C. merolae* 072C WT pre-mRNA was generated with the mfold server. Predicted free energy: Delta G = -115 kJ/mol (initially -157 kJ/mol).

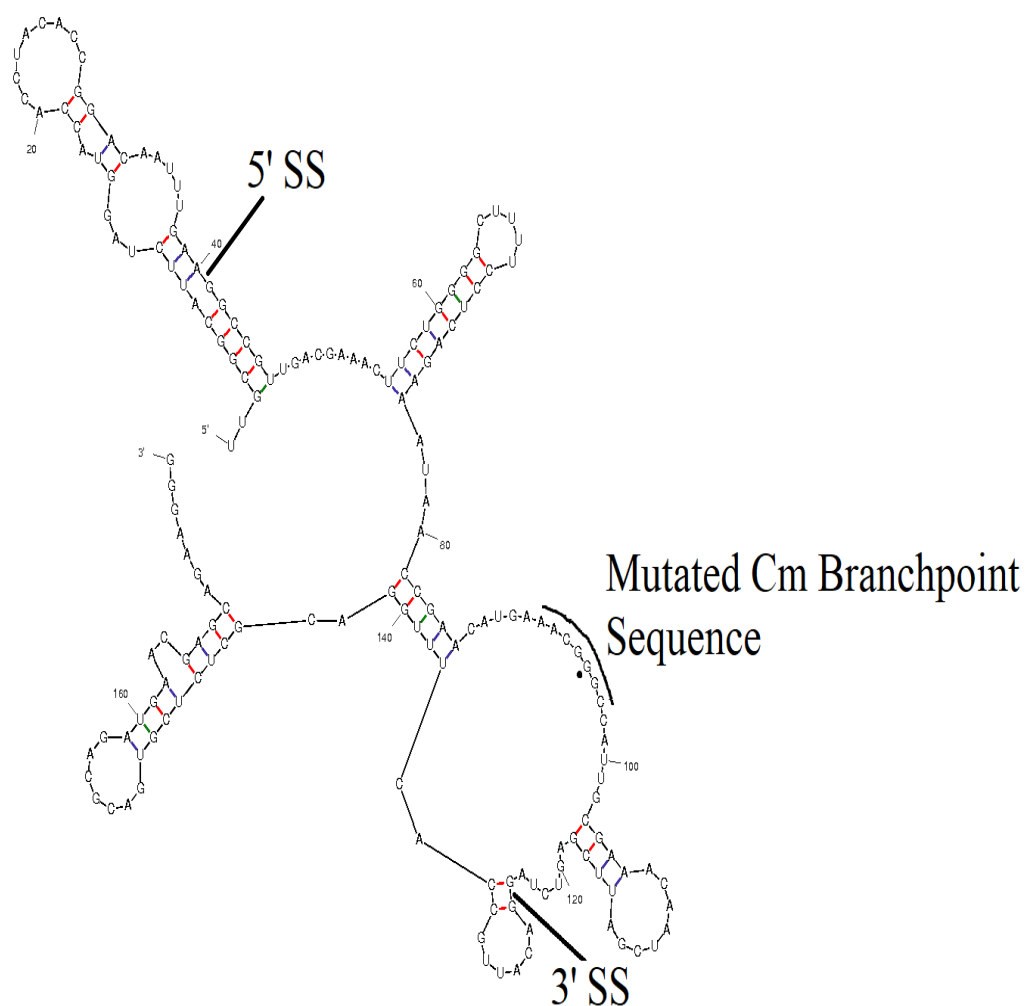


Figure 2.12. Simulated RNA folding and predicted secondary structure of mutant pre-mRNA. The predicted secondary structure of *C. merolae* 072C mutant pre-mRNA was generated with the mfold server. Predicted free energy: Delta G = -127 kJ/mol (initially -144 kJ/mol).

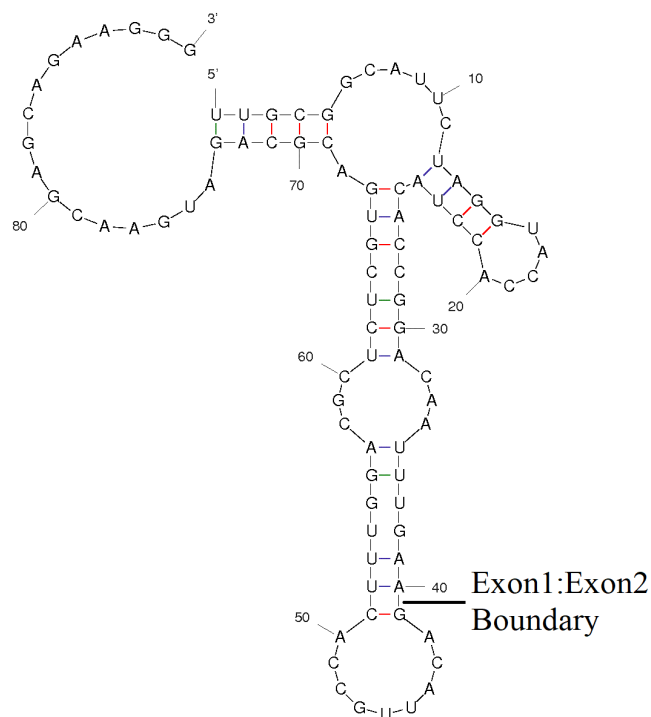


Figure 2.13. Simulated RNA folding and predicted secondary structure of WT mRNA. The predicted secondary structure of *C. merolae* 072C WT mRNA was generated with the mfold server. The predicted secondary structure was generated with the mfold server. Predicted free energy: Delta G = -57 kJ/mol (initially -64 kJ/mol).

While these models may be wildly inaccurate, they illustrate the effect of sequence on RNA structure. The 5' SS and BPS were mutated in the MUT pre-mRNA, and could easily cause structural changes that make the protein more difficult to recognize for splicing factors beyond the original sequence mutation. The lowest energy models for WT and MUT 072C transcripts seemed reasonably similar, and both had single stranded BPSs (Fig. 2.11-2.12). When splicing is complete and the mRNA product is released by the spliceosome, the transcript would have a very different structure compared to when it contained an intron, as was seen by the lowest energy model (Fig. 2.13; Warf and Berglund, 2010). Binding of Msl5 to pre-mRNA (WT and mutant), as well as mRNA was measured through EMSA in triplicate. RNA generated from IVT and

run on a native acrylamide gel often generates multiple bands due to the adoption of different secondary structure (Derrick and Horowitz, 1993). Three distinct bands are seen in the first lane of the WT pre-mRNA EMSA, and the top two have likely adopted different secondary structures while the third is degraded RNA, evidenced by band shifts (Fig. 2.14). The first noticeable shift occurs by 100 nM Msl5 (lane 8), and further shifts occur at each subsequent step (lanes 9-12), leaving a question of whether the multiple shifts were the result of different RNA conformations or cooperative binding. The answer was revealed when the data was analyzed and later fit to the Hill Equation. Gel shifts of the mutant transcript and mRNA appeared to be progressively worse (Fig. 2.15-2.16).

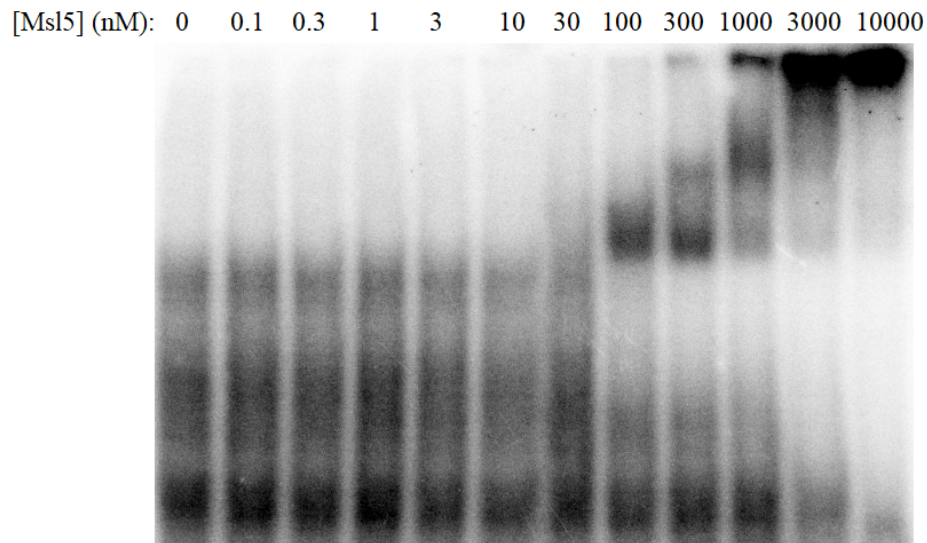


Figure 2.14. EMSA measurement of Msl5 binding the WT 072C pre-mRNA transcript. *C. merolae* WT pre-mRNA was made from IVT then [32 P]-labeled. 25 nM (24000-28000 cpm) of RNA was present in each reaction.

[Msl5] (nM): 0 0.1 0.3 1 3 10 30 100 300 1000 3000 10000

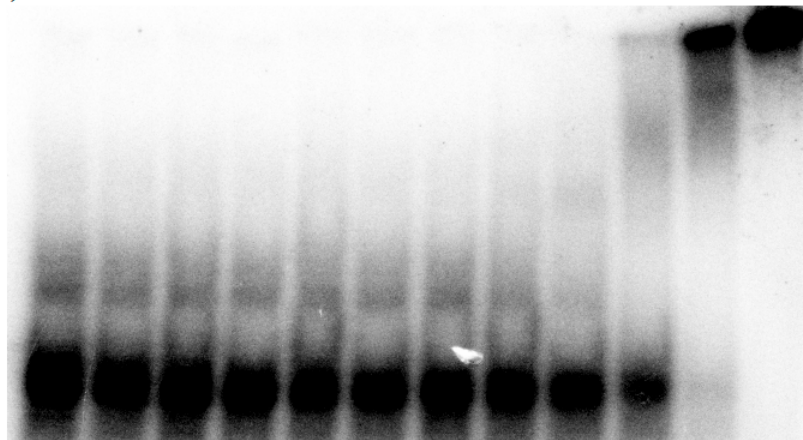


Figure 2.15. EMSA measurements of Msl5 binding the mutant 072C pre-mRNA transcript. *C. merolae* mutant pre-mRNA was made from IVT then [32 P]-labeled. 25 nM (24000-28000 cpm) of RNA was present in each reaction.

[Msl5] (nM): 0 0.1 0.3 1 3 10 30 100 300 1000 3000 10000

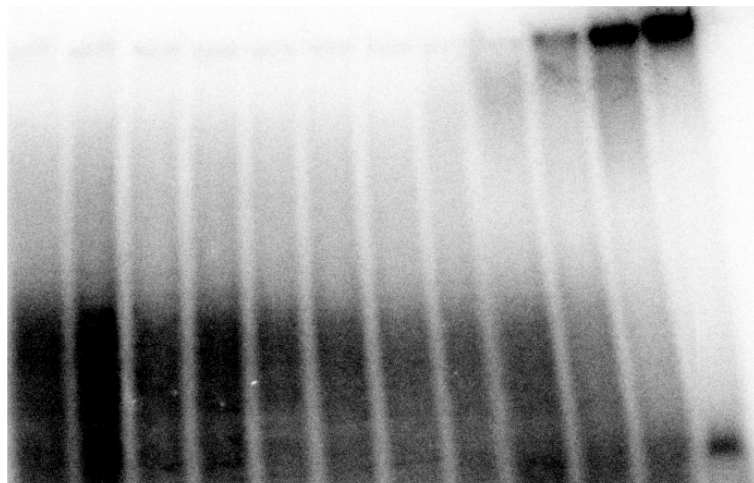


Figure 2.16. EMSA measurements of Msl5 binding the 072C mRNA transcript. The *C. merolae* mRNA was made from IVT then [32 P]-labeled. 25 nM (24000-28000 cpm) of RNA was present in each reaction.

Each of these gels, and their replicates, were transformed into respective fractions of bound RNA and plotted against Msl5 concentration (Fig. 2.17). As the RNAs were all relatively large and presented the KH-QUA2 domain the opportunity for a non-canonical (i.e. cooperative) interaction, the Hill Coefficients resulting from these EMSAs were of great interest.

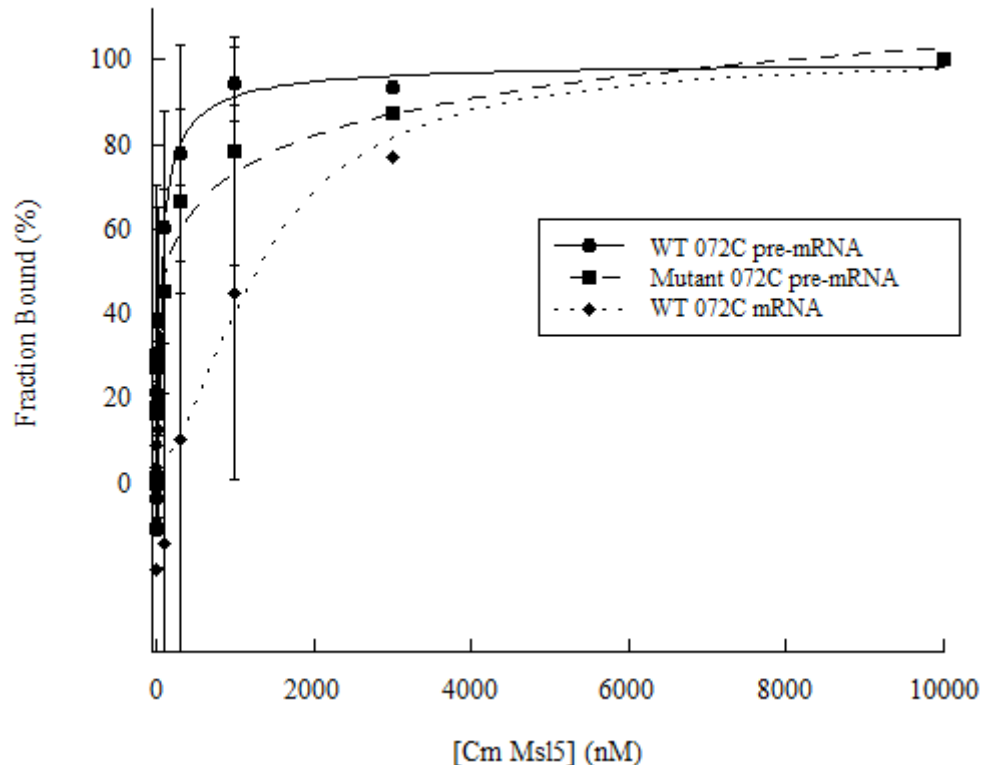


Figure 2.17. Saturation curves of *CmMsl5* against various 072C RNA transcripts: WT pre-mRNA (black); mutant pre-mRNA (grey); mRNA (dotted grey). Each point represents the average of 3 or more intensities from replicate EMSAs. Curve fitting for all 3 data sets was done using Equation 2.

The curves were fitted to the Hill Equation to generate binding affinities and respective Hill Coefficients. From this curve fit, the Msl5-072C pre-mRNA interaction had a Hill Coefficient of 0.8 ± 0.12 ($R^2 = 0.99559$), which suggested that there was no cooperativity at play (Table 2.20). Strikingly, the binding affinity of Msl5 for the 175 nt full-length intron (IVT 072C WT pre-mRNA) was identical to that of the 11 nt BPS oligo (ro64). The tight binding of Msl5 to both short and long BPS-containing RNA suggests that there were no additional contacts made between Msl5 and the full-length pre-mRNA that either promoted or inhibited the interaction.

Table 2.20. Binding affinities of Msl5 for various *Cm* 072C Nascent transcripts.

RNA target	K _D (nM)	Hill Coefficient	R ² (goodness of fit)	Fit to Equation
WT pre-mRNA	46 ± 9	0.8 ± 0.15	0.99559	2
WT pre-mRNA	43 ± 8	1	0.99390	3
MUT pre-mRNA	8100 ± 4700	0.3 ± 0.13	0.97474	2
MUT pre-mRNA	210 ± 92	1	0.96764	3
mRNA	1300 ± 670	1.7 ± 1.23	0.92841	2
mRNA	2100 ± 1600	1	0.92154	3

2.6.4. Structural Measurements of Msl5 with CD Spectroscopy

Circular dichroism (CD) was measured to assess the secondary structure content of Msl5 free and bound to BPS RNA (Fig. 2.18). CD spectra data points were taken for: buffer + RNA (ro64); buffer + protein; buffer + protein + RNA. The buffer + RNA (ro64) data was subtracted from the data from buffer + Msl5 + RNA (ro64) in order to exclude the CD effect of RNA. Using the software and reference database from the K2D3 server, the alpha helical and beta sheet content was determined (Louis-Jeune *et al.*, 2012). CD was also employed to determine the thermal stability of Msl5 (Fig. 2.19).

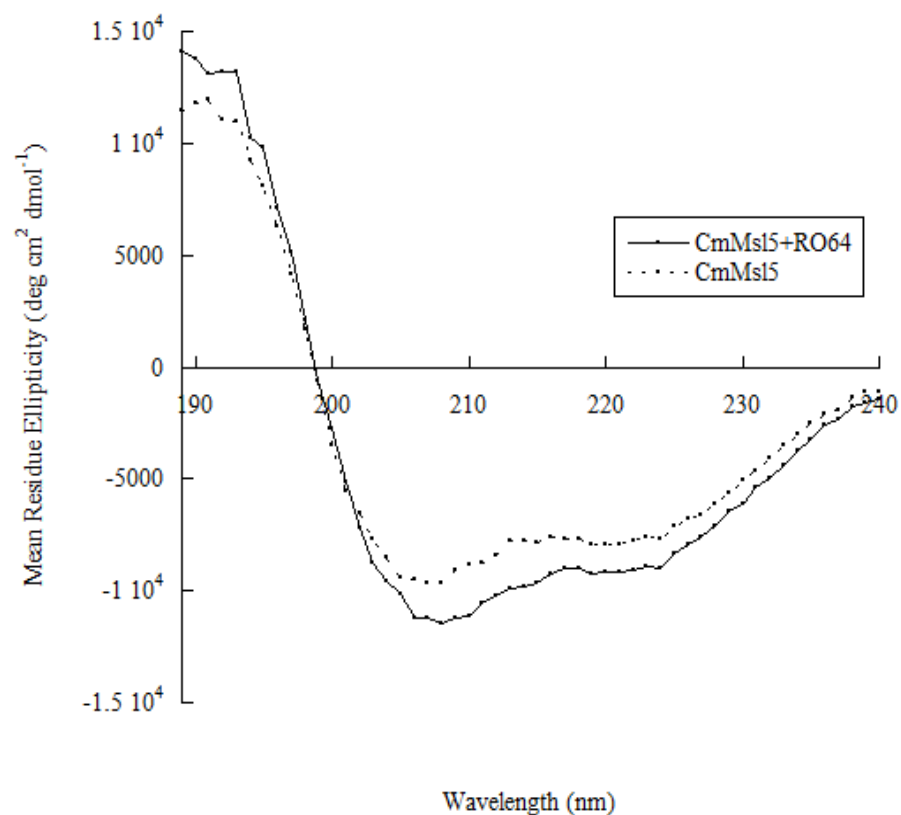


Figure 2.18. CD spectra of Msl5 and the *Cm* consensus BPS. Overlay of circular dichroism spectra of Msl5 (dotted) and Msl5 bound to BPS RNA oligo in a 1:3 molar ratio (line). All spectra data was generated from 8 accumulations using a Jasco J-815; samples were scanned in 15mM NaH_2PO_4 pH 7.5, 40mM NaF buffer. The background reference (buffer + ro64 RNA) was subtracted from the buffer + protein +RNA prior to graphing and secondary structure estimation.

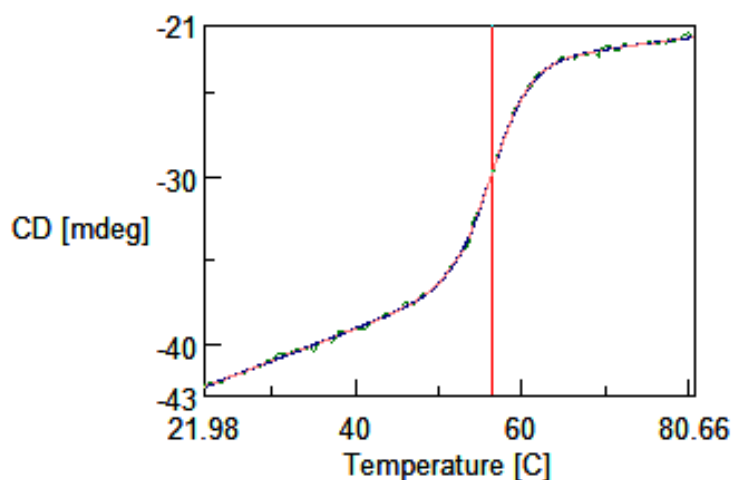


Figure 2.19. Thermal stability analysis of Msl5. Data was generated using the Peltier temperature controller attached to the Jasco J-815. The *C. merolae* Msl5 sample was scanned in 15mM NaH₂PO₄ pH 7.5, 40mM NaF buffer.

Wave-scans performed on both free and ro64-bound Msl5 were done in duplicate, from 260 nm to 189 nm (Fig. 2.18). The 189 nm cut-off was due to high-tension (HT) reaching 600V at 190 nm. Thermal stability was similarly measured in duplicate but was regrettably not measured for an Msl5-RNA complex (Fig. 2.19). The wave-scan spectra for free versus bound Msl5 appeared to be different, with free Msl5 having smaller minima and maxima (Fig. 2.18). When the data was extracted and uploaded onto the K2D3 server, secondary structure predictions were very similar to each other (Table 2.21). The measured secondary structure content disagrees with the prediction done using the PSIPRED server for α -helix and β -strand content, but not intrinsically disordered regions (Jones, 1999; Buchan *et al.*, 2013). The thermal transition of Msl5 was determined to be $56.5 \pm 1^\circ\text{C}$, which was very close to the 55°C T_m reported for yeast BBP and the limit for survivability of *C. merolae* (Matsuzaki *et al.*, 2004; Garrey *et al.*, 2006).

Table 2.21. Secondary structure of Msl5 by Circular Dichroism and predictive software.

Species	α -helix (%)	β -strand (%)	Intrinsically Disordered (%)	Method
Msl5 (Free)	77.62 ± 5	1.80 ± 2	20	CD Spec
Msl5 (ro64-bound)	77.64 ± 5	1.76 ± 2	20	CD Spec
Msl5 (Free)	35	5	19	PSIPRED

2.6.5. Recombinant Expression and Purification of Msl5/Mud2 Dimer

Co-expressed His-*Cm*Mud2 and Msl5 was purified by incubation with Ni-NTA resin, washed, and then eluted by centrifugation. The expression and elution fractions were subsequently analyzed on a 10% polyacrylamide SDS-PAGE at 22 mA for 50 min (Fig. 2.20). The His-Mud2 and Msl5 constructs are 145.25 kDa and 36.23 kDa; His-Mud2 migrated slightly below the 200 kDa marker while untagged Msl5 was thought to be a band near 45 kDa. Further analysis was needed to confirm the presence of Msl5. Several contaminant bands were visible near the 66 kDa, 31 kDa, and 21 kDa markers.

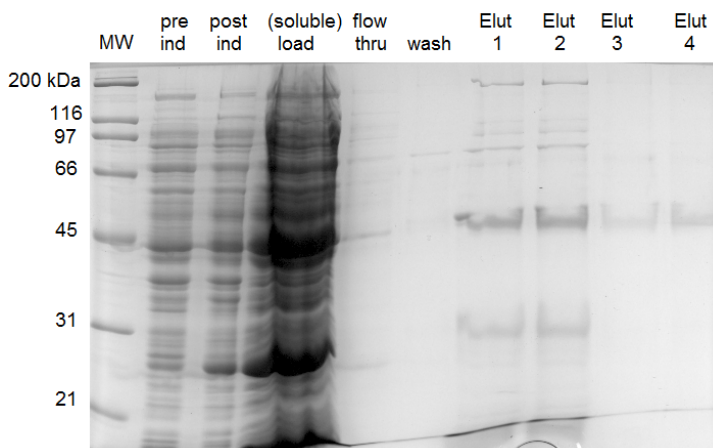
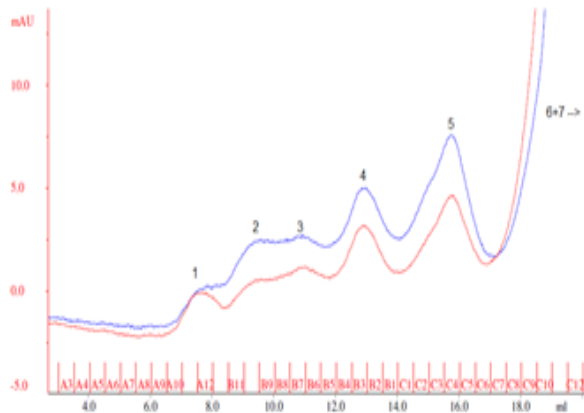


Figure 2.20. Co-expression and Ni-NTA batch purification of His-Mud2 and Msl5 analyzed on a 12% SDS-PAGE. Elut 1 and 2 refer to the first and second fractions collected, while 3 and 4 correspond to fractions 14 and 15.

Unlike Msl5 alone, the Mud2/Msl5 dimer is a large (180 kDa) complex that seemed an excellent candidate for size exclusion purification over the Superdex200 10/300 GL column (GE). Both La Verde (2017) and the chromatogram of the Gel

Filtration Standards suggested that 180 kDa proteins are separated from 150 kDa proteins without peak overlap/contamination (La Verde, 2017). Thus this step should have improved purity and removed free Mud2 and Msl5 proteins. Affinity-purified sample was injected 1 mL at a time and ran through the Superdex200/ÄKTA Purifier10 system (Fig. 2.21 A). The first peak was suspected to be within the void volume of the column: fractions that contain aggregates of misfolded or degraded protein products (Fig. 2.21 A; Carpenter *et al.* 2010). From SDS-PAGE analysis, only Peak 2 was kept in each run because it was the only peak that contained bands near the 200 kDa and 45 kDa markers (Fig. 2.21 A, B). SDS-PAGE was perhaps the best available technique because band intensities could be equated with stoichiometry, which was expected to be 1:1 (Berglund *et al.* 1998b; Selenko *et al.*, 2003). Band quantification showed that the bands near 200 and 45 kDa from the elution (lanes 7-10) had a 1:2 stoichiometry (Fig. 2.21). The stoichiometry of the proteins suggested there's a population of free Mud2 as well as dimer. It would be unlikely that *E. coli* would synthesize equal amounts of the proteins; having an affinity tag on Msl5 and doing two separate affinity purifications would have limited the amount of non-dimerized protein (Gräslund *et al.*, 2008).

