Investigations on the Role of the U1 snRNA in Pre-mRNA Splicing

by

William Martelly

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved April 2021 by the
Graduate Supervisory Committee:

Shalini Sharma, Co-Chair
Marco Mangone, Co-Chair
Kurt Gustin
Julian Chen

ARIZONA STATE UNIVERSITY

May 2021

ABSTRACT

The splicing of precursor messenger RNAs (pre-mRNAs) plays an essential role
in dictating the mature mRNA profiles of eukaryotic cells. Mis-regulation of splicing, due
to mutations in pre-mRNAs or in components of the splicing machinery, is associated
with many diseases. Therefore, knowledge of pre-mRNA splicing mechanisms is
required to understand gene expression regulation during states of homeostasis and
disease, and for the development of therapeutic interventions.

Splicing is catalyzed by the spliceosome, a dynamic and protein-rich ribozyme
composed of five small nuclear ribonucleoproteins (snRNPs) and ~170 auxiliary factors.
Early interactions that occur in prespliceosomal complexes formed by the 5′- and 3′-
splice-site bound U1 and U2 snRNPs are responsible for committing introns for removal.
However, the mechanisms underlying these early interactions remain to be fully
characterized for understanding the influence of alternative splicing factors and the
impact of recurrent disease-associated mutations in prespliceosomal proteins.

The goal of my dissertation research was to delineate the role of the U1 small
nuclear RNA (snRNA) during prespliceosome assembly. By applying a cellular minigene
reporter assay and a variety of *in vitro* techniques including cell-free protein expression,
UV-crosslinking, electrophoretic mobility shift assays, surface plasmon resonance, and
RNA affinity purification, my work establishes critical roles for the U1 snRNA stem-
loops 3 (SL3) and 4 (SL4) in formation of intron definition interactions during
prespliceosome assembly. Previously, the SL4 of the U1 snRNA was shown to form a
molecular bridge across introns by contacting the U2-specific splicing factor 3A1
(SF3A1). I identified the Ubiquitin-like domain of SF3A1 as a non-canonical RNA

binding domain responsible for U1-SL4 binding. I also determined a role for the SL3 region of the U1 snRNA in splicing and characterized the spliceosomal RNA helicase UAP56 as an SL3 interacting protein. By knocking-down the SL3- and SL4-interacting proteins, I confirmed that U1 splicing activity *in vivo* relies on UAP56 and SF3A1 and that their functions are interdependent. These findings, in addition to the observations made using *in vitro* splicing assays, support a model whereby UAP56, through its interaction with U1-SL3, enhances the cross-intron interaction between U1-SL4 and SF3A1 to promote prespliceosome formation.

DEDICATION

This dissertation is dedicated to my family. To my parents, Patricia and Christophe, every door that has ever opened to me is because of you, and with your love in my heart, I know I can face any obstacle that may lie beyond them. To my big brother Alex, you taught me how to play hard…and how to work hard too. To my sister-in-law Meghann and hers' and Alex's newborn, Luc. You are always on my mind and putting a smile on my face. And to my fiancé Sofia, you are the wind in my sails and my partner in crime, thank you for all your love and support.

ACKNOWLEDGMENTS

There are many people whose time and efforts have directly impacted and contributed to the contents of this dissertation, yet it would not have been possible without the mentorship and support from my advisor, Dr. Shalini Sharma. Your door was always open for reviewing protocols, for editing grants and presentations, and for discussing the perplexing data of the week. Your enthusiasm for RNA biology first drew me to the lab, and it was infectious. The results and observations presented in this dissertation has grown out of the strong foundation of research that has been cultivated in the Sharma Lab. It is with deep gratitude that I thank you for giving me the opportunity to contribute to these projects throughout my graduate studies, and for all your endless support and patience along the way.

Thank you to the members of my committee who have all invested time and energy into my development as a graduate student and thank you for our discussions and for your feedback which have always served to improve the quality of my research. Thank you, Dr. Marco Mangone, for always making sure I was first on the list to personally meet with visiting Nobel laureates like Dr. Andy Fire and Dr. Marty Chalfie. Thank you, Dr. Julian Chen, for always asking the tough questions during our annual meetings that encouraged me to think deeper. And a big thanks to you Dr. Kurt Gustin for your time and guidance while serving as chair for my comprehensive examination.

I would also like to thank our collaborators like Dr. Tim Marlowe who personally trained me in surface plasmon resonance which allowed me to apply the technique for investigating RNA-protein interactions. Thank you to Paul Kang for assistance with developing the statistical analysis applied for assessing the synergistic impacts of U1

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

xii

CHAPTER 1

INTRODUCTION

Information flow and the discovery of discontinuous genes

Modern molecular biology and genetics has expanded from a central dogma that describes how inherited genetic information stored in DNA is converted into the protein that ultimately defines many of the physical and biochemical properties of the cell (Figure 1.1). This general description of the flow of genetic information, first posited in 1958 by Francis Crick five years after the molecular structure of DNA was published, provided a theoretical framework for gene expression that became widely adopted [1, 2]. It was a prevailing assumption that genes in all organisms were organized in large blocks of contiguous protein-coding sequence directly transcribed and translated as observed in bacteria; a sentiment expressed by the well-known declaration of Jacques Monod and François Jacob that "anything found to be true of *E. coli* must also be true of Elephants" [3-6]. However, early experiments studying mRNA synthesis in eukaryotic cells produced unexplained findings; mRNA was found to be rapidly degraded in the nucleus

Figure 1.1 – Transcription and translation converts genotype to phenotype. This diagram depicts the flow of genetic information as described by the central dogma. The protein-coding information safely stored in the DNA-based genome is activated or expressed by the production of messenger RNA (mRNA) in a process called transcription. The newly synthesized single-stranded mRNA harboring a precise nucleotide sequence is shipped to ribosomes that pair amino-acids to the genetic triplet code, producing a peptide chain and generating protein in the final step of gene expression known as translation. Protein ultimately performs many of the structural and enzymatic functions that defines the physical and biochemical properties of the cell, or phenotype.

1

of the eukaryotic cell and tended to be four times the size of mRNA exported to the cytoplasm for translation [7].

The mystery of the long nuclear mRNA precursors would not be resolved until scientists directly mapped the sequences of mRNA produced in eukaryotic cells back to that of their DNA templates using R-loop mapping [8] (Figure 1.2 Top Panel). In 1977, researchers using adenovirus infection as a model for eukaryotic mRNA synthesis, applied this new technique to map mRNA from the adenovirus hexon gene produced during HeLa cell infection back to the viral genome. Instead of observing a single continuous R-loop as would be expected if genes in eukaryotes were transcribed as a single uninterrupted block of protein-coding sequence, several labs observed an unhybridized region of adenovirus mRNA protruding from the 5′ ends of the R-loop [9, 10]. Further investigation in the laboratories of Phil Sharp and Richard Roberts demonstrated that these 5′ mRNA tails could hybridize and form three additional R-loops with complementary regions of the viral genome at distal locations upstream from the initial R-loop site, correctly positing that long precursor mRNA molecules containing the complementary and intervening genomic sequences are transcribed but later must be processed, or "spliced", to form a final mRNA product with the non-contiguous, intragenic regions removed (Figure 1.2 Bottom Panel) [11, 12].
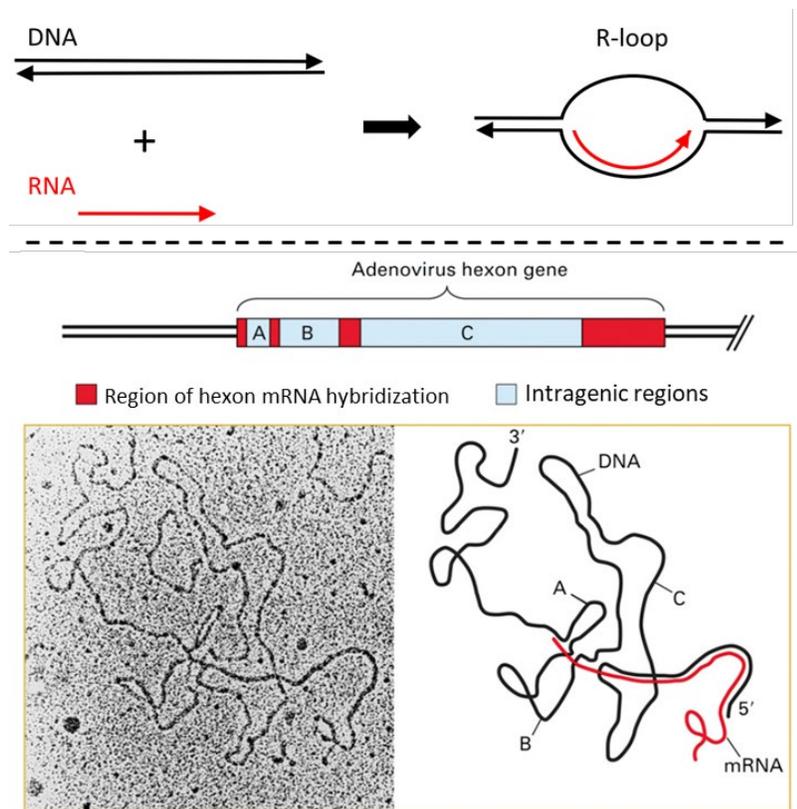
Figure 1.2 – R-loop mapping of the adenovirus hexon mRNA revealed the organization of split genes in eukaryotes. Top panel – R-loop mapping is performed by mixing purified RNA (red) with purified double-stranded DNA (black) under high formamide conditions that favor RNA-DNA hybrids. Regions of complementarity between the RNA and DNA form characteristic R-loops as one of the strands of DNA is displaced by the stabilized RNA-DNA hybrid duplex. Bottom panel – A schematic of the hexon gene and the organization of its intragenic regions is displayed. Regions of mRNA hybridization are depicted as red blocks and formed the characteristic R-loops observed in the electron micrograph (EM) displayed below the schematic and shown more clearly in the trace to the right of the EM. These observations were the first to demonstrate that eukaryotic mRNA is transcribed as a long precursor molecule that is later processed to remove intragenic regions producing the mature mRNA. Figure adapted from Berka A.J. 2016.

We now understand that most eukaryotic genes are first transcribed in the form of long precursor messenger RNAs (pre-mRNAs) harboring non-coding intragenic regions called introns between blocks of protein coding sequences called exons. Pre-mRNA splicing is the process by which introns are removed and exons are fused together in the nucleus of the cell to produce a chain of consecutive exon sequences (mature mRNA)

3

that is ultimately exported to the cytoplasm and expressed into protein (Figure 1.3). The

discovery of discontinuous genes in eukaryotes not only solved the mystery of the long

nuclear mRNA precursors observed in earlier studies, it also shed light on a novel system

of gene regulation in eukaryotes and provided some explanation for a longstanding

paradox in molecular genetics.
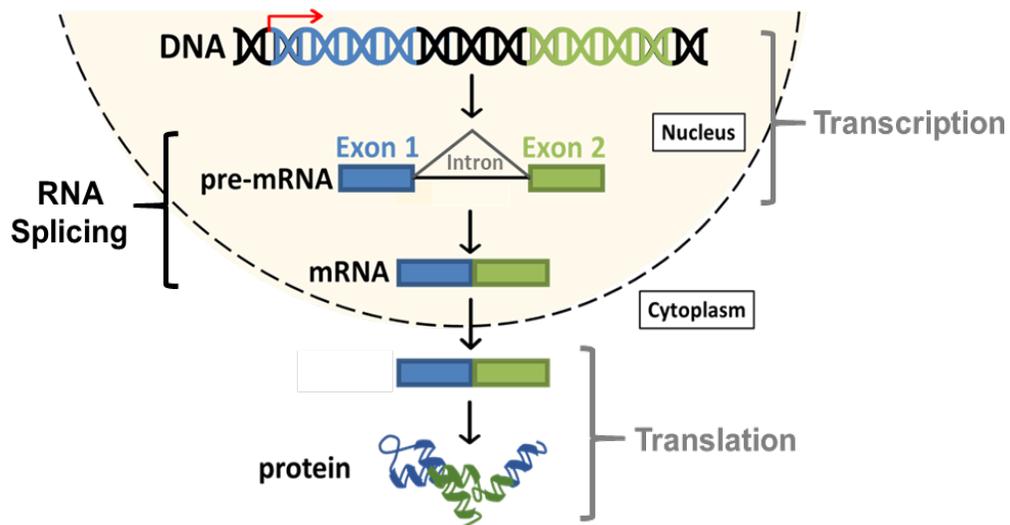
Figure 1.3 – Pre-mRNA splicing is a critical stage of gene expression in eukaryotes.
Nascent transcripts from actively transcribing genes in humans and other higher eukaryotes
produce a premature mRNA molecule containing non-protein coding sequences within
regions called introns that are removed during the splicing process. Splicing ultimately
produces a consecutive chain of protein coding exons in the mature mRNA which can be
subsequently exported to the cytoplasm and translated into protein.

Split genes expand the coding and regulatory capacities of eukaryotic genomes

The organization of split genes allows for discrete regions of protein-coding

sequence to be rearranged and linked together in novel ways, facilitating the production

of multiple mRNA isoforms that can each direct synthesis of a unique protein, all from a

single gene [13]. For a simple gene containing three exons and two introns, the cell can

fuse all three exons into one full-length mRNA isoform via constitutive splicing (Figure

1.4). However, not all exons have to be represented in the final mRNA and an important process called alternative splicing (AS) can produce many mRNA isoforms from a single pre-mRNA. There are five canonical alternative splicing patterns cells can perform by linking together exons non-constitutively [14]. The most common form of AS in humans is exon skipping in which a specific cassette exon is differentially included or excluded in the final spliced mRNA [15]. Additionally, there are mutually exclusive exons that are rarely spliced together into the same isoform, the selection of alternative 5′- or 3′-splice site (ss) sequences to produce exons of different lengths, and finally entire introns that can be retained (Figure 1.4).
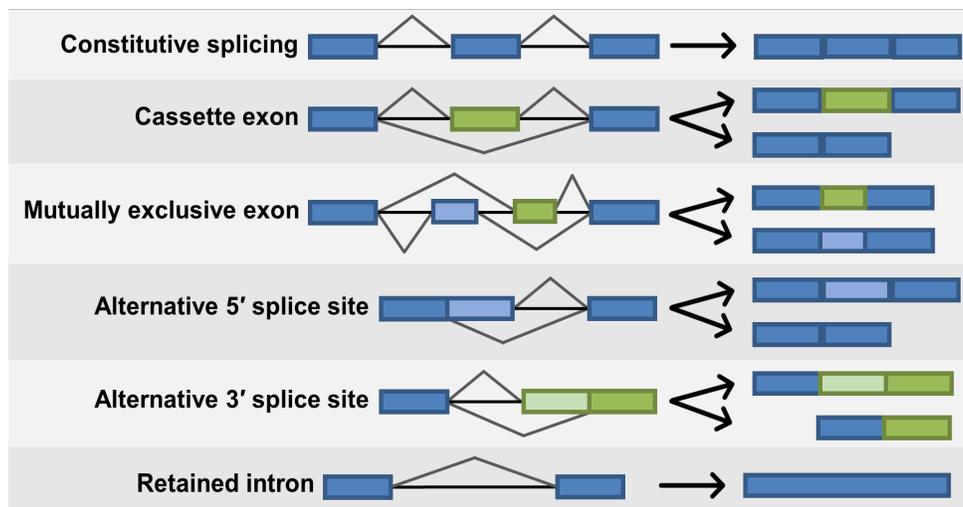
Figure 1.4 – Alternative splicing can generate many different isoforms of mRNA from a single pre-mRNA transcript. Eukaryotic cells can leverage a variety of alternative splicing patterns to expand the diversity of coding sequence produced from a single protein coding gene. The five canonical alternative splicing events commonly employed by eukaryotic cells are depicted in the schematic.

Only approximately ~3% of human genes are intronless [16, 17] and on average those that are split contain ~8-10 exons [18, 19]. Initial studies estimated that AS occurs in only 40–60% of human genes [20, 21], however, the development and application of

next-generation RNA-sequencing (RNA-seq) technologies has revealed that virtually all multi-exon genes in humans produce 2 or more mRNA isoforms [22, 23]. In some extreme cases as in *Drosophila melanogaster*, the *DSCAM* gene which codes for a receptor that guides axon growth in fly neurons has the potential to encode 38,016 unique proteins (~3-fold more mRNA isoforms than there are total unique genes in the fly genome) due to AS of numerous mutually exclusive exons at exons 4, 6, 9, and 17 of the gene [24]. In humans, the neurexin family of proteins involved in neuronal cell-to-cell adhesion at synapses are encoded from 3 independent *NRXN* genes that are extensively processed by AS and have the collective potential to produce 2,250 unique mRNA isoforms [25].

This expansion of mRNA isoform diversity made possible by AS has in part addressed the "gene number paradox" resulting from the observation that organismal complexity does not easily correlate with raw number of protein coding genes [26, 27]. Even with modern tools, estimating the number of protein coding genes in the human genome is challenging due to the organization of small protein coding sequences separated by large introns. The percentage of sequence that is ultimately translated represents only ~5% of any given pre-mRNA, resulting in a low signal-to-noise problem that affects computational predictions [28]. For this reason, even after completion of the first human reference genome, the estimated number of protein-coding genes has varied substantially [29]. One estimate, derived from the average number of human genes annotated in the major online databases like the UCSC genome browser, NCBI, and Ensemble, calculated approximately 22,500 ± 2,000; placing the number of protein-coding genes in humans somewhere between that of the chicken (16,736) and the grape

(30,434) [30]. This estimate of human genes is not only less than that of grapes, but also comparable to significantly less complex eukaryotes like the plant *Arabidopsis thaliana* which has some 25,000 genes [31], or that of the microscopic roundworm *Caenorhabditis elegans* with ~20,000 genes [32].

It is clear that the raw number of genes within an organism is not sufficient to explain its complexity, however, eukaryotes with more diverse tissues and complex behaviors, specifically vertebrates, also have higher rates of alternative splicing [15]. For example, only 25% of all protein coding genes in *C. elegans* undergo AS [33] compared to ~95% in humans [22]. Therefore, split genes and the ability to recombine them to produce novel rearrangements, in part, facilitates higher order gene regulation producing the more intricate phenotypes observed in complex organisms like humans, while exponentially increasing the amount of protein-coding information that can be stored by any given gene. Alternative splicing is a highly regulated process that facilitates regulation of gene expression in response to environmental [34, 35], developmental [36, 37] and tissue-specific [38, 39] cues. Therefore, a complete mechanistic understanding of pre-mRNA splicing, and the processes underlying its regulation, is critical to accurately predict the flow of genetic information from transcription to translation in eukaryotes.

Anatomy of the metazoan intron and its evolutionary origin

In humans, exons are relatively short with the vast majority (80-85%) tending to be less than 200 nucleotides (nts) long, while introns are large and can vary dramatically in size, with average lengths of approximately 5,419 nts [18]. Introns contain a series of consensus sequences which define their boundaries [40]. These sequences include the 5′-

splice site (ss), the branch point sequence, the polypyrimidine tract, and the 3′-ss, and are

very highly conserved among eukaryotes from yeast to humans (Figure 1.5). Although

the lengths between the 5′- and 3′-ss can vary drastically, the polypyrimidine (Py) tract

always immediately precedes the 3′-ss while the branch point sequence is found within

~50 nts upstream of the 3′-ss [41, 42]. While more degenerate and variable in humans,

these consensus splice sites are stronger in yeast (98% of yeast introns match the

consensus) which also differ from humans in that they lack a strong Py tract [43, 44].

There are however, three invariant features conserved and present in nearly all introns

including the presence of a 'GU' dinucleotide at the 5′-ss immediately marking the start

of the intron, a single adenosine 'A' in the branch point sequence, and an 'AG'

dinucleotide at the 3′-ss marking the end of an intron.
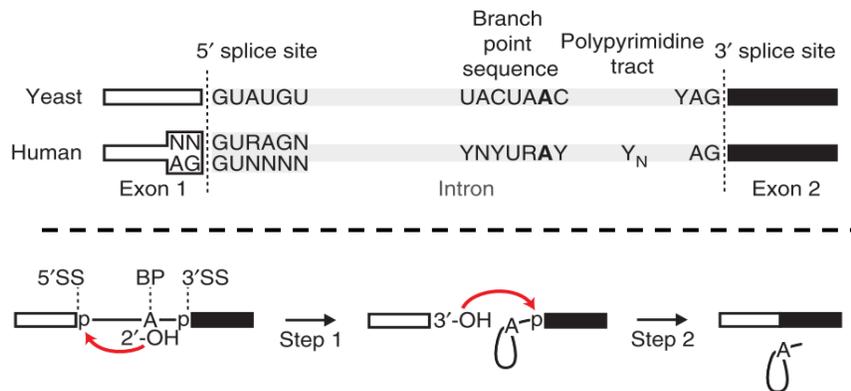
Figure 1.5 – Introns are defined by the presence of conserved splice-site sequences that also participate in the two-step splicing reaction. Top panel – A representative intron is displayed in the schematic and the consensus splice site sequences present in yeast and humans are displayed (Y represents any pyrimidine and R represents any purine). Bottom panel – The two transesterification reactions involved in pre-mRNA splicing are diagrammed. Figure adapted from Plaschka et al. 2019.

These splice site sequences define introns and are also reactive sites that directly participate in a two-step splicing reaction that removes introns from pre-mRNA and fuses exons consecutively (Figure 1.5). The first step begins when the 2′ hydroxyl group at the branchpoint adenosine attacks the 5′-ss sequence forming a 2′-5′ phosphodiester bond between the branchpoint adenosine and the 5′-ss yielding two products, a looped intron structure followed by the downstream exon called the lariat intermediate, and a now freed, upstream exon. The second step involves the 3′ hydroxyl group of the liberated exon attacking the phosphodiester bond at the 3′-ss effectively joining upstream and downstream exons and producing the intron lariat [45, 46].

The presence of 2′ hydroxyl groups in the ribose sugar backbone means that pre-mRNA harbors the reactive groups necessary to act on itself, however, pre-mRNA alone cannot catalyze the consecutive transesterification reactions need to generate mature mRNA. This is not the case for a large family of self-splicing ribozymes found in bacteria and in eukaryotic chloroplasts and mitochondria, called group II self-splicing introns [47]. These ribozymes utilize extensive *cis*-acting secondary and tertiary structures to catalyze their own splicing reactions which can occur efficiently in the absence of protein *in vitro* [48, 49]. In bacteria, group II introns do rely on intron encoded protein (IEP) to splice efficiently *in vivo* whereas in eukaryotic organelles, group II introns are more degenerate and no longer produce functional IEP, relying on host encoded factors acting in *trans* to promote self-splicing [50]. Group II introns are considered to be evolutionary precursors to split genes in eukaryotes due to the striking similarities between splice-site sequences and reaction intermediates formed during the splicing of both group II introns and pre-mRNA [47, 51]. Pre-mRNA splicing however requires intervention by numerous

9

trans-acting factors, consisting of both RNA and protein, and it is speculated that components of group II self-spicing introns were deconstructed over time and now function as individual modules repurposed for pre-mRNA processing in eukaryotes [19, 52, 53]. These discrete components and *trans*-acting factors leverage many of the diverse functions made possible by RNA to recognize pre-mRNA splice site sequences, and to juxtapose them within catalytic centers. These components are collectively referred to as the spliceosome.

The spliceosome is a dynamic and multi-component, deconstructed ribozyme

Although RNA itself displays a wide array of useful biochemical and enzymatic activities, it rarely functions in the absence of protein. Proteins can process RNA into mature molecules and facilitate proper folding of higher order structures [54, 55], regulate cellular localization and shield RNA from degradation [56], and enhance the catalytic activity of ribozymes [57]. Bioactive RNAs are better understood as ribonucleoprotein (RNPs), RNA complexed tightly to protein partners functioning as distinct units that participate in a variety of diverse cellular processes [58, 59].

A specific class of RNPs carry out their functions in the nucleus of the cell and are called small nuclear RNPs (snRNPs), five of which form the core components of the spliceosome (Figure 1.6). The major spliceosome is composed of five snRNPs called U1, U2, U4, U5, and U6 and are responsible for removal of 99.6% of all introns in the human genome [60]. Given the prefix "U" due to their uridine-rich content and numbered according to their relative abundance (U1 is the most abundant snRNP followed by U2, and so on), these snRNPs are at the core of the eukaryotic intron excision machinery [61].