A



B

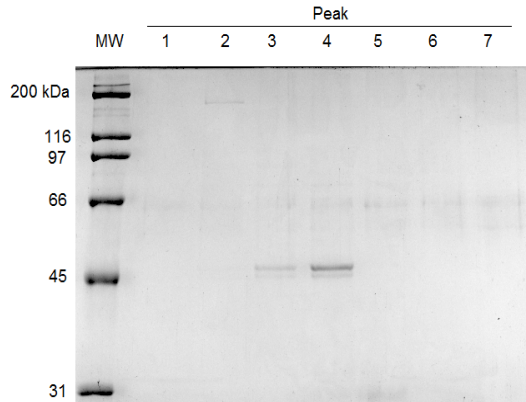


Figure 2.21. Size exclusion chromatography purification of the Mud2/Msl5 heterodimer. Red is 260nm and blue is 280 nm. A) Representative chromatogram of the size exclusion purification and B) corresponding SDS-PAGE analysis of the fractions. Peaks 1-5 are labeled on the gel and the chromatogram, corresponding to the various elution fractions of the size exclusion chromatogram. Peaks 6 and 7 were not shown on the chromatogram and corresponded to later (smaller) fractions. Mud2 and Msl5 were recombinant *C. merolae* proteins.

While the presence of His-Mud2 was evidenced by the band slightly below 200 kDa in the lysate of induced cells and subsequent purifications, I wanted to verify the presence of Msl5. The band near 45 kDa was suspected to be Msl5 because His-Msl5 migrated similarly (Fig. 2.1), but DnaJ is a 41 kDa *E. coli* chaperone known to co-purify during immobilized metal affinity chromatography (Gräslund *et al.*, 2008). With a polyclonal rat α -Msl5 (gift of Dr. Robert Burke's lab, UVIC) and monoclonal goat α rat HRP IgG (Santa Cruz Biotechnology sc-2065), I performed a Western Blot to verify that Msl5 was present in the His-Mud2 pull-down. Previously purified Msl5 (lane 1) was compared with the Mud2 pull-down (lane 3), revealing that the brightest bands of each sample appear near the 37 kDa marker (Fig. 2.22). This confirmed that Msl5 is present in the elution and bound to Mud2. When comparing lanes 1 and 3, it appeared that Msl5 from the dimer ran slightly higher (Fig. 2.22). Given that both samples were untagged

and very similar in size, three factors likely contributed to this: i) the pure Msl5 sample was older and may have been exposed to proteases; ii) the gel ran on a slant; iii) different quantities were loaded, obscuring the analysis.

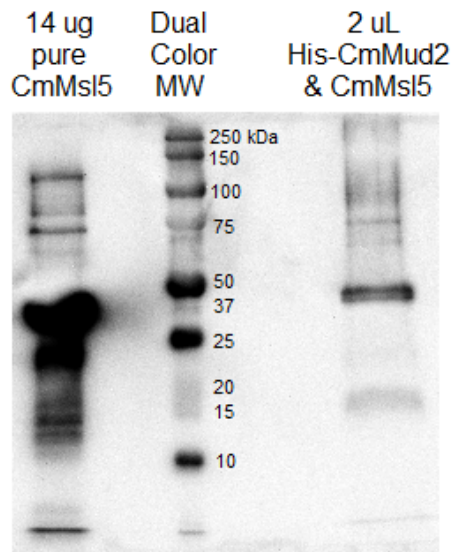


Figure 2.22. Western Blot of His-*CmMsl5* and His-*CmMud2* affinity pull-downs. The primary antibody was polyclonal rat α *CmMsl5* and the secondary was goat α rat.

With reasonable evidence that there was a population of Mud2 and Msl5 heterodimer, the concentration needed to be determined. Assays that measure total protein, such as spectrophotometry (260/280 nm wavelengths or the Bicinchoninic acid assay), were not appropriate since there were 2 different proteins. Standardization was done through SDS-PAGE by titrating known amounts of previously purified Msl5 against 2 uL of the purified dimer (Fig. 2.23). Band quantification revealed that Msl5 of the dimer, and Mud2 by association, was approximately 2000 uM (Fig. 2.23). Interestingly, the stoichiometry of the Mud2:Msl5 was 1:1.3 (lane 10, Fig. 2.23).

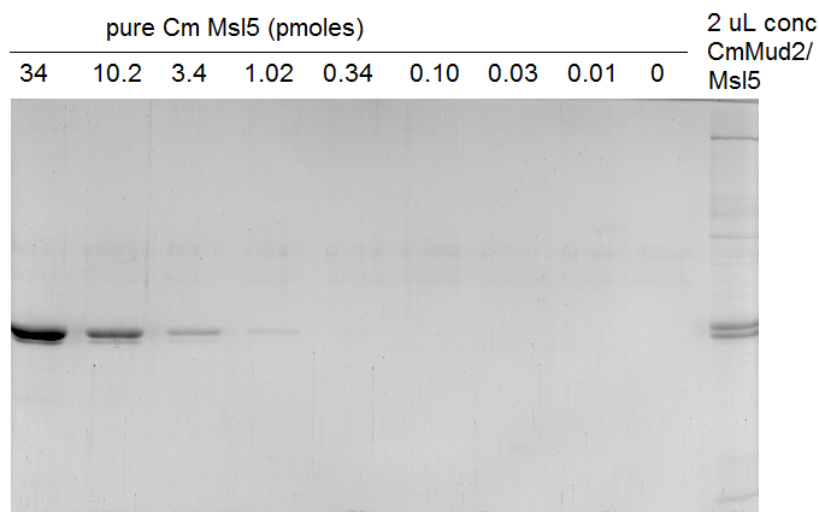


Figure 2.23. The standardization of *Cm*Mud2/Msl5 by comparison against known amounts of *Cm*Msl5. Quantification of band intensities showed that the concentration is approximately 2000 nM (or 2.5 pmol per uL).

2.6.6. Fluorescence Polarization and Bioinformatics of Msl5/Mud2 Dimer

The ideal assay to measure the effect of *Cm*Mud2 on the Msl5-BPS interaction is the gel shift. EMSA allows for the use of large RNA sequences, whereas FP is limited to RNA of approximately 40 nucleotides in size due to tumbling and polarization of the fluorophore (Prystay *et al.*, 2001). The *Cm* Mud2/Msl5 heterodimer was titrated against the 11-nucleotide long BPS (ro64) in FP due the lack of available IVT RNA and a series of troubleshooting problems that occurred during that time. Given those constraints, this experiment intended to address the question of whether Mud2 affects the RNA binding ability of Msl5. The anisotropy values were transformed and plotted against dimer concentration to generate saturation curves (Fig. 2.24). During curve fitting, The Hill Coefficient was set to 1 (A), 2 (B), 3 (C) and variable/float (D). Unlike the Msl5-BPS interaction, forcing the Hill Coefficient to 1 severely distorted the shape of the curve (Fig. 2.8 and 2.24 A). Indeed, from comparison of the various Hill Coefficients and corresponding R^2 values, it's clear the value is 2-3 (Table 2.22). This indicates the

interaction between Mud2, Msl5, and the RNA is cooperative and similar to the cooperative interaction in yeast and humans (Berglund *et al.*, 1998b; Zhang *et al.*, 2012). Unfortunately the amount and concentration of Mud2-Msl5 dimer after purification was only sufficient for a single triplicate experiment to a maximum of 250 nM (Fig. 2.24). For that reason, this experiment provided an insight into cooperativity of the Mud2-Msl5-RNA interaction, but not binding affinity (Table 2.22).

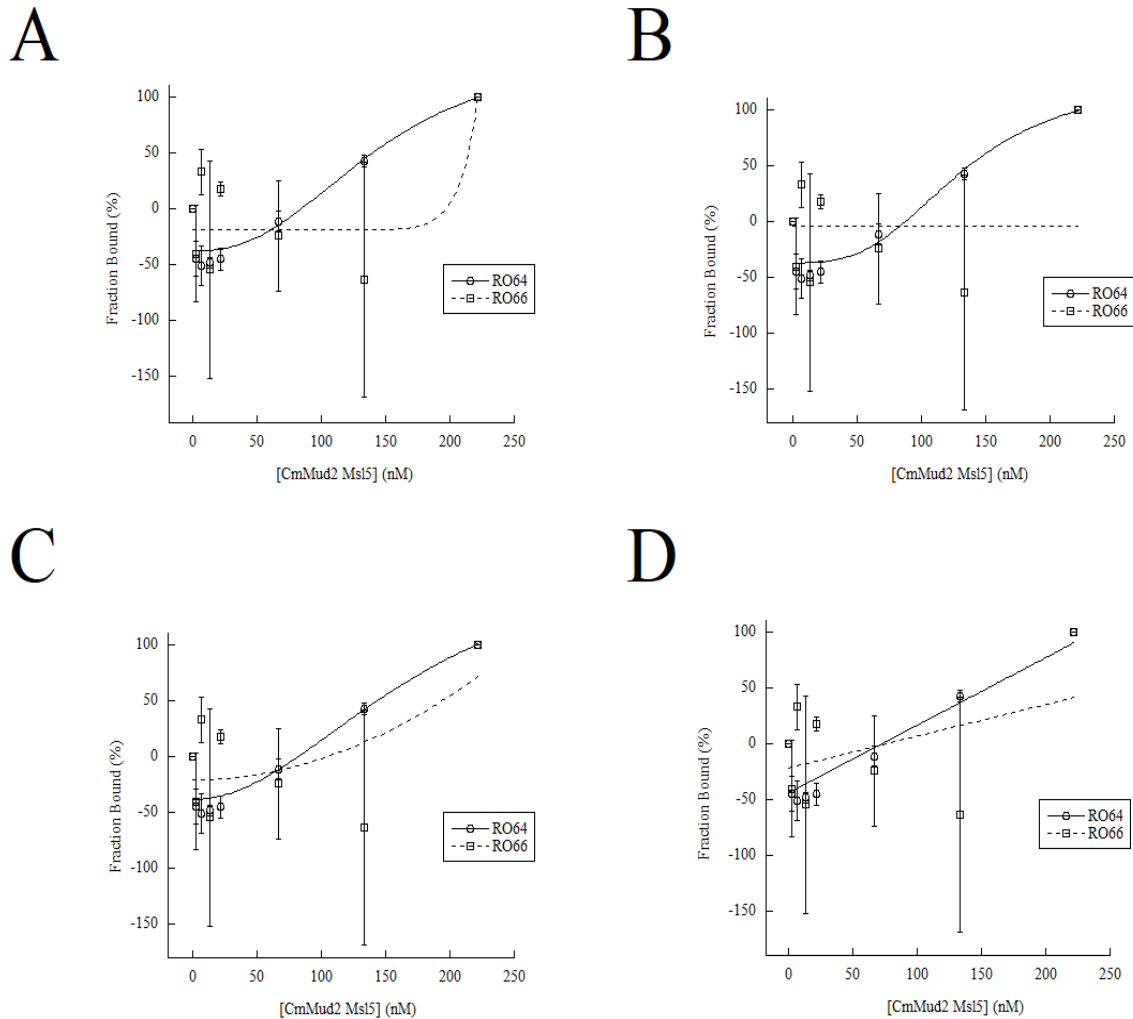


Figure 2.24. Fluorescence polarization of the *Cm*Mud2/Msl5 heterodimer titrated against the *Cm* BPS RNA. The saturation curves represent the binding of *Cm*Msl5 to the *Cm* BPS (solid black) and U4 3' end (dotted). (A-C) Curve was fitted with the Hill Coefficient set from 1-3, respectively, using Equation 3. (D) Curve was fitted with the Hill Coefficient floated using Equation 2.

Table 2.22. Dissociation constants and Hill Coefficients for FP binding assays of the Msl5/Mud2 dimer against the *Cm* BPS.

RNA Target	Apparent K_D (nM)	Hill Coefficient	R^2 (goodness of fit)	Fit to Equation
ro64	$1 \times 10^{13} \pm 3 \times 10^{-10}$	1	0.92532	3
ro66	$3 \times 10^{15} \pm 2 \times 10^{13}$	1	0.45138	3
ro64	180 ± 90	2	0.95218	3
ro66	1900 ± 1000	2	0.60416	3
ro64	132 ± 35	3	0.95240	3
ro66	850 ± 1000	3	0.68398	3
ro64	147 ± 122	2.5 ± 2.5	0.95312	2
ro66	$315 \pm 1 \times 10^5$	16 ± 170	0.77679	2

While the *Cm* Mud2-Msl5-RNA interaction has been shown to be likely cooperative, many interesting questions are not directly addressed with a binding assay using 11 nucleotides of the BPS. These questions include how Mud2 affects binding of Msl5 to a full-length intron, or what sequence Mud2 recognizes. This experiment wasn't expected to have allowed for a Mud2-RNA interaction, because the short RNA would likely be sequestered in a groove of Msl5 and not extend enough to contact Mud2 (Liu *et al.*, 2001; Zhang *et al.*, 2012; Jacewitz *et al.*, 2015). Furthermore, footprinting studies of SF1/BBP and U2AF65/Mud2 proteins have shown that SF1/BBP protects the 7 nucleotides of the BPS and two nucleotides on each end (11 total), while U2AF65/Mud2 begins contact with the Py-tract 3 nucleotides downstream of the BPS (Berglund *et al.* 1998a). RNA oligonucleotide ro64 represents the 11 nucleotides of the BPS, providing no ligand for Mud2. Since the KH-QUA2 domain and BPS affinity of Msl5 seem highly conserved, it seems unlikely Mud2 would contact the 11 nucleotide-protected region of the BPS (Berglund *et al.*, 1997). Structurally, the U2AF65-SF1 human interaction involves rearrangements of both proteins (Zhang *et al.*, 2012). While the concentration was quite small in this experiment, the binding affinity appears to be 5-fold lower, from about 40 nM to 180 nM (Tables 2.17 and 2.21). Perhaps the Mud2-Msl5 interaction in

Cm modifies the topology of the KH-QUA2 domain and thus the affinity of Msl5 for the BPS. This alteration would be unlikely and unique for Msl5 and Mud2, but they are quite divergent from yeast and human homologs. Despite sequence differences, branchpoint selection is likely driven by Msl5 and stabilized by Mud2, as occurs in yeast.

In order to get some insight into *Cm*Mud2 besides the binding experiment, I performed a series of BLAST searches and multiple sequence alignments. *Cm*Mud2 is 1316 amino acids in size, while human and yeast homologs are 475 and 527 respectively. Amazingly, no BLAST searches (of varying constraints) produced a single result that aligned to the first 800 residues. Searches that include only the first half of the protein will not produce a single result other than itself. Interestingly, both human *Sc*Mud2/U2AF65 and U2AF35 homologs are identified in the BLAST search, but occur in overlapping regions (residues 1200-1316). Multiple sequence alignments with human and budding yeast (*S. cerevisiae*) homologs revealed one particular region to be conserved: RRM3, or the U2AF Homology Motif (Fig. 2.25). The UHM interacts with the ULM of SF1/BBP (Fig. 2.5; Henscheid *et al.*, 2008; Wang *et al.*, 2013). *Sc*Mud2/U2AF65 have been shown to contain 3 RRMs, of which RRM1 and 2 are implicated in binding the Py-tract/Poly-U region, while RRM3/UHM is involved in binding SF1/BBP (Selenko *et al.*, 2003; Zhang *et al.*, 2012).

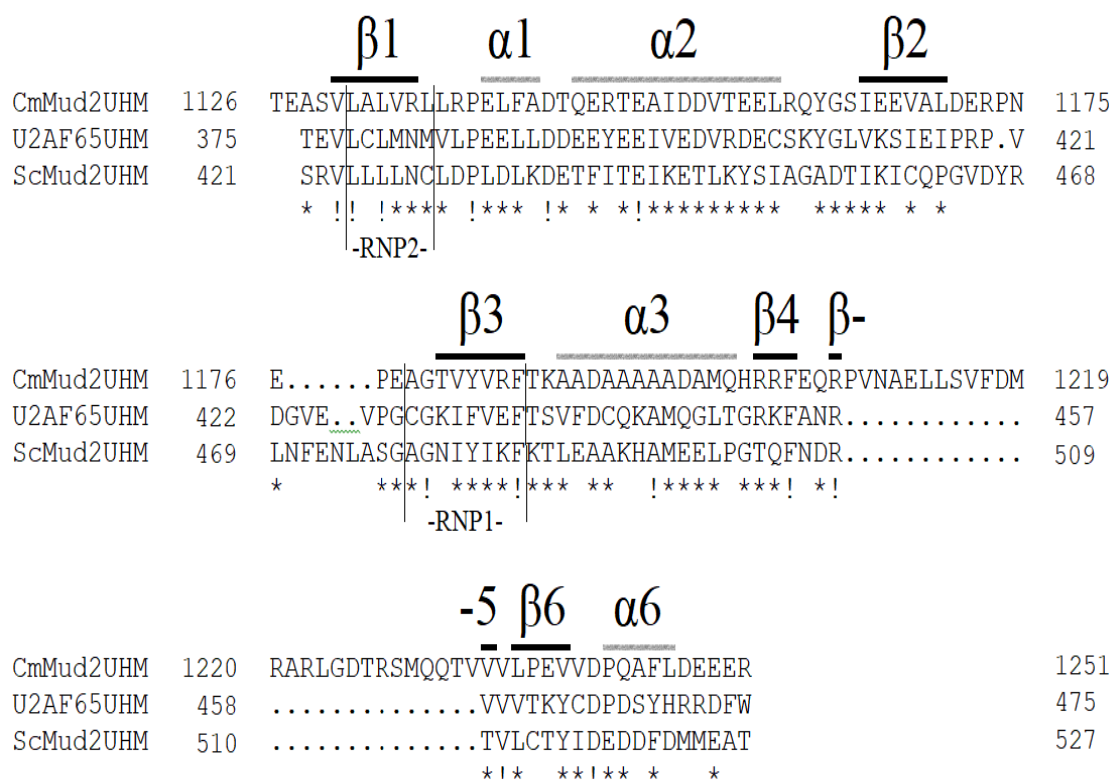


Figure 2.25. Alignment and predicted secondary structure of the U2 Homology Motif (UHM) of Mud2 against human (SF1) and yeast (BBP) homologs. Similarity was shown with “*” and identity with “!”. The UniProt Identifiers for *C. merolae* Mud2, U2AF65, and ScMud2 were CMS438C, P26368, and P36084 respectively.

The UHM of Mud2 is 45.5% similar and 26.7% identical to human, and 31.7% similar and 20.2% identical to yeast (Fig. 2.25). Human and yeast are 40.6% similar and 25.7% identical. Interestingly, the UHMs of Msl5 and U2AF65 are the most similar based on sequence. Furthermore, each of the RNP motifs (RNP2 and RNP1) agrees with the consensus sequences (Fig. 2.25; Wang *et al.*, 2013). This alignment, together with the results showing both Msl5 and Mud2 were able to co-purify, make a strong argument for an Msl5-Mud2 interaction in *C. merolae*.

Chapter 3

Functional and Structural Characterization of Snu13

This chapter follows the structural and functional investigation into Snu13, a double-stranded RNA binding protein associated with the assembly of the U4/U6 di-snRNP. The structure of the protein was determined by X-ray crystallography; to verify the structure was that of a functional Snu13, fluorescence polarization was used to measure the binding affinity of the protein to its target, the U4 5' stem loop.

3.1. Methods - Expression and Purification of *C. merolae* Snu13 Protein

3.1.1. Recombinant Expression of Snu13 from Rosetta pLysS (DE3) Cells

Dr. Andrew M. MacMillan from the University of Alberta generously provided the plasmid vector pMCSG7, which features an N-terminal 6xHis-tag and a T7 promoter. Dr. Martha Stark cloned Snu13 from the *C. merolae* genome into the pQLinkH vector (Addgene plasmid 13667) using primers 1 and 2 (see sections 2.1 and 4.4). The gene was then cloned into pMCSG7 from pQLinkH using primers 3 and 4 (Table 3.1).

Table 3.1. Primers used to clone Snu13 from the *Cm* genome into pMCSG7.

Primer	Plasmid	Sequence
1 (fwd)	pQLinkH	ATG CTT GGA TCC ATG GAA CCT TTG AGC TCC AC
2 (rev)	pQLinkH	GGT ACG CTC TCG GAT CGG CCT GAT TCG GCG
3 (fwd)	pMCSG7	TAC TTC CAA TCC AAT GCA GAA CCT TTG AGC TCC ACT GAA GCG
4 (rev)	pMCSG7	TTA TCC ACT TCC AAT G TTA TCA GAG CAG AAG CTG TTC GAT CTT TGT TCG

103ng of pMCSG7-Snu13 was transformed into 50 uL of Rosetta pLysS (DE3) competent cells by means of heat shock. 100 uL of LB was added and the mixture was plated on LB agar supplemented with ampicillin and chloramphenicol, each at a 1:1000 dilution. The plate was incubated at 37°C for 18 hours. 2 colonies were selected to inoculate 90mL of LB/AMP/Chlor broth. The broth was incubated in a shaker set at 250

RPM at 37 °C for 18 hours. 10 mL of the broth was added to 1 L LB/AMP/Chlor broth for induction. The cells were induced at an OD₆₀₀ of 0.65 with 0.1 mM IPTG. Induction was carried out for 3.5 hours before centrifugation for 12 min at 3500 RPM and re-suspension in 10 mL Snu13 Wash Buffer (Table 3.2). The ~10 mL cell suspension was centrifuged for 15 min at 3200 RPM, before the supernatant was discarded and the cell pellet was flash-frozen in liquid nitrogen before storage at -80°C.

Table 3.2. Complete list of the buffers used in the purification of Snu13.

Buffer	Components
Snu13 Wash Buffer	20 mM HEPES-Na-OH pH 7.5, 300 mM NaCl, 30mM Imidazole
Benzonase Buffer	50mM Tris-HCl pH8, 1.5mM MgCl ₂ , 50mM NaCl
Benzonase Removal Buffer	20mM HEPES-NaOH pH 7.5, 5mM β-mercaptoethanol, 500mM NaCl
Snu13 Elution Buffer	20mM HEPES-Na-OH pH 7.5, 300mM NaCl, 275mM Imidazole
Snu13 Storage Buffer	20mM HEPES-Na-OH pH 7.5, 85mM NaCl, β-mercaptoethanol

3.1.2. Lysate Preparation and Affinity Chromatography

In order to purify Snu13 from bacterial lysate, batch binding affinity chromatography was performed. Each pellet of Snu13-induced Rosetta cells was treated with Protease Inhibitor Cocktail Tablets (Sigma-Aldrich), EDTA-Free, then re-suspended with 11 mL Snu13 Wash Buffer. After re-suspension, the sample was sonicated 13X using 10 second bursts and 1 minute rests. Streptomycin sulfate was added as a nucleic acid precipitant and the lysate was cleared via centrifugation at 25000 XG for 12 min. The supernatant was collected and mixed with equilibrated nickel resin (Ni-NTA Resin) in a 50 mL conical tube. The tube was rotated end- over-end for one hour at 4°C. The resin was pelleted by centrifuging the tube for 2 min at 700 Xg and 4°C. Resin was washed twice with 10mL Snu13 Wash Buffer, and again centrifuged for 2 min at 700 g

and 4°C. After the second wash was aspirated off, the resin was treated with the RNA/DNA endonuclease Benzonase (Sigma-Aldrich) to remove all nucleic acid that co-purified with the RNA-binding protein. 35 units of Benzonase in 1mL Benzonase Buffer were added to the resin, followed by agitation every 5 minutes for 1 hour (Table 3.2). Benzonase was cleared from the Snu13-bound resin by washing twice with 10mL of Benzonase Removal Buffer (Table 3.2). One additional wash was done using 10mL of Snu13 Wash Buffer. 1mL elution fractions were collected using Snu13 Elution Buffer (Table 3.2). Based on OD_{260/260} values and protein concentration (via NanoDrop Spectrophotometer), only the first 10 elution fractions were kept.

3.1.3. His-Tag Cleavage And Removal

1mg of TEV protease (Addgene pRK792) was mixed with the pooled eluate (100 mg). The protein/TEV mixture was dialyzed for 18 hours in Snu13 Storage Buffer (Table 3.2). Following cleavage/dialysis, the sample was mixed with the Ni-NTA resin, rotated end-over-end for an hour, and centrifuged for 2 min at 700 g. The flow through (cleaved protein) was collected for downstream purification/testing.

3.1.4. Size Exclusion Chromatography Purification

Snu13 was further purified via size exclusion chromatography (SEC) using a Superdex200 10/300 GL column in-line with the ÄKTA Purifier 10 system (GE Healthcare Life Sciences). For optimal resolution on this column, 5 mg of protein was loaded per run. From the chromatogram, the 260 nm and 280 nm peaks were used to determine where the protein eluted. After running post-SEC samples on a 12% SDS-PAGE gel ((25 mA, 60 min) and taking OD_{260/280} readings, SEC elution fractions were pooled and concentrated to 10mg/mL using Millipore Amicon® Ultra 10K 15 mL

Centrifugal Filters. The protein was used for crystal trials, fluorescence polarization, and small angle X-ray scattering.

3.2. Methods - Fluorescence Polarization of Purified Snu13

3.2.1. Set-up of the Fluorescence Polarization Binding Assay

Various concentrations of Snu13 and RNA oligonucleotide together and the corresponding anisotropy signal was measured. The two probes used are detailed in Table 3.3. RNA oligonucleotides ro52 and ro26 correspond to the *Cm* U4 5' stem loop (SL) and the *Sc* U6 stem loop, both of which were ordered from Integrated DNA Technologies as 5'-fluorescein-labeled probes (Table 3.3).

Table 3.3. Fluorescein-labeled RNA oligonucleotide probes used in FP Binding Assays.

Identity	Sequence
ro52: <i>Cm</i> U4 5' SL	(6-FAM)5'-UUGCCCAGAUGAGGUUCUCCGAUGGGUAA3'
ro26: <i>Sc</i> U6 SL	5'-UUCCCCUGCAUAAGGAU-(6-FAM)-3'

Reactions were set-up for Fluorescence Polarization (FP) binding assays as per Table 3.4 in 90 uL. Reactions were mixed in 1.5 mL Eppendorf microtubes and then transferred to Nunc Thermo Scientific black 384-well Microplates. The FP Reaction Buffer 2 was 25 mM HEPES-NaOH pH 7.5, 85 mM NaCl, and 5 mM β -mercaptoethanol. Anisotropy was measured using a BioTek© Synergy 2 Multi-Mode reader. Concentration of RNA probe was 5 nM for the measurement of binding affinity (K_D) and ranged from 0 to 12000 nM for the measurement of active protein fraction.

Table 3.4. FP reagents used for the binding experiments of Msl5 with BP RNA probe.

FP Reaction Reagents	Volume	Final Concentration
Snu13	40 uL	0 to 100000 nM
FP Reaction Buffer 2	45 uL	1 X
Fluorescein-labelled probe (ro52, ro26)	5 uL	0 to 12000 nM

3.2.2. Analysis of FP data

Fluorescence polarization measurements (parallel and perpendicular) were converted into anisotropy values with Gen5 Data Analysis Software (Biotek Instruments) and subsequently into percent of RNA probe bound by protein. RNA % bound values were plotted with their respective protein concentration using Kaleidagraph Software against a Hill Equation (see below) to determine binding affinity (K_D). As described in section 2.2.3, Equation 1 was used to convert anisotropy to percent of RNA probe bound, Equation 2 was used as the Hill Equation for a floating Coefficient, and Equation 3 was used as the Hill Equation for a fixed Coefficient.

3.3. Methods - Bioinformatics and Structural Modeling

3.3.1. Multiple sequence alignment and modeling

Protein sequences were identified from NCBI BLAST searches, taken from the UniProt database, and aligned with STRAP software (Gille *et al.*, 2014). Secondary structure predictions for Snu13 and its homolog were done using the PredictProtein server (Yachdav *et al.*, 2014). The PDB files for *S. cerevisiae* Snu13p and *H. sapiens* 15.5K were taken from the RCSB Protein Data Bank (Table 3.5).

Table 3.5 UniProt and RCSB Accession numbers for Snu13 homologs.

Organism	UniProt Accession Number	RCSB PDB File
<i>C. merolae</i>	CMP335C	
<i>G. sulphuraria</i>	M2XBTS	
<i>H. sapiens</i>	P55769	2JNB, 3SIU, 2OZB
<i>S. cerevisiae</i>	P39990	2ALE

3.4. Methods - X-ray Crystallography of Snu13

3.4.1. Hanging-drop/vapor diffusion crystallography of purified Snu13

Purified protein was concentrated to 10mg/mL in Snu13 Storage Buffer and sent 1 mL to Dr. Andrew M. MacMillan's lab (Table 3.2). Several crystal trial attempts were

made in-house using Crystal Screens 1 and 2 from Hampton Research, but were unsuccessful. Dr. Erin L. Garside (then a PhD student of the MacMillan Lab) successfully crystallized Snu13. The drop/reservoir buffer used was 31% PEG 3350 with 100 mM sodium acetate pH 4.4. The drops were 2 uL in volume; 1uL protein to 1uL well solution. To each well, 500 uL of reservoir buffer was added. Vaseline was used as a sealant. Crystallography was done at room temperature (approx. 20°C) in the dark. Drops were analyzed via light microscopy daily.

3.4.2. X-Ray diffraction of Snu13 crystals and data analysis

The crystals were collected with loops, coated in cryo-protectant (20% v/v glycerol), and flash-frozen in liquid nitrogen. The samples were mailed-in to the Stanford Synchrotron Radiation Light Source and diffraction data was collected remotely. Dr. Garside initially used the HKL 2000 and CCP4 software suites for auto-indexing, integration, scaling, and merging (Otwinowski and Minor, 1997; Potterton *et al.*, 2003). PHENIX was used for structure refinement (Adams *et al.*, 2010). With Dr. Garside's assistance, the entire process was replicated. The crystallographic phase problem was solved by molecular replacement using yeast Snu13p (PDB ID 2ALE) as the search model. Structural alignment and visualization were done with PyMOL for Mac.

3.5 Results and Discussion

3.5.1. Recombinant Expression and Purification of Snu13

His-Snu13 was affinity purified by incubation with Ni-NTA resin, washed, and then eluted by centrifugation. The expression and elution fractions were subsequently analyzed on a 12% polyacrylamide SDS-PAGE (Fig 3.1 A, B). The His-Snu13 construct was 21.3 kDa; on 12% polyacrylamide SDS-PAGE, it migrated slightly below the 21.5

kDa band of the Broad Range Molecular Weight Marker (Fig. 3.1). After batch purification, 15 mL of purified Snu13 was dialyzed with TEV protease and ran over Ni-NTA resin to separate the cleaved His tag from Snu13. Cleaved Snu13 was 15.9 kDa and ran in-between the 21.5 and 14.4 kDa markers. The protein was then concentrated to 5 mg/mL and analyzed by SDS-PAGE (Fig. 3.1 C).

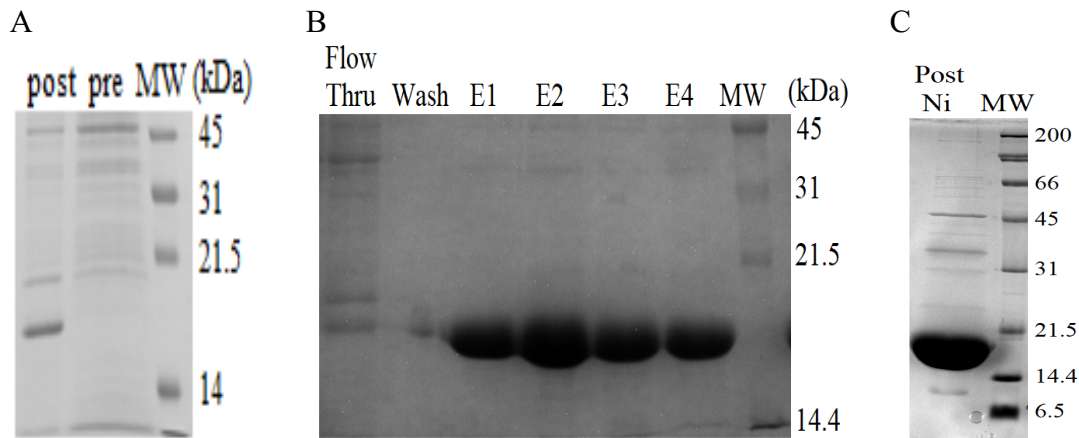


Figure 3.1. The expression and affinity purification of Snu13. (A) Expression of *C. merolae* His-Snu13 in Rosetta pLysS. (B) Ni-NTA batch purification of His-Snu13. (C) Analysis of Snu13 after His-tag removal and concentration. SDS-PAGE of 12% polyacrylamide was used for analysis.

3.5.2. X-ray Crystallography of Snu13

X-ray crystallography demands high sample purity, often upwards of 97% (Vidovic *et al.*, 2000). To remove contaminants present after affinity purification, size exclusion chromatography was performed (Fig. 3.2 A). After the sample was concentrated to 10 mg/mL, 30 ug was run on a 12% polyacrylamide gel to reveal 99% purity (Fig. 3.2 B). Hanging-drop vapor diffusion using purified Snu13 performed by Dr. Garside yielded beautiful looking rod-shaped crystals (Fig. 3.2 C).

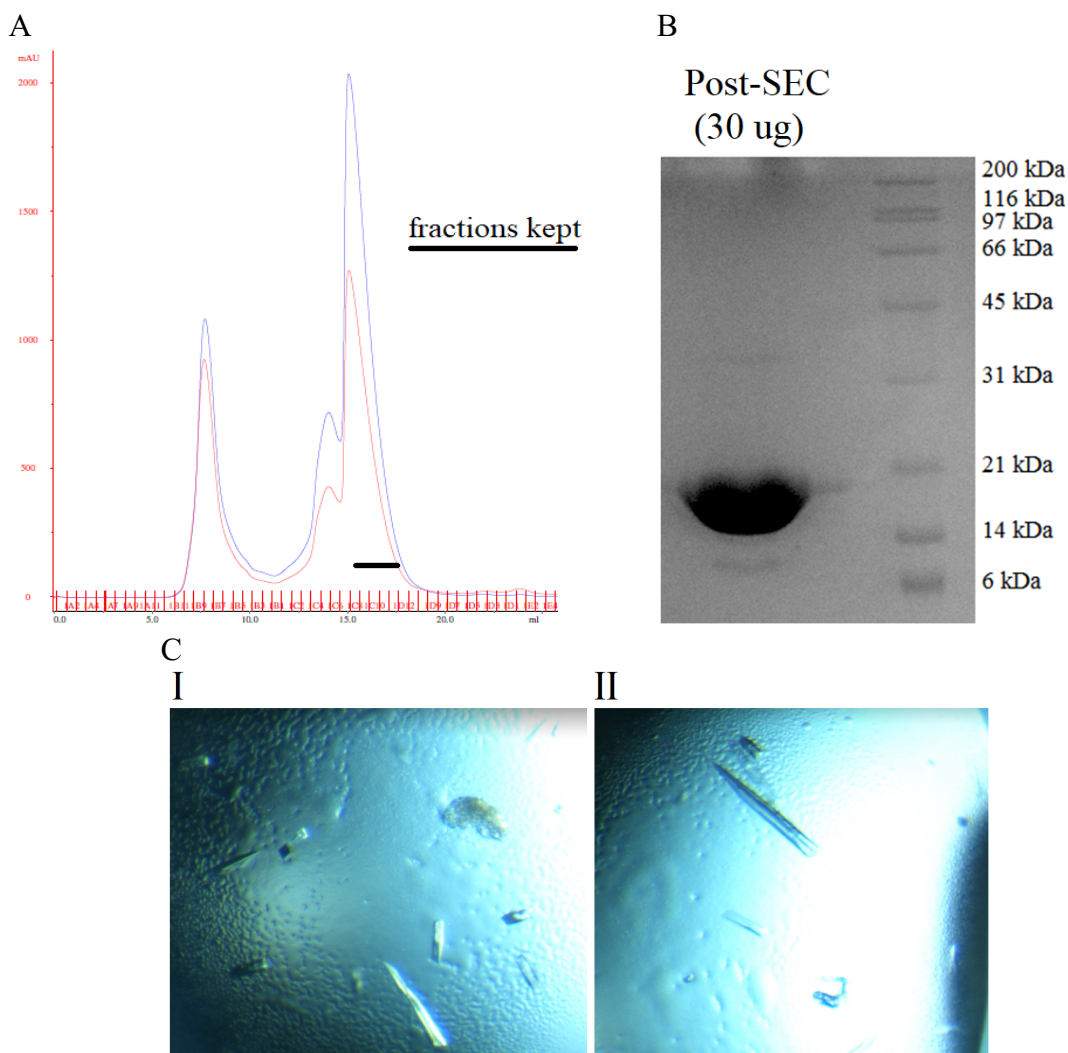


Figure 3.2. Purification and Crystallography of Snu13. (A) Chromatogram of *C. merolae* Snu13 after size exclusion (Superdex200 10/300 GL) purification using the AKTA Purifier 10. Red and blue lines relate to 260 nm and 280 nm wavelengths, respectively. (B) SDS-PAGE analysis of the purified and concentrated sample. (C) Rod-shaped Snu13 crystals generated and photographed by Dr. Erin L. Garside at the MacMillan Lab (University of Alberta).

Due to the small size of the crystals, Dr. MacMillan recommended that Dr. Garside collect diffraction data from the Stanford Synchrotron Radiation Light Source. The protein crystals were mounted on a small loop, coated with 20% glycerol, frozen in liquid nitrogen, and mounted onto the beamline (Fig. 3.3 A). Exposure to photons from the beamline generated a diffraction pattern that was later used to construct an electron density map (Fig. 3.3 B).

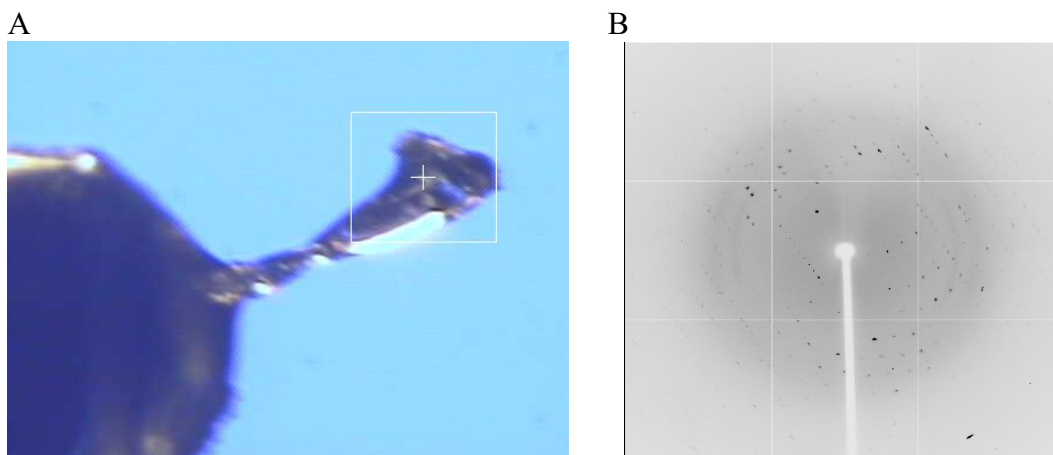


Figure 3.3. X-ray diffraction of Snu13 crystals at the Stanford Synchrotron Light Source. (A) Picture of a *C. merolae* Snu13 Crystal mounted on a loop in-line with the beamline. (B) Raw image of a diffraction pattern created by exposing the crystal to photons. Dr. Erin L. Garside provided both images.

3.5.3. Data Analysis and Crystal Structure of Snu13

A phase is the position of a point in time along a wave. The phases of the photons that encountered the crystal relate to the position of the electrons in the crystal. The phase problem of X-ray crystallography describes the loss of phase that occurs when the intensity of a reflection is measured. All of the dots on the diffraction image are reflections (Fig. 3.3 B). Molecular replacement was the best option for solving the crystallographic phase problem of Snu13 because the resolution range of the data set was 43.21-2.35 Å (Angstroms), multiple isomorphous replacement was not performed, and a good search model was available. If the resolution of the dataset was high enough (1.5 Å is the rule of thumb), and the molecule was small enough, statistics could have been used to predict the phase for all of the reflections of the dataset. Multiple isomorphous replacement would have allowed for the phase to be measured from the effect of the heavy atoms. Molecular replacement only required that a nearly identical structure be oriented correctly into the unknown unit cell of Snu13. Advances in computation have

made the model orientation trivial, but the importance of the search model remains paramount (Evans and McCoy, 2008).

To determine the best search model for molecular replacement, a multiple sequence alignment was performed using STRAP and PredictProtein software programs (Gille *et al.*, 2014; Yachdav *et al.*, 2014). High sequence similarity is a good indicator of low rms deviation (root-mean-square deviation) between two models, which means the structures are highly similar (Evans and McCoy, 2008). A multiple sequence alignment of Snu13 against homologs of human, budding yeast, and the red alga *Galdieria sulphuraria* revealed high sequence conservation among all of the sequences (Fig. 3.4). This suggested that the structure of Snu13 was highly conserved and likely important to the *Cm* splicing pathway. Snu13 shared the highest similarity with yeast Snu13p at 57%, which made it one of the most-conserved proteins among the entire *Cm* splicing suite (Stark *et al.*, 2015). Thus the model for yeast Snu13p (PDB ID 2ALE) was selected as the search model for molecular replacement (Black *et al.*, 2016).

	α1	
C.merolae	MEPLSSTEAPSRPMDVVTAPNQADPRAYPFAPADLVVEILDVLVQQASHYK	50
G.sulphuria	-----MAQNSGIHPKAYPLADAQLTVQILDLIRQASNYK	34
H.sapiens	-----MTEAD-VNPKAYPLADAHLTKKLLDLVQQSCNYK	33
S.cerevisiae	-----MSAP---NPKAFPLADAALTQQILDVVQQAANLR	30
	:: .*:~::~* * *. ::**::~*:: :	
	<div style="display: flex; justify-content: space-around; align-items: center;"> β1 α2 β2 α_{3-10} α3 β3 </div>	
C.merolae	QIKKGLEVLKSMNRGLAEFVVLAADTQPLEILLSAPLVAEDKAVPYVFV	100
G.sulphuria	QLKKGANEATKTNLNGISEFVVLAADTEPLEILLHLPLLCEDKNPVYVYV	84
H.sapiens	QLRKGANEAATKTNLNGISEFIVMAADAEPELIIHLPLLCEDKNPVYVFV	83
S.cerevisiae	QLKKGANEATKTNLNGISEFIIMAADCEPIELLHLPLLCEDKNPVYVFV	80
	*::~** ** . *:~::~*~::~*** :*:~::~* ~*:~::~*** ****:*	
	<div style="display: flex; justify-content: space-around; align-items: center;"> α4 β4 α5 </div>	
C.merolae	PSKAALGRACGVSRPVIAACAVLRADMSQLRNQITALRTKIEQLLL	145
G.sulphuria	PSKQALGRACGVSRPVIACSITVNEGSQTKTQIRNIRDQIERLLL	129
H.sapiens	RSKQALGRACGVSRPVIACSVTIKEGSQKQIQSIQSISIERLLV	128
S.cerevisiae	PSRVALGRACGVSRPVIAASITTNDASAIAKTQIYAVKDKIETLLI	126
	*: ***** :: : * :: ** :: ~* ~*:	

Figure 3.4. Sequence alignment with sequences from *G. sulphuraria* (Snu13g), (Snu13g), *Homo sapiens* (15.5K), and *S. cerevisiae* (Snu13p). Sequence identity is indicated by: sequence identity (*), conservation (:), and partial conservation (.). The predicted secondary structures of α -helices and β -strands overlay the alignment.

The X-ray crystal structure of Snu13 was solved to 2.35 Å with a completeness of 99.39% (Table 3.6; Black *et al.*, 2016). The model was refined until the working R-factor was 0.1662 and R_{FREE} was 0.2200 (Table 3.6). If a perfect search model were used, the R-factor would have been 0; if a completely random model were used, it would be about 0.63 (Kleywegt and Jones, 1997). The closer the R_{FREE} value is to the R-factor, the better the result of modeling. The X-ray structure of Snu13 was considered more than acceptable and has lower R-values than some structures of human and yeast homologs (Vidovic *et al.*, 2000; Dobbyn *et al.*, 2007). The space group of the crystal was P2₁2₁2₁

which describes an orthorhombic lattice and rectangular prism as the unit cell (Heller, 1952). There are currently 12 submissions in the RCSB PDB (Research Collaboratory for Structural Bioinformatics Protein Data Base) for *Cyanidioschyzon merolae* proteins and 10 of those are unique. The crystal structure of Snu13 is currently the only splicing protein from *C. merolae* in the PDB, under the ID of 5EWR (Black *et al.*, 2016). It is also worth noting that Snu13/5EWR is the first crystal structure of a protein to be published from a lab at the University of Northern British Columbia.

Table 3.6. Data collection, phasing, and refinement statistics.
Adapted from Black *et al.* 2016.

Data Collection	CM-Snu13 ¹
Space group	P2 ₁ 2 ₁ 2 ₁
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	30.33, 57.58, 65.38
<i>a</i> , <i>b</i> , <i>c</i> (°)	90, 90, 90
Wavelength (Å)	1.1271
Resolution (Å)	2.35
<i>I</i> / <i>sI</i>	21.36 (11.85)
Completeness (%)	99.77 (99.39)
Redundancy	7.7 (7.6)
Refinement	
Resolution (Å)	43.21-2.35
No. reflections	158498
<i>R</i> _{work} / <i>R</i> _{free}	0.1662/0.2200
No. atoms	
Protein	933
Water	42
<i>B</i> -factors (Å ²)	
Protein	49.5
Water	40.0
R.m.s deviations	
Bond lengths (Å)	0.016
Bond angles (°)	1.52

[1] Statistics for highest resolution shell (2.434-2.35 Å) are shown in parenthesis.

The structure of Snu13 was highly conserved when compared to human 15.5K and yeast Snu13p (Fig. 3.5 A). It maintained the α - β - α sandwich of the L7 family of

proteins (Fig. 3.5 A). Several ribbon structures were overlaid: Snu13 (green); yeast Snu13p (cyan); human 15.5K (yellow); human 15.5K bound to U4atac RNA and Prp31 (purple); human 15.5K bound to U4 snRNA and Prp31 (pink). The rms deviation was very low (0.014 Å) for all of the alpha carbons of the overlaid structures, evidence of their similarity (Fig. 3.5 A; Black *et al.*, 2016). The structural comparisons also provide links to the various binding functions of Snu13 in *C. merolae*. Overlay of the Snu13 with yeast revealed 5 residues hypothesized to interact with either Prp3 or Prp4 during U4/U6 di-snRNP assembly (Fig. 3.5 B; Dobbyn *et al.*, 2007). The U4/U6 di-snRNP comprises Snu13, Prp3, Prp4, Prp31, the Lsm and Sm complexes, as well as U4 and U6 snRNAs (Nottrott *et al.*, 2002; Will and Lührmann, 2011). When Snu13 was superimposed onto the human homolog (15.5K) bound to the U4atac kink-turn, the RNA-binding residues were determined (wireframe, Fig. 3.5 C).

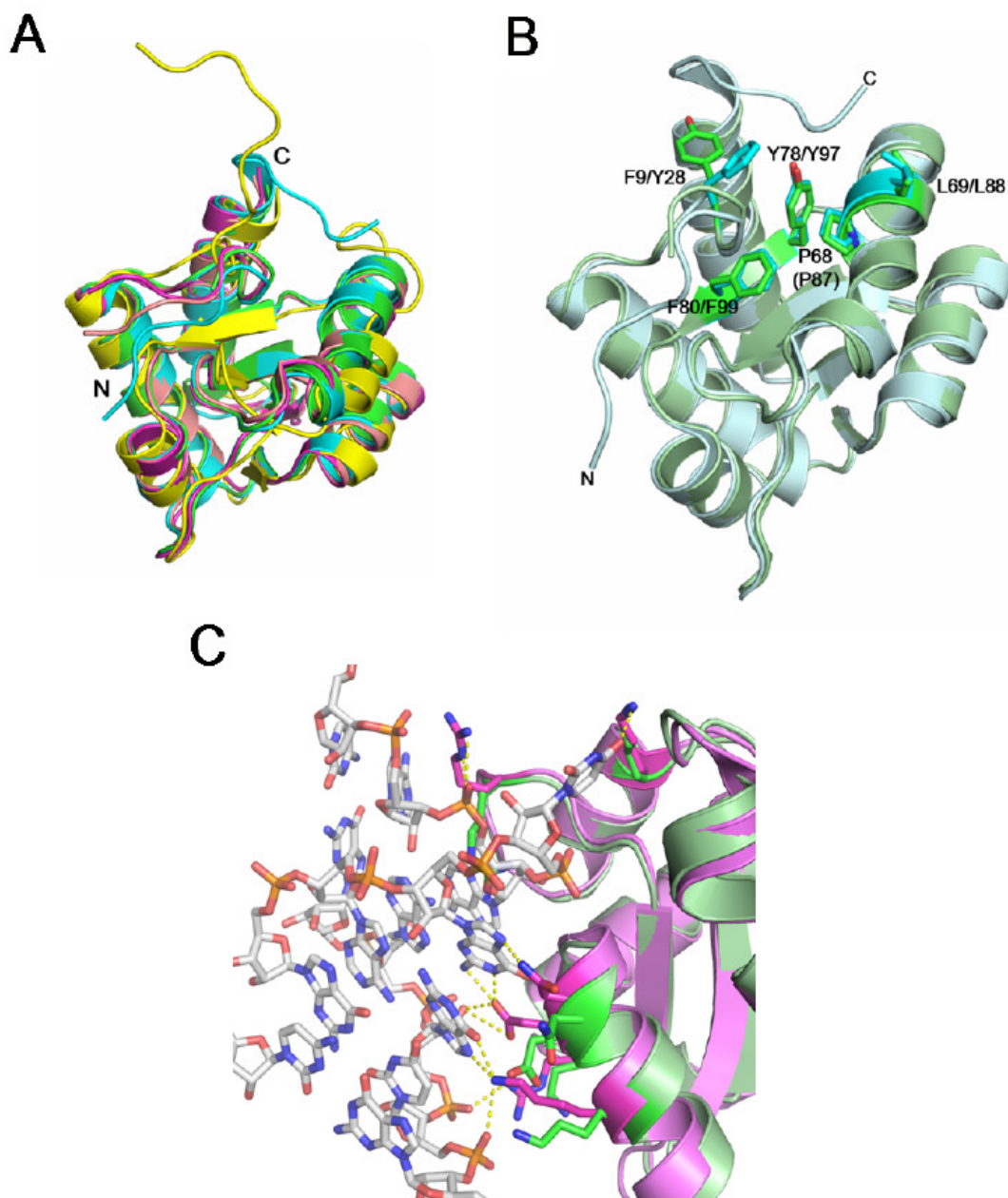


Figure 3.5. Structural comparison of *C. merolae* Snu13 with orthologs. Images were adapted from Black *et al.* and arranged using PyMOL (2016). (A) Structural overlay of ribbon diagrams of Snu13 (*C. merolae*; PDB: 5EWR; green), Snu13p (*S. cerevisiae*; PDB: 2ALE; cyan), 15.5K (human; PDB: 2JNB; yellow), 15.5K (human bound to U4atac RNA and prp31; PDB: 3SIU; magenta), 15.5K (human bound to prp31 and U4 snRNA; PDB: 2OZB; light pink) (B) Conservation of the hydrophobic pocket proposed to interact with Prp3. Shown are ribbon diagrams of *CmSnu13* (green; PDB: 5EWR) overlaid on Snu13p (*S. cerevisiae*; PDB: 2ALE) highlighting the residues lining the hydrophobic pocket: F9 (Y28), P68 (P87), L69 (L88), Y78 (Y97), F80 (F99). Yeast identification followed by algal in parentheses. (C) Detail of overlay of *CmSnu13* (*C. merolae*; PDB: 5EWR; ribbons, green) with 15.5K bound to U4atac RNA (human; PDB: 3SIU; ribbons, magenta; RNA in stick representation).