10

small nuclear ribonucleo proteins (snRNPs)

12S	17S	25S
U1	U2	U6 U4 U5
U1 snRNA	U2 snRNA	U6 U4 U5

Sm	Sm	Sm/LSm
70K	A'	hPrp8
A	B''	hBrr2
C	Snu114	Snu114
	SF3a120	hPrp6
	SF3a 66	hPrp28
	SF3a 60	40K
	SF3b155	hDib1
	SF3b145	hPrp3
	SF3b130	hPrp31
	SF3b49	hPrp4
	SF3b14a/p14	CypH
	SF3b14b	15.5K
	SF3b10	hSnu66
		hSad1
		27K

Auxiliary Proteins

U1 related	U2 related	Prp19 complex	Complex C proteins	Complex B proteins
FBP11	U2AF	Prp19	Abstract	THRAP3
S164	PUF60	CDC5	Cactin	MFAP1
p68	SPF30	SPF27	DDX35	UBL5
CA150	SPF31	PRL1	Q9BRR8	RED
Complex A proteins	SPF45	AD-002	**PPlase**	hSmu1
BUB3	CHERP	Catenin βl1	Cyp64	**Other proteins**
MGC2803	SR140	Hsp73	PPIL3b	FRA10AC1
SF4	hPrp43	NPW38	PPIG	FRG1
FLJ10839	hPrp5	NPW38BP	NY-CO10	DGCR14
CDC2L2	**Step II factors**	**Prp19 related**	**Tri-snRNP related**	NOSIP
Tat SF1	hPrp17	SKIP	MGC20398	FLJ35382
TLS/FUS	Slu7	hSyf1	PPIL2	FAM50A+B
RMB5	hPrp22	hSyf3	HSP27	FAM32A
RBM10	hPrp18	CypE	Hskin17	RNF113A
SF1	hPrp16	MGC23918	Sklv2L2	CXorf56
RNP proteins	**RES complex**	RBM22	hPrp4 Kin	C9orf78
CBP80+20	MGC13125	G10	KIAA1604	CDCC130
hnRNPs 8	SNIP	KIAA0560	hPrp2	Srm300
SR 9	GI-79	GCIP p29	GPKOW	
mRNP 8				

RNA Helicases

DEAD-box
UAP56
Prp5
Prp28

Ski2

Brr2

DEAH-box
Prp2
Prp16
Prp22
Prp43

Figure 1.6 – The major spliceosome is composed of a core set of five small nuclear RNAs and ~200 proteins. Left panel – A cartoon of the secondary structures of the five snRNAs U1, U2, U4, U5, and U6 is shown. The common snRNP proteins are listed below and blocked out in grey and the snRNP specific proteins are listed and color coated accordingly. The U4/U6.U5 snRNPs form a tri-snRNP and are shown complexed together. Right panel – In addition to the snRNP proteins, approximately 170 auxiliary proteins and 8 conserved RNA helicases contribute to the spliceosome formation to some degree and are listed. Figures adapted from: Left panel – Will L. C, and Lührmann R. 2011. Right panel – Wahl C. M, Will L. C, and Lührmann R. 2009.

All spliceosomal snRNPs contain a <200 nts long small nuclear RNA (snRNA) that folds into distinct secondary structures and are bound by a set of snRNP specific proteins [62]. The U1 and U2 snRNPs function as individual particles, while U4, U5, and U6 are active in the form of a heterotrimeric complex called the U4/U6.U5 tri-snRNP; held together by extensive base-pairing between the U4 and U6 snRNAs and a series of protein-protein interactions with U5 [63]. A common set of shared proteins called Sm, originally discovered as antigens of autoreactive antibodies involved in the pathogenesis of systemic lupus erythematosus [64, 65], form a seven-membered ring around conserved binding sites at single-stranded regions in four of five of these snRNAs and has come to define this class of spliceosomal snRNPs [61]. U6, by contrast, contains a related but distinct heptameric structure formed by Sm-like (LSm) proteins [66]. The spliceosome is

a very protein-rich RNP and in addition to the ~45 snRNP specific proteins, approximately 170 spliceosome-associated factors, consisting of a variety of RNA binding proteins and eight conserved RNA-dependent ATPases/helicases, also participate in and regulate the splicing cycle [67, 68].

Insight into spliceosome assembly, composition, and structure has progressed significantly over the past 40 years and has been achieved by the development of *in vitro* splicing assays using active nuclear extract preparations and a variety of model pre-mRNA substrates [69, 70], the affinity-purification of intermediate complexes followed by mass spectrometry [71, 72], and through the application of X-ray crystallography and cryogenic electron microscopy (cryo-EM) techniques [73-75]. Collectively, findings from these studies have revealed that the spliceosome is dynamic and assembles onto introns stepwise, forming discrete complexes that are broadly conserved and proceed through four major stages: assembly, activation, splicing, and disassembly (Figure 1.7). During assembly, splice-site sequences are recognized and paired together for splicing. Activation generates the catalytic centers required for catalysis, and splicing generates the intron lariat and mRNA products. After splicing, the snRNPs and other splicing factors are disassembled, releasing the mature mRNA, and recycling the spliceosomal components for further rounds of intron processing.
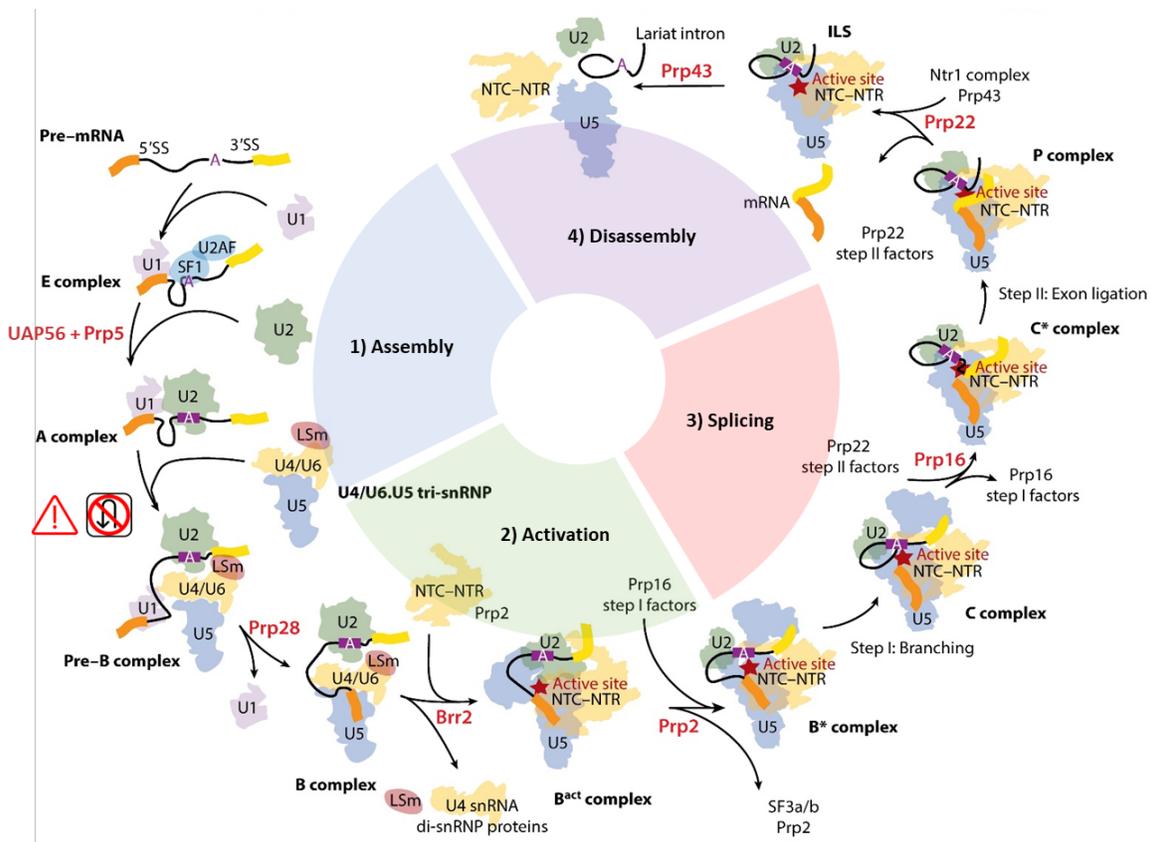
Figure 1.7 – Schematic of step-wise spliceosome assembly and the splicing cycle.
Assembly begins with splice-site recognition by the U1 and U2 snRNPs in the E and A
complex, respectively. Once the U4/U6.U5 tri-snRNP associates with the A complex to form
the pre-B complex the intron is committed to splicing as denoted by the caution signage at this
step in the cycle. Activation begins after the formation of the pre-B complex starting with
displacement of U1 and replacement by U6 at the 5′-ss forming the B complex. The catalytic
center is formed after U4 is displaced causing further hybridization between the U2 and U6
snRNAs in the B^{act} complex which is fully activated by the displacement of the SF3A and
SF3B U2-specific proteins forming the B* complex. The first step of splicing occurs during
conversion of the B* complex to the C complex which is further rearranged to place the 3′-ss
into the catalytic center near the 5′ end of the liberated exon in the C* complex which
catalyzes the second step of splicing and converts to the P complex. Disassembly begins with
release of the mature mRNA from the P complex forming the intron lariat spliceosome (ILS)
that is further taken apart to recycle the U2,U5, and U6 snRNPs and release the intron lariat.
Figure adapted from Wilkinson, M. E., Charenton, C. & Nagai, K. 2020.

Assembly begins with the formation of the E complex produced by binding of the

U1 snRNP to the 5′-ss sequence and by association of the proteins splicing factor 1 (SF1),

and U2 snRNA auxiliary factors 65 (U2AF65) and 35 (U2AF35) to the branchpoint, Py

tract, and 3′-ss sequences, respectively [69, 76] (Figure 1.7). The E complex can form in the absence of ATP and the U1 snRNA binds the 5′-ss via a conserved sequence of 8 nts at its 5′ end that is complementary to the 5′-ss consensus allowing for formation of an RNA duplex that is stabilized by the U1 specific protein U1C [77, 78]. The stable association of the U2 snRNP to the branchpoint occurs in the next stage of prespliceosome assembly and is the first ATP-dependent step in the splicing cycle requiring the actions of the RNA helicases Prp5 and UAP56 [79, 80]. The A complex is formed after SF1 is displaced and replaced by the U2 snRNA which directly base pairs with the branchpoint sequence via a conserved sequence of its own, in coordination with U2 specific proteins [81, 82] and U2AF65 [83]. The direct hybridization of U2 to the branch point is like that of U1 at the 5′-ss and is a common mechanism of snRNP mediated splice-site selection observed throughout the splicing cycle. Splice-site sequences in the A complex are thought to be paired for splicing via cross-intron interactions between the 5′-ss and branchpoint bound U1 and U2 snRNPs [84-86]. The A complex then recruits the U4/U6.U5 tri-snRNP by interactions between the U2 specific splicing factor 3B1 (SF3B1) set of proteins and the U6 LSm-ring, in addition to base-pairing between the 5′ end of U2 and the 3′ end of U6 snRNAs [87, 88]. With the loose association of the tri-snRNP to the A complex, the pre-B complex is formed and contains all five spliceosomal snRNPs forming the earliest detectable spliceosomal complex.

Activation of the spliceosome requires the stable association of the U4/U6.U5 tri-snRNP with the intron, and the formation of the catalytic centers needed for splicing (Figure 1.7). This is achieved by release of the U1 snRNP from the 5′-ss mediated by the activity of the Prp28 helicase delivered to the U1/5′-ss duplex by the tri-snRNP [89, 90].

The U1 snRNP is replaced by non-Watson Crick base-pairing between the 5′-ss and a region of U6 snRNA called the ACAGAGA box [91, 92]. This 5′-ss transfer from U1 to U6 produces the B complex which is converted to the B^{act} complex via unwinding of the U4/U6 duplex by Brr2 helicase activity leaving behind the minimal U2/U6.U5 spliceosome [93, 94]. Formation of the B^{act} complex induces folding of the U6 snRNA that produces the reactive internal stem-loop (ISL) structure, in addition to further base-pairing between U2 and U6 that collectively juxtapose the 5′-ss and branch-point sequence within a catalytic center primed for the first step of splicing [95, 96]. Conversion to the catalytically active spliceosome B* requires Prp2 helicase activity which displaces U2 specific-proteins from the branchpoint sequence thereby freeing the branchpoint adenosine to attack the 5′-ss [97-99].

The first step of splicing occurs during the B* to C transition where the U6-ISL coordinates two Mg^{2+} metal ions in the catalytic center which stabilizes the leaving groups of the transesterification reaction [95, 100] in a fashion identical to that of group II self-splicing introns [101]. The first step of splicing produces the intron-lariat intermediate and liberates the upstream exon which is kept in proximity to the catalytic center by base-pairing to the single-stranded loop in stem-loop 1 of the U5 snRNA [102-104]. Prp16 hydrolyzes ATP and translocates the 3′-ss into the catalytic center while dissociating step I-associated splicing factors, converting the C complex into the step II activated C* complex [103, 105]. The second transesterification step occurs and links upstream and downstream exons forming mature mRNA which is released from the post-splicing P complex by the action of helicase Prp22 [106, 107]. In the final stage of the splicing cycle, the U2/U6.U5 spliceosome is disassembled by the helicase Prp43 which

resets the U2/U6 duplex and the U6-ILS [108-110]. Thus, the spliceosome is a bonafide ribozyme, catalyzing the splicing reaction via RNA-mediated transesterification at catalytic sites supported by many RNA and protein-based structures formed through numerous and complex helicase-driven rearrangements. Understanding how the interactions between the various spliceosomal complex transitions are established and reformed is therefore critical in understanding pre-mRNA splicing and gene expression in eukaryotes.

Prespliceosome assembly is targeted during regulation of alternative splicing

Precise splice-site recognition and inter-snRNP interactions that pair splice-site sequences together during prespliceosome formation commit an intron for removal prior to spliceosome formation and are critical for accurate pre-mRNA splicing (Figure 1.7) [111-113]. Therefore, splicing regulation primarily occurs at the prespliceosomal stages during E and A complex formation [114, 115]. Alternative splicing is regulated largely by the presence of intronic and exonic *cis*-acting sequences recognized by *trans*-acting RNA binding proteins (RBPs) which influence the recognition and pairing of 5′- and 3′-ss sequences (Figure 1.8). Sequences in pre-mRNA that stimulate splice site selection are called intronic or exonic splicing enhancers (ISE or ESE, respectively) and are recognized by SR proteins harboring characteristic RS domains rich in arginine and serine residues [116]. These proteins also have canonical RNA recognition motifs (RRMs) used to bind ISEs and ESEs and their activities are regulated by phosphorylation of their RS domains [117, 118].

Pre-mRNA

Intron ....   Exon....   Intron ....

SR proteins (+)
hnRNP proteins (−)

Figure 1.8 The splicing code includes splicing enhancer and silencer sequences in introns and exons that influence splice-site recognition by snRNPs. SR proteins are commonly recruited to exon splicing enhancers (ESE) and intron splicing enhancers (ISE) to promote splicing at specific splice sites (green plus signs). By contrast exon splicing silencers (ESS) and intron splicing silencers (ISS) recruit hnRNP-type proteins to repress splice-site recognition (red minus signs). Figure adapted from Lee, Y, and Rio, D. C. 2015.

SR proteins promote splicing by recruiting or stabilizing prespliceosomal components to splice-site sequences in a fashion that is highly conserved among eukaryotes [119-121]. By contrast, the *cis*-acting sequences that repress splice-site recognition are called intronic or exonic splicing silencers (ISS or ESS) and recruit a different class of splicing factor referred to as heterogeneous ribonucleoproteins (hnRNPs). Binding of hnRNPs to ISS and ESS can competitively mask recognition of splice-sites sequences or interfere with the communication and pairing between the 5′- and 3′-ss complexes. (Figure 1.8).

For example, in the brain of humans and mice, a specific isoform of the tyrosine kinase encoding *c-Src* gene differentially splices a neuronal specific, 18 nts long cassette exon called N1 between constitutive exons 3 and 4 which alters the kinase's network of available protein-protein interactions leading to brain specific activities [122, 123]. Splicing of the *c-Src* transcript is under regulation of the poly-pyrimidine tract binding protein (PTB or hnRNP I) that binds CU-rich sequences upstream and downstream of the

17

N1 exon [124]. Non-neuronal cells express PTB at high levels which binds to these flanking ISS sequences causing exon skipping. In neurons however, PTB expression is low and surpassed by that of an inert isoform called neuronal PTB (nPTB) which may compete for the ISS sites without suppressing N1 exon recognition, facilitating N1 inclusion in neural *c-Src* pre-mRNA [125]. PTB binding can interfere with splice-site pairing interactions during two stages of prespliceosome formation. In the E complex, PTB can block the communication between 5′-ss bound U1 and the U2AF65/35 complex at the 3′-ss [125]. Additionally, PTB can interfere with progression past the spliceosome A complex by blocking interactions between the U1 and U2 snRNPs across the intron [126]. Interfering with splice-site pairing independent of prespliceosome formation inhibits intron commitment diverting spliceosome assembly to alternative sites and is a mechanism observed by other splicing factors that bind ESS in their target substrates like hnRNP L and RBM5 [113, 115, 127, 128].

Alternative splicing is therefore a highly regulated process that integrates information from a larger splicing code to shift the profile of splicing patterns in eukaryotic cells. The prespliceosome offers a short window of opportunity by which SR proteins and hnRNPs can intervene before the spliceosome is formed and introns are committed to splicing. Elucidating the interactions involved in prespliceosome formation is therefore vital for understanding mechanisms of AS regulation.

Splicing dysregulation in the pathology of disease and myelodysplastic syndromes

The presence of split genes imparts remarkable benefits to eukaryotes, but not without a cost. Splicing dysregulation is a common feature of many diseases and it has

been estimated that approximately 15% of all disease-causing point mutations occur at 5′- and 3′- splice sites [129]. Mutations that affect splicing can also occur in *cis*-regulatory enhancer/silencer sequences, disrupting existing sites or even generating novel sites which alter normal splicing pathways [130]. Generation of novel splice-sites are observed in β-thalassemia for example. People with β-thalassemia completely lack, or have reduced expression of, functional β-globin; a protein that forms two of the four subunits in hemoglobin required for oxygen transport by red blood cells. Many mutations in the β-globin gene *HBB* can cause β-thalassemia, however, numerous ethnically-linked mutations that generate novel splice sites have been observed [131]. In a Chinese population, a C to T point mutation introduces a novel 5′-ss in the second intron of *HBB* increasing the length of exon 2 by 73 nts [132]. Additionally, for those with β-thalassemia in the Mediterranean regions and Middle East, a G to A point mutation is disproportionately enriched in the first intron of *HBB* that generates a novel 5′-ss [133]. Both mutations introduce premature stop codons in the mis-spliced isoforms that terminate β-globin expression [134].

Mutations that affect splicing can also occur in the *trans*-acting components of the spliceosome and have been detected in snRNPs and auxiliary factors that can lead to gain- or loss-of-function. This has been well documented to occur in another disease of the blood; myelodysplastic syndromes (MDS). MDS is a disease of aging, occurring most commonly in individuals over the age of 65, and is the result of aberrant hematopoiesis due to dysregulated differentiation of hematopoietic stem cells (HSCs) (Figure 1.9) [135, 136]. In MDS, an accumulation of non-differentiated HSCs and a reduction in the levels of mature myeloid cells leads to cytopenia-related symptoms like hypoxemia, fatigue,

bruising, and an increased risk of cancer. Approximately one third of those with MDS

progress to more severe, secondary acute myeloid leukemia [137].
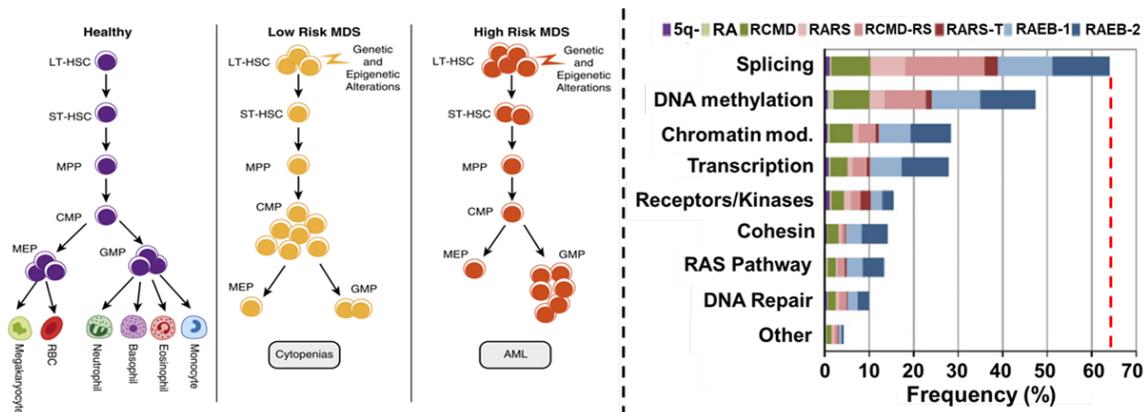
Figure 1.9 – Myelodysplastic syndromes (MDS) are a result of aberrant hematopoiesis and a high frequency of splicing factor mutations are observed in this disease. Left panel – Normal hematopoietic stem cells (HSCs) differentiate into mature red and white blood cells. In MDS however, HSCs that develop into the myeloid lineages (far left) acquire genetic mutations that disrupt differentiation and leading to accumulation of HSCs and cytopenia (middle). Those affected by MDS are also at an elevated risk of cancer and frequently develop acute myeloid leukemia (AML, right). Right panel – The frequency at which mutations are detected in certain cellular pathways in patients with MDS. Splicing is one of the most frequently mutated pathways with 50-60% of those affected harboring at least one mutation in a splicing factor gene. Figures adapted from: Left panel – Shastri et al. 2017. Right panel – Haferlach et al. 2014.

In addition to various genetic lesions and alterations in epigenetic regulation

found in MDS patient-samples, there is also a disproportionately high rate of mutations

detected in proteins involved in prespliceosome formation, specifically in those factors

involved in 3′-ss recognition (Figure 1.10) [138]. A majority of MDS patients (~75-90%)

have some form of genetic mutation thought to be acquired *de novo* in their HSCs, and

splicing factor mutations have been estimated to occur earliest, indicating that they may

serve as drivers of the disease [139, 140]. Studies applying next-generation sequencing

technologies were the first to identify enrichment of splicing factor mutations in MDS

patient samples with the most frequent mutations consistently observed in the U2 snRNP

specific protein SF3B1, the SR protein SRSF2, and the 3′-ss binding factor U2AF35 [140, 141]. Mutations in other U2 proteins like SF3A1 were also observed, in addition to the alternative splicing factor ZRSF2, and the branchpoint and Py binding proteins SF1 and U2AF35 (Figure 1.10). The disproportionate rate of mutations acquired in early spliceosomal factors in MDS highlights how vital proper assembly of the prespliceosomal complex is for accurate splicing regulation and maintenance of healthy cells. Mutations in SF3B1, SRSF2 and U2AF35 are also recurrently mutated in other diseases like breast cancer [142, 143], pancreatic cancer [144] and related blood disorders like chronic lymphocytic leukemia [145].
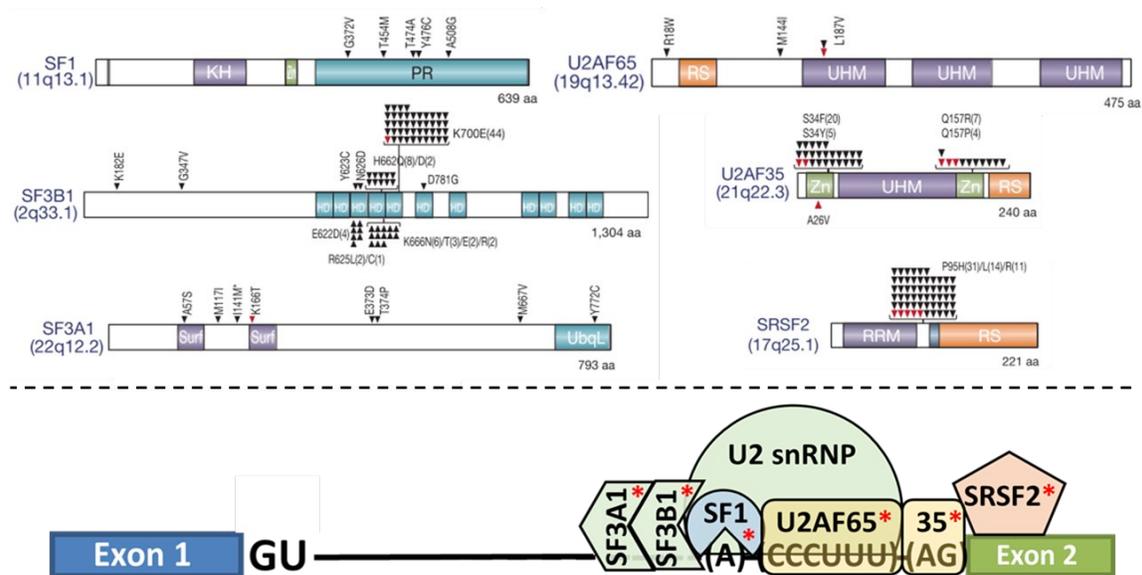
Figure 1.10 – MDS-associated mutations in splicing factors are disproportionately represented in proteins involved in 3′-ss recognition. Top panel – Schematics of the domain organization in the splicing factors frequently mutated in MDS. The positions of point mutations detected in patients are indicated by arrows. Bottom panel – Cartoon diagram showing the general organization of the MDS-associated proteins forming the 3′ ss complex. Mutations in these proteins can influence formation of this complex leading to splicing dysregulation and disrupted hematopoiesis. Figure adopted from: Top panel – Yoshida et al. 2011.