Pull-downs, mutational studies, binding assays, and structural studies have revealed Snu13 interacts with the U4 5' stem, Prp31, and the Prp3/Prp4 heterodimer (Fig. 3.6; Nottrott *et al.*, 1999; Vidovic *et al.*, 2000; Nottrott *et al.*, 2002; Hamma *et al.*, 2004; Dobbyn *et al.*, 2007; Liu *et al.*, 2011; Liu *et al.*, 2015; Hardin *et al.*, 2015; Nguyen *et al.*, 2015). An overlay of Snu13 with human 15.5K suggested Snu13 would be capable of the same interactions (Fig. 3.5). This also demonstrated how minimal (if any) conformational changes are required of the globular Snu13 to accommodate multiple protein and RNA interactions (Fig. 3.6).

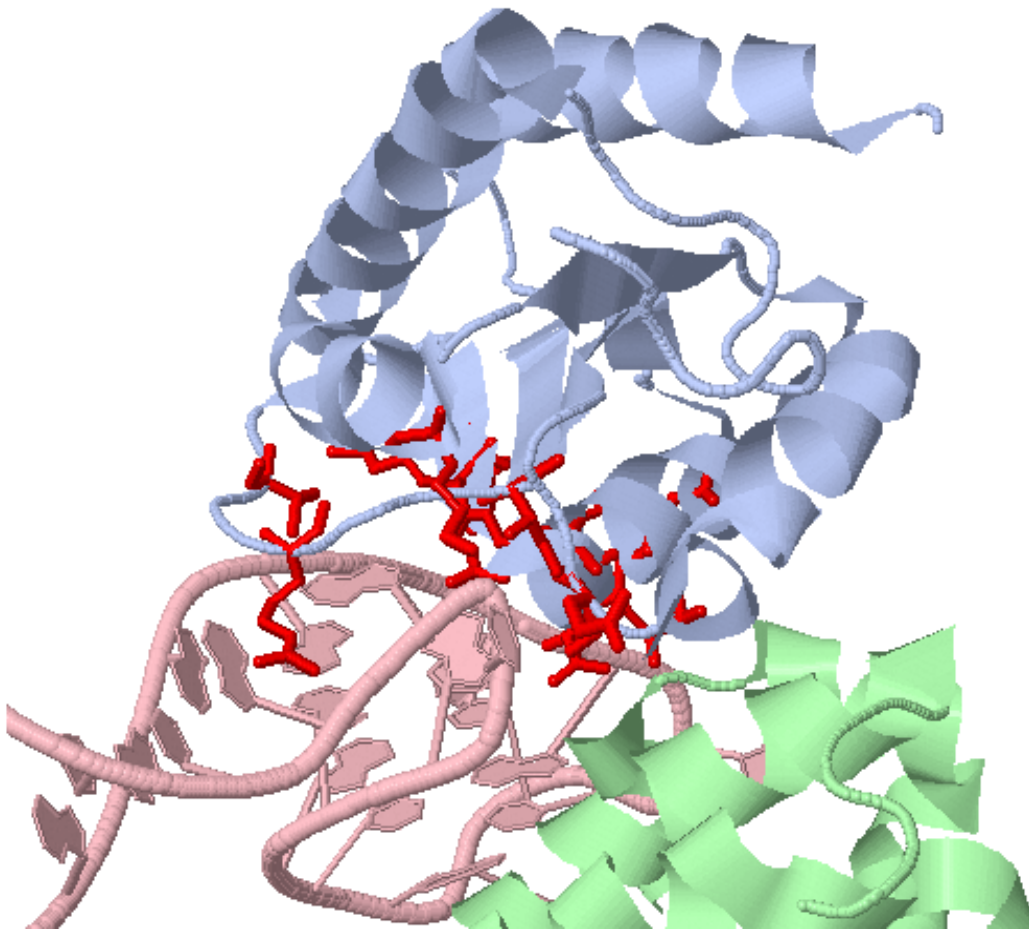


Figure 3.6. Close-up of the RNA binding domain of human 15.5K (light blue) associated with the kink-turn of human U4 snRNA (pink) and Prp31 (human 61K, green). RNA binding residues are shown in red wireframe. The PDB file 3SIU (human 15.5K/U4/Prp31) was used as the model and was visualized using PyMOL.

A hydrophobic pocket of Snu13p is theorized to interact with Prp3 (human 90K) or Prp4 (human 60K) during U4/U6 assembly (Dobbyn *et al.*, 2007). Cryogenic electron microscopy structures of the tri-snRNP particle in yeast and humans suggest that Prp4 is the most likely candidate (Nguyen *et al.*, 2015; Agafonov *et al.*, 2016). *Cm* residues Y28, P87, L88, Y97, and F99 form this putative protein-binding pocket and were based on an alignment with yeast Snu13p (Fig. 3.7; Dobbyn *et al.*, 2007).

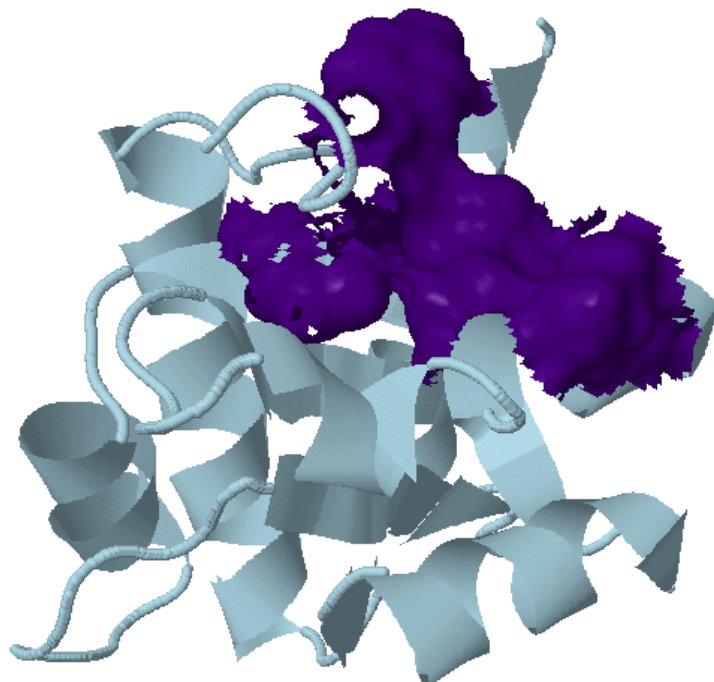


Figure 3.7. Hydrophobic pocket of Snu13 is clearly seen from the molecular surface, while the rest of the protein is shown in ribbon. The pocket comprises residues Y28, P87, L88, Y97, and F99. This region is hypothesized to interact with U4 protein Prp3 (90K) or Prp4 (60K) during di-snRNP assembly. Snu13 from *C. merolae* was used for the modeling, which was done with PyMOL [PDB ID 5EWR].

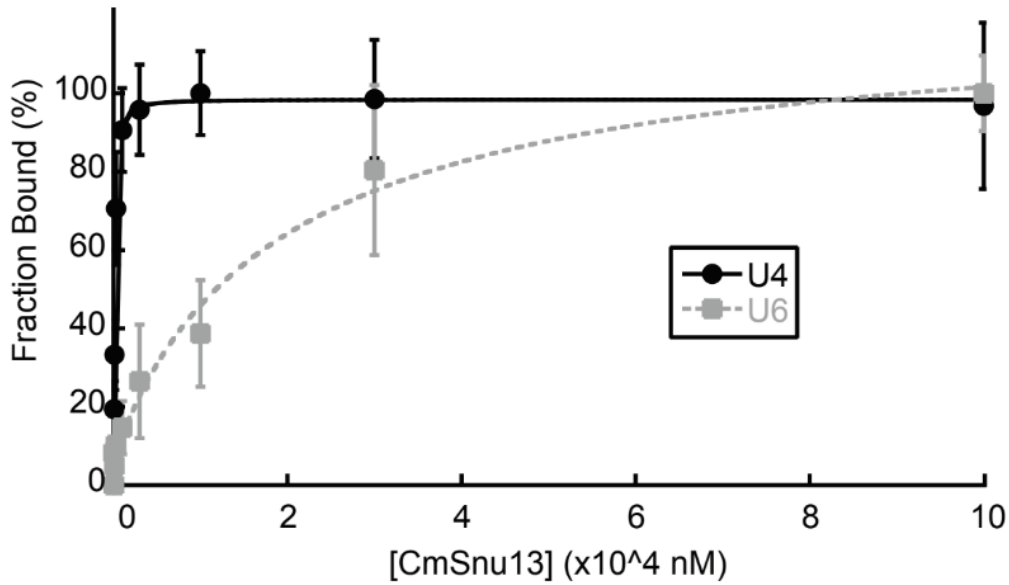
Alignment of Snu13 with human 15.5K in complex with U4 snRNA reveals binding likely occurs between β -strands 1-2 and 3-4 (Fig. 3.5 C and 3.6; Vidovic *et al.*, 2000). Furthermore, modeling of Snu13 with all human and yeast structures (free or bound to proteins/RNA) suggests all of the interactions require minimal changes (Fig. 3.5-3.7). Solution and crystal structures of human 15.5K in complex with Prp31 suggest α -helices 2 and 3 would interact with Prp31 (Fig. 3.6; Liu *et al.*, 2011; Liu *et al.*, 2015).

3.5.4. Fluorescence Polarization of Snu13 and the U4 5' Stem Loop

In order to demonstrate that Snu13 was a conserved protein, it was critical to prove that the same construct used to generate a crystal was also functional. Fluorescence polarization was used to measure the RNA binding ability of Snu13 against an RNA

oligonucleotide of the *Cm* U4 5' stem loop (Fig. 3.8 A, B middle structure). Several experiments were performed in triplicate and two of those included the same sample that crystallized. The average from 12 replicates of Snu13 against U4 5' SL and 6 replicates against the control (*Sc* U6 SL) was used to generate a saturation curve (Fig. 3.8 A, B left structure). The standard deviation of each data set was used for error.

A



B

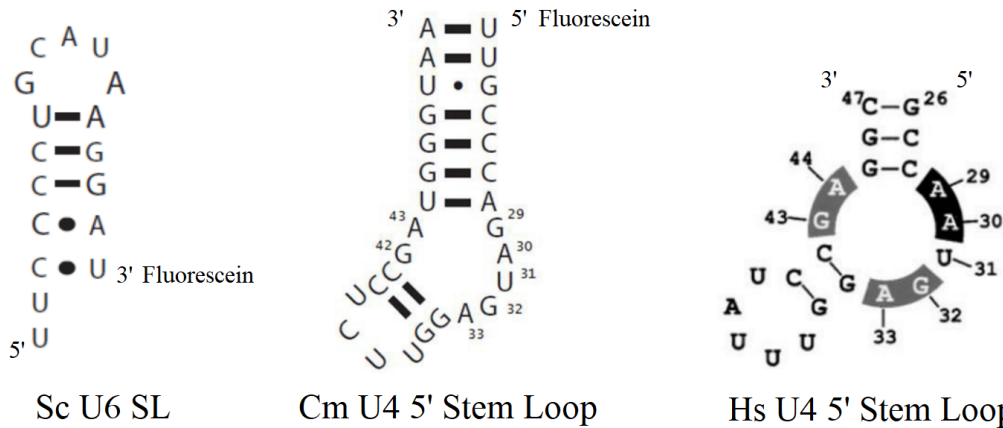


Figure 3.8. Saturation curve of Snu13 against the U4 5' SL. Adapted from Black *et al.* (2016). (A) Fluorescence polarization was measured for *C. merolae* Snu13 titrated against 5'-fluorescein-labeled RNAs of U4 (*C. merolae* U4 5' stem loop) and U6 (*S. cerevisiae* U6 RNA stem loop). Dissociation constants (K_D) were determined by fitting this data to the Hill equation. (B) Secondary structures of the two RNA oligos used in the binding experiment (*Cm* U4 and *Sc* U6) as well as *Hs* U4 5' SL for comparison. The human structure is adapted from Vidovic *et al.* (2000).

The binding affinity of Snu13 was 160 ± 10 nM for the *Cm* U4 5' SL and approximately 16 μ M for the control oligo (Table 3.7). The measured K_D was similar to values reported in literature, demonstrating that Snu13 is structurally and functionally conserved (Table 3.7). The U4 5' stem loop has been defined as having a purine-rich

asymmetrical loop with 5 nucleotides on one side, and 2 on the other (5 + 2). An extensive bioinformatics search done of U4 snRNAs by Nottrott *et al.* revealed that the U4 5' stem loop is highly conserved (1999). The canonical U4 5' SL has human-equivalent positions of U31, G32, A33, G43, and A44 that are identically conserved, while positions A29 and A30 are purines (Fig. 3.8 B; Stevens and Abelson, 1999). *Cm* 5' SL is identical to human except for G29, which is a purine and still fit the canonical model (Fig. 3.8 B). X-ray crystallography and FRET experiments of human 15.5K bound to the U4 5' SL revealed that the flanking stems of the bound asymmetrical loop bend to 65° to each other and U31 is sequestered between β -strands 1-2 and 3-4 (Fig. 3.5 C and 3.6; Vidovic *et al.*, 2000). The folded conformation of the stem loop is described as a “kink-turn” and U4 has been shown to adopt the kink-turn in presence of metal ions alone (Goody *et al.*, 2004). Accordingly, Snu13 belongs to a novel class of kink-turn binding proteins (Nottrott *et al.*, 1999; Kühn *et al.*, 2002; Marmier-Gourier *et al.*, 2003; Soss and Flynn, 2007). Given that the Snu13-RNA interaction was within the range of literature values (8 – 150 nM), it was not surprising the U4 5' SL was also similar (Table 3.7).

Table 3.7. Comparison of Protein-RNA Affinities. Adapted from Black *et al.* (2016).

Protein	RNA target	Method	Apparent K_D (nM)	Literature
15.5K	U3 snoRNA	EMSA	30	Watkins <i>et al.</i> , 2000
15.5K	U8 snoRNA	EMSA	40	Watkins <i>et al.</i> , 2000
15.5K	U14 snoRNA	EMSA	8	Watkins <i>et al.</i> , 2000
15.5K	U4 snRNA	EMSA	20	Watkins <i>et al.</i> , 2000
15.5K	U3 snoRNA	EMSA	130 ± 13	Szewczak <i>et al.</i> , 2002
L7Ae	Kt-7 23S RNA	FRET	0.9 ± 0.2	Turner <i>et al.</i> , 2005
Snu13p	U3 snoRNA	EMSA	75	Dobbyn <i>et al.</i> , 2007
Snu13p	U4 snRNA	EMSA	150	Dobbyn <i>et al.</i> , 2007
15.5K	U4 5' stem loop	SPR	27	Soss and Flynn, 2007
CmSnu13	U4 5' stem loop	FP	160 ± 10	Black <i>et al.</i> , 2016

EMSA: electrophoretic mobility shift assay; SPR: surface plasmon resonance; FP: fluorescence polarization

Recombinant proteins do not have native conditions during their expression and may not have the appropriate chaperones or chemical modifications necessary to create an identical version (Weinstein *et al.*, 2002; Gräslund *et al.*, 2008). Furthermore, *in vitro* studies are never able to reproduce the environment in which an interaction occurs. The functionally active population of a recombinantly expressed and purified protein is therefore likely below 100%. A thorough review article on the interpretation of binding data provided a mathematical model that predicted the amount of complex formed if the K_D and amounts of ligand and receptor were known and the receptor (protein) population was assumed to be 100% (Hulme and Trevethick, 2010).

$$\frac{[RL]}{[R][L]} = \frac{k_{on}}{k_{off}} = K \quad \text{Equation 5}$$

$$\frac{[R][L]}{[RL]} = \frac{k_{off}}{k_{on}} = K_d \quad \text{Equation 6}$$

$$[RL]^2 - [RL]([R_T] + [L_T] + K_d) + [R_T][L_T] = 0 \quad \text{Equation 7}$$

$$[RL] = \frac{([R_T] + [L_T] + K_d) - \sqrt{([R_T] + [L_T] + K_d)^2 - 4[R_T][L_T]}}{2} \quad \text{Equation 8}$$

Using Equation 8, the expected fraction of Snu13 bound by U4 at a 1:1 molar ratio would be expected to be 38% given the K_D of 150 nM and concentration of 150 nM (Hulme and Trevethick, 2010). In an effort to quantify the fraction of recombinant Snu13 that was functionally active, the U4 5' SL (ro52) was titrated against a fixed concentration of Snu13 (150 nM). The fraction of protein bound was plotted against the ratio of RNA to protein (Fig. 3.9). The measured fraction of Snu13 bound by U4 was approximately 0.325 or 32.5% (Fig. 3.9). Based on this model, the active fraction of Snu13 is about 85% and the adjusted K_D of Snu13 for the U5 SL would be approximately 135 nM.

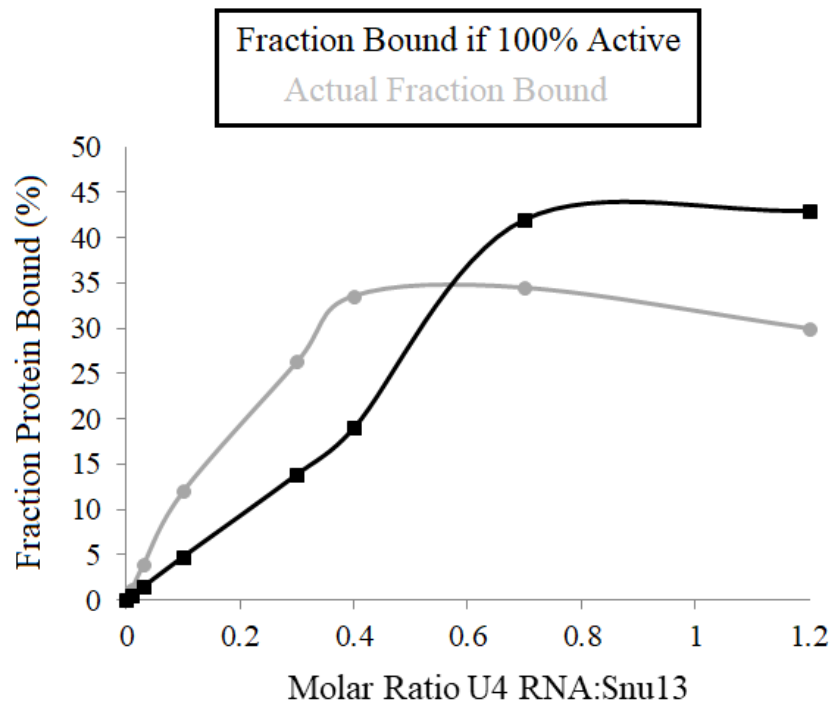


Figure 3.9. Functionally active fraction of Snu13 determined by RNA titration. *Cm* U4 5' SL (ro52) was titrated against 150 nM *Cm* Snu13. The resulting anisotropy measurements were converted into fraction of protein bound (%) and plotted against the molar ratio of probe to protein (grey line). The theoretical values assume the active fraction was 100% and were based on Equation 8 (black line).

Chapter 4

Assembly of the *Cm* U4/U6 di-snRNP

This chapter covers all of the experiments done to assemble the *Cm* U4/U6 di-snRNP in a step-wise fashion. The majority of the *Cm* spliceosome has not been studied, and it is very interesting to parse-out which mechanisms are conserved and which are novel. The U4/U6 di-snRNP assembly pathway has been studied in humans and yeast, and based on the *C. merolae* di-snRNP genes present, the assembly is theoretically similar. Gel shift assays were performed to biochemically probe di-snRNP assembly using *in vitro*-transcribed U4 and U6 snRNA and recombinant proteins. Of the 5 *Cm* U4/U6-associated proteins discovered (Snu13, LSms, Sms, Prp3, and Prp4), Snu13, LSm, and Sm proteins were used in the assay. Expression and purification conditions of Prp3, Prp4, and Prp31 require further optimization.

4.1 Methods – Preparation of Proteins Used in U4/U6 di-snRNP Assembly

4.1.1. Expression and Purification of *Cm* Snu13 and LSm/Sm Complexes

Cm Snu13 was expressed and purified as described in Chapter 3. Rader lab members Kirsten Reimer and Fatimat Shidi prepared the LSm and Sm complexes respectively. The LSm and Sm proteins were expressed in Rosetta pLysS (DE3) cells then purified with both affinity (1 mL HisTrap) and size exclusion (Superdex200 10/300 GL) column chromatographic techniques (ÄKTA Purifier 10, GE Healthcare Life Sciences).

4.2 Methods – *In vitro* Transcription of *Cm* U4 and U6 snRNA

4.2.1 Preparation of U4 And U6 DNA Templates For *In vitro* Transcription

In order to generate DNA templates for the U4 and U6 snRNAs, the polymerase chain reaction (PCR) was performed using Taq Polymerase (New England Biolabs).

Primers for the U4 and U6 DNA templates were designed and ordered from Integrated DNA Technologies. PCR reactions and primer sequences are outlined below (Table 4.1).

<u>Components of Reaction</u>		<u>Thermocycler Program</u>	
<i>Cm</i> Genomic DNA (262 ng/uL)	1 uL	95°C	2 minutes
Taq Polymerase (5000 units/mL)	0.4 uL	95°C	20 seconds
Taq Reaction Buffer (10x)	5 uL	56°C	20 seconds 35 cycles
Forward primer (10 uM)	2.5 uL	68°C	45 seconds
Reverse primer (10 uM)	2.5 uL	68°C	5 minutes
dNTPs (10 mM)	1 uL		
Water	up to 50 uL		

Table 4.1. PCR primers used to generate *Cm* U4 and U6 DNA templates for *in vitro* transcription and their respective annealing temperatures.

Identity	Sequence
U4 RNA (fwd)	AAATTAATACGACTCACTATAGGGATACTTGCGCAGTGTCTG GTTG
U4 RNA (rev)	CTTTCCAAAAATTTCCACCAAGCGC
U6 RNA (fwd)	AAATTAATACGACTCACTATAGGGTGCGCCTTTATCGGC
U6 RNA (rev)	AAAAAGGTATACCTCGAGACGATTGTCC

4.2.2. *In vitro* Transcription of *Cm* U4 and U6 RNA Transcripts

In vitro transcription (IVT) of *Cm* U4 and U6 snRNA was done cold as per the HiScribe T7 High Yield RNA Synthesis kit (New England Biolabs). Each IVT reaction was set-up according to Table 4.2 in 0.2mL clear PCR tubes (Corning) and was left for 12 hours at 37°C.

Table 4.2. Composition of the IVT reactions that generated *Cm* U4 and U6 RNA transcripts. The reactions were carried out in 20 uL for 12 hours at 37°C.

IVT Reaction Reagents	<i>Cm</i> U4 snRNA Reaction	<i>Cm</i> U6 snRNA Reaction
Template DNA	4 uL PCR Product (<i>Cm</i> U4 DNA, 250 ng)	4.4 uL PCR Product (<i>Cm</i> U6 DNA, 250 ng)
T7 RNA Polymerase Mix	1.5 uL	1.5 uL
T7 Reaction Buffer (10x)	1.5 uL	1.5 uL
ATP (100 mM)	1.5 uL	1.5 uL
UTP (100 mM)	1.5 uL	1.5 uL
GTP (100 mM)	1.5 uL	1.5 uL
CTP (100 mM)	1.5 uL	1.5 uL
Water	7 uL	6.6 uL

Each IVT product was mixed with equal volume formamide, heated at 65°C for 3 min then purified on a 6% polyacrylamide (19:1) 7M urea gel which ran in 1 x TBE at 400 V for 60 min. RNA gel bands were visualized by UV shadowing, excised into a 1.5 mL Eppendorf tube and crushed with a micro pestle (DWK Life Sciences). The crushed gel was mixed with 400 uL water and heated for 10 min at 70°C. The gel slurry was subsequently loaded onto a Performa Spin Column (Edge Bio) and spun for 3 min at 850 xg. RNA precipitation was done by adding 1.5 uL glycogen (20 mg/mL), 1/10 volume 3M sodium acetate, and 1 mL cold ethanol (100%) and subsequently being stored for 18 hours at -20°C. The precipitated RNA was pelleted by centrifugation at 17000 xg for 30 min at 4°C. Each RNA pellet was washed twice with 70% cold ethanol before being resuspended in 50 uL water and stored at -20°C. The full-length sequences of the products are presented in Table 4.3. For reference purposes, the Accession Number and Chromosomal locations for U4 and U6 are #AP006487 Chromosome 5/222390- 222566, and #APOO6501 Chromosome 19/483497- 483364, respectively.

Table 4.3. Sequences of the full-length U4 and U6 snRNA generated from IVT.

snRNA	Sequence
U4	AUACUUGCGCAGUGUCGGUUGUUGCCCAGAUGAGGUUCUCCGAUG GGUAAACGGUAUGCU GAACACAAACAAUUUACCACAGUGGACUUU UUAACCGGUCCGUCUGUGACGGGCAUGCC UCAAGCCCAGCCACGUGCGAAGAACACGUG UUCUGCGCUUGGUGGAAAUUUUUGGAAAG
U6	GGUGCGCCUUUAUCGGCGAUUUGCCCCGAAUCGUCGGUAGGGGUAC GCGCCGUCCAUCCAUGGAUAACAAACUGAGAUGAUCAGCUUCCG CACUGCGCAAGUAUGCGGACAAUCGUCU CGAGGUAUACCUUUUU

4.3 Methods – Assembly of the U4/U6 di-snRNP

4.3.1. U4 and U6 Annealing Assays

A 9% acrylamide/TBE gel (native) was made and cooled to 4°C. Every reaction was then set-up in a 1.5 mL microtube (Eppendorf) and included Buffer D, 5X Splicing Mix, and one or both RNA (Table 4.4). Dye was added immediately prior to injection into the polyacrylamide gel. For heated reactions, the mixture was heated at 65°C for 3 min (VWR Standard Heatblock) before the block was removed from the apparatus and placed on the bench for 1 hour to gently cool. After 30 min of cooling, the non-heated reactions were set-up and incubated at room temperature (on bench) for 30 min. For samples with one RNA, there was no incubation.