Mutations in splicing factors are also thought to lead to different gain-of-function activities that can have unique effects on basal splicing patterns resulting in distinct disease progression and characteristics [146]. Patients with a subtype of MDS characterized by a specific type of dysplastic cell called ring sideroblasts are heavily associated with mutations in SF3B1 for example [147]. In MDS, these point mutations are largely heterozygous and mutually exclusive, producing cells with one mutant copy of the gene rarely in combination with other splicing factor mutations [141]. It is thought that cancer cells harboring homozygous mutations or alterations in multiple splicing factor genes are too destabilizing and non-viable and are therefore not selected for during cancer growth. This is the basis for the use of spliceosome targeting small molecules which may induce synthetic lethality in those cells with already disrupted spliceosomes like the HSCs in MDS [146]. A deeper understanding of prespliceosome formation and the interactions involved is therefore critical to better understand the consequences of splicing factor mutations, and to identify novel interactions that may yield viable targets for the development of therapeutics to effectively treat diseases like MDS and other cancers.

U1 snRNP as a hub for early spliceosomal interactions

The U1 snRNP, in addition to being a critical regulator of prespliceosome formation, is also involved in other transcription related RNA processing. Because of this, U1 is frequently the target of *trans*-acing factors and studying these interactions has revealed a lot about the nature of RNA-protein interactions and regulation of gene expression in eukaryotes. The first RNA recognition motif to be solved by x-ray

crystallography was that of the U1 specific protein U1A [148] and revealed that these domains contain two-α helices packed against a four membered β-sheet in a rather simple $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ topology [149]. Four years later, the structure of U1A bound to its natural ligand revealed how two conserved ribonucleoprotein (RNP) motifs that define classical RRMs bind RNA through base stacking and ionic interactions with target sequence [150]. Initial attempts to solve the structure of the fully intact U1 snRNP purified from nuclear extract had limited resolution [151] however X-ray crystal structures of functional U1 snRNPs assembled *in vitro* have revealed the 3-dimensional organization of U1 in great detail [77, 152].

The U1 snRNA folds into a relatively basic structure containing four stem-loops with the first three assembled together at a tight four-way junction towards the 5′ end of the RNA (Figure 1.11). The seven-membered Sm-ring assembles on a short single-stranded region of RNA just downstream from stem-loop 3 (SL3) at the conserved Sm site and this structure separates stem-loop 4 (SL4) from the body of the snRNP. The RRM of U1-70k binds stem-loop 1 (SL1) and a long N-terminal alpha helix projects into the body of snRNP contacting the Sm ring. Both U1-70k and the Sm ring act as a scaffold for the association of U1C near the 5′ end of the U1 snRNA that serves to stabilize the U1-5′-ss interaction [153, 154]. The final U1 snRNP specific protein U1A binds to the loop of stem-loop 2 (SL2) via its N-terminal RRM. SL3 and SL4 are oriented opposite to one another and are the only regions of the snRNA not bound by U1 specific protein, remaining available for binding by *trans*-acting factors (Figure 1.11). Nearly all the U1-specific proteins and stem-loop structures play a role in the regulation of splicing and transcription.

23
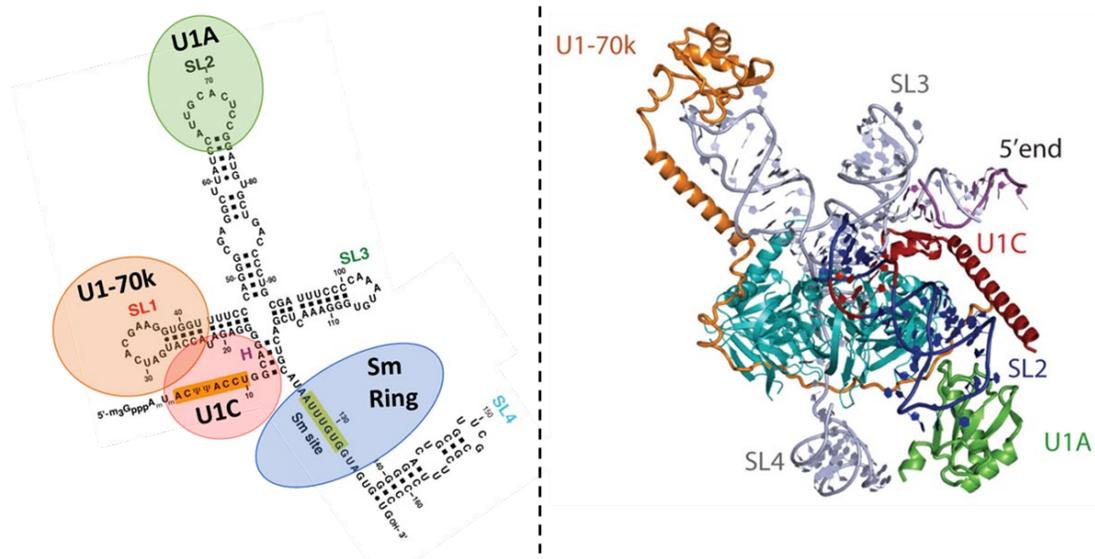
Figure 1.11 – The crystal structure of the mature U1 snRNP. Left panel – A cartoon depiction of the U1 snRNA secondary structure along with the general positions of the three U1-specific proteins and the Sm ring. Right panel – The X-ray crystal structure of the in vivo reconstituted U1 snRNP confirms the general organization of the schematic shown in the left panel. The snRNA is shown in gray and protein components are color coated as in the left panel. Importantly, the U1 SL3 and SL4 structures are confirmed to be un-bound by any U1-specific protein, allowing access to binding by trans-acting factors. Figures adopted from: Left panel – Kondo et al. 2015. Right panel – Krummel et al. 2009.

In the cell nucleus, splicing is a co-transcriptional process that functions in tandem with transcription, removing introns during synthesis of nascent pre-mRNA [155]. These processes are mutually reinforcing and the U1 snRNP plays a central role in coupling splicing with transcription via direct and indirect interactions with RNA polymerase II (RNAP II) [156, 157]. Mass spectrometry of RNAP II complexes affinity purified from nuclear extract detected co-purification of U1 and high levels of SR proteins, but no other spliceosomal snRNPs [158]. A recent cryo-EM structure of the U1 snRNP in complex with RNAP II assembled onto a DNA-RNA scaffold revealed that the positively charged α-helices of U1-70k directly contact a negative pocket formed by RNAP II domains RPB2 and RPB12, tethering U1 to transcription [159].

The U1 snRNP has also recently been observed to suppress premature cleavage and poly-adenylation (PCPA) in a process called telescription. During standard transcription termination, *cis*-acting polyadenylation signals (PAS) in the 3′ untranslated regions of pre-mRNA recruit numerous cleavage and polyadenylation factors (CPAFs) including an endonuclease and a poly(A) polymerase that release nascent pre-mRNA once fully transcribed. PCPA however occurs at PAS sites found typically within the first intron of genes leading to rapidly aborted transcription and the production of truncated pre-mRNA transcripts [160]. In addition to 5′-ss recognition, the U1 snRNP functions to suppress PCPA in a fashion dependent on binding of U1 to pre-mRNA, serving as a green light that allows transcription to proceed past PCPA checkpoints [161]. Cells transfected with an anti-sense morpholino oligonucleotide (AMO) complementary to the 5′ end of the U1 snRNA block the U1-pre-mRNA interaction and induces PCPA in a dose dependent manner that is conserved in mouse and fruit fly systems [162]. All U1 snRNP-specific proteins U1A, U1-70k, and U1C can interact directly with various CPAFs, however U1A appears to be the functional CPA-inhibiting factor that is also disrupted in the presence of U1 AMO, derepressing PCPA which subsequently aborts transcription [163, 164]

The SL3 and SL4 structures of the U1 snRNA also interact with *trans*-acting factors. U1-SL3 has been reported as a target of the fused in sarcoma (FUS) protein [165]. This interaction may contribute to the aberrant accumulation and mis-localization of the U1 snRNP in cytoplasmic stress-granules frequently observed in FUS-linked amyotrophic lateral sclerosis, and may also act as a splicing enhancer by mediating recruitment of U1 to the 5′-ss [165]. The FUS-SL3 interaction may also be involved in

co-transcriptional splicing since FUS interacts with RNAP II and has been reported to mediate the interaction of U1 with the polymerase [166]. Similarly, U1-SL4 has been found to be targeted by the hnRNP protein PTB in the regulation of *c-Src* pre-mRNA. Intron bound PTB does not repress splicing by interfering with U1 hybridization to the 5′-ss [125]. Instead, intron bound PTB binds to U1-SL4 by either one of its two N-terminal RRM domains which in-turn interferes with splice-site pairing interactions across the intron between the 5′- and 3′-ss bound U1 and U2 snRNPs, causing skipping of N1 [167]. Further studies into the function of U1-SL4 during prespliceosome formation revealed that the U2-specific splicing factor 3A1 (SF3A1) contacts U1-SL4 during A complex formation and this SL4-SF3A1 contact was proposed to serve as a cross-intron bridge that stabilizes splice-site pairing [168].

Hypothesis and Specific Aims

The U1 snRNP therefore serves as an important platform for a variety of factors that target U1 to promote or inhibit the transcription and splicing of pre-mRNA. The interactions made by the U1 snRNP are especially important during splice-site pairing as they influence prespliceosome formation by communicating directly and indirectly with complexes formed at the 3′-ss during assembly of the spliceosome [112, 125, 168, 169]. Even after 40 years of research, the early stages of spliceosome assembly remain to be the least well characterized, and novel U1 snRNP interactions continue to be elucidated. Identifying U1 snRNP binding proteins, and characterizing the nature of these interactions, will provide critical insight into the formation and regulation of splice-site pairing, a vital stage of prespliceosome assembly that commits introns for splicing.

I hypothesize that the U1 snRNA plays an active role in splice-site pairing via its interactions with splicing factors that are vital for functional prespliceosome assembly. To test this hypothesis, I developed and applied numerous *in vitro* biochemical assays to characterize and identify U1-SL3 and U1-SL4 interacting proteins, and leveraged a splicing reporter gene assay to assess the extent to which these structures contribute to U1 splicing activity *in vivo*. The techniques and methods used to investigate this hypothesis are described briefly in the aims below, and the results and analysis from the associated experiments are reported in the following chapters.

Aim #1: Identify and characterize the RNA binding domain of SF3A1. (Chapter 2)

The U2-specific protein SF3A1 was identified as a U1-SL4 interacting splicing factor involved in the early stages of spliceosome assembly. However, without an obvious RNA binding domain, it is unclear how SF3A1 binds U1-SL4. Using a number of SF3A1 truncated proteins expressed *in vitro*, I isolated the essential U1-SL4 RNA binding domain and applied various biochemical techniques to quantify the strength, and assess the specificity of this RNA-protein interaction. Pursuing this aim led to the identification of the Ubiquitin-like domain as a non-canonical RNA binding domain in SF3A1 and the finding that an MDS-associated point mutation (Y772C) in this region reduces SF3A1 RNA binding activity.

Aim #2: Assess the contribution of U1-SL3 towards U1 splicing activity and identify SL3 interacting factors. (Chapter 3)

The U1-SL3 is accessible for binding by splicing factors and, like U1-SL4, may also play a role in splicing. To test this, the impact of U1-SL3 mutations on U1 splicing activity was determined by applying a minigene splicing reporter assay performed in transfected HeLa cells. Additionally, U1-SL3 interacting factors were identified by UV crosslinking experiments in HeLa cell nuclear extract in addition to RNA affinity purification of U1-SL3 bound protein complexes followed by mass spectrometry. Mutations in U1-SL3 were found to reduce U1 splicing activity and the ATP-dependent DExD/H box RNA helicase UAP56 that is involved in prespliceosome assembly, was found to preferentially bind the U1-SL3 structure with remarkable specificity.

Aim #3: Elucidate the functions of U1-SL3 and U1-SL4 during the early steps of spliceosome assembly. (Chapter 4)

It was observed that mutations in either SL3 or SL4 do not completely abolish U1 splicing activity, potentially indicating that these structures alone still support splicing via interactions with their respective binding partners characterized in Aims #1 and #2. In this Aim, I examined the consequences of introducing double mutations in both SL3 and SL4, and applied an siRNA-based approach, in combination with the minigene splicing reporter, to validate the functional significance of the SL3- and SL4-specific binding proteins. These experiments revealed that the functions of SL3 and SL4 are interdependent since double mutations have more than additive effects on U1 activity, and that this effect can be recapitulated by combining protein knockdown with single

28

stem-loop mutations. Additionally, *in vitro* splicing assays and UV-crosslinking

experiments performed using HeLa nuclear extract revealed a potential mechanism by

which U1-SL3 enhances prespliceosome assembly via U1-SL4.

CHAPTER 2

IDENTIFYING THE U1 STEM-LOOP 4 BINDING DOMAIN OF
SPLICING FACTOR SF3A1

Publication Note

The research reported in this chapter was previously published in *RNA*. William

Martelly, Bernice Fellows, Kristen Senior, Tim Marlowe and Shalini Sharma.

Identification of a noncanonical RNA binding domain in the U2 snRNP protein SF3A1.

RNA 2019 25: 1509–1521. All co-authors have granted permission for this work to be

included in this dissertation.

Overview

Removal of introns and ligation of exons in pre-mRNAs is catalyzed by the

spliceosome, a dynamic complex comprising five small nuclear ribonucleoproteins

(snRNPs) (U1, U2, U4, U5, and U6) and many auxiliary proteins [170]. Spliceosome

assembly occurs *de novo* onto each intron and proceeds through a series of intermediate

complexes. Stable binding of the U1 snRNP to the 5′ splice site and of the U2 snRNP to

the branch point sequence forms the prespliceosomal A complex that interacts with the

preformed U4/U6.U5 tri-snRNP to generate the spliceosomal pre-B complex. Subsequent

to this, extensive structural and conformational remodeling leads to formation of at least

six distinct complexes that are referred to as B, B^{act}, B*, C, C*, and P [87, 88, 94, 96,

171-174]. The catalytic steps of splicing occur during the B*→C and C*→P transitions.

Cryo-EM analyses have revealed in great detail the compositional and structural changes

that occur in spliceosomal complexes after binding of the tri-snRNP and during the

transitions that accompany the catalytic steps. However, interactions that occur during the very early steps of spliceosome assembly, prior to binding of the tri-snRNP, remain to be elucidated.

Spliceosome assembly begins with binding of the U1 snRNP to the 5′ splice site, splicing factor 1 (SF1) to the branch-point, and the U2 auxiliary factor 65 (U2AF65 or U2AF2) and U2AF35 (U2AF1) to the polypyrimidine tract and the 3′ splice site, respectively [170]. The U2 snRNP has been reported to associate with this early (E) complex via interactions between SF3B1 and U2AF65 [175, 176]. Stable binding of the U2 snRNP to the branchpoint forms the prespliceosomal A complex. Cryo-EM analysis of the yeast prespliceosome has provided some insight into the U1-U2 interface and identified two regions of contact between components of pre-mRNA bound U1 and U2 snRNPs [86]. The first interface forms after a stable interaction occurs between yeast U1 specific protein Prp39 and the core U2 protein U2A′ (Lea1 in yeast). The second interface involves interactions of yeast U1 snRNA stem-loop 3 (SL3) with the SF3B complex protein SF3B3 (SF3B130; Rse1 in yeast) and with the SF3A complex protein SF3A3 (Prp9 in yeast). However, the human U1 and U2 snRNPs differ significantly from those of yeast [152, 177]. The human U1 (164 nts) and U2 (188 nts) snRNAs are considerably shorter than their yeast orthologues, which are 568 and 1175 nts long, respectively. The human U1 snRNP consists of three particle specific proteins, U1-70k, U1C and U1A. On the other hand, the yeast U1 contains seven additional particle specific proteins, namely Prp39, Prp40 (human Prp40 or FBP11), Prp42, Nam8 (human TIA-1), Snu56, LUC7 (human LUC7L), and Snu71 (human RBM25). The human and yeast U2 snRNPs have similar numbers of particle specific proteins, 7 and 6, respectively.

31

However, the primary structures of some of these proteins are significantly different. For example, the SF3A1 protein that is relevant to this study is 793 amino acids (aa) in humans, and its yeast counterpart, Prp21, is only 280 aa [178]. Thus, it is likely that the U1-U2 interface contacts during the early steps of human spliceosome assembly differ from those observed in yeast.

Previously, we reported that during spliceosome assembly, the SF3A complex engages in direct contact with the U1 snRNP via interactions between the SF3A1 protein and stem-loop 4 of the U1 snRNA (U1-SL4). This interaction occurs between pre-mRNA bound U1 and U2 snRNPs and was found to be critical for formation of the prespliceosomal A complex [168]. The SF3A1 protein (120 kDa) interacts with SF3A2 (66 kDa) and SF3A3 (60 kDa) to create the SF3A complex, which is necessary for formation of the mature 17S U2 snRNP and for pre-mRNA splicing *in vitro* and *in vivo* [179-181]. Notably, SF3A1 lacks a conventional RNA-binding domain. At its N-terminus, SF3A1 contains two suppressor-of-white-apricot (S1 and S2) domains and a short segment of charged residues (Figure 2.1). SF3A1 interacts with SF3A3 through a region harboring the S2 domain (aa 145-243) and with SF3A2 via a 26 residue region (aa 269-295) [182]. The C-terminal region of SF3A1 harbors a nuclear localization signal
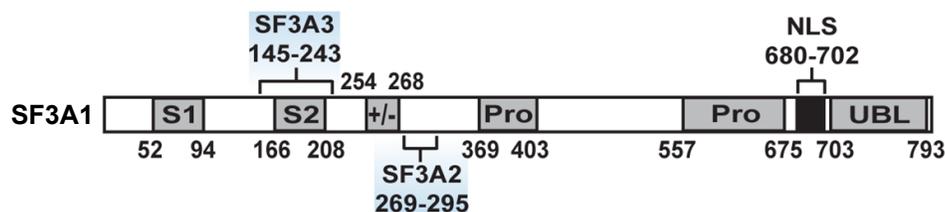
Figure 2.1 – The U2-specific splicing factor SF3A1 is a 120 kDa protein with no obvious RNA binding domain. Schematic diagram of human SF3A1 protein depicting the known domain organization, and the regions that interact with SF3A2 and SF3A3.

(NLS) and a Ubiquitin-like (UBL) domain. The N- and C-terminal domains are separated by a long central region that contains two proline-rich segments (Pro). Integrity of all these regions is important for normal physiology as several point mutations that occur across the entire length of SF3A1 are known to be associated with hematological diseases including myelodysplastic syndromes (MDS), chronic myelomonocytic leukemia (CMML), and acute myeloid leukemia (AML) [139, 141].

In this study, we report that the UBL domain of SF3A1 is a non-canonical RNA-binding domain. Applying a combination of techniques including UV-crosslinking, electrophoretic mobility shift assays (EMSAs), and surface plasmon resonance (SPR), we demonstrate that the SF3A1-UBL domain binds U1-SL4 with high affinity ($K_D = \sim 97$ nM). Investigations of the impact of an MDS-linked residue Y772, and an adjacent highly conserved residue Y773, revealed that Y772C and Y773C mutations decrease the affinity of SF3A1-UBL for U1-SL4 and reduced its ability to interact with the U1 snRNP. Since Prp21, the yeast ortholog of SF3A1, lacks the C-terminal region that harbors the UBL domain [178], we propose that the U1-SL4/SF3A1 contact may be unique to the U1-U2 interface during the early steps of human spliceosome assembly.

Results

SF3A1 interacts with U1-SL4 through the C-terminal Ubiquitin-like domain

Human SF3A1 is a 793 aa protein that lacks a canonical RNA binding domain (Figure 2.1). To identify the U1-SL4 binding domain in SF3A1, we created C-terminally 6xHis-tagged constructs for expression of full-length (FL) SF3A1, and a series of N-terminal (ΔN) and C-terminal (ΔC) deletions (Figure 2.2). These constructs were
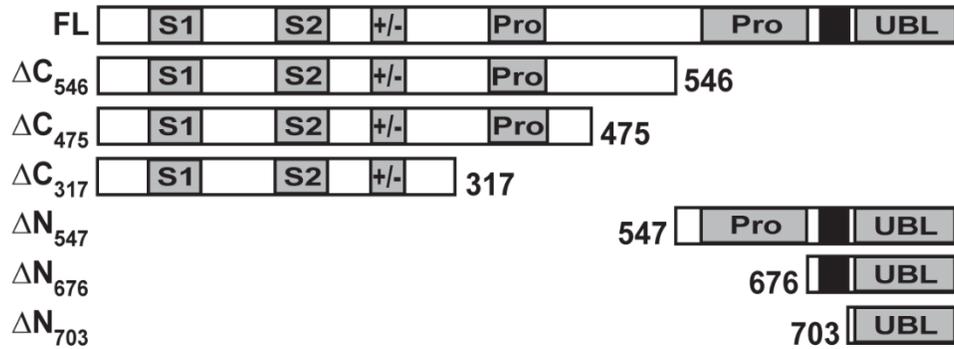
Figure 2.2 – Full-length SF3A1 and truncated proteins were expressed using a HeLa cell based, cell-free expression system. Schematic of all N- and C-terminal deletion constructs tested for U1-SL4 binding activity.

expressed by the HeLa cell lysate based in vitro cell-free expression (CFE) system

(Thermo Fisher Scientific) that uses coupled transcription-translation reactions. Western

analysis with an anti-6xHis antibody demonstrated efficient expression of all FL, ΔN, and

ΔC SF3A1 proteins (Figure 2.3 Left Panel).
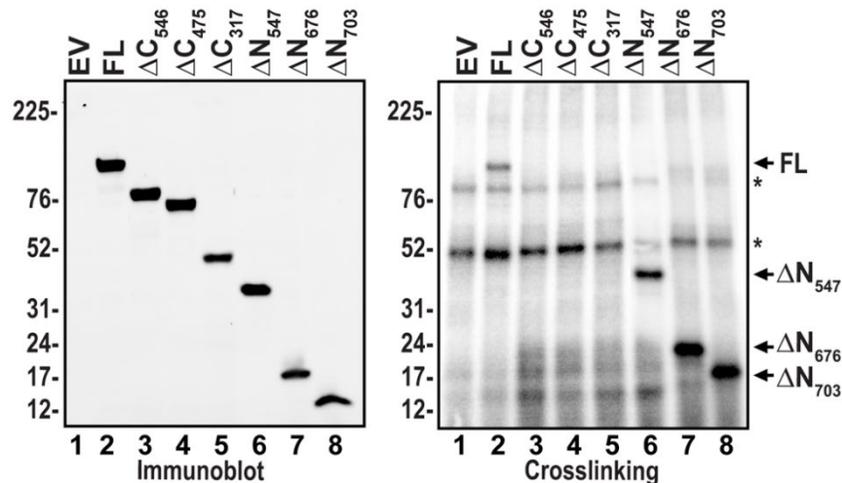
Figure 2.3 – U1-SL4 RNA cross-links to the C-terminal Ubiquitin-like domain of SF3A1.
Left panel – Western blot of 6xHis-tagged protein expressed in CFE reactions containing the control empty vector (EV), or expression vectors with full-length (FL) protein, C-terminal truncations (ΔC_{546}, ΔC_{475}, ΔC_{317}), and N-terminal truncations (ΔN_{547}, ΔN_{676}, ΔN_{703}). Proteins were detected using anti-6xHis primary antibody. Right Panel – ^{32}P-U1-SL4 was added to CFE extracts containing SF3A1 proteins, UV-crosslinked, and then separated on SDS-PAGE gels and visualized by phosphor imaging. Arrows indicate specific crosslinked products while asterisks indicate non-specific products that were observed in all reactions including the empty vector control.

34

To identify the U1-SL4 interacting domain in SF3A1, binding reactions were

assembled by adding uniformly ^{32}P-labeled U1-SL4 RNA to the expressed proteins under

splicing conditions. The reactions were UV-crosslinked and separated by SDS-PAGE.

This analysis showed that U1-SL4 crosslinked to FL SF3A1, but not to C-terminal

truncations, ΔC_{546}, ΔC_{475}, or ΔC_{317} (Figure 2.3 Right Panel – compare lane 2 with lanes

3-5). All N-terminal truncations, however, retained the capacity to crosslink to U1-SL4,

including ΔN_{703} that only contains the UBL domain (lanes 6-8). There was some non-

specific crosslinking in all reactions, which was also observed in the empty vector (EV)

control (lane 1). However, specific binding of SF3A1-FL and ΔN proteins to ^{32}P-U1-SL4

was clearly identifiable by the appearance of appropriate size bands. To check the

specificity of binding, we employed the U1-SL4/M10h mutant that was previously found

to reduce U1 snRNP splicing activity (Figure 2.4). Crosslinking analysis with FL protein

demonstrated binding of SF3A1 to wildtype U1-SL4, but not to M10h (Figure 2.4,

compare lanes 3 and 4). Similarly, all N-terminal truncations ΔN_{547}, ΔN_{676}, and ΔN_{703}

crosslinked to wildtype U1-SL4 (Figure 2.5 lanes 3, 5, and 7), but not to the mutant

M10h (lanes 4, 6, and 8). Thus, the crosslinking analysis demonstrated that the U1-SL4

interacting domain resides in the C-terminal region of SF3A1, from residues 703-793 that

harbors the UBL domain.
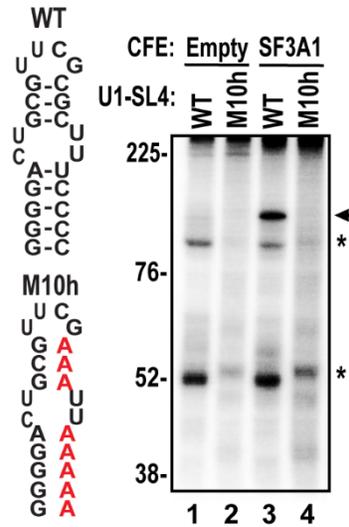
Figure 2.4 – FL SF3A1 expressed in CFE lysate crosslinks WT U1-SL4 but not to a mutant RNA. To the left of the gel, the schematic of wildtype (WT) and mutant (M10h) U1-SL4 RNAs is displayed. Crosslinking of ^{32}P-U1-SL4-WT and -M10h RNAs in CFE extracts expressing EV or FL SF3A1. Arrows indicate specific crosslinked products while asterisks indicate non-specific products that were observed in all reactions including the empty vector control.
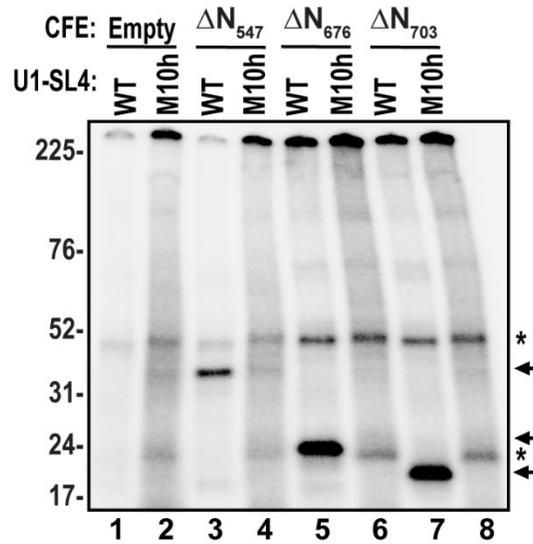
Figure 2.5 – Truncated SF3A1 expressed in CFE lysate crosslinks WT U1-SL4 but not to a mutant RNA. Crosslinking of WT or M10h ^{32}P U1-SL4 in CFE extracts expressing EV or ΔN SF3A1 constructs. Arrows indicate specific crosslinked products while asterisks indicate non-specific products that were observed in all reactions including the empty vector control.

36

Ubiquitin is a 76 aa protein consisting of two alpha helices and five beta-sheets that fold into a ββαββαβ topology [183]. A previous analysis showed that the C-terminus of SF3A1 (aa 714-790) shares ~29.6% identity and ~54.9% similarity with Ubiquitin [184]. However, the region of SF3A1 that folds into a Ubiquitin-like motif is 90 aa, encompassing residues 703-793 (Figure 2.6). A comparison of the structures and sequences of Ubiquitin and SF3A1-UBL revealed that the first beta strand in SF3A1-UBL is located further upstream starting at aa P704. In SF3A1-UBL, the linker between the β1 and β2 strands is longer (15 aa) than the 4 aa linker in Ubiquitin. Additionally, at its C-terminal end, SF3A1-UBL contains three positively charged residues (RKK) downstream of an RGG sequence, which is present in Ubiquitin (Figure 2.6), and also conserved in SF3A1 UBL domains from other species (see Figure 2.17).
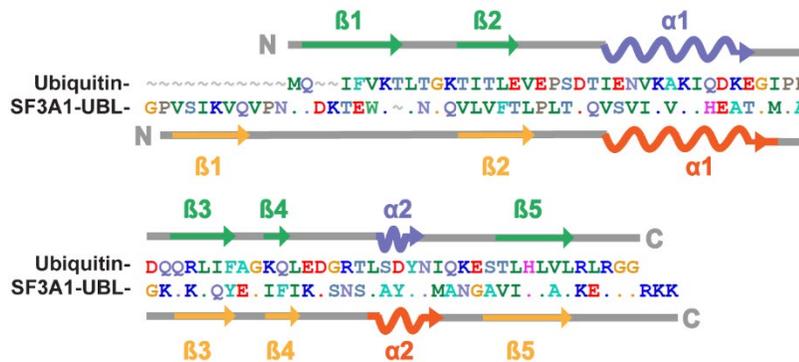
Figure 2.6 – The SF3A1 UBL domain and Ubiquitin share a similar secondary structure organization. Alignment of the SF3A1 UBL domain (aa 703-793) to Ubiquitin. Graphical representations of secondary structure in relation to primary structure were created based on PDB entries for Ubiquitin (PDB ID: 1UBQ) and SF3A1-UBL (PDB ID: 1ZKH). In the alignment, a dot indicates the presence of an identical residue to the reference sequence used in the alignment (Ubiquitin) and a tilde indicates a gap.

To determine if these unique terminal features of the SF3A1-UBL domain are required for RNA binding, we created three additional constructs that involved deleting the first beta strand in UBL$_{714-793}$, the RKK motif in UBL$_{703-790}$, and the RGGRKK

sequence in UBL$_{703-786}$ (Figure 2.7). Expression of 6xHis-tagged UBL domain deletion

constructs in CFE reactions was confirmed by immunoblotting (Figure 2.8 Left Panel).

UV-crosslinking analysis demonstrated that the UBL$_{703-793}$ construct bound ^{32}P-U1-SL4

RNA, whereas UBL$_{714-793}$, UBL$_{703-790}$, and UBL$_{703-786}$ did not (Figure 2.8 Right Panel).

Thus, these results confirmed that the unique N- and C-terminal features of the SF3A1-

UBL domain are required for RNA binding.
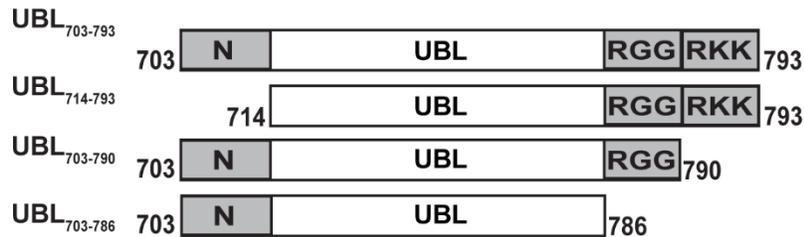
Figure 2.7 – Full-length SF3A1 UBL and truncated UBL proteins were expressed using the HeLa CFE lysate. Schematic of all N- and C-terminal deletion constructs made to study the SF3A1 UBL domain.
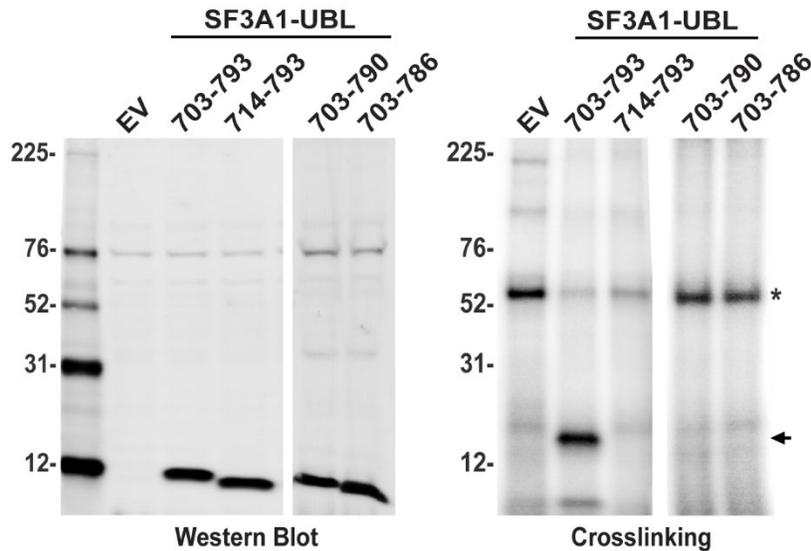
Figure 2.8 – Only the full-length SF3A1 UBL domain crosslinks to U1-SL4 RNA. Left panel – Western blot of 6xHis-tagged proteins in CFE reactions containing the empty vector (EV) control, or expression vectors for full-length (UBL703-793), N-terminal truncated (UBL714-793), and C-terminal truncated (UBL703-790 and UBL703-786) SF3A1-UBL. The expressed UBL domain proteins were detected with anti-6xHis primary antibody. Right Panel – UV-crosslinking of ^{32}P-U1-SL4 in CFE reactions expressing EV or UBL proteins. Arrows indicate specific cross-linked products and asterisks indicate non-specific products that were also observed in reactions containing the empty vector control.

SF3A1-UBL domain binds G-C rich stem-loop RNA

To characterize the affinity and specificity of RNA binding by the SF3A1-UBL domain, we created a fusion construct of GST and the UBL domain (GST-UBL). GST and GST-UBL proteins were expressed in *Escherichia coli*, purified using glutathione agarose beads, and confirmed by Coomassie blue staining, and Western blotting using antibodies to both GST and SF3A1 (Figure 2.9).
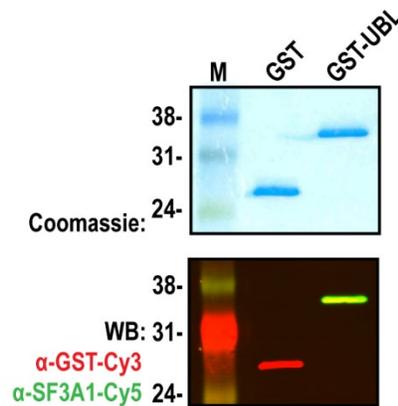
Figure 2.9 – GST-UBL fusion proteins were expressed and purified from *Escherichia coli* and used in a variety of assays to quantify the SL4-SF3A1 interaction. Expression and purification of GST alone and GST-UBL fusion proteins were confirmed by Coomassie blue staining and Western blotting using anti-GST and anti-SF3A1 antibodies.

To determine the dissociation constant (K_D) for the interaction between SF3A1-UBL and U1-SL4, we applied two independent quantitative methods: electrophoretic mobility shift assay (EMSA) and surface plasmon resonance (SPR). For EMSAs, binding reactions consisting of varying concentrations of GST or GST-UBL protein and Cy5-labeled U1-SL4 RNA were prepared and then separated on native gels as described previously (Figure 2.10) [185]. Dose response curves created from the fraction of RNA-protein complexes formed indicated dose-dependent assembly of GST-UBL/U1-SL4 complexes with a K_D = 96.93±10 nM, whereas GST alone did not exhibit any RNA binding (Figure 2.10 and 2.11).
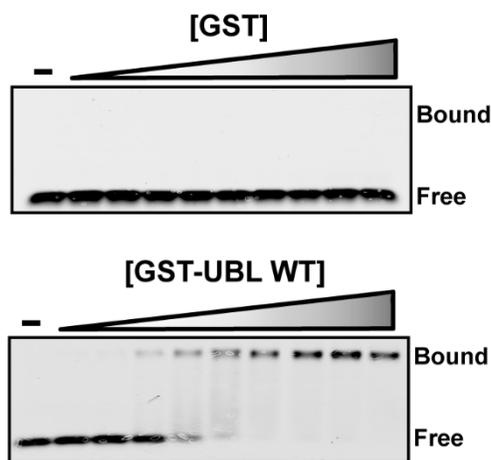
Figure 2.10 – Purified GST-UBL protein forms stably bound RNA-protein complexes with U1-SL4 labeled with a fluorophore. Electrophoretic mobility shift assays for binding of GST and GST-UBL proteins to 5′-Cy5-labeled U1-SL4 RNA. The concentrations of GST and GST-UBL were 0, 0.02, 0.03, 0.06, 0.13, 0.25, 0.5, 1, 2, and 4 μM.

Figure 2.11 – Quantification of the UBL-SL4 interaction by EMSA. Dose response curves generated by plotting the fraction of Cy5-U1-SL4 bound versus GST or GST-UBL protein concentration. The K_D value was determined from triplicate experiments.

For SPR experiments, biotinylated U1-SL4 RNA was immobilized on Neutravidin coated biosensors. Sensorgrams recorded during association and dissociation phases demonstrated a concentration dependent response upon injection of GST-UBL and lack of a response with GST (Figure 2.12). Binding parameters were calculated by non-linear regression analysis of the data assuming a 1:1 stoichiometry by the Langmuir binding model. Kinetics of the SF3A1-UBL/U1-SL4 interaction are characterized by an

association rate of $9.27 \pm 0.01 \times 10^5$ M^{-1} s^{-1} and a dissociation rate of $9.56 \pm 0.008 \times 10^{-2}$ s^{-1}, resulting in a $K_D = 103.1 \pm 0.1$ nM. The agreement of the K_D values obtained by EMSA and SPR indicates that immobilization of the U1-SL4 RNA on the biosensor does not perturb the binding reaction and that dissociation constants for the UBL-SL4 interaction can be reliably determined by either technique.

k_a ($M^{-1}s^{-1}$)	k_d (s^{-1})	K_D (nM)
$9.27 \times 10^5 \pm 0.01$	$9.557 \times 10^{-2} \pm 0.0008$	103.1 ± 0.1
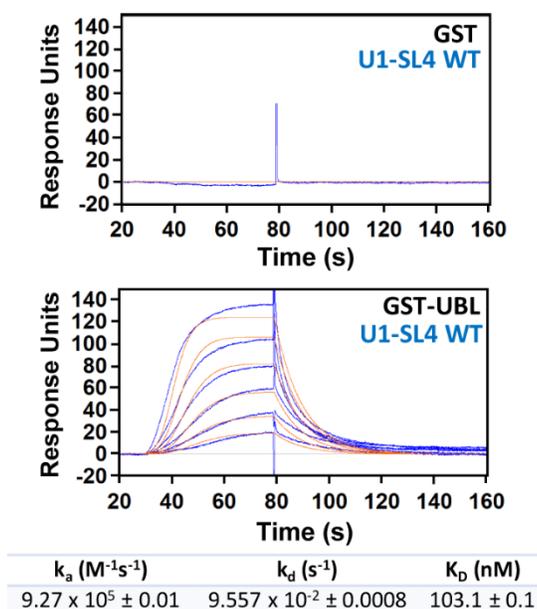
Figure 2.12 – Quantification of the UBL-SL4 interaction by surface plasmon resonance. SPR analysis for binding of GST and GST-UBL to immobilized 5′-biotinylated U1-SL4. Blue lines in the sensorgrams represent the raw curve generated by association and dissociation of GST injected at 1 µM and GST-UBL injected at 0.02, 0.03, 0.06, 0.13, 0.25, and 0.5 nM. Orange traces represent the global fit to the raw data by the Langmuir binding model assuming a 1:1 stoichiometry. Binding kinetics and dissociation constant values are shown below the sensorgrams.

To determine the specificity of the UBL-SL4 interaction, we performed competitive EMSAs and SPR. To reactions containing preformed GST-UBL/Cy5-U1-SL4 complexes, increasing concentrations of competitor RNAs were added. The RNA-protein complexes were separated, quantified, and the dissociation constants for competitor RNAs (K_C) were determined from plots of fraction of Cy5-U1-SL4 RNA bound versus competitor RNA concentration (Figure 2.13 and 2.14) [186].

Figure 2.13 – SF3A1-UBL binds to G-C rich stem-loop RNA. All competitive EMSA reactions contained 500 nM GST-UBL and 10 nM WT Cy5-U1-SL4. Final concentrations of competitor RNAs for WT U1-SL4, Drosophila melanogaster (Dm), and two mutant RNAs, M10 and M10h, were 0.02, 0.03, 0.06, 0.13, 0.25, 0.5, and 1 µM. The dissociation constants of competitor RNAs (K_C) that competed out Cy5-U1-SL4 are provided (n=3).

Figure 2.14 – Quantification of the affinities of a variety of competitive RNAs for SF3A1-UBL estimated by competitive EMSA. Dose-response curves plotted from the fraction of Cy5-U1-SL4 bound versus the concentration of competitor RNA are shown.

As expected, U1-SL4/WT strongly competed out Cy5-U1-SL4 with a K_C = 140.8±40.6 nM. Like human SL4, the *Drosophila melanogaster* U1-SL4 RNA (U1-SL4/Dm) forms a G-C rich stem ending in a tetra-loop. We previously demonstrated that U1-SL4/Dm can be functionally substituted for human U1-SL4 in U1 complementation assays and that it binds SF3A1 [168]. U1-SL4/Dm was able to compete out GST-UBL/U1-SL4 complexes as efficiently as WT human U1-SL4 with a similar K_C = 143.3±33.3 nM and also exhibited the capacity to bind the SF3A1-UBL protein by SPR with an association rate of $6.86±0.02 \times 10^5$ M^{-1} s^{-1} and a dissociation rate of $5.92±0.0001 \times 10^{-2}$ s^{-1}, resulting in a K_D = 86.3±0.3 nM (Figure 2.14 and 2.15).

k_a ($M^{-1}s^{-1}$)	k_d (s^{-1})	K_D (nM)
$6.86 \times 10^5 ± 0.02$	$5.92 \times 10^{-2} ± 0.0001$	$86.3 ± 0.3$
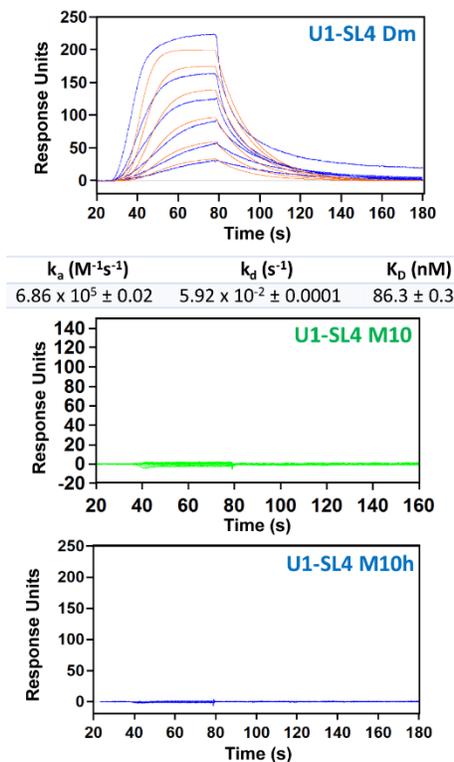
Figure 2.15 – SPR confirms the binding activity of SF3A1-UBL to the stem-loop RNAs investigated by competitive EMSA. Sensorgrams from SPR experiments for binding of GST-UBL to Dm, M10, and M10h RNAs. GST-UBL concentrations were 0.02, 0.03, 0.06, 0.13, 0.25, and 0.5 μM. Orange traces represent the global fit to the raw data by the Langmuir binding model assuming a 1:1 stoichiometry. Binding kinetics and dissociation constant values are summarized below sensorgrams where binding was detected.

On the other hand, minimal to no competition was observed when using two mutant RNAs. In the M10 mutant, all G-C basepairs were replaced with A-U to retain the secondary structure of U1-SL4 while lowering the G-C content, and in the M10h mutant, basepairing was disrupted to obtain a single stranded RNA (ssRNA) by changing guanine or cytosine nucleotides of one of the strands to adenines. The M10 RNA did not exhibit any UBL-binding by competitive EMSAs or SPR (Figure 2.14 and 2.15). While the M10h mutant did not crosslink to SF3A1 (Figure 2.4 and 2.5) or demonstrate binding to GST-UBL by SPR (Figure 2.15), it showed some capacity to compete out GST-UBL/SL4 bound complexes by EMSA but the binding did not reach saturation (Figure 2.13). This non-specific binding might be because of a short stem-loop structure that the M10h mutant may form as predicted by the mfold algorithm (Figure 2.16) [187]. Taken together, EMSA and SPR analyses demonstrate that the UBL domain of SF3A1 binds to a G-C basepaired stem-loop RNA, but not to an A-U rich stem-loop or to ssRNA.

Figure 2.16 – Secondary structure prediction of U1-SL4 M10h mutant. The two lowest free-energy (ΔG) structures for the M10h RNA as predicted by mfold are shown. The non-specific binding observed for this ligand in competitive EMSA experiments may be explained by the capacity to form short stem-loops in vitro.

Tyrosines 772 and 773 are important for U1-SL4 binding by SF3A1-UBL

The UBL domain of SF3A1 shares a ~98% and ~93% sequence identity with the mouse and the zebra fish domains, respectively, and <50% with flies and worms (Table 2.1). Multiple alignment of these sequences identified several highly conserved residues (Figure 2.17). Notable amongst these are tyrosine residues 772 (Y772) and 773 (Y773). Y772 has been reported to be mutated to a cysteine in individuals with MDS and CMML [141]. It is conserved between humans, mice, and fish, and has a conservative substitution to phenylalanine in flies and worms, while being replaced by an aspartate in Ubiquitin (Figures 2.6 and 2.17). Y773, on the other hand, is conserved from humans to worms and is even present in Ubiquitin.

Identity Matrix	H. sapiens	M. musculus	D. rerio	D. melanogaster	C. elegans
H. sapiens	1.0	0.978	0.934	0.472	0.373
M. musculus		1.0	0.912	0.461	0.373
D. rerio			1.0	0.450	0.351
D. melanogaster				1.0	0.406
C. elegans					1.0

Table 2.1 – Identity matrix comparing SF3A1 UBL domains in higher eukaryotes. The SF3A1 UBL domain is conserved among higher eukaryotes. SF3A1 UBL domains and their sequence identity to one another are scored from lowest to highest identity on a 0-1 scale and color coated accordingly.
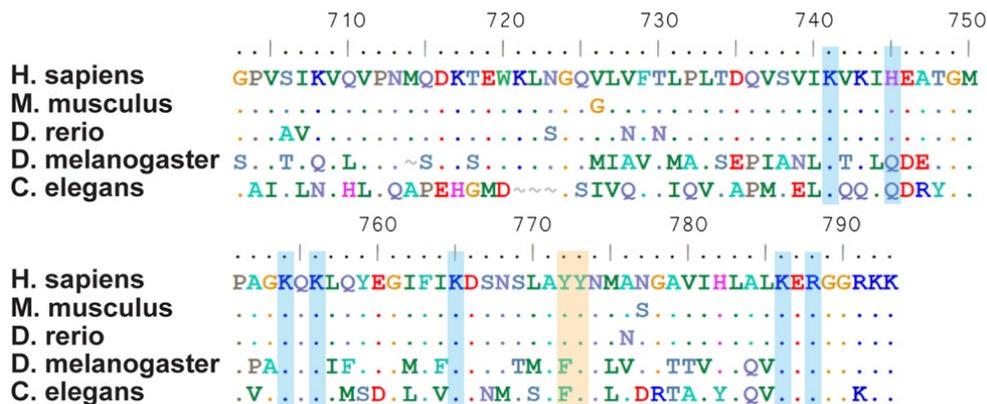
Figure 2.17 – There are many residues in the SF3A1-UBL domain that are conserved across species. Alignment of SF3A1 UBL domains from human, mouse, zebrafish, flies, and worms. In the aligned sequences, a dot indicates the presence of an identical residue and a tilde indicates a gap.

To examine if these residues are involved in U1-SL4 binding, we created SF3A1 plasmid constructs carrying the Y772C and Y773C mutations and expressed them by the HeLa CFE system. We made an additional mutant, SF3A1-R511Q, that has been reported to be associated with AML [139]. Western analysis confirmed the expression of WT and mutant full-length SF3A1 proteins in the CFE reactions (Figure 2.18 lanes 2-5). UV crosslinking of binding reactions containing the SF3A1 proteins and ^{32}P-U1-SL4 revealed that the Y772C and Y773C mutations reduced the binding capacity of SF3A1 by approximately 2-fold in comparison to SF3A1-WT, whereas the R511Q mutation did not have any effect (Figure 2.19).
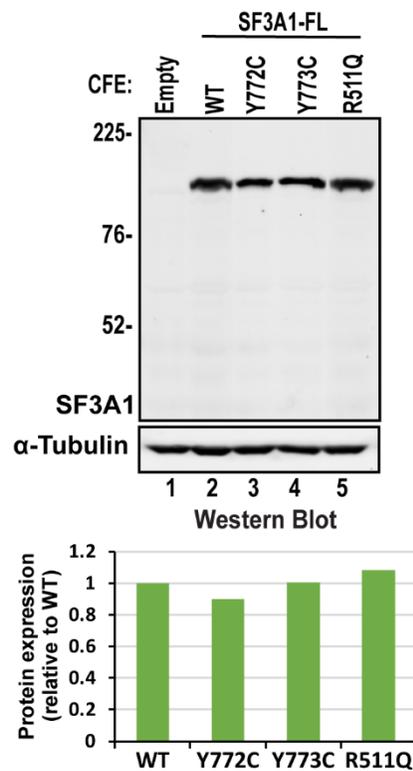
Figure 2.18 – FL SF3A1 harboring a series of point mutations were expressed in HeLa cell CFE lysate. Western analysis using anti-SF3A1 primary antibody demonstrating expression of SF3A1-WT and point-mutants Y772C, Y773C, and R511Q in CFE reactions. The expression of mutant proteins relative to WT, normalized to α-Tubulin, is graphed below the Western blot.
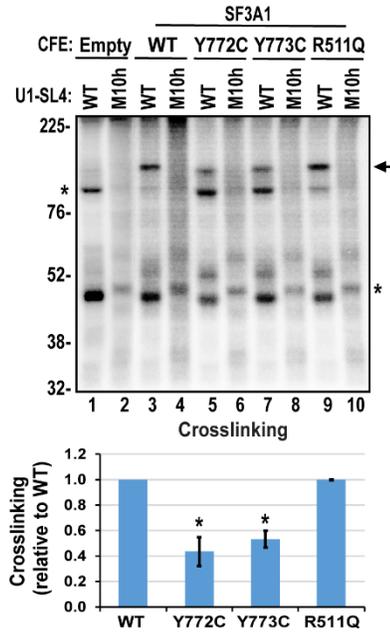
Figure 2.19 – Mutation of conserved tyrosine residues affects U1-SL4 binding by SF3A1. Crosslinking of WT and mutant SF3A1 proteins to ^{32}P-U1-SL4-WT and -M10h RNAs. Crosslinking efficiency, relative to WT SF3A1, is represented in the graph below (* = p < 0.05, Student's t-test). In the crosslinking gel, arrows indicate specific crosslinked products and asterisks indicate non-specific products that were seen in reactions containing the empty vector control.