Table 4.4. Buffer components used in U4 and U6 snRNA annealing reactions.

Buffer	Components
5X Splicing Mix	300 mM KPO ₄ , 12.5 mM MgCl ₂ , 15% PEG 8000
Buffer D	20 mM HEPES-NaOH pH 7.9, 50 mM KCl, 0.2 mM EDTA, 20% Glycerol

The Annealing Reaction I is as follows:

5X Splicing Mix	2 uL
Buffer D	4 uL
U4 RNA (8 pmol per reaction)	2 uL
U6 RNA (11 pmol per reaction)	2 uL
Water	to 10 uL
6X Native Dye	2 uL

Once the reactions were complete, native dye was added and the samples were injected into the pre-chilled 9% native polyacrylamide gel. The gel was then ran at 300V for 1 hour at 4°C. Once ran, the gel was placed in 100 mL of 0.5 ug/mL ethidium bromide and stained for 30 min before it was rinsed with water and exposed with the Fluor Chem Q imager (Protein Simple).

4.3.2. U4/U6 di-snRNP Assembly

With the annealing reaction successful, the next steps were optimization and the addition of U4/U6 di-snRNP associated proteins (*Cm* Snu13, LSm, Sm). The di-snRNP Assembly Reaction I was set-up as follows:

Di-snRNP:	
5X Splicing Mix	2 uL
Buffer D	4 uL
U4 RNA (2 pmol per reaction)	1 uL
U6 RNA (2.5 pmol per reaction)	2 uL
Water	to 9 uL
12 uM Protein (Snu13, LSm or Sm)	1 uL

Reactions were set-up similar as previously described (section 4.3.1). A 9% acrylamide/TBE gel was poured and cooled to 4°C. Heated reactions were mixed, heated to 65°C, and cooled for 1 hour prior to mixing unheated reactions. Once the reactions were ready, protein was added (when applicable) and incubated at room temperature (on bench) for 30 min. Samples with only one RNA (and no protein) were not incubated. Once the reactions were complete each sample was injected into a pre-chilled 9% native polyacrylamide gel. The gel was then ran at 300V for 1.5 hour at 4°C. When finished the gel was put into 100 mL of 0.5 ug/mL ethidium bromide and stained for 30 min before exposure with with the Fluor Chem Q imager (Protein Simple).

4.4. Methods – Recombinant Expression of *Cm* Proteins Prp3, Prp4, and Prp31

4.4.1. Cloning of *Cm* Prp3, Prp4, and Prp31

The pQLink H and N plasmid vectors (H denotes N-terminal His-tag, N is untagged) were purchased from Addgene (plasmids 13670 and 13667) and modified by Rader lab member Fatimat Shidi to allow for ligase-independent cloning (LIC), previously discussed in section 2.1.1. LIC using these plasmids and the *Swa*I and *Pac*I restriction enzymes has been described in detail (Scheich *et al.*, 2007). The figure below illustrates the use of LIC to construct a plasmid containing both *Cm* Prp3 and Prp4 (Fig. 4.1).

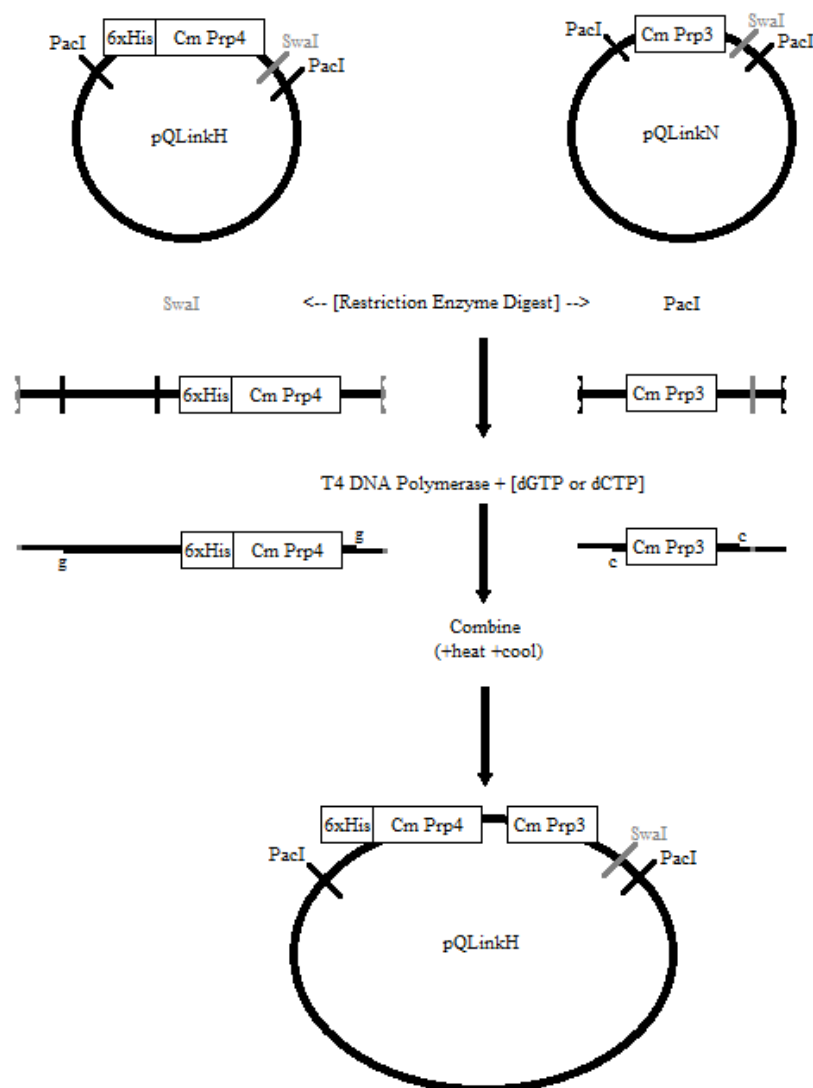


Figure 4.1. Overview of ligase-independent cloning to produce a single plasmid with two genes.

Having expressed and purified several proteins, it became apparent that an expression vector was needed that contained a C-terminal affinity tag. Thus the pQLinkC plasmid was constructed by modifying the pQLinkN (untagged) plasmid vector. A cleavable C-terminal 6x His tag downstream of the LIC sites SwaI and PacI was added by including the TEV protease recognition site upstream of the 6x His sequence. The forward and reverse primers are listed below, with the gene-specific region in capitals and the TEV recognition site in bold (Table 4.5).

Table 4.5. Primers used to modify pQLinkN into pQLinkC, as well as to subclone *Cm* Prp4 and Prp31 into pQLinkH and C plasmid vectors.

Gene	Target Plasmid	Sequence
Top strand	pQLinkN (becomes C)	AGCTTAGAGAACCTGTACTTCCAATCCGGTAC CCACCATCACCATCATCATTAGGC
Bottom strand	pQLinkN (becomes C)	TCAGCCTAATGATGATGGTGGTGGGTACC GGATTGGAAGTACAGGTTCTCTA
<i>Cm</i> Prp4 fwd	pQLinkH	TAC TTC CAA TCC CAC CGA AGT CAC GAA TTT CGA GCA GAG
<i>Cm</i> Prp4 rev	pQLinkH	TTA TCC ACT TCC CAC G CTA TGC GTA TAA CCT GTG GAC
<i>Cm</i> Prp31 fwd	pQLinkH	TAC TTC CAA TCC CAC GCA AGC ACT GGC GTT TCT GC
<i>Cm</i> Prp31 rev	pQLinkH	TTA TCC ACT TCC CAC G TTA TGC TTC ATC CTC ATC TAC ACC
<i>Cm</i> Prp31 fwd	pQLinkC	TAC TTC CAA TCC CAC G AGG AGA AAT TAA CT ATG AGC ACT GGC GTT TCT G
<i>Cm</i> Prp31 rev	pQLinkC	TTA TCC ACT TCC CAC G TGC TTC ATC CTC ATC TAC ACC

Three plasmid constructs were made for the remaining 3 proteins of the di-snRNP: N- and C-terminally labeled Prp31 variants, as well as a construct with both tagged Prp4 and untagged Prp3. The rationale was that co-expression of a predicted heterodimer should produce the most stable product. The *Cm* genes for Prp4 and Prp31 were cloned from genomic DNA into pQLinkH and pQLinkC plasmids using the primers listed in Table 3.3. Once *Cm* Prp4 was cloned into pQLinkH, Prp3 was then also subcloned from a pQLinkN plasmid using LIC. Prp3 was subcloned into the pQLinkN plasmid by Dr. Martha Stark. The Prp4 and Prp31 genes were amplified from genomic DNA using Q5 High Fidelity (HF) DNA Polymerase (New England Biolabs). Based on the predicted annealing temperatures of primers for Prp4 and Prp31 (66°C and 65°C respectively), 66°C was used as the annealing temperature and both PCR reactions were executed simultaneously.

The following plan includes the details of the PCR reaction:

Q5 HF Polymerase (2000 units/mL)	0.5 uL			
Q5 Reaction Buffer (5x)	10 uL	98°C	30 seconds	
Genomic DNA	250 ng	98°C	15 seconds	
Forward primer (10 uM)	2.5 uL	66°C	30 seconds	35 cycles
Reverse primer (10 uM)	2.5 uL	72°C	15 seconds	
dNTPs (10 mM)	1 uL	72°C	5 minutes	
Water	to 50 uL			

All 3 genes were ready for entry into their respective vectors (*Cm* Prp4 and Prp31 were PCR-amplified and *Cm* Prp3 was in pQLinkN). LIC requires T4 DNA Polymerase, a 5'-3' polymerase with 3'-5' exonuclease activity. The target pQLink plasmids (H and C) were linearized with *pmlI* (NEB) then treated with T4 polymerase and one deoxyribose nucleoside triphosphate (dNTP) species. T4 chews backward until it reaches the same dNTP that was added, where it stops due to constant removal and addition. The insert is treated similarly, with a dNTP added that's complementary to the one used for the plasmid. So long as the "chewed-back" regions have been designed complementarily, ligation will occur without an enzyme. A flowchart was included to outline the details of LIC (Fig. 4.2).

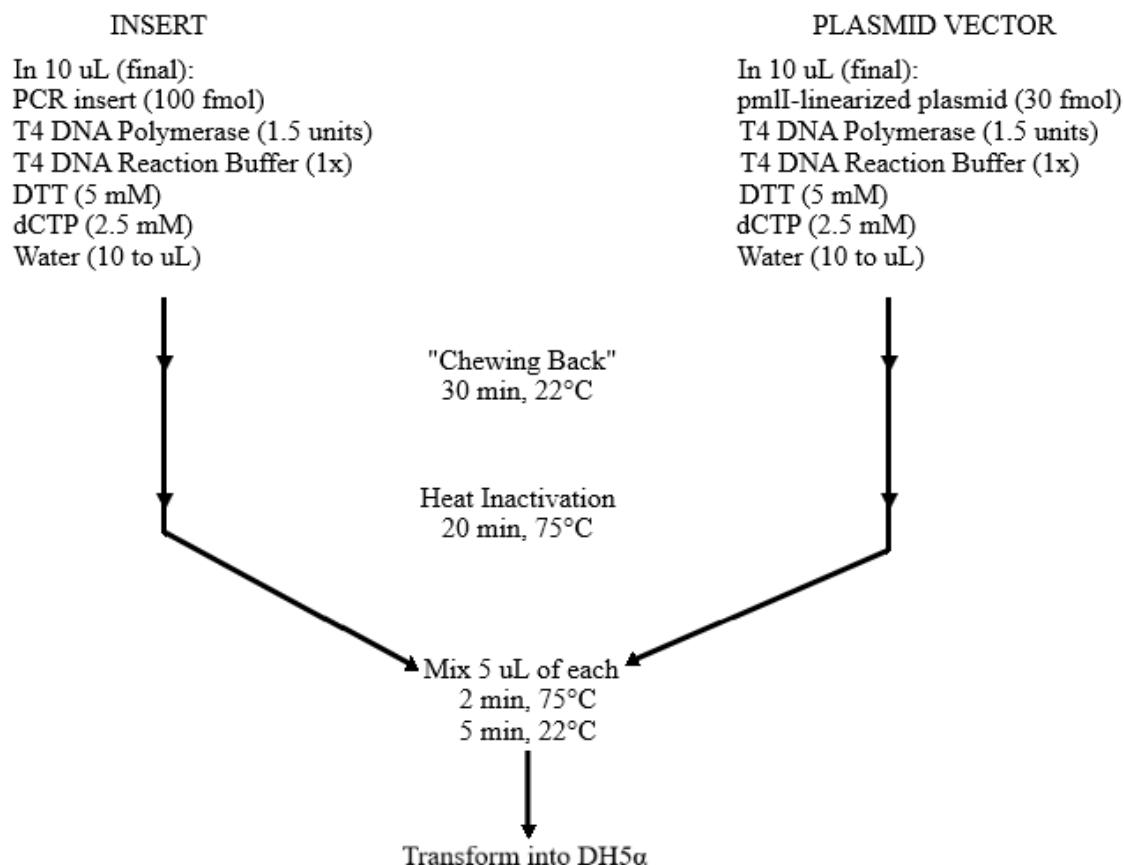


Figure 4.2. Overview of the reactions and incubations used in LIC to build the various *Cm* Prp3/4- and Prp31-containing constructs.

4.4.2. Glycerol Stocks of DH5α Transformed with Prp3+Prp4 and Prp31

5 uL of each LIC reaction was mixed (pQLinkH-Prp3+Prp4; pQLinkH-Prp31; pQLinkC-Prp31) with 50 uL thawed DH5α cells in 1.5 mL microtubes (Eppendorf) on ice for 25 min. The cells were heat-shocked at 42°C for 45 seconds followed by ice for 5 min. 300 uL L.B. was added to each tube and the cells recovered for 20 min in a shaker (New Brunswick Scientific, model Innova 40) set at 37°C and 300 RPM. The cells were then centrifuged for 10 seconds, resuspended in 100 uL fresh L.B. and plated on L.B. agar supplemented with 20 ug/mL ampicillin (IBI Scientific) followed by an 18 hour incubation at 37°C.

Several colonies from each plate (DH5 α with either Prp3 and Prp4, or Prp31) were selected to inoculate 2 mL L.B. broth aliquots (20 ug/mL ampicillin) that were placed in a shaker at 37°C and 300 RPM for 16 hours. Their plasmids were purified using the EZNA Mini Kit (Omega) and restriction enzyme digests (EcoNI and SapI, NEB) were performed on the plasmids for 1 hour at 37°C. Each sample was injected into a lane of a 1% agarose gel and ran at 120 V for 105 min. All 3 constructs were successfully created and their corresponding DH5 α transformants were stored as glycerol stocks. 1 mL of each overnight broth was mixed with 80 uL DMSO (Sigma Aldrich) into 2 mL microtubes (Corning) and stored at -80°C.

4.4.3. Transformation of Rosetta pLysS Cells with Prp3, Prp4, And Prp31

3x 50 uL of Rosetta pLysS (DE3) cells were thawed on ice in 1.5 mL Eppendorf tubes (Fisher Scientific) for 15 min before adding 150 ng of either pQLinkH-Ppr3/4, pQLinkH-Prp31, or pQLinkC-Prp31. The tubes sat on ice for 30 min before being heat-shocked for 45 seconds at 42°C and returned them to ice for another 5 min. 250 uL of L.B. broth (Fischer Scientific) was added and the cells were placed in a shaker at 37°C and 300 RPM for 30 min. The cells were centrifuged (Eppendorf, model Centrifuge 5417 C) for 10 seconds then resuspended in 100 uL fresh L.B. broth and plated on L.B. agar supplemented with 20 ug/mL ampicillin (IBI Scientific) and chloramphenicol (Sigma-Aldrich) antibiotics. The plates were incubated at 37°C for 18 hours and each produced between 30 and 100 transformants.

4.4.4. IPTG-Induction of Prp3, Prp4, And Prp31 with Rosetta pLysS Cells

One transformant colony from each plate (Rosetta pLysS cells transformed with pQLinkH-Ppr3/4, pQLinkH-Prp31, or pQLinkC-Prp31) was used to inoculate 10 mL

L.B. broths supplemented with 20 ug/mL each of ampicillin and chloramphenicol. Each culture was put in a shaker for 18 hours at 37°C and 300 RPM. 2mL of each overnight broth was added to 3x 20 mL L.B. broths (20 ug/mL ampicillin and chloramphenicol) for small scale expression (9 cultures in total). All of the cultures were placed in a shaker for about 3.7 hours at 300 RPM and 37°C until they reached an OD₆₀₀ of approximately 0.5 (Beckman Coulter, model DU 800). All of the cultures were induced with 1 mM Isopropyl-β-D-1-thiogalactopyranoside (IPTG, VWR) before being split-up for induction under 3 different temperatures: 16°C, 25°C, and 37°C (for 18 hours, 12 hours, and 4 hours respectively). After induction under each condition was complete, 15 mL of each culture was centrifuged for 30 min at 3200 RPM and 4°C (Beckman Coulter, model Avanti HP-20 XPI) and the supernatant discarded. All cell pellets were flash-frozen in liquid nitrogen then stored at -80°C. Both before and after induction, 1 mL of each culture was taken and resuspended in SDS loading buffer to a final concentration of 1 OD unit per 100 uL (A600) before being storage at -20°C. Expression was checked via 12% polyacrylamide (37.5:1 acrylamide/Bis, Bio-Rad) gel via sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Each sample at was heated at 95°C for 5 minutes then centrifuged at 20000 Xg for 3 minutes. 10 uL of each sample was injected into the gel ran at 22 mA for 50-60 minutes.

4.4.5. Auto- Induction of Prp3, Prp4, And Prp31 with Rosetta pLysS Cells

Auto-induction requires two types of media: a non-inducing media for growth and short-term storage, as well as an inducing media for expression (Studier, 2005). MDG and ZYM-5052 are the non-inducing and auto-inducing media, respectively. For convenience, 3 stocks were made ahead of time: 50x M, 50x 5052, and 1000x Trace

Metals. The final concentrations used in the growth and expression of transformed Rosetta cells are shown below (Table 4.6).

Table 4.6. Components of the MDG and ZYM-5052 media used in Auto-Induction.

MDG (non-inducing) components	ZYM-5052 (auto-inducing) components
2 mM MgSO ₄	2 mM MgSO ₄
0.5% w/v Glucose (Sigma Aldrich)	10 mg/mL Tryptone (Fischer Scientific)
0.25% w/v L-Aspartate (Sigma Aldrich)	5 mg/mL Yeast Extract (Fischer Scientific)
1x M	1x M
25 mM Na ₂ HPO ₄	25 mM Na ₂ HPO ₄
25 mM KH ₂ PO ₄	25 mM KH ₂ PO ₄
50 mM NH ₄ Cl	50 mM NH ₄ Cl
5 mM Na ₂ SO ₄	5 mM Na ₂ SO ₄
1x Trace Metals	1x 5052
10 uM FeCl ₃	0.5% v/v Glycerol
4 uM CaCl ₂	0.05% w/v Glucose
2 uM MnCl ₂	0.2% w/v Lactose
2 uM ZnSO ₄	
0.4 uM CoCl ₂	
0.4 uM CuCl ₂	
0.4 uM NiCl ₂	
0.4 uM Na ₂ MoO ₄	
0.4 uM Na ₂ SeO ₃	
0.4 uM H ₃ BO ₃	

Prior to auto-induction of *Cm* Prp3 and Prp4, Rosetta pLysS (DE3) cells were transformed with pQLinkH-Prp3/4 as described above but plated on MDG agar supplemented with 20 ug/mL carbenicillin (IBI Scientific) and chloramphenicol. The plates were incubated at 37°C for 24 hours, as transformants were slower to grow. One transformant colony was selected from each plate to inoculate 2 mL MDG broth supplemented with 20 ug/mL each of carbenicillin and chloramphenicol. The cultures

were placed in a shaker for 18 hours at 37°C and 300 RPM. Both the MDG plate and culture are able to be stored at 4°C for several weeks (Studier 2005). The overnight MDG culture was used to inoculate 2x 25 mL cultures to a starting OD₆₀₀ of 0.01 (in 250 mL Erlenmeyer flasks for aeration). One culture was placed at 16°C for 48 hours and the other at 37°C for 24 hours, both in shakers at 300 RPM. Samples were taken before and after induction and analyzed via SDS-PAGE.

4.5 Methods – Protein Disorder Predictions

4.5.1 *In silico* Modeling of Intrinsic Disorder

Amino acid sequences for Snu13, Prp3, Prp4, and Prp31 were taken from NCBI and UniProt databases (Table 4.7). Each protein sequence was uploaded to the server DisEMBL Intrinsic Protein Disorder Prediction 1.5 (Linding *et al.*, 2003; Iakoucheva and Dunker, 2003).

Table 4.7. NCBI and UniProt accession numbers of U4/U6 di-snRNP proteins.

Protein	NCBI Accession Number	UniProt Identifier
Snu13		CMP335C
Prp3		CMT170C
Prp4	XP_005538590.1	
Prp31	XP_005539062.1	

4.6 Results and Discussion

4.6.1. Preparation of di-snRNP Components

In the human and yeast models of the spliceosome, the U4 and U6 small nuclear RNA (snRNA) base-pair to form the U4/U6 di-snRNA (Bringmann *et al.*, 1984; Hashimoto and Steitz, 1984). The association of related proteins to the RNA create the U4/U6 di-small nuclear ribonucleoprotein complex (di-snRNP). The predicted secondary structure of the *Cyanidioschyzon merolae* U4/U6 di-snRNA revealed 3 regions of base pairing, denoted by Stems I-III (Fig. 4.3; Stark *et al.*, 2015). Importantly, the canonical

binding regions of the U4/U6 associated proteins are conserved which suggests a similar model for di-snRNP assembly as compared to human and yeast (Nottrott *et al.*, 2002; Hardin *et al.*, 2015). The predicted secondary structure of the yeast U4/U6 di-snRNP is based on known interactions between U4/U6 associated proteins and RNA (Fig. 4.3; Hardin *et al.*, 2015). There are *Cm* homologs present for every U4/U6-related protein; conservation would suggest this complex is important for splicing (Fig. 4.3).

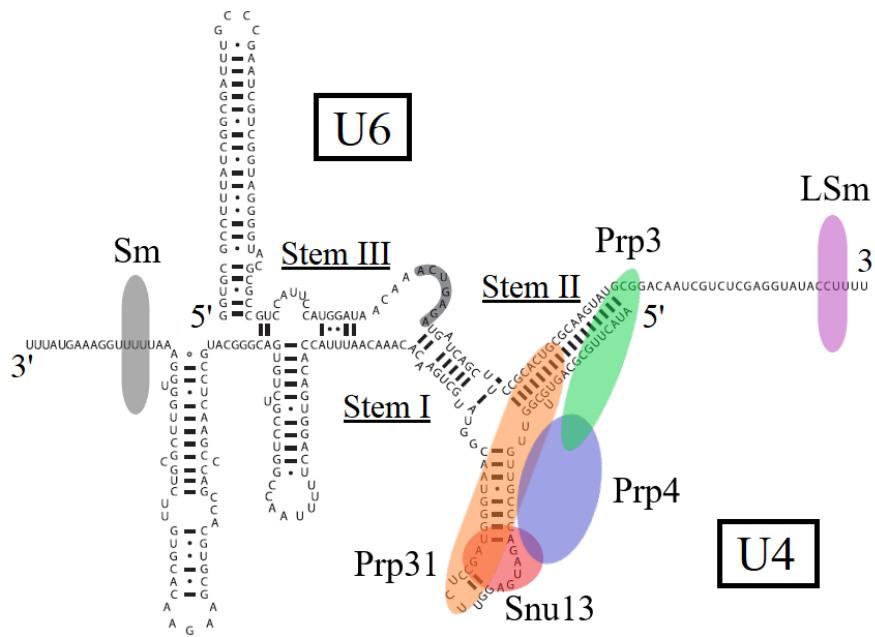


Figure 4.3. The *Cyanidioschyzon merolae* U4/U6 di-snRNP, based on human and yeast modeling. Placement of the proteins was guided by a cryo-EM yeast structure (Nguyen *et al.*, 2015) and the RNA secondary structure was adapted from Stark *et al.*, 2015.

The assembly pathway of the U4/U6 di-snRNP was previously determined through a series of gel shifts (Nottrott *et al.*, 2002; Hardin *et al.*, 2015; Liu *et al.*, 2015; Theuser *et al.*, 2016). The Sm and LSm proteins bind the 3' ends of U4 and U6, respectively, which likely occurs before RNA base pairing (Karaduman *et al.*, 2006). Snu13 binds the U4 5' stem loop and acts as a nucleating factor for the assembly of

remaining proteins Prp3, Prp4, and Prp31 (Nottrott *et al.*, 2002). The Prp3/4 dimer and Prp31 both make contact with Snu13, the U4 5' stem loop, and Stem II (Nguyen *et al.*, 2015; Agafonov *et al.*, 2016). In yeast and likely humans, Prp3/4 and Prp31 bind independently of each other (Hardin *et al.*, 2015). Here, assembly of the *Cm* U4/U6 di-snRNP was similarly demonstrated by EMSAs (electrophoretic mobility shift assays). Base-paired U4/U6 snRNAs would migrate slower than either of the two alone. Bound proteins would cause the RNA to migrate slower still. Of the U4/U6 associated proteins, only Snu13 and the two heteroheptameric rings (LSm and Sm) were available due to complications with expression and purification. The available proteins were checked on a polyacrylamide gel prior to use in gel shift assays (Fig. 4.4). Snu13 and the two heteroheptameric rings (LSm and Sm) were about 90%, 60%, and 80% pure respectively (Fig. 4.4).