To further characterize the impact of the Y772C and Y773C mutations on U1-SL4 binding, we introduced these mutations into the GST-UBL fusion construct. Wildtype and mutant GST-UBL proteins were purified using glutathione agarose beads and their purity was confirmed by Coomassie blue staining and Western blotting (Figure 2.20).
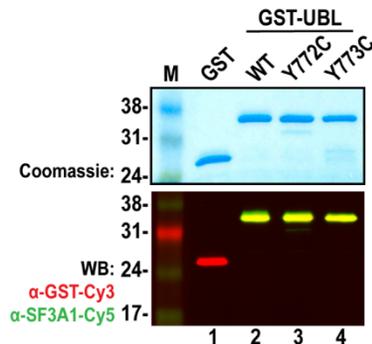
Figure 2.20 – The Y772C and Y773C mutations were introduced into GST-UBL fusion proteins expressed and purified from E. coli and used in a variety of assays to quantify the impact of these mutations on RNA binding activity. Expression and purification of WT and Y772C/Y773C GST-UBL variants as assessed by Coomassie blue staining and Western analysis.

Initial analysis by UV-crosslinking indicated that the Y772C mutation caused a moderate reduction in U1-SL4 binding, whereas the Y773C mutation led to a drastic reduction in crosslinking efficiency (Figure 2.21 compare lanes 7-9 and 10-12 to 4-6).
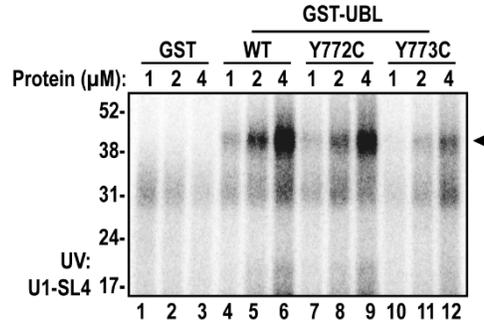
Figure 2.21 – Cross-linking efficiency of mutant GST-UBL proteins is reduced. Cross-linking of ^{32}P-U1-SL4 RNA to WT and mutant GST-UBL proteins.

Binding analysis by EMSAs supported these initial observations; GST-UBL proteins carrying either mutation formed complexes with Cy5-U1-SL4, however, not as efficiently as the WT protein (Figure 2.22). The dose response curves exhibited a clear rightward shift and quantification revealed ~2-fold (K_D = 190.5±35.1) and ~5-fold (K_D = 458.2±101.4) increase in K_D values for Y772C and Y773C mutations, respectively (Figure 2.23). Thus, Y772C and Y773C mutations cause a decrease in binding affinity. These results indicate that Y772 and Y773 may have a role in U1-SL4 binding by the SF3A1-UBL domain.
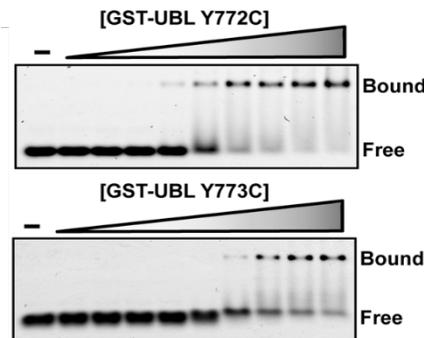
Figure 2.22 – Purified mutant GST-UBL proteins form stably bound RNA-protein complexes with fluorescent U1-SL4. EMSA analysis for binding of GST-UBL carrying Y772C and Y773C mutants to Cy5-U1-SL4. The final concentrations for mutant GST-UBL proteins were 0, 0.02, 0.03, 0.06, 0.13, 0.25, 0.5, 1, 2, and 4 µM.

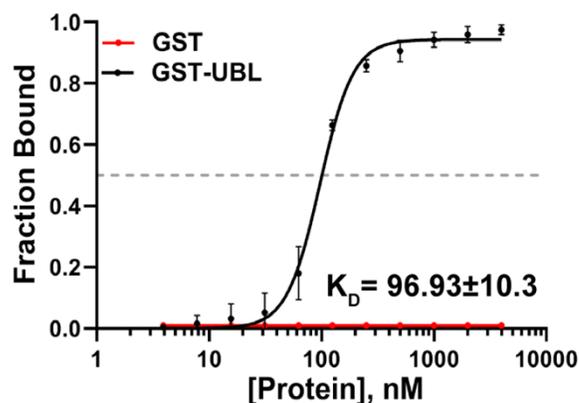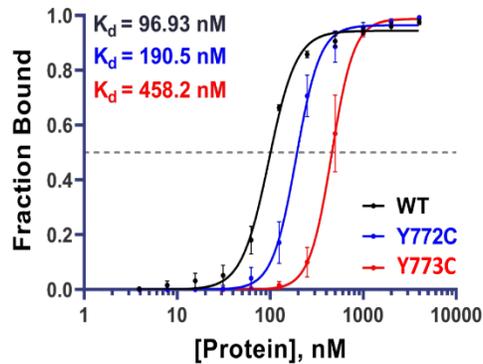**Figure 2.23 – Quantification of the mutant UBL-SL4 interactions by EMSA.** Dose response curves were created by plotting the fraction of bound Cy5-U1-SL4 versus GST-UBL/WT, Y772C, and Y773C protein concentrations, and K_D values were determined from triplicate experiments.

GST-UBL protein selectively interacts with the U1 snRNP in nuclear extract

To confirm that the UBL-domain of SF3A1 can interact with U1-SL4 in the context of an intact U1 snRNP, we performed GST affinity pull-down assays. First, HeLa cell nuclear extracts were pre-cleared of GST-binding proteins, as described [188]. Then, GST and GST-UBL/WT proteins were added in the absence or presence of U1-SL4/WT or U1-SL4/M10 competitor RNAs, followed by pull-down using glutathione agarose beads. Spliceosomal components associated with bound complexes were analyzed by Northern and Western blotting (Figure 2.24 and 2.25). Northern analysis confirmed the enrichment of the U1 snRNA, but not the U2, U4, U5 or U6 snRNAs, in GST-UBL/WT complexes (Figure 2.24 lane 2). Preincubation with the U1-SL4/WT RNA competed out the U1 snRNA from the GST-UBL/WT complexes (Figure 2.24 compare lanes 2 and 4). Whereas preincubation with the U1-SL4/M10 RNA did not have any impact (Figure 2.24 lane 6). Western analysis confirmed the presence of U1 snRNP specific proteins U1C and U1-70k in the pull-down complexes and the absence of U2 specific protein SF3A3

49

(Figure 2.25 lane 2). The U1 proteins were similarly competed out by preincubation with

the U1-SL4/WT, but not the U1-SL4/M10 RNA (Figure 2.25 lanes 4 and 6).
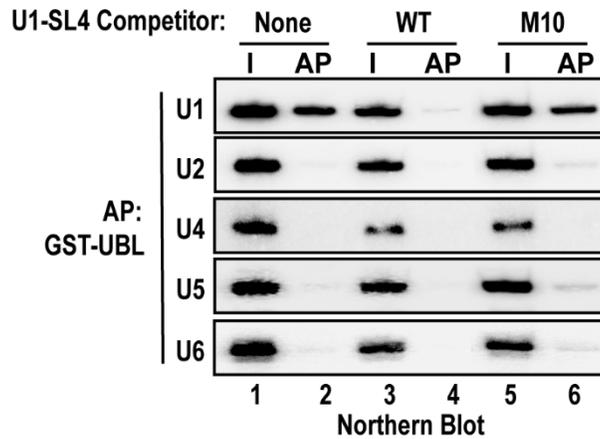
Figure 2.24 – U1 snRNA is enriched in GST-UBL complexes affinity purified from HeLa nuclear extract which can be competed out with the addition of free WT U1-SL4 RNA.
Northern analysis for spliceosomal snRNAs in GST-UBL pull-down complexes. GST-UBL/WT protein was incubated in HeLa cell nuclear extracts in the absence or presence of U1-SL4/WT or M10 RNAs as competitors, followed by pull-down using glutathione agarose. I and AP indicate input and affinity pull-down complexes, respectively.
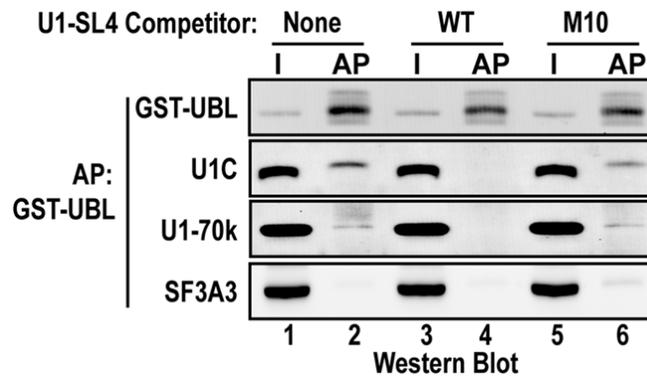
Figure 2.25 – GST-UBL specifically binds to the U1 snRNP via U1-SL4. Western blot analysis of proteins in the pull-down complexes using antibodies to U1C, U1-70k, and SF3A3 proteins. I and AP indicate input and affinity pull-down complexes, respectively.

GST alone did not pull-down any snRNP specific RNAs or proteins under any of these

conditions (Figure 2.26 and 2.27).
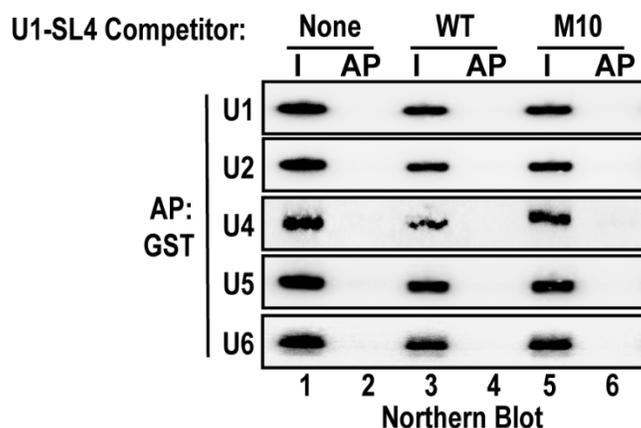
Figure 2.26 – Analysis of spliceosomal RNAs present in control GST pulldown complexes. Northern blotting of U1, U2, U4, U5, and U6 snRNA show no enrichment of any snRNAs in GST alone affinity-purification fractions. I and AP indicate input and affinity pull-down complexes, respectively.
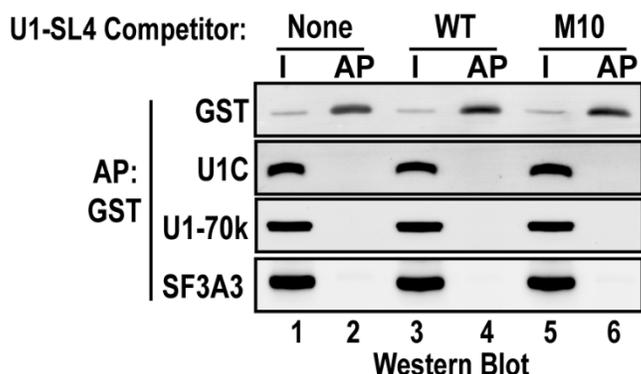
Figure 2.27 – Analysis of U1 and U2 proteins present in control GST pulldown complexes. Immunoblotting of proteins present in the GST affinity pull-down complexes demonstrating the lack of enrichment of U1C, U1-70k or SF3A3. I and AP indicate input and affinity pull-down complexes, respectively.

To examine the impact of the tyrosine mutations, pull-down assays were performed with GST-UBL/Y772C and GST-UBL/Y773C fusion proteins. Western analysis confirmed the presence of the GST fusion proteins and revealed enrichment of the U1 snRNP specific protein U1-70k in the pull-down complexes. In agreement with the mild impact of the Y772C mutation on binding affinity for U1-SL4, there was a ~2-fold reduction of U1-70k in the GST-UBL/Y772C complexes in comparison to the

51

wildtype complexes. In the case of the more severe Y773C mutation, there was a drastic

reduction of U1-70K in the pull-down complexes (Figure 2.28 compare lanes 6 and 8 to

lane 4). U2 specific proteins SF3A1 and SF3A3, were not present in the complexes for

wildtype or mutant GST-UBL proteins. Thus, these results demonstrate that the UBL-

domain of the SF3A1 protein is sufficient for binding to the U1 snRNA in the context of

the U1 snRNP, and this interaction occurs through SL4. Results also indicate a more

significant role for Y773 than Y772 in the interaction between SF3A1-UBL and the U1

snRNA.
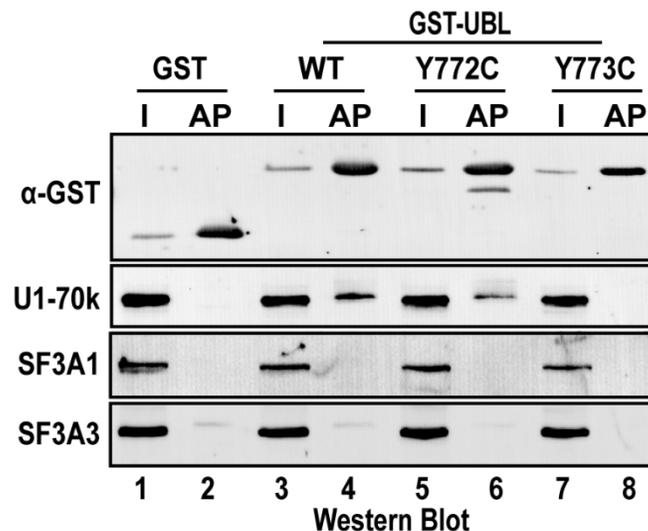
Figure 2.28 The Y772C and Y773C mutations interfere with GST-UBL mediated U1 snRNP affinity purification. Western analysis of proteins present in GST-UBL/WT, Y772C, Y773 pull-down complexes. I and AP indicate input and affinity pull-down complexes, respectively.

Other human UBL domains with potential to bind nucleic acids

To identify other UBL domains with potential to bind RNA in humans, we

searched databases of UBL domain containing proteins for the salient features identified

in SF3A1. A total of 955 proteins were scanned for the presence of a conserved tyrosine

at positions analogous to 772 and/or 773, the C-terminal RGG motif, or positively

charged residues at the C-terminus (see Figure 2.6). This search identified thirty-eight

UBL domain containing proteins that harbored one or more of these features. Alignment

of the identified protein sequences is represented in Figure 2.29. Although not identified

in our search, the UBL domains of transactivation response element (TAR) DNA-binding

protein 43 (TDP-43), and the small Ubiquitin-related modifier 1 (SUMO-1) are also

included in the alignment as they have previously been demonstrated to bind nucleic

acids [189-191].

In the newly identified sequences, a tyrosine at position analogous to 773 is the

most conserved feature and is present in 29 UBL domains (highlighted blue) while three

proteins (PCGF1, TBK1, and MIDN) have a conservative phenylalanine substitution at

this position (highlighted green) (Figure 2.29). A tyrosine or a conservative

phenylalanine substitution at position analogous to 772 is present in three proteins-

MAP2K5, PARD6B, and UBL7 (highlighted purple). Of the 29 domains harboring a

conserved tyrosine, four (Ubiquitin, RPS27A, NEDD8, and ZFAND4) also have the C-

terminal RGG motif (indicated by an asterisk). However, of these four, only RPS27A and

NEDD8 contain positively charged residues downstream of RGG. Despite lacking the

RGG motif, many UBL proteins have arginines and/or lysines at their C-terminus. Five

proteins (PARK2, UBAC1, PARD6A, BAG1, and NUB1) lack a conserved tyrosine(s),

but are enriched in arginines and/or lysines at their C-terminus (highlighted yellow).

Although the residues involved in RNA-binding in the N-terminal extension of SF3A1-

UBL have yet to be determined, we observed that many of the UBL domains had similar

N-terminal extensions when compared to Ubiquitin. Like SF3A1, the N-termini of most

of these UBL domains contained at least one arginine or lysine residue. Notably, TDP43-

UBL and SUMO-1 lack Y772 and Y773, but do have positively charged residues at their C- and N-terminus, respectively.

A few proteins identified in our search are known to either bind RNA or be involved in an RNA processing step. These include seven members of the ATG8 family of UBL proteins (MAP1L3A, MAP1L3B, MAP1L3B2, MAP1L3C, GABARAP, GABARAPL1, and GABARAPL2), RPS27A, and SNRNP25. The ATG8 family proteins have a conserved tyrosine and an N-terminal extension. Although lacking a positively charged C-terminus, the N-terminal half of these proteins are rich in arginine and lysine, and one of them, MAPILC3B, has been reported to bind RNA [192, 193]. RPS27A is a ribosomal protein that is cleaved post-translationally to produce a free UBL-domain and S27a, and may be involved in ribosome biogenesis [194, 195]. SNRNP25 is of particular interest, as it is a unique component of the U11/U12 di-snRNP of the minor spliceosome. It is essential for cell viability and appears to be more associated with the U11 snRNP than the U12 snRNP [196]. Since the SF3A1-UBL directly binds the U1 snRNA via SL4, a similar function may be involved in the association of SNRNP25 with the U11 snRNA which is structurally similar to U1 and folds into four distinct stem-loops. Thus, the sequence analysis indicates that many other UBL-domains in humans have salient features that were found to be important for binding RNA by SF3A1-UBL, and the proteins harboring these domains may also have the capacity to bind RNA.
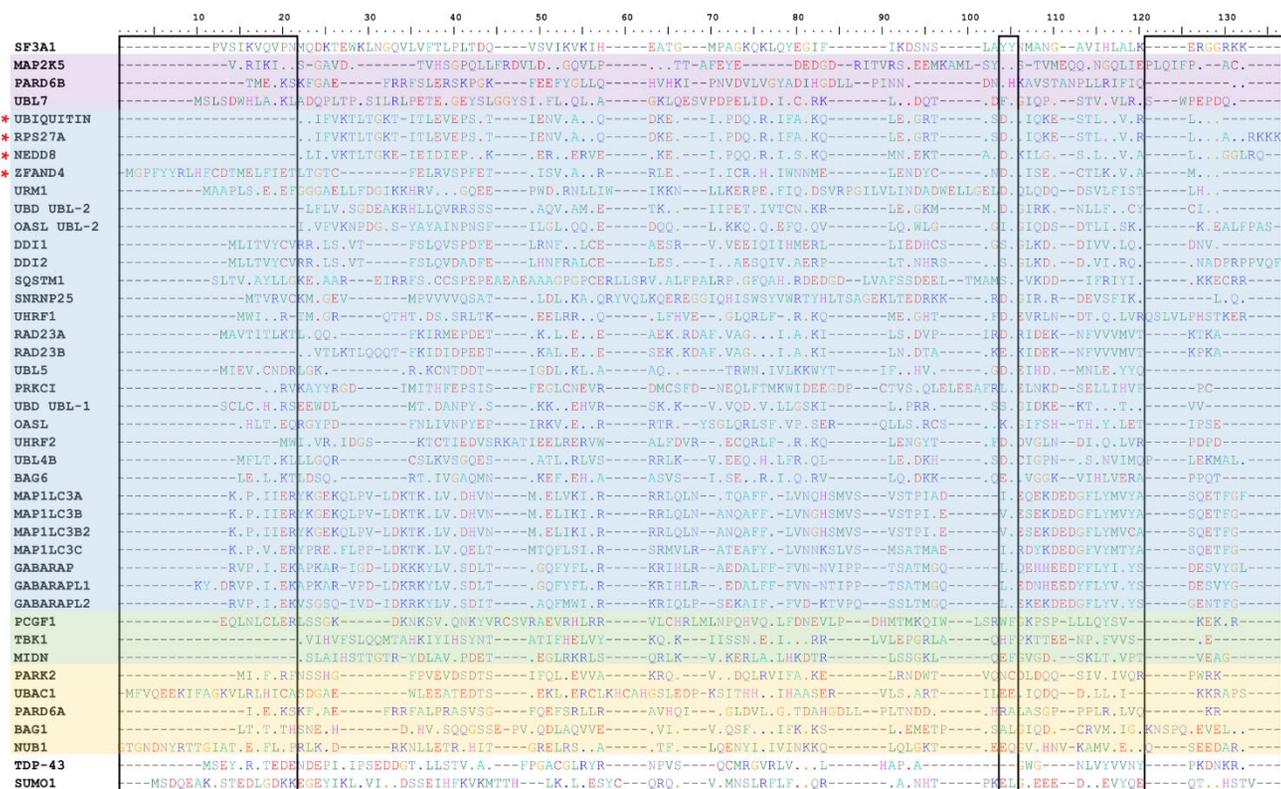
Figure 2.29 – Alignment of human UBL domain sequences harboring features similar to SF3A1-UBL. Sequences were aligned by ClustalW and key features that were used as criteria for performing the search for UBL domains are indicated by boxes. These features include the N-terminal extension, conserved tyrosine residues at or near position 772 and/or 773 relative to SF3A1, and C-terminal lysine and/or arginine residues. UBL domains with a tyrosine or a conservative phenylalanine substitution at position analogous to 772 are highlighted in purple. UBL domains with a tyrosine at position analogous to 773 are highlighted in blue and domains with sequences containing a conservative phenylalanine substitution at this position are highlighted in green. Domains that lack a conserved tyrosine/phenylalanine or the RGG motif, but contain C-terminal lysine and/or arginine residues are highlighted in yellow. Sequences of UBL domains containing a conserved tyrosine and the RGG motif are indicated by an asterisk. In the aligned sequences, a dot indicates the presence of an identical residue to reference (SF3A1-UBL) and a dash indicates a gap. The UniProt accession numbers for all proteins included in the alignment are listed in Appendix C.

A positively charged patch on SF3A1-UBL may mediate U1-SL4 binding

The TDP-43 N-terminal domain (NTD) and SUMO-1 have been shown to bind single stranded (ss) and double stranded (ds) DNA, respectively, but their capacity to bind RNA has not been reported [189-191]. SUMO-1, but not TDP-43-NTD which was initially characterized as a UBL domain, appears to fold into the Ubiquitin-like $\beta\beta\alpha\beta\beta\alpha\beta$ topology [197, 198]. Structural analysis of TDP-43-NTD in the presence of ssDNA, and SUMO-1 in the presence of dsDNA have identified clusters of positively charged surface residues that facilitate DNA binding in both proteins. Electrostatic surface potential analysis of the available structure of the SF3A1 UBL domain (PDB ID:1ZKH) by PyMOL demonstrates the presence of a similarly highly positively charged surface that is formed by five lysine residues (K741, K754, K756, K765, and K786), histidine H745, and arginine R788 (Figure 2.30).
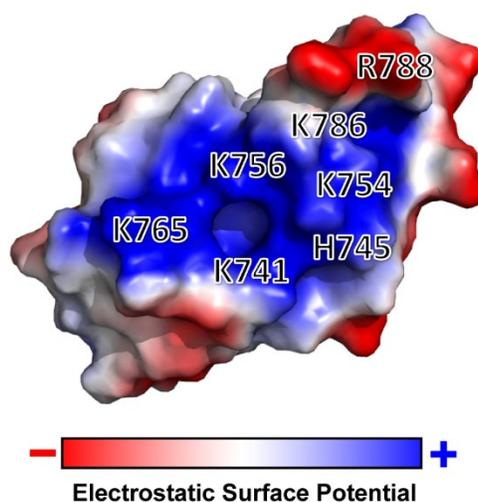
Figure 2.30 – The SF3A1 UBL domain has a positively charged surface that could potentially facilitate interactions with U1-SL4 RNA. Electrostatic potential analysis of the SF3A1-UBL domain surface performed using the available NMR structure (SF3A1 aa 704-789; PDB ID: 1ZKH). General positions and identity of residues that contribute to the positively charged surface (blue) are highlighted.

56
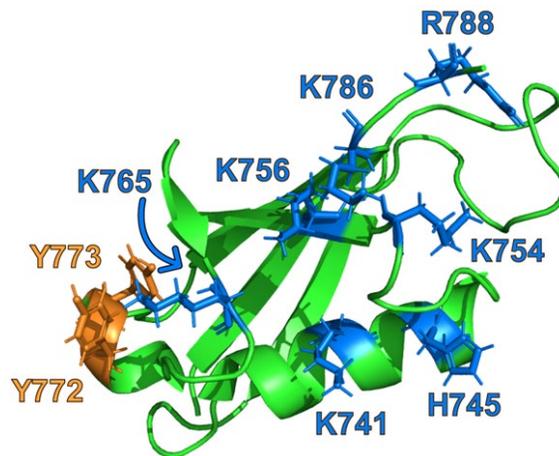
SF3A1-UBL (PDB ID: **1ZKH**)

Figure 2.31 – The conserved tyrosine residuesY772 and Y773 are in the same plane as the positively charged surface of the SF3A1 UBL domain. Ribbon diagram of SF3A1-UBL represented in the same orientation as in Figure 2.31. All residues highlighted in the electrostatic surface potential analysis, and tyrosines Y772 and Y773, are shown.