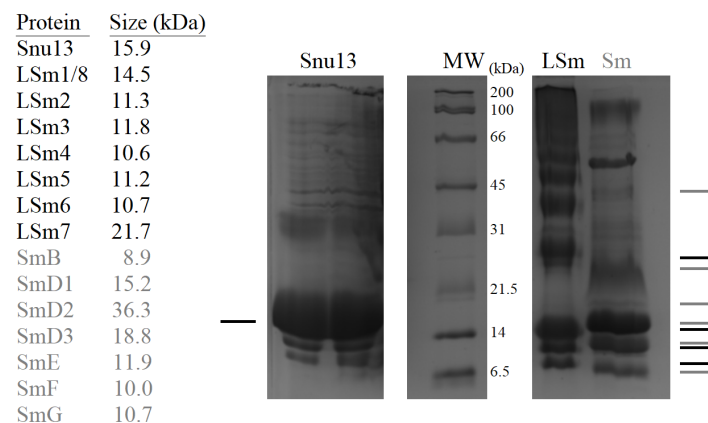


Figure 4.4. SDS-PAGE analysis of purified *Cm* Snu13, LSm, and Sm proteins for use in the assembly assays. The black bars show the expected sizes of Snu13 (left) and the LSm proteins (right), while the grey bars show the expected sizes of the Sm proteins (right).

All proteins are recombinantly expressed and purified *C. merolae* proteins.

4.6.2. U4 and U6 snRNA Base-Pairing Assay

In vitro transcribed *Cm* U4 and U6 snRNAs were mixed together at two different temperatures to determine base-pairing efficiency (Fig. 4.5 A). Base pairing between U4

and U6 was confirmed by the presence of bands that ran slower (lanes 3-4) than those of either snRNA alone (lanes 1-2, Fig. 4.5 A). Heating to 65°C and allowing a 1 hour cool-down yielded 71% of base-paired U4 as compared to only 17% if the snRNAs were left on ice for the hour (Fig. 4.5 B).

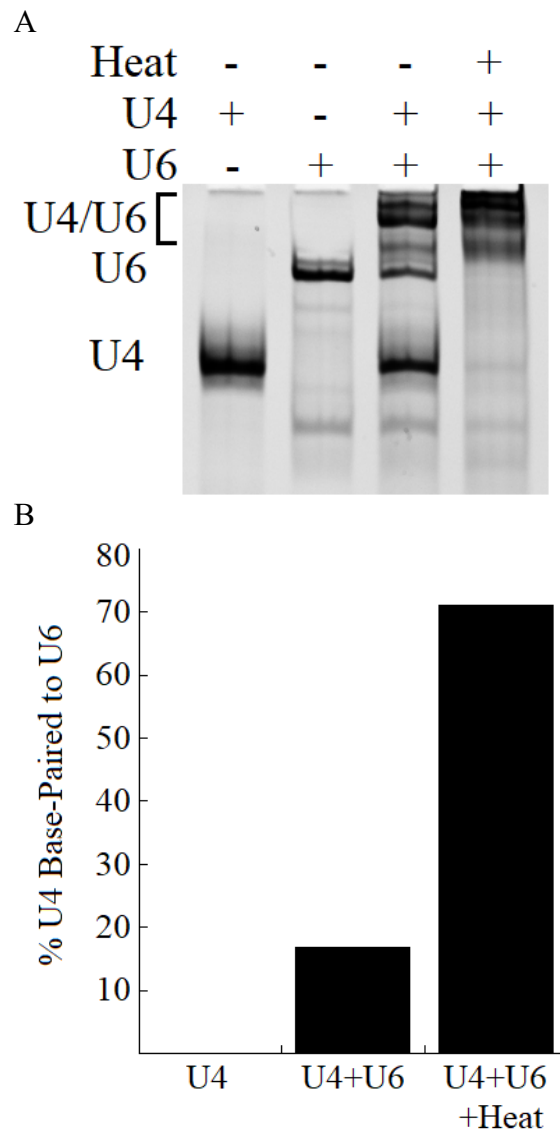


Figure 4.5. U4 and U6 RNA annealing assay. (A) Electrophoretic mobility shift assay of *C. merolae* U4 and U6 snRNA. (B) Bar graph representing the amount of U4 base-paired to U6, expressed as percentage.

4.6.3. U4/U6 di-snRNP Assembly Assay

The first U4/U6 di-snRNP assembly assay was done using U4, U6, and Snu13. Snu13 was selected because EMSAs and single molecule FRET (fluorescence resonance energy transfer) previously showed that Snu13 was required to interact with U4/U6 before either Prp3/4 or Prp31 were able to bind (Nottrott *et al.*, 2002; Karaduman *et al.*, 2006; Hardin *et al.*, 2015). All combinations of RNA and protein were mixed with and without pre-heating of RNA. The first 6 lanes re-confirmed previous findings that heat improved base-pairing efficiency (Fig. 4.6). Base-paired U4/U6 produced several bands, the highest of which was slightly above the band corresponding to Snu13 and either U4 or U6 (lanes 3, 6, 7-8, 10-11, Fig. 4.6). The highest band present was produced when Snu13 was mixed with base-paired U4/U6 (lanes 9, 12, Fig. 4.6). This assay showed that Snu13 was able to interact with 3 different RNA species: U4, U6, and U4/U6. Given that the U4 5' stem loop would be accessible regardless of whether or not U4 is base-paired to U6, the affinity of Snu13 should be very similar for both U4 and U4/U6 (Fig. 4.3). Snu13 was in 10-fold excess to U4 and U6 snRNAs because the intention of this assay was to show U4, U6, and Snu13 were able to assemble into a single complex. While Snu13 shifted both U4 and U4/U6 (lanes 7, 9, 10, 12), Snu13 also shifted U6 alone (lanes 8 and 11, Fig. 4.6). The Snu13-U6 interaction is likely non-specific and a result of the molar excess of protein to RNA. Optimization of this experiment would include stoichiometric ratios of protein to RNA, which should eliminate non-specific interactions such as Snu13-U6.

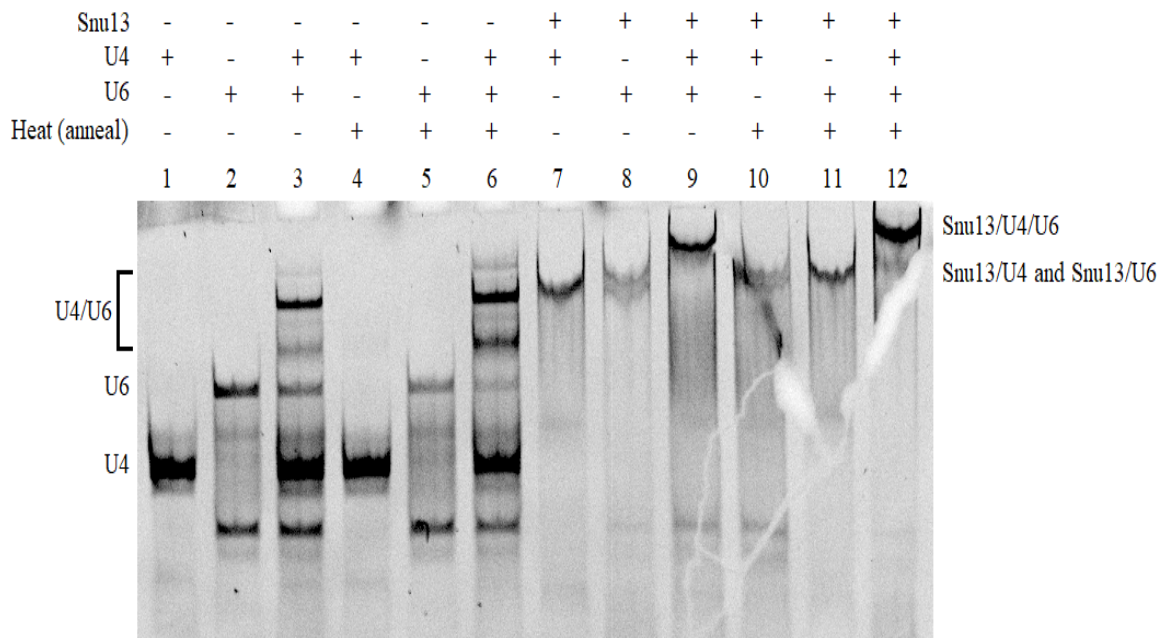


Figure 4.6. U4 and U6 *Cm* di-snRNP assembly assay with Snu13 done by electromagnetic mobility shift with 9% polyacrylamide gels.

Next, assembly of Snu13, Sms, and LSms onto the U4/U6 di-snRNA was observed (Fig. 4.7). This experiment was successfully repeated twice (5 attempts were made) and unfortunately the RNA bands without protein were difficult to distinguish from those with protein. The first 3 lanes show snRNA alone and base-paired (lanes 1-3, Fig. 4.7). Snu13 and the Sm proteins both produced gel shifts that confirmed those proteins were able to bind the di-snRNA (lanes 4 and 6, Fig. 4.7). Unfortunately, the LSm proteins did not seem to interact with U4/U6 (lane 5, Fig. 4.7). It was later discovered that the pI of the LSms was very high, and it is possible that the basic protein complex migrated upward. If the positively charged LSms had interacted with the RNA and migrated upward, there should not be bands visible. In comparing lanes 3 and 5, the RNA bands appear to be similar (Fig. 4.7). This suggests that the LSms did not interact with the RNA. Overall, native gel electrophoresis has shown that: i) U4 and U6 are able

to base-pair; ii) in excess, Snu13 is able to interact with either snRNA alone; iii) Snu13 and the Sm proteins are each able to assemble onto the U4/U6 di-snRNA.

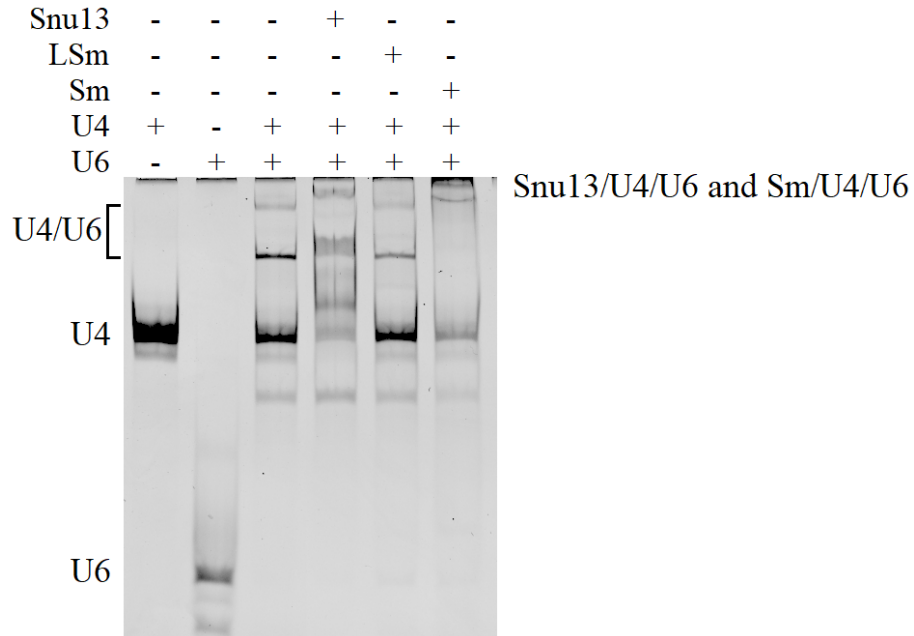
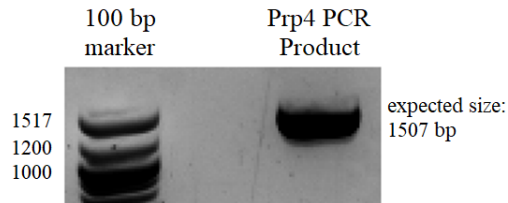


Figure 4.7. U4 and U6 di-snRNP assembly assay with Snu13, LSm, and Sm proteins. Electromagnetic mobility shift assay for assembly of the *C. merolae* di-snRNP was done with a 9% polyacrylamide gel.

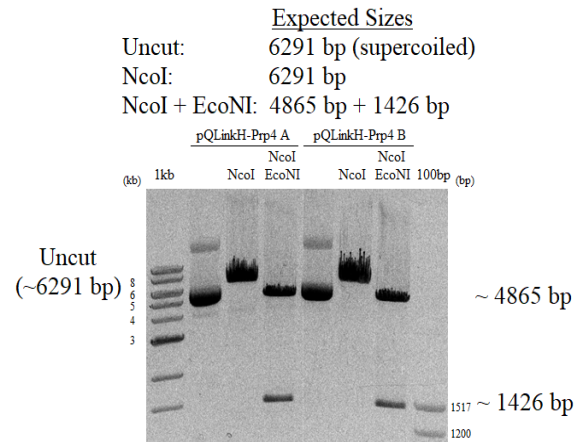
As mentioned above, availability of di-snRNP proteins was limited to Snu13, Sms, and LSms because Prp3, Prp4, and Prp31 could not be expressed and purified. In fact, during the U4/U6 EMSA experiments, only Prp3 was known to exist. Expression and purification attempts of Prp3 yielded small quantities of protein (data not shown). Prp3 was expressed as a fusion protein with Maltose Binding Protein (MBP), a well-studied protein useful for both affinity purification as well as improving solubility of its fusion partner (Sachdev and Chirgwin, 2000). After purification over both nickel- and maltose-bound resin, cleavage of the MBP tag revealed Prp3 was not very soluble: over half of the protein was lost as a precipitant (data not shown). Several months later, Dr. Martha Stark discovered Prp4 and Prp31 when she performed additional pull-down

assays and mass spectrometry analyses (XP_005538590.1 and XP_005539062.1, respectively). Prp31 was subcloned into two separate constructs: pQLinkH and pQLinkC, featuring N- and C-terminal TEV-cleavable 6x His-tags. Prp4 was subcloned into pQLinkH, and Prp3 was later added to the same plasmid for co-expression of tagged Prp4 and untagged Prp3. While sub-cloning of Prp31 was straightforward, ligase-independent cloning (LIC) was employed for Prp3 and Prp4 (Fig. 4.1 and 4.2). Agarose gel electrophoresis revealed successful amplification of the Prp4 gene from the *Cm* genome (Fig. 4.8 A). Using LIC, which utilizes long complementary overhangs to replace the need for ligase, Prp4 was cloned into pQLinkH and verified by EcoNI and NcoI restriction enzyme (RE) digests (Fig. 4.8 B). Lanes 2-3 and 5-6 showed the supercoiled and NcoI-linearized plasmid (Fig 4.8 B). If Prp4 was successfully inserted into the plasmid, dual RE digest would have produced two double-stranded (base-paired, bp) fragments of 4865 and 1426 bp in size. RE digest and DNA sequencing confirmed that the gene was properly inserted (lanes 4 and 7, Fig 4.8 B). As explained in detail (section 4.4.1), LIC allows for unrestricted addition of genes into a single plasmid using the SwaI and PacI restriction enzymes (Scheich *et al.*, 2007). Untagged Prp3 was excised from pQLinkN with PacI and added to SwaI-linearized pQLinkH-Prp4. Since the large plasmid contained multiple cut sites for each RE, EcoNI and SapI RE digests were individually performed to verify both genes in pQLinkH (Fig. 4.8 C). Both digests produced 2 dsDNA fragments as expected, with some uncut plasmid left in the SapI-digested lane (lanes 3-4, Fig. 4.8 C).

A



B



C

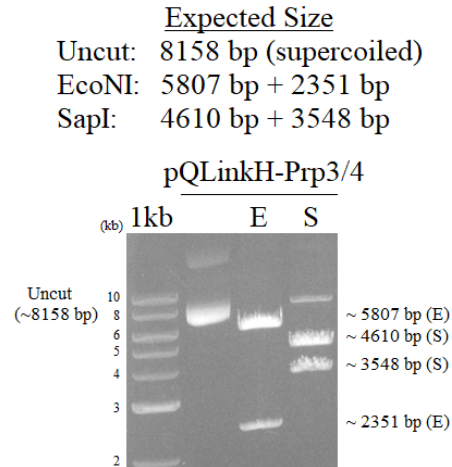


Figure 4.8. Cloning of Prp4 and Prp3 into pQLinkH. (A) Amplification of *Cm* Prp4 gene by PCR. (B) Restriction digest check of *Cm* Prp4 insertion into pQLinkH. (C) Restriction digest check of *Cm* Prp3 insertion into pQLinkH-Prp4.

N-terminally His-tagged Prp4 was available both on its own and in the same plasmid as Prp3. This was important for a control to ensure both proteins could be visualized by SDS-PAGE (Fig 4.9). Prp4 was clearly expressed when Prp4- and Prp3/Prp4-transformed *E. coli* cells were induced with IPTG at 37°C (lanes 3 and 5, Fig. 4.9). While both proteins are similar in size, there was no difference between lanes 3 and 5, suggesting that Prp3 was either not expressed or expressed poorly (Fig. 4.9).

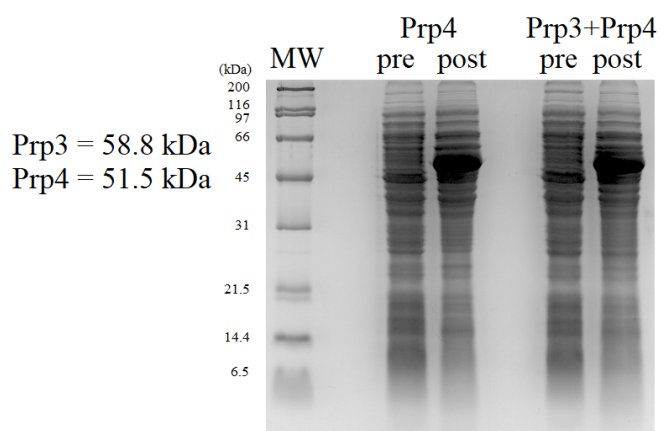


Figure 4.9. SDS-PAGE analysis of *Cm* Prp3 and Prp4 co-expression through IPTG induction at 37°C.

Using both IPTG as well as auto-induction (sections 4.4.4 and 4.4.5), several co-expression trials were done at 16°C, 25°C, and 37°C (Fig. 4.10). A single band above the 45 kDa marker appeared after co-expression for all 3 temperatures, but whether or not it contained both Prp3 and Prp4 was unknown (lanes 3-4, 6, 8, Fig. 4.10). Auto-induction worked at 16°C but remained ambiguous for 37°C (lanes 4, 9, Fig. 4.10). Co-expression attempts of Prp3/4 thus far have only confirmed expression of Prp4. Restriction enzyme digest and DNA sequencing confirmed that both genes were present and properly oriented (Fig. 4.8 C). Both genes contained a T7 promoter sequence, but transcription efficiency is nonetheless lower for downstream genes (Kim *et al.*, 2004).

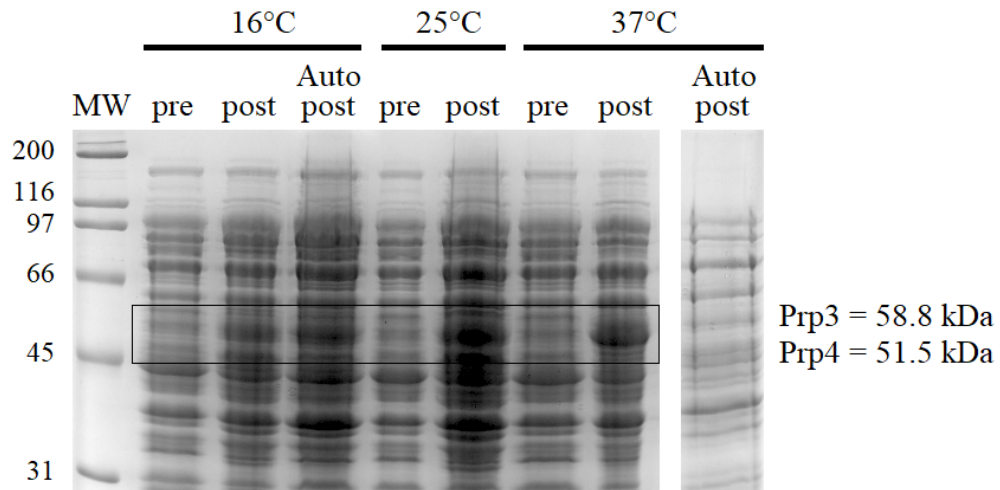


Figure 4.10. SDS-PAGE analysis of *Cm* Prp3 and Prp4 co-expression through IPTG and auto-induction at various temperatures.

N- and C-terminal 6x His-tag versions of Prp31 were cloned. Both constructs were transformed into *E. coli* and induced with IPTG at 37°C (Fig. 4.11). N-terminally tagged Prp31 appeared to weakly express at 37°C (lanes 2-3), but no convincing bands appeared for the C-terminal tagged version (lanes 4-5, Fig. 4.11).

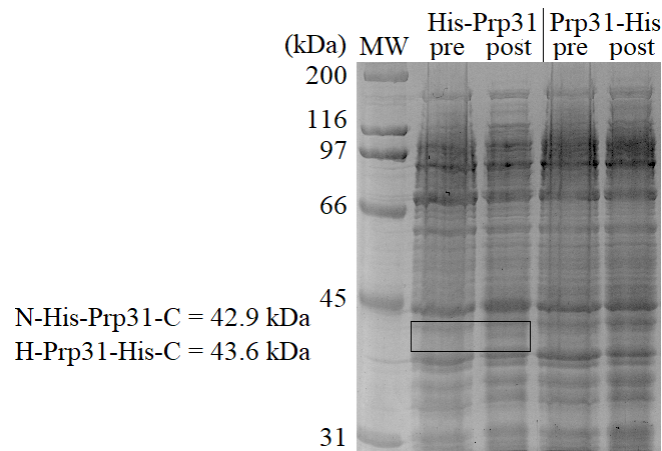


Figure 4.11. SDS-PAGE analysis of *Cm* Prp31 expression through IPTG induction at 37°C.

Further attempts were made to check the expression of Prp31 at 16°C, 25°C, and 37°C (Fig. 4.11). Unfortunately there was no evidence by SDS-PAGE that Prp31 was successfully induced at any temperature using IPTG (Fig. 4.12).

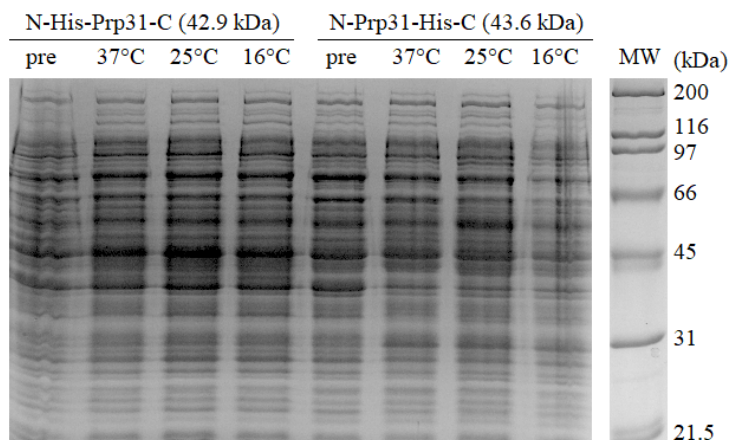


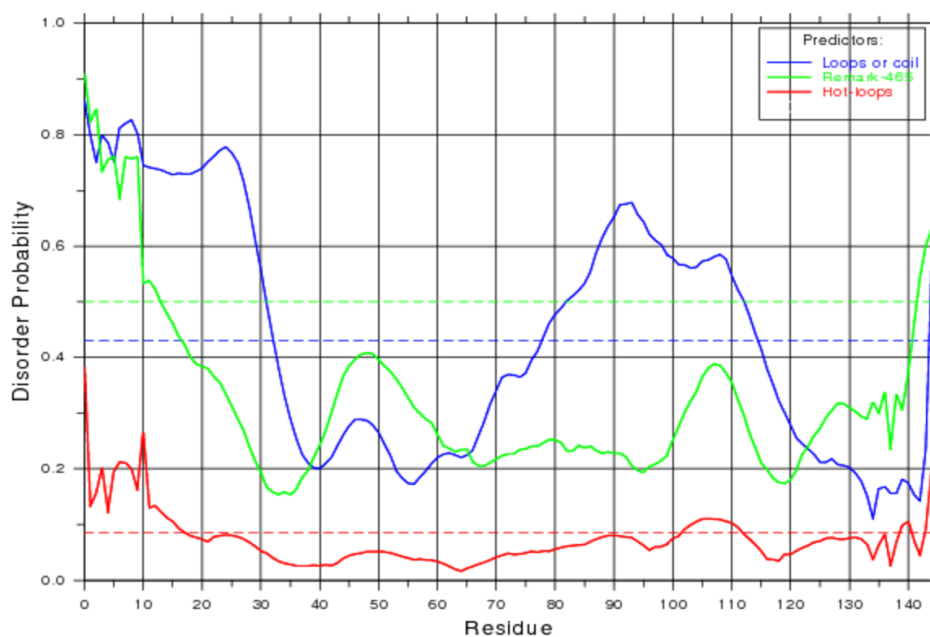
Figure 4.12. SDS-PAGE analysis of *Cm* Prp31 expression through IPTG induction at various temperatures.

4.6.4. Prediction of Intrinsically Disordered Regions

X-ray crystal structures exist for the complete sequences of Snu13 (PDB ID 5EWR), LSm (4M77), and Sm (3CW1) proteins. X-ray crystallography demands that the protein be highly structured so that the electron density is high enough for mapping (Dunker *et al.*, 2001). Indeed, a protein that is highly dynamic or flexible would be difficult to even crystallize, as crystallization occurs when a homologous sample creates a uniform lattice. Crystal structures of Prp3, Prp4, and Prp31 are all incomplete; the only full-length structures of these proteins have been obtained through cryogenic electron microscopy (Nguyen *et al.*, 2015; Agafonov *et al.*, 2016). The inability of these proteins to readily crystallize suggests there are flexible regions that allow for flip-flopping in solution. A region that is flexible is often defined as being intrinsically disordered (Iakoucheva and Dunker, 2003). Intrinsically disordered proteins (IDPs) are regularly

difficult to recombinantly express and purify (Dunker *et al.*, 2001). In an effort to determine why Prp3, Prp4, and Prp31 were difficult to work with, *in silico* mapping of intrinsic disorder was performed. Using the DisEMBL Intrinsic Protein Disorder Prediction 1.5 server, intrinsically disordered regions were first predicted for Snu13 (Fig. 4.13).