These seven positively charged amino-acids are conserved in SF3A1-UBL from humans to worms except for H745, that is replaced by a glutamine, in flies and worms. These residues are in the same plane as tyrosines 772 and 773 that were found to be important for binding to U1-SL4 (Figure 2.23 and 2.31). The C-terminal RKK motif that was found to be critical for U1-SL4 binding could also contribute to the positively charged surface in SF3A1-UBL as it is immediately downstream of R788. Thus, this positive surface, in coordination with Y772 and Y773, likely confers the SF3A1-UBL with RNA-binding properties.

Structure of the SL4-bound SF3A1 Ubiquitin-like domain

Since the identification that the Ubiquitin-like domain is the U1-SL4 interacting region of SF3A1, the structure of the SL4-bound UBL domain (SL4-UBL) has been

solved in collaboration with the laboratory of Frédéric Allain along with Sébastien

Campagne and Tebbe de Vries at ETH Zürich in Switzerland. Using a combination of

techniques including cross-linking of isotopically labeled RNA combined with tandem

mass spectrometry (CLIR-MS/MS), nuclear magnetic resonance (NMR), and X-ray

crystallography, key nucleotides and residues were identified, and the SL4-UBL structure

was determined.
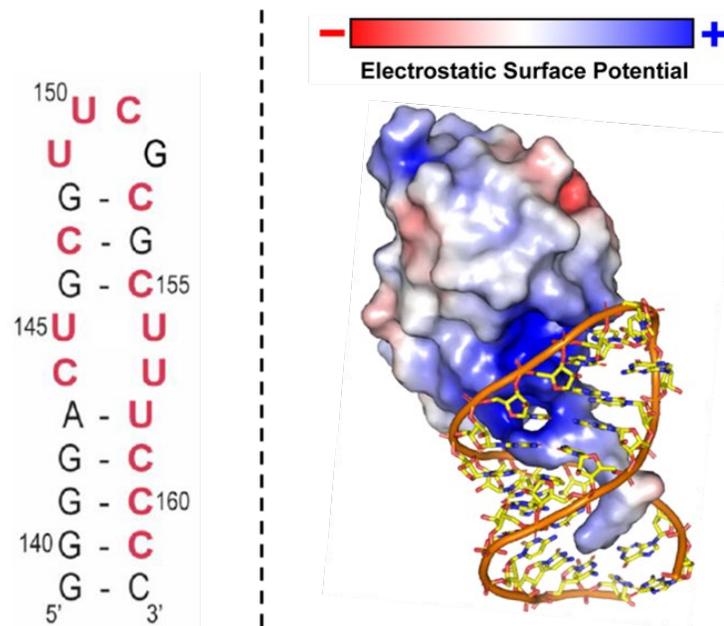
Figure 2.32 – A broad negatively charged surface of the SF3A1-UBL protein binds U1-SL4.
Left Panel – A schematic of the secondary structure of U1 stem-loop 4 RNA. Nucleotides colored
red experience chemical shift perturbations in response to the addition of UBL protein as detected
by nuclear magnetic resonance. Right Panel – The structure of the UBL domain (surface model)
bound to U1-SL4 RNA solved by X-ray crystallography is shown. The C-terminal tail of SF3A1-
UBL that contains the conserved RGGRKK motif protrudes into the SL4 major groove.

NMR identified that a large portion of the U1-SL4 is contacted by UBL protein

when present at a 1:1 ratio (Figure 2.32 left panel) with most of the stem and three

nucleotides of the tetranucleotide loop experience chemical shift perturbations in the

presence of UBL protein (red nucleotides). X-ray crystallography, informed by results

from CLIR-MS/MS and NMR, demonstrated that a large positively charged surface of

SF3A1-UBL contacts U1-SL4 via many of the residues noted in Figure 2.30 (Figure 2.32 right panel). The C-terminal tail was found to be stabilized upon binding to U1-SL4 and buries itself into the major groove of the stem-loop structure. A number of residues were found to directly interact with the nucleotides in the loop and stem of SL4, while others form hydrogen bonds and salt bridges with the phosphate backbone of the stem. The evolutionarily conserved phenylalanine residue F763 (Figure 2.17) was found to interact with C151 in the loop of U1-SL3 via a base-stacking interaction (Figure 2.33). The lysine residue K765 also interacts with the loop region by hydrogen bonding to U150. The residues that primarily interact with the stem-region of SL4 are present in the C-terminal tail of the UBL protein which hydrogen bond with the phosphate backbone down the length of the 3′-half of U1-SL4. The final two lysine residues that complete the tail were unable to be definitively modeled in the structure but likely continue further down the stem through the major groove.
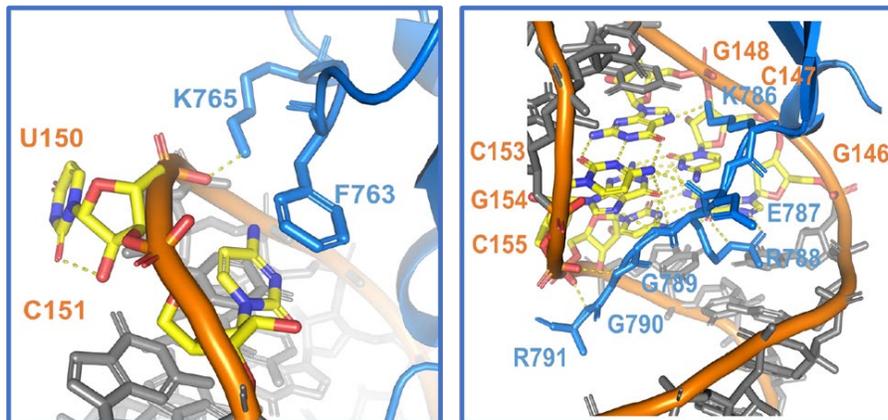
Figure 2.33 – A detailed view of interactions between U1-SL4 and SF3A1-UBL. The conserved phenylalanine residue F763 in the fourth β-sheet of the UBL domain engages C151 in the loop of SL4 via a base-stacking interaction and the lysine residue K765 hydrogen bonds to U150. The positively charged C-terminal tail of the UBL domain fits into the major groove of U1-SL4 making numerous hydrogen bonds with the phosphate backbone at nucleotides in the 3′ half of the stem-loop RNA, in addition to base-specific interaction with G154 and C155.

A summary of the key residues of SF3A1-UBL identified as involved in direct interactions with U1-SL4 is shown in Table 2.2 along with their proposed functions in the interaction.

	Residue	Function
1	K717	Salt bridge – phosphate backbone of U145
2	K754	Salt bridge – phosphate backbone of C147
3	F763	Base-stacking - C151 tetraloop UUCG
4	K765	Hydrogen bonding - phosphate backbone of U150
5	K786	Hydrogen bonding – nucleotides G148 and U149
6	R788	Hydrogen bonding – nucleotide G146
7	G789	Hydrogen bonding – nucleotides G154 and C155
8	G790	Hydrogen bonding - phosphate backbone of G154
9	R791	Hydrogen bonding - phosphate backbone of C155
10	KK(792-793)	Residues not visible but experience strong chemical shifts in response to UBL protein

Table 2.2 – List of SF3A1-UBL residues involved in interactions with U1-SL4 as determined by crosslinking assays, NMR, and X-ray crystallography. Alanine point mutations were introduced at each of these residues to test their contribution to UBL RNA binding activity determined by surface plasmon resonance.

Alanine mutations were introduced into each of these residues, and WT and mutant UBL protein were expressed in *E. coli* using a bacterial vector, allowing for the purification and isolation of tag-free UBL protein after His-tag affinity purification and TEV protease cleavage. These purified WT and mutant proteins were tested for U1-SL4 RNA binding activity by SPR as in Figures 2.12 and 2.15. WT-UBL protein had a K_D = 332±5 nM for U1-SL4. All mutants had drastically reduced binding affinities for U1-SL4 with the mildest mutation, K717A, causing an 8-fold increase in the dissociation constant of the UBL protein (Table 2.3). All other point mutations had varying degrees of impact more severe than K717A, with the most detrimental effects occurring in the RGG(788-790) and KK(792-793) motifs (Table2.3 and Appendix A).

Protein	K_D (M)	Fold Increase
UBL-WT	0.332 ± 5 µM	1.00
UBL-K717A	2.66 ± 0.05 µM	8.01
UBL-K754A	20.6 ± 0.6 µM	62.05
UBL-F763A	4.06 ± 0.08 µM	12.23
UBL-K765A	13.1 ± 0.2 µM	39.46
UBL-K786A	91 ± 3 µM	274.10
UBL-RGG2AAA	1,700 ± 0.6 µM	5120.48
UBL-G789I	38.6 ± 0.7 µM	116.27
UBL-G790I	11.4 ± 0.2 µM	34.34
UBL-R791A	9.0 ± 0.1 µM	27.11
UBL-KK2AA	124 ± 4 µM	373.49

Table 2.3 – Summary of dissociation constants of WT and mutant UBL protein for U1-SL4 RNA as estimated by surface plasmon resonance. All point mutations reduced the affinity of the SF3A1-UBL domain for U1-SL4 by ≥ 8-fold with the most severe mutations occurring in the RGG (788-790) and KK (792-793) motifs. See Appendix A for the saturation binding curves where these affinity constants (K_D) have been derived.

To determine the impact of these mutations on splicing *in vivo*, an siRNA-based rescue assay was optimized. First, an RNAi resistant (RNAiR) FLAG-tagged SF3A1 clone was generated that carried silent point mutations at an SF3A1-targeting siRNA (siSF3A1) binding site. The FLAG-RNAiR-SF3A1 clone expresses well in transfected HeLa cells when co-transfected with either negative control non-targeting siRNA (siNT) or with siSF3A1 which efficiently knocks down endogenous SF3A1 expression (Figure2.34). All point mutations tested by SPR were introduced into the RNAiR SF3A1 clone in addition to several double mutants harboring mutations in residues that interact with the upper stem-loop region of SL4 (K754 and K765) and the RGG(788-790) motif. All mutants are efficiently expressed under endogenous SF3A1 knockdown conditions (Figure 2.35).
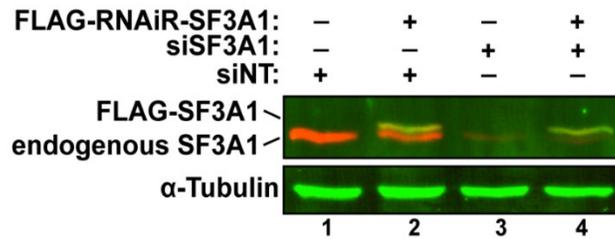
Figure 2.34 – An siRNA resistant, FLAG-tagged SF3A1 clone can be expressed under endogenous SF3A1 knockdown conditions in transfected HeLa cells. Western blotting of transfected HeLa cells demonstrate efficient expression of an RNAi-resistant SF3A1 clone (FLAG-RNAiR-SF3A1) under control siRNA treatment (lane 2) or when cells are treated with endogenous SF3A1 targeting siRNA (siSF3A1, lane 4).
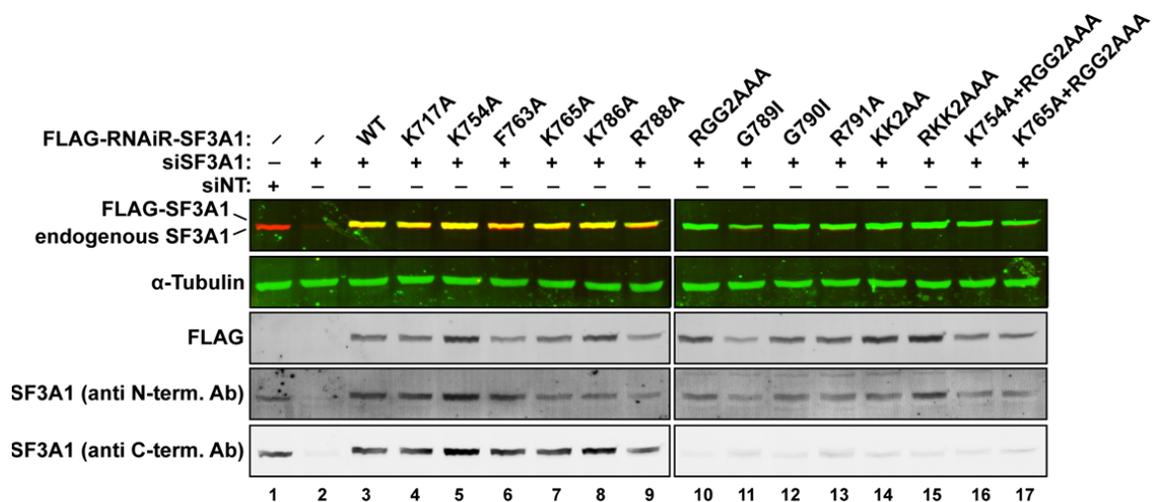
Figure 2.35 – WT and mutant FLAG-RNAiR-SF3A1 protein expression in transfected HeLa cells. Several point mutations were introduced into the RNAi resisted clone and Western blotting confirms expression of all proteins after 48-hours co-transfection with siSF3A1. A merged image of immunoblots probed with anti-FLAG antibody and anti-SF3A1 C-terminus targeting antibody (anti C-Term. Ab) is shown in the top panel. Endogenous SF3A1 (red) is efficiently silenced and many of the mutants are detectable using both antibodies (yellow). Point mutations introduced closer to the C-terminus of SF3A1 interfere with SF3A1 (anti C-term. Ab) recognition but are detectable with anti-FLAG antibody (green). Western blots were stripped and re-probed with N-terminus targeting SF3A1 antibody (anti N-term. Ab) and confirm that all mutant FLAG-RNAiR-SF3A1 proteins are detected using an SF3A1-specific antibody. The individual scans from each antibody used in the merged image is shown in the lower panels and α-Tubulin was used as a loading control.

To test the impact of UBL-point mutations on splicing in HeLa cells, a 3-exon, 2-intron minigene reporter (Dup51p) was transfected into HeLa cells 48-hours post-transfection with siSF3A1 and plasmid encoded WT and mutant FLAG-RNAiR SF3A1.

62

The splicing of the reporter is dependent on a variant of U1 snRNA called U1-5a which is encoded on a plasmid that is co-transfected with the minigene (Chapter3) [168]. In these experiments, a mutant of U1-5a was applied called U1-5a/SL3-M1g which contains three G→A nucleotide mutations in the upper region of U1 stem-loop 3 (SL3) to make the spliceosome more susceptible to mutations in SF3A1 (see Chapter 3, page 81 for more details about the reporter and stem-loop 3 mutant used). Under SF3A1 knockdown conditions, baseline Exon 2 inclusion levels of the minigene are low (~25-30%) (Figure 2.36). Co-transfection of siSF3A1 with the WT FLAG-RNAiR-SF3A1 clone rescues full-length splicing of the minigene reporter with exon 2 inclusion reaching (~58%). All mutant FLAG-RNAiR-SF3A1 clones could increase the levels of full-length Dup51p splicing above baseline, although most mutants only rescued exon 2 inclusion to ~40-50%, significantly less than that of the WT clone (Figure 2.36). These findings demonstrate that mutations in SF3A1 at residues identified as interacting directly with U1-SL4 by CLIR-MS/MS, NMR, and X-ray crystallography, significantly reduce the strength of the SL4-UBL interaction and reduce SF3A1 splicing activity *in vivo*.
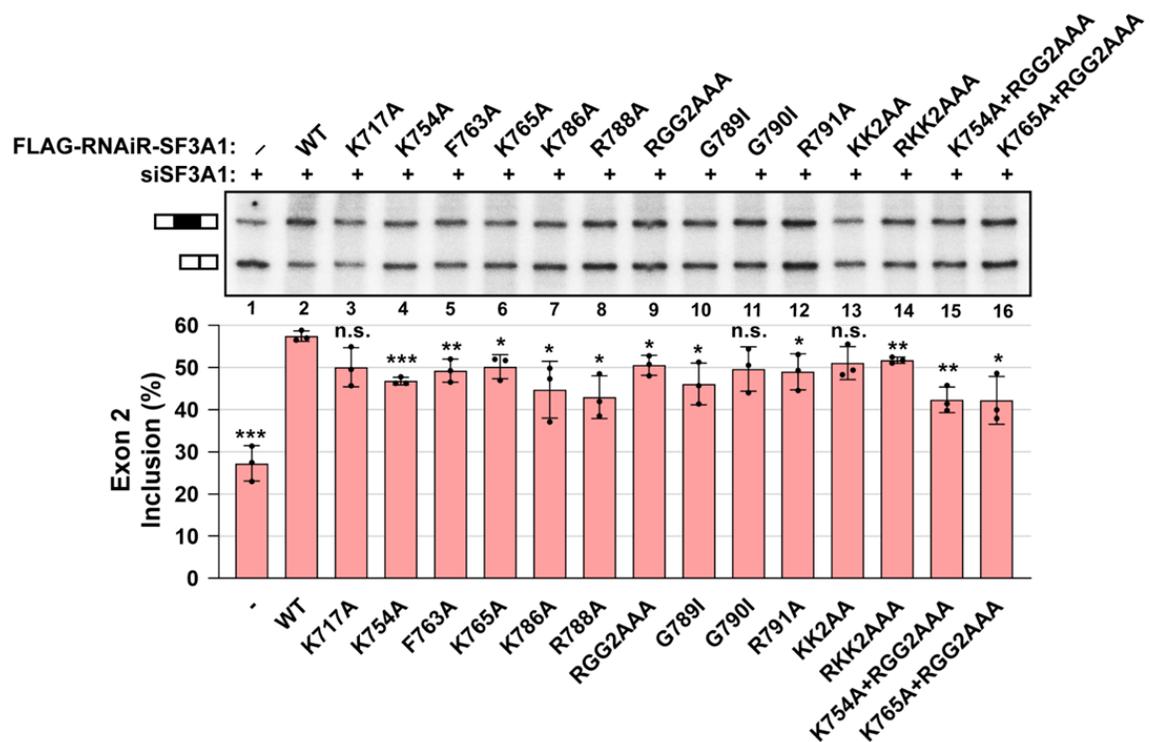
Figure 2.36 – Mutations in SF3A1-UBL interfere with rescue of the Dup51p minigene reporter under SF3A1 knockdown conditions. Primer extension analysis monitors the splicing of full-length and exon 2 skipped isoforms of the Dup51p minigene reporter (products are diagrammed to the left of the gel image). All cells were treated with siSF3A1 and plasmid harboring WT or mutant FLAG-RNAiR-SF3A1. In the absence of the RNAi resistant clone, full-length splicing of the reporter is inhibited (lane 1). Co-transfection with the WT RNAi resistant SF3A1 clone rescues exon 2 inclusion under siSF3A1 treatment (lane 2) which is reduced if splicing rescue is performed using mutant RNAi resistant SF3A1 (lane 3-16). Percent Exon 2 inclusion (n =3; * = p < 0.05, ** = p < 0.01, *** = p < 0.001) is graphed below the gel.

The impact of the Y772C and Y773C mutations was also tested using the minigene reporter assay. The Y772C and Y773C mutations that were found to reduce the RNA binding activity of GST-UBL for U1-SL4 by UV crosslinking and EMSA (Figures 2.21 and 2.23) were introduced into the FLAG-RNAiR-SF3A1 clone. RNAiR SF3A1 plasmids harboring these mutations express well after 48-hours post-transfection in HeLa cells (Figure 2.37). Both tyrosine mutations reduced the capacity for SF3A1 to rescue splicing under siSF3A1 treatment, compared to WT (Figure 2.38).
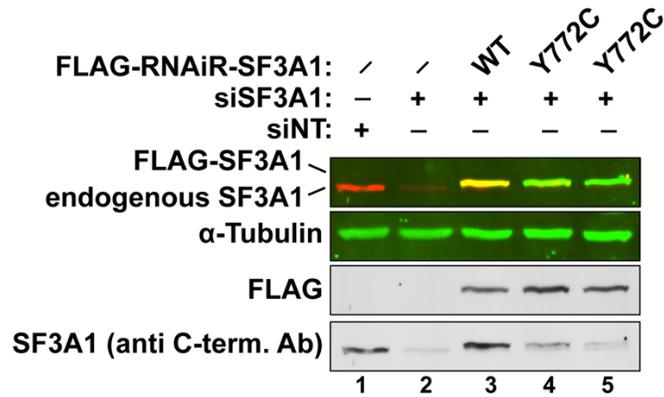
Figure 2.37 – FLAG-RNAiR-SF3A1 protein harboring Y772C and Y773C mutations are efficiently expressed in transfected HeLa cells. A merged Western blot image probed with anti-FLAG antibody and anti-SF3A1 C-terminus targeting antibody (anti C-Term. Ab) is shown in the top panel. Endogenous SF3A1 (red) is efficiently silenced and WT RNAiR SF3A1 is detectable using both antibodies (yellow). The Y772C and Y773C mutations interfere with SF3A1 (anti C-term. Ab) recognition but are detectable with anti-FLAG antibody (green). The individual scans from each antibody used in the merged image is shown in the lower panels and α-Tubulin was used as a loading control.
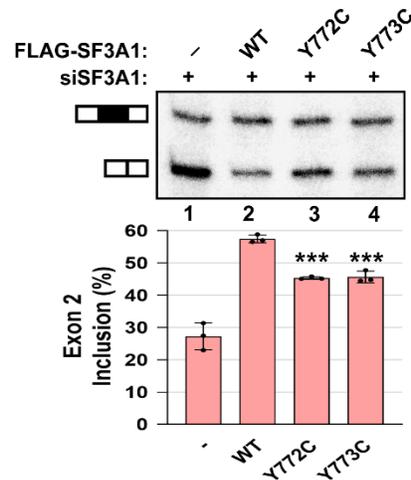
Figure 2.38 – The Y772C and Y773C mutations interfere with rescue of the Dup51p minigene reporter under SF3A1 knockdown conditions. Primer extension analysis monitors the splicing of full-length and exon 2 skipped isoforms of the Dup51p minigene reporter (products are diagrammed to the left of the gel image). Percent Exon 2 inclusion (n =3; * = p < 0.05, ** = p < 0.01, *** = p < 0.001) is graphed below the gel.

Although not identified as residues which directly contact U1-SL4 in the SL4-UBL structure, Y772 and Y773 are in proximity to the SL4 tetraloop and might provide structural support to the UBL domain as both tyrosines appear to engage in a T-shaped base-stacking interaction with one-another. Tyrosine residues dispersed throughout canonical RRMs have been found to be involved in maintaining the structural integrity of these domains and can interfere with RNA-binding activity when mutated to hydrophilic or non-aromatic residues [199]. Therefore, tyrosine to cysteine mutation at the highly conserved residues 772 and 773 of the UBL domain (Figure 2.17) may destabilize the overall stability of the fold leading to inefficient U1-SL4 binding.

Discussion

In this study, we identify a non-canonical RNA binding domain in the U2 snRNP specific protein SF3A1. Results from experiments employing UV-crosslinking, EMSA, and SPR demonstrate that the C-terminal Ubiquitin-like domain of SF3A1 (aa 703-793) is necessary and sufficient for binding to the stem-loop 4 of the U1 snRNA (Figure 2.3 and 2.5). The UBL domain of SF3A1 specifically recognizes the double-stranded, G-C rich stem-loop features of this RNA substrate (Figure 2.10 and 2.13). Also, this domain specifically binds U1-SL4 in the context of the U1 snRNP in HeLa cell nuclear extracts (Figure 2.24 and 2.25). Additionally, tyrosine residues Y772 and Y773 in the UBL domain likely play an important role in binding to U1-SL4 (Figures 2.19, 2.23 and 2.28). Finally, our search of human UBL domains for features that were found to be important for U1-SL4 binding by SF3A1-UBL, identified thirty-eight UBL domains which could potentially bind RNA (Figure 2.29). The C-terminal UBL domain of SF3A1 is unique to

66

higher eukaryotes and is not present in the yeast ortholog, Prp21, which is only 280 aa. Prp21 is homologous to the N-terminal region of SF3A1 that harbors the two SURP domains and the short charged region (Appendix B) [178]. Thus, the absence of a UBL domain in Prp21 suggests that the potential role of this SF3A1 domain (and its capacity to bind U1-SL4) during the early steps of spliceosome assembly may be unique to higher eukaryotes.

The binding affinity of SF3A1-UBL for U1-SL4 is high (K_D = ~97 nM) in comparison to the canonical RNA recognition motifs (RRMs) of another U1-SL4 interacting protein, the polypyrimidine tract-binding protein P1 (PTBP1). PTBP1 contains four RRMs, and RRM1 and RRM2 bind U1-SL4 with dissociation constants of 850 nM and 390 nM, respectively, and to single stranded CU-rich RNAs with K_D values ≥ ~1 μM [167, 200]. However, the affinity of the SF3A1-UBL/U1-SL4 interaction is significantly weaker than binding of the core U1 protein, U1A, to SL2 of the U1 snRNA. U1A very stably binds to SL2 via its N-terminal RRM with a K_D of ~32 pM [201]. Since, the SF3A1-U1 snRNA interaction would need to be disrupted during progression of spliceosome assembly (see discussion below), the intermediate nM range affinity of the interaction reported here might be reflective of the transient nature of this prespliceosomal RNA-protein contact.

Mutation analysis of conserved tyrosines in SF3A1-UBL has revealed a critical role for Y773 in binding to the U1-SL4 RNA and the intact U1 snRNP. However, the MDS-associated mutation Y772C was found to have a weaker effect, and reduced U1-SL4 binding and U1 snRNP association by only ~2-fold. This small effect is similar to the MDS-mutation induced change in RNA-binding affinity observed for other splicing

factors such as SRSF2 and U2AF1. SRSF2 mutations P95H/L/R were reported to cause ~1.2- to ~2.1-fold decrease in affinity for RNA sequences harboring the 5′-GGAG-3′ consensus [202, 203]. Similarly, the S34F mutation was found to decrease the binding affinity of U2AF1 for consensus 3′ splice site sequences by ~1.3- to ~4-fold [204, 205]. In both SRSF2 and U2AF1, the MDS-mutations also changed sequence specificity by causing an increase in the affinity for alternative sequences. However, any impact of the Y772C mutation on binding specificity of SF3A1 remains to be determined. Overall, the analysis showed that the magnitude of decrease in binding affinity caused by the SF3A1-Y772C mutation is similar to that observed for other MDS-associated mutations and may be sufficient for induction of pathogenic molecular effects.

The SL4-UBL structure confirmed that many of the conserved residues present on the positive-charged surface of the UBL domain (Figure 2.31) interact with U1-SL4 like K754, K765, K786, and R788. Additionally, the UBL-SL4 structure identified a base-stacking interaction between F763 and C151 in the loop of U1-SL4 that is important for binding (Figure 2.33 and Table 2.3). Base stacking interactions between cyclic amino acids and exposed nucleotides in the loop of stem-loop RNA is a common mechanism employed by canonical RRMs in binding RNA and F763 is part of a conserved RNP2-like motif (I-F-I) on the fourth β-sheet of SF3A1-UBL (Figures 2.6 and 2.17) [149]. The structure also revealed the importance of the C-terminal RGGRKK motif identified to be essential for binding SL4 by UV-crosslinking (Figure 2.8). The UBL tail sticks into the major groove of U1-SL4 and makes a number of base-specific, and phosphate backbone interacting, hydrogen bonding with the stem (Figure 2.33). This binding mechanism is also observed in RRMs that contain RGG motifs such as the protein FUS, where repeats

68

of C-terminal RGG motifs enhance the affinity of the protein to RNA substrates by engaging the minor groove of stem-loop RNA [206, 207]. Additionally, RGG motifs demonstrate some preference for G-C rich RNA and this may explain the inability for GST-UBL to bind the SL4-M10 mutant (Figures 2.13 and 2.15) [208-210]. Although the Y772 and Y773 residues do not contact directly with U1-SL4 as determined by the structure, Y→C mutations reduce binding by UV crosslinking (Figures 2.21 and 2.19), EMSA (Figure 2.23), affinity pull-down assay (Figure 2.28), and interfere with SF3A1 splicing activity *in vivo* (Figure 2.39). Although the mechanism underlying the impact of these tyrosine mutations cannot be explained by the contacts observed in the UBL-SL4 structure, their impacts in UBL RNA binding and SF3A1 splicing activity are clear and might be due to structural destabilization of the UBL fold, rather than direct inhibition of RNA-protein contacts. The UBL-SL4 structure is the first to capture the binding of a Ubiquitin-like domain to an RNA substrate and reveals that the interaction is facilitated by modes of RNA binding found in traditional RRMs like aromatic base stacking and the use of an RGG motif that stabilizes the complex via interaction with the SL4 major grove.

Our biochemical analyses support the occurrence of SF3A1 contact with U1 snRNA during the transition of the E to A complex [168]. These observations are supported by other studies that have demonstrated an essential role for SF3A1 in early prespliceosome formation and the potential association of SF3A1 with prespliceosomal proteins such as SF1 [180, 211]. How long the SF3A1-UBL/SL4 interaction may persist after A complex formation is not clear. Structural and compositional analyses of higher order human pre-B, B^{act}, and B* complexes demonstrate the presence of SF3A1 in pre-B

and B^{act} complexes, but not in B*, and that the SF3A complex is released through the action of the Prp2 helicase during the B^{act}→B* transition [96, 212]. Prior to the release of SF3A, the U1 snRNP is released by the Prp28 helicase during conversion of the pre-B to B^{act} complex [87]. The cryo-EM structures of the human pre-B complex do not demonstrate any contacts between the U1-U2 snRNPs, indicating that the SF3A1/U1 snRNA contact may be disrupted during the A→pre-B complex transition [88, 90]. However, the C-terminal half of SF3A1 is not visible in either of the pre-B complex structures, and therefore its orientation relative to U1-SL4 cannot be determined definitively. This difficulty in resolving the C-terminal region of SF3A1 is probably due to the central, proline-rich region that separates the UBL domain from the N-terminus, and is likely unstructured (Figure 2.1). Thus, this study provides relevant functional insight into a critical prespliceosome interaction involving SF3A1 and U1 snRNA that has been challenging to evaluate by structural analyses, and may be uniquely occurring during human spliceosome assembly. In addition, it identifies UBL domains with RNA binding potential in several other proteins that are involved in RNA processing steps in essential cellular pathways.

Experimental

Cell-free protein expression and UV crosslinking

Full-length and truncated SF3A1 proteins were expressed using the 1-Step Human Coupled in-vitro Translation Kit (Thermo Fisher Scientific). SF3A1 cDNA was cloned into NdeI/SalI restriction sites of pT7CFE-CHis expression vector and truncations were made by whole plasmid amplification via PCR. Constructs were added to cell-free

expression (CFE) reactions and incubated for six hours according to manufacturer's instructions. CFE reactions were then dialyzed against one liter of buffer DG (20 mM HEPES pH 7.9, 80 mM K-glutamate, 0.1 mM EDTA, 1 mM DTT, 1 mM PMSF, 20% glycerol) using Slide-A-Lyzer dialysis cassettes (Thermo Fisher Scientific) overnight at 4°C. Dialyzed reactions were aliquoted and stored at -20°C.

Radioactive ^{32}P-UTP labeled wild-type and mutant U1-SL4 RNAs were in vitro transcribed by T7 RNA polymerase from annealed DNA templates, gel purified, and ethanol precipitated. Crosslinking reactions consisted of 100 nM ^{32}P-U1-SL4, 2.2 mM MgCl$_2$, and 60% dialyzed CFE reactions. After a thirty-minute incubation at room temperature (RT), reactions were UV crosslinked with 1800 mJ total energy in a GS Gene Linker (Bio-Rad Laboratories). Reactions were separated on an SDS-PAGE gel that was dried and visualized by the Typhoon FLA 9500 Phosphor Imager.

Western and Northern analysis

For protein analysis, all samples were boiled in 1x SDS-PAGE sample buffer, separated on a 10 or 12% SDS-PAGE gel and analyzed by Coomassie blue staining or Western blotting using PVDF membranes. To analyze all SF3A1 truncations on single gels, precast 4-20% gradient gels were used (Bio-Rad Laboratories). Polyclonal antibodies against SF3A1 have been described previously [168]. Anti-α-Tubulin mouse monoclonal antibody (Calbiochem; CP06-DM1A), anti-6x His rabbit monoclonal antibody (Thermo Fisher Scientific; MA1-21315), anti-GST monoclonal antibody (Abcam; Ab92), and anti-FLAG monoclonal M2 antibody (Sigma-Aldrich; F1804) were

obtained commercially. Secondary anti-mouse and anti-rabbit antibodies conjugated to Cy3 and Cy5 fluorophores were purchased from GE Healthcare.

For Northern blotting, RNA samples were separated on 10% Urea-PAGE gels and transferred onto Amersham Hybond Nylon Membrane (GE Healthcare) for 1 hour at 15V and 400 mA using the Trans-blot Turbo Semi-dry Transfer System (Bio-Rad). Transferred RNA was UV crosslinked to nylon membranes for 10 min. and pre-hybridized in 15 ml of ULTRAhyb Hybridization Buffer (Thermo Fisher Scientific). The membranes were probed with ^{32}P-labeled oligo probes at 1.0 x 10^6 cpm/ml hybridization buffer overnight at 42°C. Sequences of oligonucleotides used to probe for target snRNAs are summarized in Table 2.4. Membranes were washed at the temperature of hybridization once with 2x saline-sodium citrate buffer (SSC) containing 0.1% SDS for 10 min. and twice with 2x SSC for 10 min. Northern blots were visualized using the Typhoon FLA 9500 Imager.

Oligo Name	Sequence (5′→3′)
U1$_{27-46}$R	TGATCACGAAGGTGGTTTTC
U2$_{114-135}$R	AGATGGAATAGGAGCTTGCTCC
U4$_{86-106}$R	CCGTGACGACTTGCAATATAG
U5$_{50-68}$R	GATTTCCGTGGAGAGGAAC
U6$_{36-57}$R	ACGATACAGAGAAGATTAGCA

Table 2.4 – Sequences of oligonucleotide probes used for Northern blotting. Primers displayed were ^{32}P-labeled and used to detect the major snRNAs in GST and GST-UBL affinity-purified complexes.

Glutathione agarose affinity chromatography and pull-down assay

GST-UBL fusion constructs were created by cloning the cDNA for the UBL domain of SF3A1 (aa 704-793) into the BamHI and XhoI restriction sites of plasmid pGEX-5x. GST and GST-UBL fusion proteins were expressed in *Escherichia coli*

(BL21-DE3) by induction with 0.5 mM IPTG for 4 hours at 37°C. Induced proteins were isolated using glutathione agarose beads (Thermo Fisher Scientific) according to manufacturer's protocol, separated by SDS-PAGE, and analyzed by Coomassie blue staining and Western blotting. Purified proteins were dialyzed against two liters of buffer DG and stored at -80°C.

For GST pull-down assays, HeLa cell nuclear extracts were pre-cleared of GST-binding proteins, as described previously [188]. Purified GST and GST-UBL proteins were added to 100 μl of nuclear extract at a final concentration of 2-5 μM. After a 30 min. incubation at 4°C, the reactions were added to a 30 μl packed volume of glutathione agarose beads and allowed to incubate for an additional 30 min. at 4°C with rotation. Beads were washed four times in buffer DG and bound proteins were eluted by boiling in 1x SDS-PAGE sample buffer. Total RNA was extracted with the TRIzol reagent.

Electrophoretic mobility shift assays

Binding reactions were prepared in 10 μl total volume and contained 10 nM 5′-Cy5-labeled U1-SL4 RNA (Integrated DNA Technologies, San Jose, CA), 2.2 mM MgCl$_2$, 60% buffer DG, and varying concentrations of GST-UBL protein. After incubation for one hour at RT, the binding reactions were loaded onto a horizontal native-PAGE gel (6% 29:1 Bis-Acrylamide, 0.1% ammonium persulfate, and 1:1000 dilution of TEMED in 0.5X TBE buffer) and run at 100V for 2 hours at 4°C [213]. The gels were visualized using the Typhoon FLA 9500 Imager and bands were quantitated with ImageQuant Software (GE Healthcare). For competition EMSAs, first, binding reactions containing 10 nM 5′-Cy5-U1-SL4 and 500 nM GST-UBL protein were prepared [186].

Then the competitor RNAs were added at varying concentrations and incubation was continued at RT for 1 hour. The complexes were separated and visualized as above. Dose response curves for fraction of Cy5-U1-SL4 bound versus log of protein or competitor RNA concentrations were generated by a non-linear regression analysis of the data and assuming one binding site using GraphPad Prism Software v8.1.0.

Surface plasmon resonance

SPR experiments were performed on the Pioneer FE Surface Plasmon Resonance System (ForteBio) and largely based on parameters described by Katsamba and colleagues [214]. 5′-biotinylated RNAs (Integrated DNA Technologies) were immobilized on streptavidin-coated biosensor chips (SADH Biosensors). Prior to immobilization, the RNAs were diluted to 1 µM in base SPR buffer (10 mM Tris-HCL, pH 8.0, 150 mM NaCl) heated to 65°C for 5 min, and cooled slowly to RT to allow folding. Folded RNA substrates were diluted 100-fold in SPR running buffer (10 mM Tris-HCl, pH 8.0, 150 mM NaCl, 5% glycerol, 62.5 µg/ml bovine serum albumin, 125 µg/ml tRNA, 1 mM dithiothreitol, and 0.05% tween-20), and injected at 10 µl/minute until 45-50 response units of RNA were captured on the SADH biosensor. For testing the impact of point mutations in tag-free UBL protein, 150 response units of biotinylated-U1-SL4 RNA was immobilized. GST-UBL protein stocks were diluted to a 500 nM concentration in SPR running buffer and ten to twelve, 2-fold serial dilutions were prepared and injected using the One-Step Kinetics assay setting. For tag-free WT and mutant UBL proteins, lyophilized proteins were reconstituted in SPR running buffer and WT UBL was injected at a top concentration of 1.0 µM while all mutants were injected at

top concentrations of 10 or 20 µM followed by a series of four to five 2-fold serial dilutions. Binding experiments were performed at 20°C at a flow rate of 150 µl/min. Under the conditions employed, >95% of the protein was removed from the surface during the dissociation phase and therefore a regeneration step was not required. Background signal from a streptavidin-only reference flow cell was subtracted from all data sets. Data were fit assuming a 1:1 stoichiometry to the Langmuir binding model to obtain kinetic parameters and dissociation constant.

Sequence analysis

Annotated human Ubiquitin-like domains and/or proteins were collected from the integrated annotations for Ubiquitin and Ubiquitin-like conjugations database (iUUCD) as well as from the InterPro application from EMBL-EBI [215, 216]. A total of 955 proteins were manually scanned for the presence of a conserved tyrosine at positions analogous to 772 and/or 773, the C-terminal RGG motif, and C-terminal positively charged residues. This search narrowed down the number of analyzed UBL domains to approximately 146 proteins. From this initial group, redundant protein sequences were purged, and sequences representative of UBL families were selected, yielding 38 UBL domains, which were aligned to SF3A1-UBL. The alignment of this final collection of UBL domains are displayed in Figure 2.29 and UniProt accession numbers are reported in Appendix C. All protein sequence alignments were performed using ClustalW [217]. Electrostatic surface potential analysis for SF3A1-UBL domain (PDB ID: 1ZKH) was performed in PyMOL [218]. RNA secondary structure predictions were performed in the mfold online web server [187].

Plasmid constructs and transfections

Cloning and purification of tag-free WT and UBL proteins, and downstream CLIR-MS/MS, NMR, and X-ray crystallography was performed by Tebbe de Vries and Sébastien Campagne. The SF3A1 UBL open reading frame was cloned into the vector pET24b to generate a GB1-6xHis-TEV-UBL fusion protein that could be efficiently expressed in *E. coli*. Expressed protein was then purified by Ni^{2+} affinity chromatography and tag-free UBL protein was isolated after TEV protease cleavage and size exclusion chromatography.

The three-exon/two-intron reporter pDUP51p and the U1 snRNA expression plasmid pNS6U1 have been described previously [168]. The construct expressing U1-5a snRNA carrying SL3 mutation M1d were generated by PCR mutagenesis using oligonucleotides and was verified by DNA sequencing. The sequences of the oligonucleotides used for U1-5a mutagenesis are provided in Appendix D. The full-length SF3A1 gene containing an upstream Kozak sequence followed by the FLAG-tag sequence (DYKDDDDK) was cloned into the HindIII and NotI restriction sites of the mammalian expression vector pcDNA3.1; allowing for the expression of N-terminal FLAG-tagged SF3A1 protein when transfected into HeLa cells. Primers used to introduce silent point mutations in SF3A1 generating the RNAi resistant SF3A1 clone, in addition those used to introduce point mutations into the C-terminal UBL domain, are reported in Appendix D.

HeLa cells, originally purchased from ATCC, were a gift from Kurt Gustin (University of Arizona, College of Medicine-Phoenix). They were cultured in DMEM

containing 10% FBS and antibiotics (100 U/mL penicillin, 100 μg/mL streptomycin).

Culture supernatants were tested for mycoplasma by PCR using a pool of six

mycoplasma specific primers, and found to be negative [219]. For the siRNA knockdown

and FLAG-RNAiR-SF3A1 rescue experiments using the Dup51p reporter minigene, 0.2

x10^5 cells per well of a 12-well plate were first transfected with 50 nM siSF3A1 and 125

ng of pcDNA3.1 plasmid harboring WT and mutant FLAG-RNAiR-SF3A1 constructs

using the Lipofectamine 2000 reagent (Thermo Fisher Scientific). After 24 hours, cells

were transfected with the Dup51p reporter and U1 snRNA expression plasmids and total

RNA was extracted after an additional 24-hour incubation using TRIzol reagent. In the

rescue experiments, a pNS6U1-5a/SL3-M1d to Dup51p ratio of 7.5:1 (1.5 μg pNS6U1-

5a/SL3-M1d and 0.2 μg Dup51p) was used along with the addition of pNS6U1-WT (0.3

μg) to maintain the total levels of U1 expression plasmid consistent to that used in

previous experiments (ratio of pNS6U1:Dup51p = 9:1; see Chapter 3). All siRNAs,

including the non-targeting control (siNT; siGENOME Non-Targeting Pool #1) were

purchased from Horizon Discovery; sequences are provided in Appendix E. 2-3 μg of

total RNA harvested from transfections were used in primer extension reactions using the

^{32}P-labeled Dup3r primer and exon 2 inclusion of Dup51p was monitored by separation

of reaction products on 10% Urea-PAGE gels (Appendix E).

CHAPTER 3

DETERMINING A ROLE FOR U1 STEM-LOOP 3 IN SPLICING

Publication Note

The research reported in this chapter has been prepared in a manuscript that has been submitted for peer review at the journal *RNA Biology*. William Martelly, Bernice Fellows, Paul Kang, Ajay Vashisht, James A. Wohlschlegel and Shalini Sharma. All co-authors have granted permission for this work to be included in this dissertation.

Overview

The spliceosome assembles from sequential binding of five small nuclear RNPs (snRNPs: U1, U2, U4, U5, and U6) and many auxiliary proteins [170]. First, binding of the U1 snRNP to the 5′ splice site (5′-ss), splicing factor 1 (SF1) to the branch-point sequence, and the U2 auxiliary factor (U2AF) 65/35 dimer to the polypyrimidine tract and the 3′ splice site (3′-ss), respectively, forms the early (E) complex. The U2 snRNP loosely associates with the E complex by a U2AF65-SF3B1 contact [175, 176]. Following this, RNA helicases UAP56 and Prp5 facilitate conversion of the E to A complex, which involves stable binding of U2 to the pre-mRNA by basepairing of the U2 snRNA to the branchpoint sequence [79, 220, 221]. The U4/U6.U5 tri-snRNP is then recruited to form the pre-B complex, and the U1 and U4 snRNPs are released by the actions of helicases Prp28 and Brr2, respectively [87, 94, 96, 171, 172]. In subsequent steps, complexes containing the U2/U6/U5 snRNPs perform splicing catalysis [103, 173,

174, 222]. Recent cryo-EM studies of the budding yeast and human spliceosomal complexes have revealed the nature of molecular contacts and transitions in the later complexes. However, information on early interactions that lead to formation of a stable A complex is lacking, especially for the human spliceosome.

The human and budding yeast U1 snRNPs differ significantly in their composition. Human U1 small nuclear RNA (snRNA) is 164 nucleotides (nts) long and folds into a structure consisting of four stem-loops. It interacts with the seven-member (B/B′, D1, D2, D3, E, F, and G) Sm-ring and three U1-specific proteins (U1-70k, U1C, and U1A). In the mature U1 particle, the first three stem-loops are separated from the terminal stem-loop 4 (SL4) by the Sm-ring [77, 223]. The yeast U1 snRNA is much longer (568 nts) than its human paralog and lacks a structure analogous to SL4 downstream of the Sm ring [177, 224, 225]. The yeast stem-loop 3 (SL3) region is 15 times the size of SL3 in human U1, and folds into seven stem-loops. Yeast U1 contains seven additional particle specific proteins, namely Prp39, Prp40 (human Prp40), Prp42, Nam8 (human TIA-1), Snu56, LUC7 (human LUC7L), and Snu71 (human RBM25). In humans, Prp40, TIA-1, LUC7L, and RBM25 act independently as alternative splicing factors and orthologs for the other proteins have not been identified [226-228].

Cryo-EM structures of the human E and A complexes are not available, but structures of the yeast complexes provide some insight into early spliceosomal interactions. In the yeast E complex, Prp40 bridges the 5′- and 3′-ss complexes via interactions with U1-70k and SF1 (yeast MSL5) [229]. Other biochemical studies have also reported the occurrence of this contact [230]. The yeast A complex structure identifies two regions of contact between the pre-mRNA bound U1 and U2 snRNPs [86].

The first interface forms from a stable interaction between the U1 protein Prp39 and the core U2 protein U2A′ (yeast Lea1). The second interface involves interactions of SL3 of the U1 snRNA with the U2 proteins SF3B3 (yeast Rse1) and SF3A3 (yeast Prp9). The Prp39-U2A′ contact was found to be preserved in the yeast pre-B complex [231]. Interestingly, proteins SF1, U2AF65 (yeast MUD2), and the RNA helicases UAP56 (yeast SUB2) and Prp5 were not found in the yeast A complex; Prp5 was detected at sub-stoichiometric levels, but not observed in the structure. Thus, the associations of these proteins (and possibly other early factors) with the pre-mRNA are likely transient and precede formation of the stable interfaces observed in the A complex. The dynamic nature of these interactions may present challenges in the structural analysis of the very early spliceosomal complexes. Additionally, differences in compositions of yeast and human U1 snRNPs suggest that the contacts made by human U1 during spliceosome assembly may be different from those observed in yeast.

Previously, we reported a human spliceosome specific contact between pre-mRNA bound U1 and U2 snRNPs that is crucial for splicing. We found that SL4 of the 5′-ss bound U1 snRNA interacts with the 793 amino acids (aa) long protein SF3A1 of the 3′-ss bound U2 snRNP during the E to A complex transition [168], and identified the C-terminal Ubiquitin-like (UBL) domain of SF3A1 (aa residues 703-793) as the SL4 binding region [232]. The 280 aa long yeast ortholog of SF3A1 (Prp21) lacks a UBL domain and the yeast U1 snRNA lacks a SL4 like structure, thereby explaining the absence of this U1-U2 contact in yeast spliceosomal complexes [86, 178, 184, 231]. Here, we demonstrate that SL3 of the U1 snRNA is also important for U1 function in splicing and identify the ATP dependent DExD/H box RNA helicase UAP56 as a SL3

interacting spliceosomal protein. The spliceosomal interactions of SL3 and SL4 are distinct since UAP56 binds to SL3 but not to SL4 and conversely, SF3A1-UBL binds to SL4 but not to SL3. The UAP56 interaction with U1-SL3 occurs in an ATP-dependent fashion and we also observe the co-purification of UAP56 with the U1 snRNP specifically in the presence of ATP-γ-S.

Results

Stem-loop 3 of the U1 snRNA is important for U1 function

We have developed a genetic complementation assay that uses a 3-exon/2-intron minigene reporter (Dup51p) to examine the role of U1 snRNA in pre-mRNA splicing (Figure 3.1) [168]. When expressed in HeLa cells, 5′-ss mutations in the second intron of Dup51p cause skipping of exon 2 in the mature transcript (Figure 3.3 lane 1). These 5′-ss mutations can be complemented with a compensatory U→A mutation at the 5th position in the U1 snRNA 5′ region that basepairs with the pre-mRNA. In cotransfection assays, expression of the U1-5a snRNA rescues exon 2 inclusion in the Dup51p transcript (Figure 3.3 lane 2).

Figure 3.1 – A schematic of the 3-exon 2-intron minigene reporter applied to study the impacts of stem-loop 3 mutations on U1 activity is shown. The Dup51p pre-mRNA carries a 5′-ss mutation (indicated by the red asterisk) in intron 2 that causes skipping of exon 2 in the mature transcript when transfected into HeLa cells.