DisEMBL 1.5 Prediction Results for Snu13



Disordered by Loops/coils definition

```
>Snu13_LOOPS 1-33
MEPLSSTEAP SRPMDVVtAP NQADPRAYPF APAdlvveil dlvqqashyk qikkglnevl ksmnrglaef vvlaadtqpl eillsaplv
edkavpyvfv pskaalgrac gvsrpviaca vlradsqrlr nqitalrtki eqlll
```

Disordered by Hot-loops definition

```
>Snu13_HOTLOOPS 1-17
MEPLSSTEAP SRPMDVVtAP nqadpraypf apadlvveil dlvqqashyk qikkglnevl ksmnrglaef vvlaadtqpl eillsaplv
edkavpyvfv pskaalgrac gvsrpviaca vlradsqrlr nqitalrtki eqlll
```

Disordered by Remark-465 definition

```
>Snu13_REM465 1-13
MEPLSSTEAP SRPMDVVtAP nqadpraypf apadlvveil dlvqqashyk qikkglnevl ksmnrglaef vvlaadtqpl eillsaplv
edkavpyvfv pskaalgrac gvsrpviaca vlradsqrlr nqitalrtki eqlll
```

Figure 4.13. Output for Predicted Intrinsic Disorder of the DisEMBL Intrinsic Protein Disorder Prediction 1.5 server. Example output for *Cm* Snu13. Loops/Coils are blue, Hot-Loops are red, and Remark-465 is green. Loops/Coils is defined by the “Dictionary of Secondary Structure of Proteins” and each residue is assigned a secondary structure value (alpha helix, 3-10, beta strand, or none). Hot-Loops are a subset of Loops/Coils, whereby mobility is predicted based upon C-alpha temperature B-factors. Remark-465 searches for missing coordinates in submitted X-Ray structures submitted in the RCSB PDB.

Snu13 was used as a control to adjust the parameters of the algorithm, as its crystal structure was previously characterized (Fig. 4.14 A; Black *et al.*, 2016). By altering parameters for predictions according to Loops/Coils, Hot-Loops, and Remark-465, Loops/Coils produced the most reliable prediction when compared to the crystal structure of Snu13 (Fig. 4.14 A). Loops/Coils is defined by the “Dictionary of Secondary Structure of Proteins” and each residue is assigned a secondary structure value: alpha helix, 3-10, beta strand, or none (Dunker *et al.*, 2001; Iakoucheva and Dunker, 2003). Hot-Loops are a subset of Loops/Coils, whereby mobility is predicted based upon C-alpha temperature B-factors. Remark-465 searches for missing coordinates in submitted X-Ray structures submitted in the RCSB PDB.

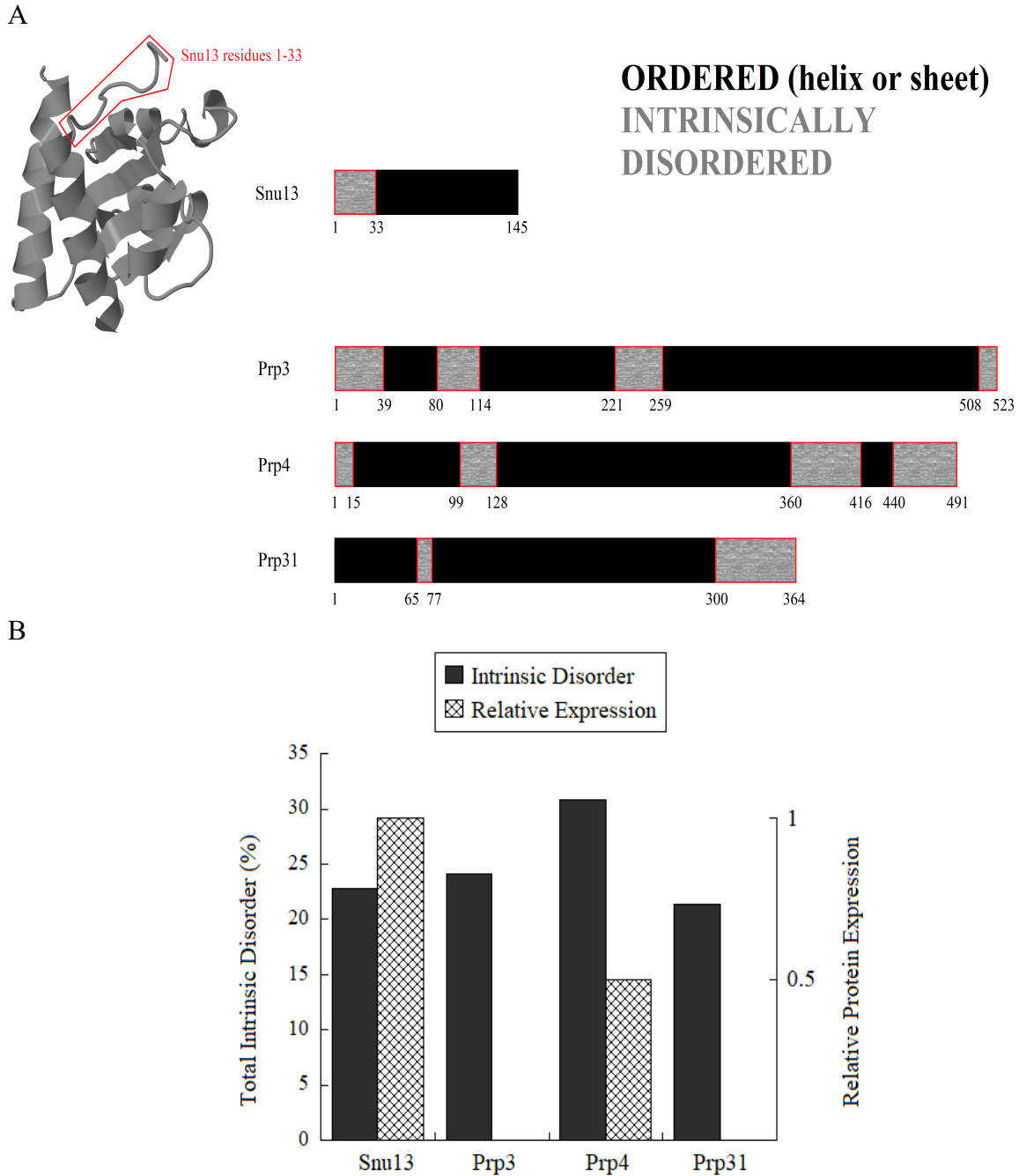


Figure 4.14. Map of predicted intrinsically disordered regions within the U4/U6 di-snRNP proteins. (A) Crystal structure of *C. merolae* Snu13 with the first 33 amino acids highlighted to show the lack of electron density (PDB ID 5EWR). (B) Bar graph of predicted disorder (left) and recombinant protein expression (right) for *C. merolae* proteins Snu13, Prp3, Prp4, and Prp31. The left vertical axis describes the predicted intrinsic disorder, denoted by black bars. The right vertical axis describes the protein expression relative to Snu13 in *E. coli*. Analysis was done using the DisEMBL Intrinsic Protein Disorder Prediction 1.5 server.

Intrinsic disorder was then predicted for Prp3, Prp4, and Prp31 (Fig. 4.14 B). Total predicted intrinsic disorder was plotted against expression of the same protein (Fig. 4.14 B). Predicted disorder and protein expression were not good indicators for each other, as Snu13 and Prp4 were 22% and 30% disordered; yet both resulted in protein expression (Fig. 4.14 B). Prp3 and Prp31, similar to Snu13 in total predicted disorder, could not be expressed (Fig. 4.14 B). It might also be that the location of intrinsic disorder affected expression success, but Prp3 and Prp4 both have flexible regions on the N- and C-terminals, yet Prp3 was not visibly expressed (Fig. 4.14 A). Failure to express Prp3 and Prp31 could still be explained by loss of transcriptional efficiency and location of disordered regions (Dunker *et al.*, 2001; Kim *et al.*, 2004).

Chapter 5

General Discussion and Conclusion

5.1. Improving the yield of the Msl5/Mud2 dimer

The expression and purification of Msl5/Mud2 dimer was inefficient. 2 L of transformed Rosetta pLysS cells were induced with 1 mM IPTG for 14 hours at 28°C. The lysate was affinity purified via batch binding followed by size exclusion chromatography. This method produced 70 uL of *Cm*Mud2/Msl5 dimer at a concentration of 2000 nM. There are several steps that could be optimized. Overnight expression conditions should involve a lower temperature (i.e. 18°C), and auto-induction should also be attempted. The volume of culture should be increased and the affinity purification should be done using the FPLC system. Using an FPLC system for the immobilized metal ion affinity chromatography (IMAC) has advantages for purifying a protein complex, namely that it is more controlled, less disruptive, and allows for an elution gradient accompanied with UV sensors.

5.2. Role of the branchpoint bridging protein Msl5

Full length Msl5 binds the *Cm* BPS/pre-mRNA with similar affinity to reported values for homologous interactions (Tables 2.17, 2.18, and 2.20). Alignment of the protein sequence against human and yeast revealed that the only conserved domain present in *C. merolae* is KH-QUA2, which is necessary and sufficient for RNA binding activity in humans (Fig. 2.4 and 2.6; Berglund *et al.* 1998a; Liu *et al.* 2001). While RNA binding is localized to KH-QUA2, human SF1 requires U2AF65 (Mud2) for optimal binding and target specificity (Berglund *et al.* 1998b). KH-QUA2 comprises the $\beta 1$ - $\alpha 1$ - $\alpha 2$ - $\beta 2$ - $\beta 3$ - $\alpha 3$ - $\alpha 4$ topology adopted by the STAR family of proteins (Vernet and Artzt,

1997; Liu *et al.* 2001). The domain folds into a hydrophobic groove that binds the single-stranded BPS (Fig. 2.6 B; Liu *et al.*, 2001). Yeast BBP diverges from SF1 and other homologs as it requires the N-terminal Zinc knuckle for RNA binding activity, but not Mud2 (Garrey *et al.* ,2006). The additional Zinc knuckle may explain why BBP prefers a BPS with an upstream stem loop (Fig. 2.4; Berkowitz *et al.*, 1995). Since Msl5 lacks a Zinc knuckle motif, KH-QUA2 is expected to be necessary and sufficient for BPS binding activity in *C. merolae*. Overall, Msl5 is expected to be similar to human SF1 in that KH-QUA2 is necessary and sufficient for BPS selection, The RNA binding domain is located in the middle of Msl5 and is the only region of Msl5 predicted to have secondary structure (Fig. 2.4). It would be interesting to probe the function of the flanking regions by incorporating truncations or mutations. Perhaps the unstructured sequences interact with Mud2 and bypass the need for a canonical ULM (Fig. 2.5).

5.3. Bridge between Msl5 and the 5' splice site

The yeast two-hybrid assay revealed that yeast BBP contacts Mud2 and Prp40, while Prp40 contacts U1, BBP and Prp8 (Abovich and Rosbash, 1997). In this way, Prp40 links the 3' and 5' splice sites by interacting with BBP and U1 (Fig. 1.3). Prp40 is missing in *C. merolae*, along with all other U1 associated proteins (Stark *et al.* 2015). One possibility is that Mud2 replaces the role of Prp40. While there is no sequence similarity between the proteins, much of the Mud2 sequence is novel and likely disordered. Intrinsically disordered proteins are evolutionarily advantageous as they can accommodate several different protein interactions (Dunker *et al.*, 2001). Rsp31 or SRSF2 could also fill the role of Prp40, or interact with Mud2 to bridge the BPS and 5' splice site. Rsp31 and SRSF2 are serine-arginine (SR) rich splicing factors; SR proteins

are postulated to facilitate the 5' SS-BPS bridge (Krainer *et al.*, 1990; Kohtz *et al.*, 1994; Li and Manley, 2005). The SR proteins do not and would not require secondary structures similar to that of Prp40. Prp40 is characterized as having a WW domain: 40 amino acids containing several proline repeats and two tryptophan residues 20 residues apart (Bork and Sudol, 1994). The resulting fold produces a curved, triple-stranded antiparallel β -sheet (Macias *et al.*, 1996). Several motifs have been recognized as ligands for WW domains, including: PPxY, LPxY, phosphoserine, and phosphothreonine (Aragón *et al.*, 2012; Bruce *et al.* 2008; Lu *et al.*, 1999). *Cm* Msl5 lacks the PPxY/LPxY motif, suggesting a novel protein-protein interaction is responsible in bridging the BPS and 5' SS (Fig. 2.4 and 2.5). To verify a novel functional interaction of Msl5 or Mud2 against Rsp31 and SRSF2, several *in vitro* (gel shift, fluorescence polarization) or *in vivo* (yeast two hybrid, fluorescence microscopy) methods are available.

5.4. Role of the U2 Auxiliary Factor Mud2 in *C. merolae*

The interaction between Msl5 and the BPS in *Cm* was confirmed by both fluorescence polarization and EMSAs, but several questions are left unanswered (Fig. 2.8-2.9, 2.14-2.17). Do Mud2 and Msl5 recognize the BPS as a dimer, or does one of the proteins bind RNA first? What is the sequence recognized by Mud2, and is there still a preference for uridine despite the lack of a Py-tract or poly-U region? The 145 kDa Mud2 is also much larger than the 54 kDa human or 60 kDa yeast homologs, and rigorous BLAST searches failed to generate a single search result for the first 800 amino acids (Stark *et al.* 2015). It would be very interesting to study the function of the N-terminal half of Mud2, as it may have evolved to replace an otherwise essential protein that is missing in *C. merolae*. Structural studies of a complex of Msl5, Mud2, and pre-mRNA

using cryo-EM or X-ray crystallography would provide insight into the mechanism of BPS recognition. Cryo-EM is able to capture different conformational states of the likely dynamic and intrinsically disordered Mud2, whereas cross-linking and/or truncations would likely be required for X-ray crystallographic studies.

5.5. Displacement of Msl5/Mud2 from the pre-mRNA

Once the BPS/3' SS has been recognized by Msl5/Mud2, the proteins must be displaced in order for U2 snRNA to base-pair with the BPS/3' SS region and drive the Commitment Complex into the Pre-Spliceosome of the splicing reaction. Two DExD/H box proteins, Sub2 and Prp5, have been implicated in the displacement of Msl5/Mud2. Genetic experiments in yeast have shown that the function of Sub2 is not needed if the Mud2 gene is deleted (Kistler and Guthrie, 2001). Prp5 drives the Commitment Complex (Complex E, Fig. 1.1; Fig. 1.2) to the Pre-Spliceosome stage (Complex A, Fig. 1.1; Fig. 1.4) in an ATP-dependent manner (Perriman and Ares, 2000). Perhaps Sub2 displaces Mud2 in *C. merolae* as well, although the mechanism for Msl5 displacement remains unknown. With so many related diseases and a spliceosome that offers complexity, pre-mRNA splicing is a competitive field of study. Elucidating the mechanism of Msl5 displacement is arguably the most urgent step forward from this research. Once understood, it would unlock new information applicable to all eukaryotes and possibly offer a pharmaceutical target for a splicing-related disease.

5.6. Prp4 connects Snu13 and the 5' Stem Loop with Stem II

Binding assays and X-ray crystallography have revealed conservation of the function and structure of Snu13. The crystal structure of Snu13 (5EWR) is the first reported structure of a *Cyanidioschyzon merolae* splicing protein (Fig. 3.5). The next step

of this project would be to confirm the Snu13-Prp4 interaction *in vitro*. It has long been known that the Snu13-Prp3/Prp4 interaction connected the U4 5' Stem Loop with Stem II of the di-snRNP (Nottrott *et al.*, 2002). The interaction between Snu13 and Prp4 specifically was suggested when high-resolution cryo-EM structures showed that they were more proximal to each other than Snu13 and Prp3 (Nguyen *et al.*, 2015; Agafonov *et al.*, 2016). A biochemical assay has yet to support the protein-protein interaction and offers an opportunity to validate the canonical di-snRNP architecture using *C. merolae*.

5.7. Assembly of the U4/U6 di-snRNP

Snu13 and Sm proteins were both able to assemble onto base-paired U4/U6 di-snRNA (Fig. 4.5 and 4.6). Thus far, U4/U6 di-snRNP assembly in *C. merolae* is similar to human and yeast models, in that Snu13 and Sm proteins require only U4 or base-paired U4/U6 (Nottrott *et al.*, 2002; Hardin *et al.*, 2015). The LSm complex did not appear to interact with the U4/U6 di-snRNA, as the lane with U4/U6 appeared to be the same as U4/U6 mixed with the LSms (Fig. 4.6 B). It is possible that the recombinant LSm complex was denatured or misfolded. Furthermore, the pI of the LSm complex (>10) gave it a global positive charge; perhaps the pI affected how the sample migrated. To address the possibility of a denatured/misfolded complex, freshly purified recombinant LSms should be centrifuged prior to use in the EMSA. To address the possibility that the positively charged LSms ran upward, EMSAs should be performed with the electrodes reversed.

5.8. Chaperones and solubility factors of thermophilic organisms

Prp3 and Prp31 were not successfully expressed. In an effort to determine the cause of poor expression, an *in silico* analysis of intrinsically disordered regions was

done. There was no obvious link between predicted disorder and expression (Fig. 4.13). Perhaps the failed expression was not due to intrinsic disorder or transcriptional inefficiency; the problem could be the thermophilic nature of *Cyanidioschyzon merolae* (Sternier and Liebl, 2001). Thermophilic organisms have adapted to survive harsh conditions. Protein synthesis is a costly event and the thermophile would have likely evolved a strategy to maintain protein levels as efficiently as possible at high temperatures. Translation and degradation is one way to regulate protein turnover. A less energetically costly method would be to have reduced synthesis coupled with repair pathways and solutes that help to keep proteins stable. It may also be that some genes from a thermophile such as *C. merolae* require chaperones (bacterial chaperones GroEL and GroES may help) or specific ions present during protein synthesis (Sternier and Liebl, 2001). The Rosetta pLysS *E. coli* strain was used throughout this thesis because it featured an additional plasmid with tRNA rare to *E. coli*, an eloquent reference to the Rosetta Stone and the ability to recognize more codons. Having rare codon tRNA should improve success with recombinant expression of eukaryotic proteins. These tRNA genes have been optimized for human expression, however, and may not be optimal for *Cm* proteins. One final thought that may be useful for future expression work is optimizing promoter strength and plasmid copy number for each specific protein/complex. Many of these *Cm* splicing factors are RNA binding proteins, which may well wreak havoc in a bacterial cell, resulting in toxicity and poor expression.

5.9. Intrinsically disordered regions and splicing proteins

BLAST searches and multiple sequence alignments suggested there were regions of Msl5 and Mud2 that lacked secondary structure or “order” (Fig. 2.4 and 2.19). CD

Spectroscopy of Msl5 revealed 20% of the protein was intrinsically disordered (Table 2.20). The crystal structure of Snu13 lacked electron density for the first 33 residues, indicative of a random coil (Fig. 3.4, 3.5, 4.13). Predictions of intrinsic disorder within di-snRNP proteins Snu13, Prp3, Prp4, and Prp31 ranged from 20-30% (Fig. 4.13). All of these results paint a picture that intrinsically disordered regions may be important for the function of *C. merolae* splicing proteins. Pre-mRNA splicing is an incredibly dynamic process where splicing factors are constantly rearranged or undergoing conformational changes. It would be evolutionarily advantageous for an organism with a reduced spliceosome to have intrinsically disordered proteins, as they are able to interact with multiple protein targets (Dunker *et al.*, 2001; Iakoucheva and Dunker, 2003).

5.10. Conclusion

Msl5 interacts with the BPS with an affinity of about 50 nM, which is similar to literature values for yeast BBP and human SF1/U2AF65. During the recognition of an intron-containing pre-mRNA, branchpoint sequence selection in *C. merolae* is likely driven by Msl5 and stabilized by Mud2. The large and unique Mud2 protein may have a novel role in *C. merolae* beyond facilitating BPS recognition. Structural and functional studies of Snu13 and Msl5 have revealed that some *Cm* splicing factors are highly conserved. Snu13 represents the first crystal structure of a *C. merolae* splicing protein to be reported (PDB ID 5EWR). While Snu13 is highly similar to human and yeast, so too is the U4 5' stem loop with the exception of a single nucleotide (G29A), which remains a purine. These results highlight the conservation and importance of the Snu13-U4 interaction during di-snRNP assembly. Several components of the U4/U6 di-snRNP, including U4, U6, Snu13, and the Sm proteins were able to assemble in a stepwise

manner. Di-snRNP proteins Prp3, Prp4, and Prp31 were not successfully expressed and purified, leading to an investigation of novel or unstructured regions of the protein. It was found that several di-snRNP proteins in *C. merolae* were extended or intrinsically disordered, suggesting that some splicing factors may be performing multiple roles. This work has demonstrated that *C. merolae* is an interesting organism in which to study pre-mRNA splicing. Conservation of a *Cm* structure or function underscores critically important aspects of splicing, while novelty reveals adaptations unique to the thermophilic red alga.

References

- Abovich, N., Liao, X.C., Rosbash, M. (1994). The yeast mud2 protein: an interaction with prp11 defines a bridge between commitment complexes and u2 snrnp addition. *Genes Dev.* **8**(7):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.
- Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Jung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C., Zwart, P.H. (2010). PHENIX: a comprehensive python-based system for macromolecular structure solution. *Acta Cryst.* **66**: 213-221.
- Agafonov, D.E., Kastner, B., Dybkov, O., Hofele, R.V., Liu, W.T., Urlaub, H., Lührmann, R., Stark, H. (2016). Molecular architecture of the human u4/u6.u5 tri-snrnp. *Science.* **351**(6280):1416-1420.
- Agrawal, A.A., Salsi, E., Chatrikhi, R., Henderson, S., Jenkins, J.L., Green, M.R., Ermolenko, D.N., Kielkopf, C.L. (2016). An extended u2af65-rna-binding domain recognizes the 3' splice site signal. *Nat. Commun.* **7**:10950.
- Aragón, E., Goerner, N., Qiaoran, Xi, Gomes, T., Gao, S., Massagué, J., Macias, M.J. (2013). Structural basis for the versatile interactions of smad7 with regulator ww domains in tgf- β pathways. *Structure.* **20**(10):1726-1736.
- Achsel, T., Brahms, H., Kastner, B., Bachi, A., Wilm, M., Lührmann, R. (1999). A doughnut-shaped heteromer of human sm-like proteins binds to the 3'-end of u6 snrna, thereby facilitating u4/u6 duplex formation *in vitro*. *EMBO J.* **18**:5789-5802.
- Baggett, N.E., Zhang, Y., Gross, C.A. (2017). Global analysis of translation termination in *E. coli*. *Plos Genet.* **13**(3):e1006676.
- Becerra, S., Andrés-León, E., Prieto-Sánchez, S., Hernández-Munain, C., Suñé, C. (2016). Prp40 and early events in splice site definition. *Wires RNA.* **7**:17-32.
- Berglund, J.A., Chua, K., Abovich, N., Reed, R., Rosbash, M. (1997). The splicing factor bbp interacts specifically with the pre-mrna branchpoint sequence uacuaac. *Cell.* **89**:781-787.
- Berglund, J.A., Fleming, M.L., Rosbash, M. (1998a). The kh domain of the branchpoint sequence binding protein determines specificity for the pre-mrna branchpoint sequence. *RNA.* **4**:998-1006.

- Berglund, J.A., Abovich, N., Rosbash, M. (1998b). A cooperative interaction between u2af65 and mbbp/sf1 facilitates branchpoint region recognition. *Genes Dev.* **12**:858-867.
- Berkowitz, R.D., Ohagen, A., Hoglund, S., Goff, S.P. (1995). Rretoviral nucleocapsid domains mediate the specific recognition of genomic viral rnas by chimeric gag polyproteins during rna packaging *in vivo*. *J.Virol.* **69**:6445-6456.
- 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**(4):911-916.
- Bork, P., and Sudol, M. (1994). The ww domain: a signaling site in dystrophin? *Trends Biochem Sci.* **19**(12):531-533.
- Bringmann P, Appel B, Rinke J, Reuter R, Theissen H, Lührmann R. (1984). Evidence for the existence of snrnas u4 and u6 in a single ribonucleoprotein complex and for their association by intermolecular base pairing. *EMBO J.* **3**:1357-1363.
- Bruce, M.C., Kanelis, V., Fouladkou, F., Debonneville, A., Staub, O., Rotin, D. (2008). Regulation of nedd4-2 self-ubiquitination and stability by a py motif located within its hect-domain. *Biochem. J.* **415**:155–163.
- Buchan, D.W.A., Minneci, F., Nugent, T.C.O., Bryson, K., Jones, D.T. (2013). Scalable web services for the psipred protein analysis workbench. *Nucleic Acids Res.* **41**(W1):W340-W348.
- Carpenter, J.F., Randolph, T.W., Jiskoot, W., Crommelin, D.J.A., Middaugh, C.R., Winter, G. (2010). *J. Pharm. Sci.* **99**(5):2200-2208.
- Chang, J., Schwer, B., Shuman, S. (2012). Structure-function analysis and genetic interactions of the yeast branchpoint binding protein msl5. *Nucleic Acids Res.* **40**(10):4539-4552.
- Chatrikhi, R., Wang, W., Gupta, A., Loerch, S., Maucuer, A., Kielkopf, C.L. (2016). SF1 phosphorylation enhances specific binding to u2af65 and reduces binding to 3' splice-site rna. *Biophys. J.* **111**:1-17.
- Colot, H.V., Stutz, F., Rosbash, M. (1996). The yeast splicing factor mudl3p is a commitment complex component and corresponds to cbp20, the small subunit of the nuclear cap-binding complex. *Genes Dev.* **10**:1699-1708.
- Consalvi, V., Chiaraluce, R., Giangiacomo, L., Scandura, R., Christova, P., Karshikoff, A., Knapp, S., Ladenstein, R. (2000). Thermal unfolding and conformational stability of the recombinant domain ii of glutamate dehydrogenase from the hyperthermophile thermotoga maritima. *Protein Engineering.* **13**(7):501-507.

Coolidge, C.J., Seely, R.J., Patton, J.G. (1997). Functional analysis of the polypyrimidine tract in pre-mrna splicing. *Nucleic Acids Res.* **25**(4):888-896.

Dalhus, B., Saarinen, M., Sauer, U., Eklund, P., Johansson, K., Karlsson, A., Ramaswamy, S., Bjørk, A., Synstad, B., Naterstad, K., Sirevåg, R., Eklund, H. (2002). Structural basis for thermophilic protein stability: structures of thermophilic and mesophilic malate dehydrogenases. *J. Mol. Biol.* **318**(3):707-721.

Derrick, W.B., and Horowitz, J. (1993). Probing structural differences between native and *in vitro* transcribed *Escherichia coli* valine transfer rna: evidence for stable base modification-dependent conformers. *Nucleic Acids Res* **21**(21):4948-4953.

Didychuk, A.L., Butcher, S.E., Brow, D.A. (2018). The life of u6 small nuclear rna, from cradle to grave. *RNA*. **24**(4):436-460.

Dobbyn, H.C., mcewan, P.A., Krause, A., Novak-Frazer, L., Bella, J., O'Keefe, R.T. (2007). Analysis of pre-mrna and pre-rRNA processing factor snu13p structure and mutants. *Biochem Biophys Res Commun.* **360**:857-62.

Dunker, A.K., Lawson, J.D., Brown, C.J., Williams, R.M., Romero, P., Oh, J.S., Oldfield, C.J., Campen, A.M., Ratliff, C.M., Hipps, K.W., Ausio, J., Nissen, M.S., Reeves, R., Kang, C., Kissinger, C.R., Bailey, R.W., Griswold, M.D., Chiu, W., Garner, E.C., Obradovic, Z. (2001). Intrinsically disordered protein. *J. Mol. Graph Model.* **19**:26-59.

Evans, P., and mccoey, A. (2008). An introduction to molecular replacement. *Acta Crystallographica Sec D*. **D64**:1-10.

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. *Mol Cell.* **36**:593-608.

Fischer, U., Sumpter, V., Sekine, M., Satoh, T., Lührmann, R. (1993). Nucleo-cytoplasmic transport of u snrnps: definition of a nuclear location signal in the sm core domain that binds a transport receptor independently of the m3g cap. *EMBO J.* **12**:573-583.

Fleckner, J., Zhang, M., Valcárel, J., Green, M.R. (1997). U2AF65 recruits a novel human dead box protein required for the u2 snrnp-branchpoint interaction. *Genes Dev.* **11**:1864-1872.

Garrey, S.M., Voelker, R., Berglund, J.A. (2006). An extended rna binding site for the yeast branch point-binding protein and the role of its zinc knuckle domains in rna binding. *J. Biol. Chem.* **281**(37):27443-27453.

- Garrey, S.M., Cass, D.M., Wandler, A.M., Scanlan, M.S., Berglund, J.A. (2008). Transposition of two amino acids changes a promiscuous rna binding protein into a sequence-specific rna binding protein. *RNA*. **14**(1):78-88.
- Gille, C., Föhling, M., Weyand, B., Wieland, T., Gille, A. (2014). Alignment-annotator web server: rendering and annotating sequence alignments. *Nucleic Acids Res* **42**:W3-W6
- Glasel, J.A. (1995). Validity of nucleic acid purities monitored by 260nm/280nm absorbance ratios. *Biotechniques*. **18**(1):62-63.
- Goody, T.A., Melcher, S.E., Norman, D.G., Lilley, D.M.J. (2004). The kink-turn motif in rna is dimorphic, and metal ion-dependent. *RNA*. **10**: 254–264
- Gräslund, S., and 85 others. (2008). Protein production and purification. *Nat. Methods*. **5**(2):135-146.
- Graveley, B. (2001). Alternative splicing: increasing diversity in the proteomic world. *TRENDS in genetics*. **17**:100-107.
- Green, M.R. (1986). Pre-mrna splicing. *Annu Rev Genet*. **20**:671–708.
- Green, M.R. (1991). Biochemical mechanisms of constitutive and regulated pre-mrna splicing. *Annual Review of Cell Biology*. **7**:559-599.
- Guth, S., and Valcárel, J. (2000). Kinetic role for mammalian sf1/bbp. *J. Biol. Chem*. **275**(48):38059-38066.
- Hálová, M., ej Gahura, O., Převorovský, M., Cit, Z., Novotný, M., Valentová, A., Abrahámová, K., Půta, F., Folk, P. (2017). Nineteen complex-related factor prp45 is required for the early stages of cotranscriptional spliceosome assembly. *RNA*. **23**(10):1512-1524.
- Hamma, T., Ferre-D'Amare, A.R. (2004). Structure of protein l7ae bound to a k-turn derived from an archaeal box h/aca srna at 1.8Å resolution. *Structure*. **12**:893–903.
- Haraguchi, N., Andoh, T., Frendeway, D., Tani, T. (2006). Mutations in the sf1-u2af59-u2af23 complex cause exon skipping in *schizosaccharomyces pombe*. *J. Biol. Chem*. **282**(4):2221-2228.
- Hardin, J.W., Warnasooriya, C., Kondo, Y., Nagai, K., Rueda, D. (2015). Assembly and dynamics of the u4/u6 di-snrnp by single-molecule fret. *Nucleic Acids Res*. **43**(22):10963-10974.
- Hashimoto, C., and Steitz, J.A. (1984). U4 and U6 rnas coexist in a single small nuclear ribonucleoprotein particle. *Nucleic Acids Res* **12**: 3283-3293.

- Heller, L. (1952). The structure of dicalcium silicate α -hydrate. *Acta. Cryst.* **5**:724-728.
- Hensheid, K.L., Voelker, R.B., Berglund, J.A. (2008). Alternative modes of binding by u2af65 at the polypyrimidine tract. *Biochemistry.* **47**:449-459.
- Horowitz, D.S., Kobayashi, R., Krainer, A.R. (1997). A new cyclophilin and the human homologues of yeast prp3 and prp4 form a complex associated with u4/u6 snrnps. *RNA.* **3**:1374–1387.
- Huang, T., Vilardeell, J., Query, C.C. (2002). Pre-spliceosome formation in *s. pombe* requires a stable complex of sf1-u2af59-u2af23. *EMBO J.* **21**(20):5516-5526.
- Hulme, E.C., and Trevethick, M.A. (2010). Ligand binding assays at equilibrium: validation and interpretation. *Br. J. Pharmacol.* **161**:1219-1237.
- Iakoucheva, L., and Dunker, A.K. (2003). Order, disorder, and flexibility: prediction from protein sequence. *Structure.* **11**(11):1316-1317.
- Jacewicz, A., Chico, L., Smith P., Schwer, B., Shuman, S. (2015). Structural basis for recognition of intron branchpoint RNA by yeast Msl5 and selective effects of interfacial mutations on splicing of yeast pre-mrnas. *RNA.* **21**:401-414.
- Jandrositz, A., Guthrie, C. (1995). Evidence for a prp24 binding site in u6 snrna and in a putative intermediate in the annealing of u6 and u4 snrnas. *EMBO J.* **14**:820-32.
- Jennings, M.J., Barrios, A.F., Tan, S. (2016). Elimination of truncated recombinant protein expressed in *E. coli* by removing cryptic translation initiation site. *Protein Expr. Purif.* **121**:17-21.
- Jones, D.T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292**:195-202
- Kao, H.Y., and Siliciano, P.G. (1996). Identification of prp40, a novel essential yeast splicing factor associated with the u1 small nuclear ribonucleoprotein particle. *Mol. Cell. Biol.* **16**(3):960-967.
- Karaduman, R., Fabrizio, P., Hartmuth, K., Urlaub, H., Lührmann, R. (2006). RNA structure and rna-protein interactions in purified yeast u6 snrnps. *J. Mol. Biol.* **356**:1248-1262.
- Keller, E.B., and Noon, W.A. (1984). Intron splicing: a conserved internal signal in introns of animal pre-mrnas. *Proc. Natl. Acad. Sci. USA.* **81**(23):7417-7420.
- Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., Sternberg, M.J.E. (2015). The phyre2 web portal for protein modeling, prediction, and analysis. *Nat. Protoc.* **10**:845-858.

- Kielkopf, C.L., Rodionova, N.A., Green, M.R., Burley, S.K. (2001). A novel peptide recognition mode revealed by the x-ray structure of a core u2af35/u2af65 heterodimer. *Cell*. **108**:595-605.
- Kielkopf, C.L., Lücke, S., Green, M.R. (2004). U2AF homology motifs: protein recognition in the rrm world. *Genes Dev*. **18**:1513-1526.
- Kim, K.J., Kim, G.E., Lee, K.H., Han, W., Yi, M.J., Jeong, J., Oh, B.H. (2004). Two-promoter vector is highly efficient for overproduction of protein complexes. *Protein Science*. **13**:1698-1703.
- Kistler, A.L., and Guthrie, C. (2001). Deletion of mud2, the yeast homolog of u2af65, can bypass the requirement for sub2, an essential spliceosomal atpase. *Genes Dev*. **15**:42-49.
- Klein, D.J., Schmeing, T.M., Moore, P.B., Steitz, T.A. (2001). The kink-turn: a new rna secondary structure motif. *RNA*. **20**(15):4214-4221.
- Kleywegt, G.J., and Jones, T.A. (1997). Model building and refinement practice. *Methods in Enzymology*. **277**:208-230.
- Kohtz, J.D., Jamison, S.F., Will, C.L., Zuo, P., Lührmann, R., Garcia-Blanco, M.A., Manley, J.L. (1994). Protein-protein interactions and 5'-splice site recognition in mammalian mrna precursors. *Nature*. **368**:119-124.
- Krainer, A.R., Conway, G.C., Kozak, D. (1990). Purification and characterization of pre-mrna splicing factor sf2 from hela cells. *Genes Dev*. **4**:1158-1171.
- Kramer, E.B., and Farabaugh, P.J. (2007). The frequency of translational misreading errors in *E. coli* is largely determined by trna competition. **13**(1):87-96.
- Kühn, J.F., Tran, E.J., Maxwell, E.S. (2002). Archaeal ribosomal protein 17 is a functional homologue of the eukaryotic 15.5k/snu13p snrnp core protein. *Nucleic Acids Res*. **30**:931-941.
- Larson, J.D., and Hoskins, A.A. (2017). Dynamics and consequences of spliceosome e complex formation. *Elife*. **6**:27592.
- Lee, K.C., Jang, Y.H., Kim, S.K., Park, H.Y., Thu, M.P., Lee, J.H., Kim, J.K. (2017). RRM domain of *Arabidopsis* splicing factor sf1 is important for pre-mrna splicing of a specific set of genes. *Plant Cell. Rep*. **36**(7):1083-1095.
- Leung, A.K.W., Nagai, K., Li, J. (2011). Structure of the spliceosomal U4 snrnp core domain and its implication for snrnp biogenesis. *Nature*. **473**:536-539.

- Li, X., and Manley, J.L. (2005). Inactivation of the sr protein splicing factor asf/sf2 results in genomic instability. *Cell*. **122**(3):365-378.
- Liang, W.W., and Cheng, S.C. (2015). A novel mechanism for prp5 function in prespliceosome formation and proofreading the branch site sequence. *Genes Dev.* **29**:81-93.
- Lin, K.T., Lu, R.M., Tarn, W.Y. (2005). The ww domain-containing proteins interact with the early spliceosome and participate in pre-mrna splicing *in vivo*. *Mol. Cell. Biol.* **24**(20):9176-9185.
- Linding, R., Jensen, L.J., Diella, F., Bork, P., Gibson, T.J., Russell, R.B. (2003). Protein disorder prediction: implications for structural proteomics. *Structure*. **11**(11):1453-1459.
- Lipp, J.J., Marvin, M.C., Shokat, K.M., Guthrie, C. (2015). SR protein kinases promote splicing of noncensus introns. *Nat. Struct. Mol. Biol.* **22**:611-617.
- Liu, S., Ghalei, H., Lührmann, R., Wahl, M.C. (2011). Structural basis for the dual u4 and u4atac snrna-binding specificity of spliceosomal protein hprp31. *RNA*. **17**:1655–1663.
- Liu, S., Mozaffari-Jovin, S., Wollenhaupt, J., Santos, K.F., Theuser, M., Dunin-Horkawisz, S., Fabrizio, P., Bujnicki, J.M., Lührmann, R., Wahl, M.C. (2015). A composite double-/single-stranded rna-binding region in protein prp3 supports tri-snrnp stability and splicing. *Elife*. **4**:e07320.
- Liu, Z., Luyten, I., Bottomley, M.J., Messias, A.C., Houngninou-Molango, S., Sprangers, R., Zanier, K., Krämer, A., Sattler, M. (2001). Structural basis for reception of the intron branch site rna by splicing factor 1. *Science*. **294**:1098-1102.
- Louis-Jeune, C., Andrade-Navarro, M.A., Perez-Iratxeta, C. (2011). Prediction of protein secondary structure from circular dichroism using theoretically derived spectra. *Proteins: Structure, Function, and Bioinformatics*. **80**(2).
- Lu, P.J., Zhou, X.Z., Shen, M., Lu, K.P. (1999). Function of ww domains as phosphoserine- or phosphothreonine-binding modules. *Science*. **283**(5406):1325-1328.
- Ma, P., and Xia, X. (2011). Factors affecting splicing strength of yeast genes. *Comp. Funct. Genomics*. **2011**(212146):1-13.
- Marmier-Gourrier, N., Clery, A., Senty-Segault, V., Charpentier, B., Schlotter, F., Leclerc, F., Fournier, R., Branlant, C. (2003). A structural, phylogenetic, and functional study of 15.5kd/snu13 protein binding on u3 small nucleolar rna. *RNA*. **9**:821–838.

- Mayes, A.E., Verdone, L., Legrain, P., Beggs, J.D. (1999). Characterization of sm-like proteins in yeast and their association with u6 snrna. *EMBO J.* **18**:4321-4331. Achsel Achsel
- Macias, M.J., Hyvönen, M., Baraldi, E., Schultz, J., Sudol, M., Saraste, M., Oschkinat, H. (1996). Structure of the ww domain of a kinase-associated protein complex with a proline-rich peptide. *Nature.* **382**:646-649.
- Manceau, V., Swenson, M.C., Le Caer, J.P., Sobel, A., Kielkopf, C.L., Maucuer, A. (2006). Major phosphorylation of sfl on adjacent ser-pro motifs enhances interaction with u2af65. *FEBS J.* **273**:577–587.
- Marmier-Gourrier, N., Clery, A., Senty-Segault, V., Charpentier, B., Schlotter, F., Leclerc, F., Fournier, R., Branlant, C. (2003). A structural, phylogenetic, and functional study of 15.5-kd/snu13 protein binding on u3 small nucleolar rna. *RNA.* **9**(7):821-838.
- Matsuzaki, M., Misumi, O., Shin-I, T., Maruyama, S., Takahara, M., Miyagishima, S.Y., Mori, T., Nishida, K., Yagisawa, F., Nishida, K., Yoshida, Y., Nishimura, Y., Nakao, S., Kobayashi, T., Momoyama, Y., Higashiyama, T., Minoda, A., Sano, M., Nomoto, H., Ogasawara, N., Kohara, Y., and Kuroiwa, T. (2004). Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10d. *Nature.* **428**:653-7.
- Moore, M.J., Query, C.C., Sharp, P.A. (1993). Splicing of precursors to mRNAs by the spliceosome. In: Gesteland RF, Atkins JF, editors. *The RNA world*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; pp. 303–357.
- Mullen, M.P., Smith, C.W.J., Patton, J.G., Nadal-Ginard, B. (1991). A-tropomyosin mutually exclusive exon selection: competition between branchpoint/polypyrimidine tracts determines default exon choice. *Genes Dev.* **5**:642-655.
- NCBI Resource Coordinators. (2017). Database resources of the national center for biotechnology information. *Nucleic Acids Res.* **45**(D1):D12-D17.
- Nottrott S, Hartmuth K, Fabrizio P, Urlaub H, Vidovic I, Ficner R, Lührmann R. (1999). Functional interaction of a novel 15.5kd [u4/u6.u5] tri-snRNP protein with the 5' stem-loop of u4 snRNA. *EMBO J.* **18**:6119–6133.
- Nottrott, S., Urlaub, H., & Lührmann, R. (2002). Hierarchical, clustered protein interactions with u4/u6 snRNA: a biochemical role for u4/u6 proteins. *The EMBO Journal*, **21** (20), 5527-5538.
- Nguyen, T.H., Galej, W.P., Bai, X.C., Savva, C.G., Newman, A.J., Scheres, S.H., Nagai, K. (2015). The architecture of the spliceosomal u4/u6.u5 tri-snRNP. *Nature.* **523**:47–52.

- Oruganti, S., Zhang, Y., Li, H. (2005). structural comparison of yeast snrnp and spliceosomal protein snu13p with its homologs. *Biochem Biophys Res Commun.* **333**(2):550-554.
- Otwinowski, Z.M., Minor, W. (1997). Processing of x-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**: 307–326.
- Patterson, B., and Guthrie, C. (1991). A u-rich tract enhances usage of an alternative 3' splice site in yeast. *Cell.* **64**:181-187.
- Perriman, R., and Ares Jr, M. (2000). ATP can be dispensable for prespliceosome formation in yeast. *Genes Dev.* **14**:97-107.
- Potterton, E., Briggs, P., Turkenburg, M., Dodson, E. (2003). A graphical user interface to the ccp4 program suite. *Acta Crystallogr. D Biol. Crystallogr.* **59**:1131–1137.
- Prystay, L., Gosselin, M., Banks, P. (2001). determination of equilibrium dissociation constants in fluorescence polarization. *J. Biomol. Screen.* **6**:141-150.
- Query, C.C., Strobel, S.A., Sharp, P.A. (1996). Three recognition events at the branch-site adenine. *EMBO J.* **15**:1392-402.
- Romfo, C.M., and Wise, J.A. (1997). Both the polypyrimidine tract and the 3' splice site function prior to the first step of splicing in fission yeast. *Nucleic Acids Res* **25**(2):4658-4665.
- Rosbash, M., and Séraphin, B. (1991). Who's on first? The u1 snrnp-5' splice site interaction and splicing. *TIBS.* **16**:187-190.
- Ruskin, B., and Green, M.R. (1985). Role of the 3' splice site consensus sequence in mammalian pre-mrna splicing. *Nature.* **317**:732-734.
- Rutz, B., and Séraphin, B. (1999). Transient interaction of bbp/scsf1 and mud2 with the splicing machinery affects the kinetics of spliceosome assembly. *RNA.* **5**:819-831.
- Sachdev, D., and Chirgwin, J.M. (2000). Fusions to maltose-binding protein: control of folding and solubility in protein purification. *Methods Enzymol.* **326**:312-321.
- Sander, B., Golas, M.M., Makarov, E.M., Brahms, H., Kastner, B., Lührmann, R., Stark, H. (2006). Organization of core spliceosomal components u5 snrna loop i and u4/u6 di-snrnp within u4/u6.u5 tri-snrnp as revealed by electron cryomicroscopy. *Mol. Cell.* **24**:267-278.
- Scheich, C., Kümmel, D., Soumailakakis, D., Heinemann, U., Büsow, K. (2007). Vectors for co-expression of an unrestricted number of proteins. *Nucleic Acids Res.* **43**(6):e43.

- Selenko, P., Gregorovic, G., Sprangers, R., Stier, G., Rhani, Z., Krämer, A., Sattler, M. (2003). Structural basis for the molecular recognition between human splicing factors u2af65 and sf1/mbpp. *Mol. Cell.* **11**:965-976.
- Sjöback, R., Nygren, J., Kubista, M. (1995). Absorption and fluorescence properties of fluorescein. *Spectrochimica Acta Part A.* L7-L21.
- Soss, S.E., and Flynn, P.F. (2007). Functional implications for a prototypical k-turn binding protein from structural and dynamical studies of 15.5k. *Biochemistry.* **46**:14979–14986.
- Spingola, M., Grate, L., Haussler, D., Ares Jr, M. (1999). Genome-wide bioinformatic and molecular analysis of introns in *Saccharomyces cerevisiae*. *RNA.* **5**:221-234.
- Staley, J. P., and Guthrie, C. (1998). Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell.* **92**:315-26.
- Stark, H., and Lührmann, R. (2006). Cryo-electron microscopy of spliceosomal components. *Annual Review of Biophysics.* **35**:435-457.
- Stark, M.R., Dunn, E.A., Dunn, W.S.C., Grisdale, C.J., Daniel, A.R., Halstead, M.R.G., Fast, N.M., Rader, S.D. (2015). Dramatically reduced spliceosome in *Cyanidioschyzon merolae*. *Proc. Natl. Acad. Sci. USA.* **112**(11):E1191-E1200.
- Sterner, R., and Liebl, W. (2001). Thermophilic adaptation of proteins. *Crit. Rev. Biochem. Mol. Biol.* **36**(1):39-106.
- Stevens, S.W., and Abelson, J. (1999). Purification of the yeast u4/u6.u5 small nuclear ribonucleoprotein particle and identification of its proteins. *Proc. Natl. Acad. Sci. USA.* **96**:7226–7231.
- Stevens, S.W., Ryan, D.E., Ge, H.Y., Moore, R.E., Young, M.K., Lee, T.D., Abelson, J. (2002). Composition and functional characterization of the yeast spliceosomal pentasnrnp. *Mol Cell.* **9**:31-44.
- Studier, F.W. (2005). Protein production by auto-induction in high-density shaking cultures. *Protein Expr. Purif.* **41**:207-234.
- Szewczak, L.B.W., degregorio, S.J., Strobel, S.A., Steitz, J. (2002). Exclusive interaction of the 15.5 kd protein with the terminal box c/d motif of a methylation guide snorncp. *Chem. Biol.* **9**:1095–1107.
- Theuser, M., Höbartner, C., Wahl, M.C., Santos, K.F. (2016). Substrate-assisted mechanism of rnp disruption by the spliceosomal brr2 rna helicase. *PNAS.* **113**(28):7798-7803.

- Turner, B., Melcher, S.E., Wilson, T.J., Norman, D.G., Lilley, D.M. (2005). Induced fit of rna on binding the l7ae protein to the kink-turn motif. *RNA*. **11**(8):1192-1200.
- Valverde, R., Edwards, L., Regan, L. (2008) Structure and function of kh domains. *FEBS J*. **275**:2712-2726.
- Vargas, W.P. (2016). Interaction of spliceosomal u2 snrnp protein p14 with its branch site rna target. *Graduate Center, City University of New York*. Phd Dissertation.
- La Verde, V., Dominici, P., Astegno, A. (2017). determination of hydrodynamic radius by size exclusion chromatography. *Bio-protocol*. **7**(8):e2230.
- Vernet, C., and Artzt, K. (1997). STAR, a gene family involved in signal transduction and activation of rna. *Trends Genet*. **13**(12):479-484.
- Vidovic, I., Nottrott, S., Hartmuth, K., Luhrmann, R., Ficner, R. (2000). Crystal structure of the spliceosomal 15.5kd protein bound to a u4 snrna fragment. *Mol Cell*. **6**:1331-1342.
- Wang, Q., Zhang, L., Lynn, B., Rymond, B.C. (2008). A bbp-mud2p heterodimer mediates branchpoint recognition and influences splicing substrate abundance in budding yeast. *Nucleic Acids Res* **36**(8):2787-2798.
- Wang, W., Maucuer, A., Gupta, A., Manceau, V., Thickman, K.R., Bauer, W.J., Kennedy, S.D., Wedekind, J.E., Green, M.R., Kielkopf, C.L. (2013). Structure of phosphorylated sf1 bound to u2af65 in an essential splicing factor complex. *Structure*. **21**:197-208.
- Warf, M.B., and Berglund, J.A. (2010). Role of rna structure in regulating pre-mrna splicing. *Trends Biochem Sci*. **35**(3):169-178.
- Watkins, N.J., Dickmanns, A., Lührmann, R. (2002). Conserved stem ii of the box c/d motif is essential for nucleolar localization and is required, along with the 15.5k protein, for the hierarchical assembly of the box c/d snornp. *Mol. Cell Biol*. **22**:8342–8352.
- Waugh, A., Gendron, P., Altman, R., Brown, J.W., Case, D., Gautheret, D., Harvey, S.C., Lenthis, N., Westbrook, J., Westhof, E., Zuker, M., Major, F. (2002). RNAML: A standard syntax for exchanging rna information. *RNA*. **8**(6):707-717.
- Will, C.L., and Lührmann, R. (2011). Spliceosome structure and function. In: *RNA worlds*. (Gesteland, R.F., Cech, T.R. & Atkins, J.F., eds), pp. 1-23, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Yachdav, G., Kloppmann, E., Kajan, L., Hecht, M., Goldberg, T., Hamp, T., Hönigsmid, P., Schafferhans, A., Roos, M., Bernhofer, M., Richter, L., Ashkenazy, H., Punta, M., Schlessinger, A., Bromberg, Y., Schneider, R., Vriend, G., Sander, C., Ben-Tal, N., Rost, B. (2014). Predictprotein – an open resource for online prediction of protein structural and functional features. *Nucleic Acids Res.* **42**:W337-W343.

Zamore, P.D., Patton, J.G., Green, M.R. (1992). Cloning and domain structure of the mammalian splicing factor u2af. *Nature.* **355**:609-614.

Zhang, Y., Madl, T., Bagdiul, I., Kern, T., Kang, H.S., Zou, P., Mäusbacher, N., Sieber, S.A., Krämer, A., Sattler, M. (2012). Structure, phosphorylation and u2af65 binding of the n-terminal domain of splicing factor 1 during 3'-splice site recognition. *Nucleic Acids Res* **41**(2):1343-1354.

Zuker, M., and Jacobson, A.B. (1998). Using reliability information to annotate rna secondary structures. *RNA.* **4**:669-679.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**(13):3406-3415.