Using this assay, we examined the role of SL3 of the U1 snRNA in pre-mRNA splicing. SL3 consists of a nine basepairs long stem with a single cytidine bulge and a seven nts long terminal loop (Figure 3.2). To evaluate the role of SL3, we created 16 variants of the U1-5a construct carrying SL3 mutations (Figure 3.2). In M1a, the cytidine bulge was deleted. In M1b, M1c, and M1d, G to A changes disrupted basepairing in the stem. Additionally, in M1e, M1f, and M1g, A to U changes were introduced. The strands of the stem were swapped in M1h and in M1i, which also had deletion of the cytidine bulge. The G-C and A-U basepairs were altered partly in M1j and M1k, and completely in M1l. The number of nucleotides in the terminal loop was reduced in M2a, M2b, M2c, and M2d.
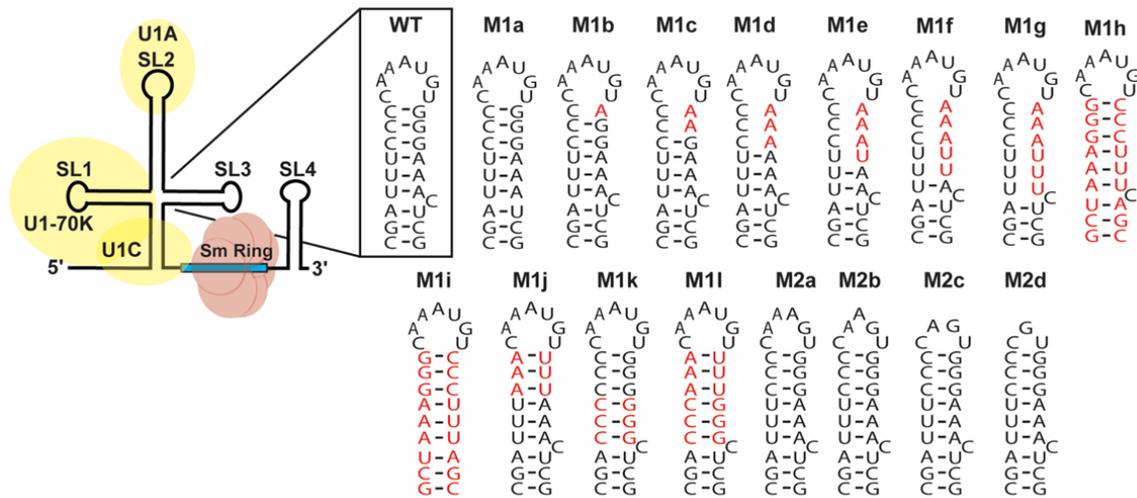
Figure 3.2 – The U1 stem-loop three structure is unbound by U1 specific proteins and forms a nine basepair stem with a seven-nucleotide loop. Schematic diagram of the U1 snRNP. Secondary structure and sequence of wildtype SL3 and mutations introduced into the U1 snRNA are depicted; nucleotide changes are shown in red.

The U1-5a variants carrying SL3 mutations were cotransfected with the Dup51p reporter and tested for their ability to rescue exon 2 inclusion (Figure 3.3). The analysis revealed that disruption of three or more basepairs in the upper region of the stem significantly ($\geq 10\%$ reduction in exon 2 inclusion and $p < 0.05$) affected the rescue of full-length Dup51p splicing (M1e, M1f, and M1g; Figure 3.3 lanes 7-9). Disruption of 1-2 basepairs in M1c and M1d had a smaller effect (lanes 5 and 6). Changes that did not exert any effect included deletion of the cytidine bulge (M1a; lane 3), swapping of the two strands of the stem by itself or in combination with deletion of the cytidine bulge (M1h and M1i; lanes 10 and 11), and reducing the size of the terminal loop from six to three nucleotides in M2a, M2b, M2c, and M2d (lanes 15-18). Changing the basepairs in the stem from A-U to G-C and vice versa also did not significantly reduce exon 2 inclusion in M1h, M1i, M1j, M1k, and M1l (lanes 10-14), thereby indicating that the sequence did not matter, as long as the basepairing was maintained.
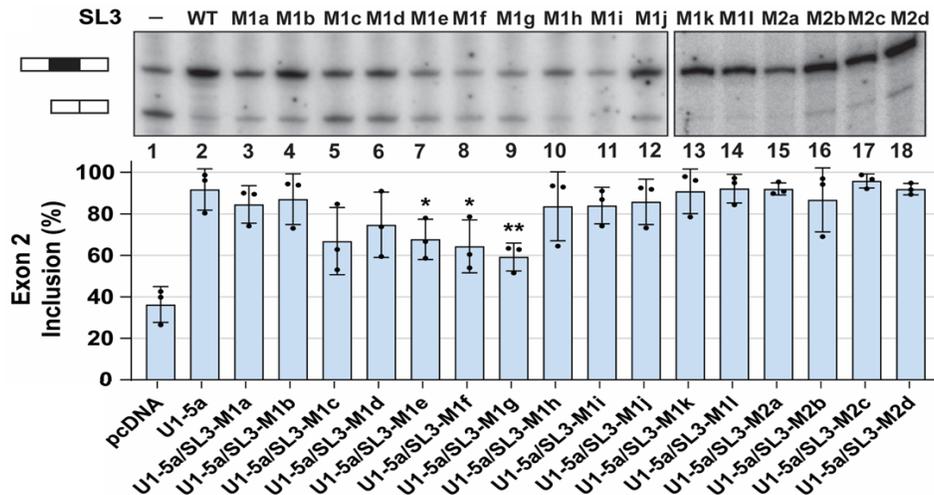
Figure 3.3 – Stem-loop 3 of the U1 snRNA is important for U1 function. Primer extension analysis to monitor splicing of the minigene reporter Dup51p after co-transfections with control (pcDNA) or U1-5a plasmids expressing wildtype or mutant U1 snRNAs. The full-length and exon 2 skipped Dup51p mRNA products are depicted. The percentage of the full-length product ($\pm$ s.d.) is represented in the graph below and statistical significance was determined by comparisons to the wildtype control (lane 2) (n = 3; * = p < 0.05, ** = p < 0.01).

83

The loss of basepairing in M1e, M1f, and M1g displayed the maximum effect and significantly reduced exon 2 inclusion from ~96% to ~60-70% (lanes 7-9). Thus, SL3 of the U1 snRNA plays an important role in U1 function. RT-qPCR quantification showed that expression of variants carrying SL3 mutations was ~4-fold more than the endogenous U1 snRNA (Figure 3.4), thereby indicating abundant expression of U1-5a variants and that loss of activity of the snRNAs is caused by the mutations in SL3.
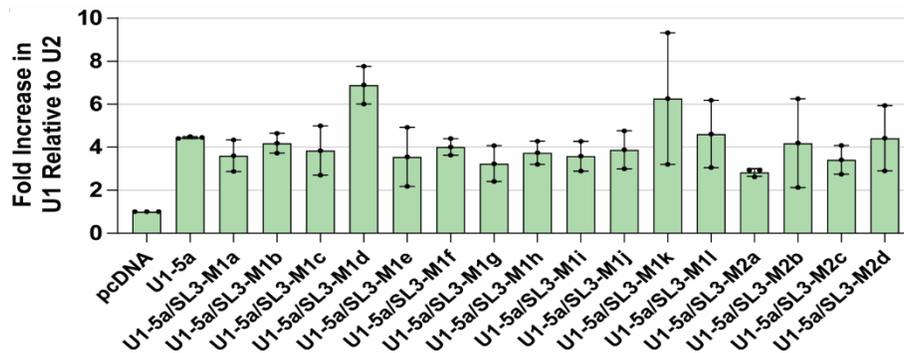
Figure 3.4 – All plasmid encoded U1-5a variants harboring mutations in SL3 are overexpressed in transfected HeLa cells. RT-qPCR analysis of U1 snRNA expression in HeLa cells co-transfected with Dup51p and U1-5a variants carrying wildtype SL3 or mutations. Fold change in U1 snRNA expression was calculated relative to the pcDNA control after normalization to U2; fold change (± s.d.; n = 3) in U1 is graphed.

U1-SL3 and U1-SL4 bind distinct spliceosomal proteins

To identify U1-SL3 interacting spliceosomal proteins, we first performed UV crosslinking using uniformly ^{32}P-labeled U1-SL3 RNA and found that in HeLa nuclear extracts, SL3 crosslinks only to a ~50 kDa protein (Figure 3.5). The efficiency of crosslinking was similar in the presence of ATP and ATP-γ-S (Figure 3.5 lanes 1-3 and 7-9), but much weaker in the absence of ATP (lanes 4-6). Crosslinking of the mutant U1-SL3-M1g RNA was different from that of U1-SL3-WT and easily disrupted upon preincubation of the nuclear extract with NaCl (Figure 3.6 lane 5-8). In the case of U1-

84

SL3-WT, significant amounts of the ~50 kDa crosslinked product formed even at higher

NaCl concentrations (Figure 3.6 lanes 2 and 3). These results indicate that wildtype U1-

SL3 specifically interacts with a ~50 kDa protein and mutations in the upper region of the

stem disrupt this interaction.
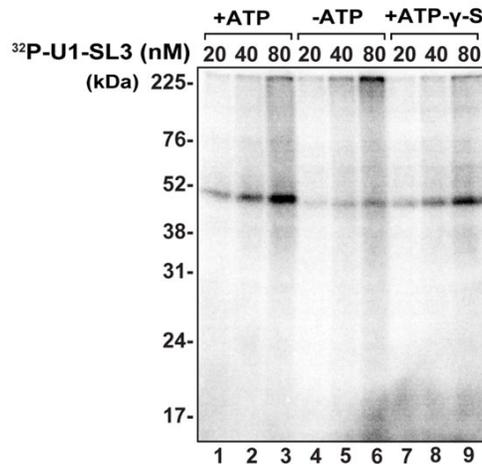
Figure 3.5 – Free U1-SL3 primarily crosslinks to a ~50 kDa protein in HeLa nuclear extract in an ATP-dependent manner. UV crosslinking analysis for U1-SL3 interacting protein(s). HeLa nuclear extracts were incubated with 20, 40, and 80 nM ^{32}P-U1-SL3 RNA in the presence or absence of ATP and ATP-γ-S.
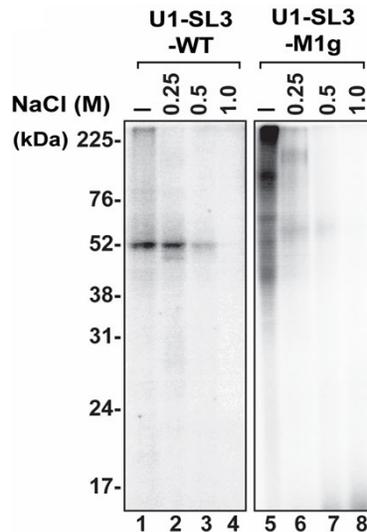
Figure 3.6 – Crosslinking of U1-SL3 to the ~50kDa protein persists even under stringent high salt conditions and is specific to WT SL3. UV-crosslinking analysis of wildtype and mutant ^{32}P-U1-SL3 RNAs in the presence of increasing NaCl concentration (0 - 1.0 M NaCl). Interacting proteins were UV crosslinked and analyzed by SDS-PAGE.

We isolated the U1-SL3 complex by RNA affinity purification (RAP) and

analyzed it for snRNA and protein composition. The snRNA analysis showed that none

of the spliceosomal snRNAs were present in either the U1-SL3-WT or the U1-SL3-M1g

complexes (Figure 3.7 lanes 3 and 4). As reported previously, the U2 snRNA was present

in the U1-SL4-WT complex (lane 2) [168].
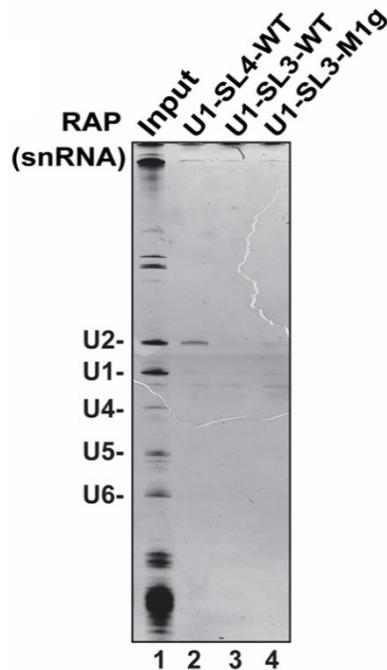
Figure 3.7 – Unlike U1-SL4 which can pull-down the U2 snRNP, U1-SL3 cannot affinity purify any snRNPs. SnRNA analysis of the complexes that bind to biotinylated U1-SL4-WT, U1-SL3-WT, and U1-SL3-M1g RNAs in HeLa nuclear extract. Total RNA was extracted from complexes, separated by urea-PAGE and visualized by GelRed total nucleic acid staining. Positions of the spliceosomal snRNAs are indicated.

To analyze protein composition, U1-SL3-WT and U1-SL3-M1g bound proteins

were subjected to mass spectrometry (MS). Proteins enriched in either the wildtype or the

mutant SL3 complexes and proteins present in both complexes were identified by

comparing the normalized spectral abundance factor (NSAF) for each protein [233]. This

identified several core spliceosomal proteins in the wildtype and mutant U1-SL3

complexes (Table 3.1). Since U1-SL3 crosslinked to a ~50 kDa protein, we focused on

validating proteins in this molecular weight range, including the RNA helicase UAP56

(DDX39B) and Prp19. Lists of peptides for U1-SL3-WT and U1-SL3-M1g complexes

are provided in Appendix F.

WT only	WT and Mutant	Mutant only
DDX39B (O00148)	LSm2 (Q9Y333)	LSm3 (P62310)
DDX17 (Q92841)	LSm6 (P62312)	LSm7 (Q9UK45)
SF3B3 (Q15393)	SmD2 (P62316)	LSm4 (Q9Y4Z0)
SF3A3 (Q12874)	SmF (P62306)	LSm1 (O15116)
Prp19 (Q9UMS4)	DDX5 (P17844)	U2AF1 (Q01081)
U5-116k (Q15029)	U2AF2 (P26368)	ISY1 (Q9ULR0)
SF3B1 (O75533)	U1-70k (P08621)	SF1 (Q15637)
Prp8 (Q6P2Q9)	SF3B1 (Q15459)	Aquarius (O60306)
U5-200k (O75643)	SF3B2 (Q13435)	SRRM2 (Q9UQ35)

Table 3.1 – Spliceosomal proteins identified in the wildtype and mutant U1-SL3 complexes. Summary of abundant proteins detected by mass spectrometry present in WT and or mutant M1g U1-SL3 RNA affinity purified complexes. Top candidates for relevant U1-SL3 interacting proteins based on results from crosslinking assays are highlighted and are both ATP-dependent helicases with masses of approximately 50 kDa.

Immunoblot analysis confirmed the presence of UAP56 in the U1-SL3-WT

complex. In comparison to mutant U1-SL3-M1g, binding of UAP56 to U1-SL3-WT was

found to be stronger (Figure 3.8 compare lanes 2 and 6). Similar to UV crosslinking

analysis, preincubation with up to 250 mM NaCl did not compete out UAP56 from the

U1-SL3-WT complex indicating strong binding (Figure 3.8 lane 3). Prp19 was not found

in either the wildtype or the mutant U1-SL3 complexes. U1 and U2 proteins, U1-70k and

SF3A1, respectively, were also not detected in either complex.
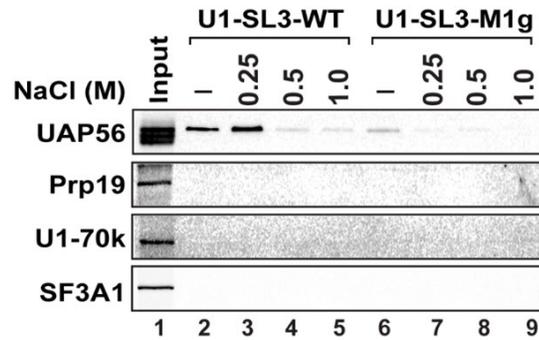
Figure 3.8 Western blotting validates results from mass spectrometry identifying the DExD/H box helicase UAP56 as a U1-SL3 interacting protein. Western analysis of proteins in wildtype and mutant U1-SL3 complexes. HeLa nuclear extracts were preincubated with 0, 0.25, 0.5, and 1.0 M NaCl prior to RNA affinity purification using biotinylated U1-SL3-WT and U1-SL3-M1g RNAs.

Previously, we demonstrated that SL4 of the U1 snRNA binds to SF3A1 via the C-terminal UBL domain [232]. To test if SL3 and SL4 can directly interact with UAP56 *in vitro*, we expressed and isolated glutathione S-transferase-UAP56 fusion protein (GST-UAP56), GST-SF3A1-UBL, and GST alone from *Escherichia coli* (Figure 3.9) and performed electrophoretic mobility shift assays (EMSAs).
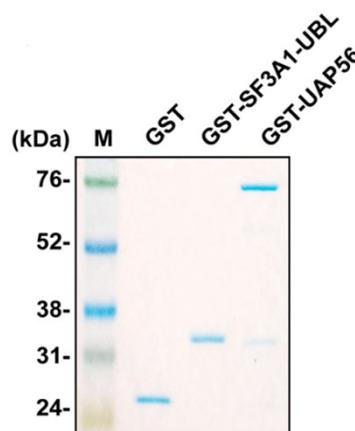
Figure 3.9 – A GST-UAP56 fusion protein was used to study the interaction of UAP56 to U1-SL3 RNA. Coomassie stained SDS-PAGE gel of purified GST, GST-SF3A1-UBL, and GST-UAP56 proteins that were used in RNA binding experiments.
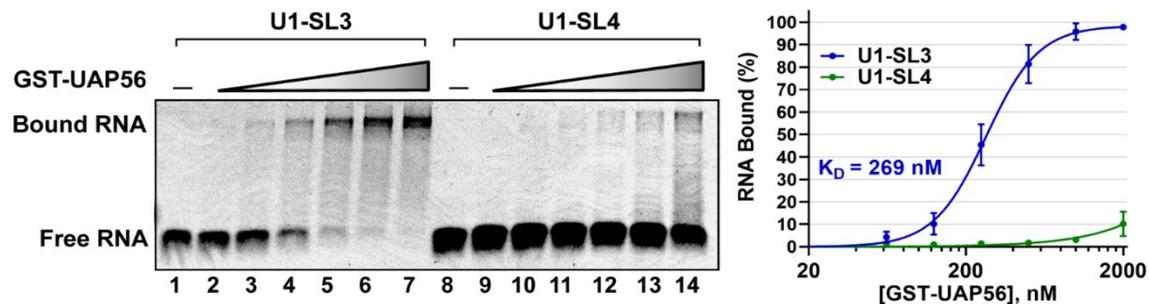
Figure 3.10 Purified GST-UAP56 protein specifically binds U1-SL3 RNA substrate. EMSAs monitoring binding of Cy5-labeled U1-SL3 (lanes 1-7) or U1-SL4 (lanes 8-14) RNAs (10 nM) in the absence and presence of GST-UAP56 (0.0625, 0.125 0.25, 0.5, 1.0, and 2.0 μM). Dose-response curves were generated from EMSA images by plotting the average percent of bound U1-SL3 and U1-SL4 RNA (± s.d.) versus GST-UAP56 protein concentration and is displayed to the right of the gel image (n = 3).

Purified GST-UAP56 was incubated with Cy5 labeled U1-SL3 and U1-SL4 RNAs (10 nM) in the presence of ATP. GST-UAP56 bound U1-SL3 with a K_D of ~269 ± 43 nM (Figure 3.10 lanes 2-7) but did not bind to U1-SL4 (Figure 3.10 lanes 9-14). An interaction between UAP56 and U1-SL4 was not detected even when a higher concentration of RNA (50 nM) was used in presence of either ATP or ATP-γ-S (Figure 3.11), or at the highest GST-UAP56 concentration (8 μM) used in this study (Figure 3.12 lane 14).
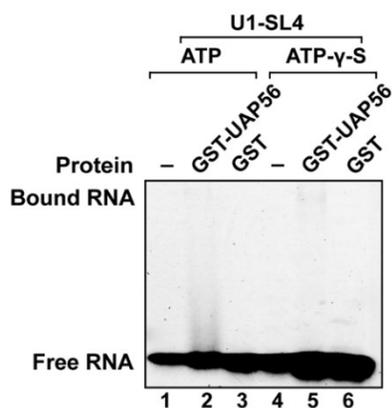
Figure 3.11 GST-UAP56 does not stably interact with U1-SL4. EMSAs monitoring binding of Cy5-labeled U1-SL4-WT RNAs (50 nM) to GST or GST-UAP56 (2.0 μM) in the presence of either ATP (lanes 1-3) or ATP-γ-S (lanes 4-6).
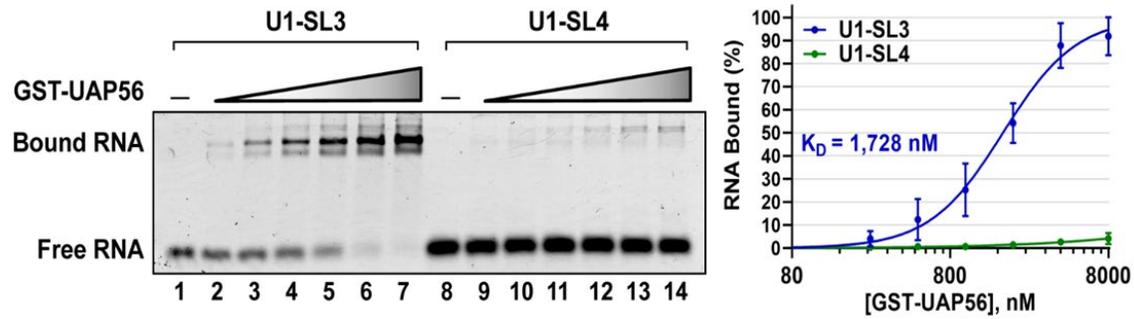
Figure 3.12 – GST-UAP56 can bind U1-SL3 in the presence of ATP-γ-S but with lower affinity. EMSA monitoring binding of Cy5-labeled U1-SL3 (lanes 1-7) or U1-SL4 (lanes 8-14) RNAs (10 nM) to GST-UAP56 (0.125, 0.25, 0.5, 1.0, 2.0, 8.0 μM) was performed in the presence of ATP-γ-S. Dose-response curves were generated from EMSA images by plotting the average percent of bound U1-SL3 and U1-SL4 RNA (± s.d.) versus GST-UAP56 protein concentration and is displayed to the right of the gel image (n = 3).

The U1-SL3 binding affinity of UAP56 in the presence of ATP-γ-S (K_D = ~1,728 ± 243 nM; Figure 3.12) was 6-fold lower than that in the presence of ATP (K_D of ~269 ± 43 nM). At higher concentration of U1-SL3 (50 nM), a smear pattern was observed in the presence of ATP, which is likely due to binding and subsequent dissociation of UAP56 from the RNA substrate (Figure 3.13 and 3.14).
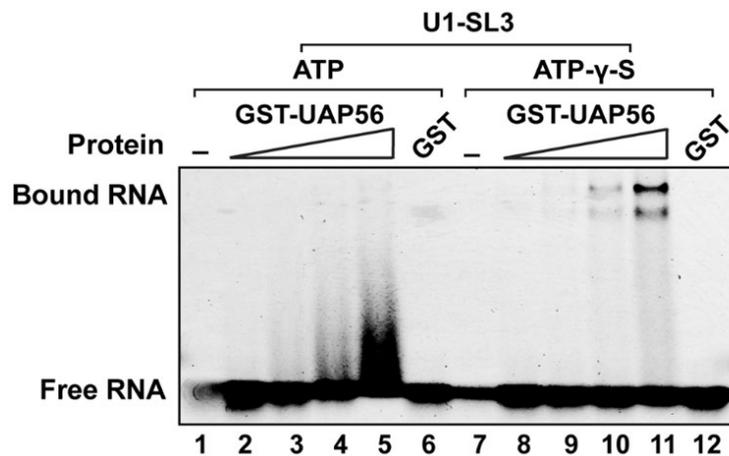
Figure 3.13 – No stable binding is detectable between UAP56 and U1-SL3 when incubated with high concentrations of RNA. EMSAs monitoring binding of Cy5-labeled U1-SL3-WT RNAs (50 nM) to GST (2.0 μM) or GST-UAP56 (0.25, 0.5, 1.0, and 2.0 μM) in the presence of either ATP (lanes 1-6) or ATP-γ-S (lanes 7-12).

90

Binding reactions in the presence of other NTPs confirmed the ATP specificity of

the UAP56-U1-SL3 interaction. Neither UTP, GTP, nor CTP supported formation of the

GST-UAP56/U1-SL3 complex (Figure 3.14 lanes 5, 6, and 7), thereby indicating that the

characteristics of this interaction are consistent with those of an ATP-dependent DEAD

box helicase. GST by itself did not bind to either U1-SL3 or U1-SL4 (Figure 3.11 lanes 3

and 6 and Figure 3.13, lanes 6 and 12).
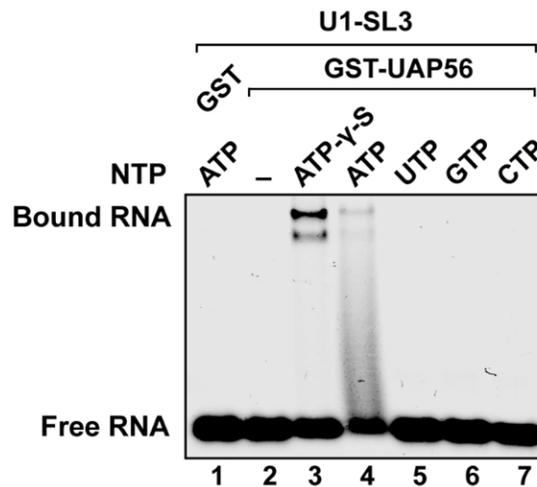
Figure 3.14 – UAP56 requires ATP to bind U1-SL3. EMSA for determining
NTP specificity of UAP56 was performed with 50 nM Cy5-U1-SL3 and 2.0 μM
GST or GST-UAP56 in the presence of ATP-γ-S, ATP, UTP, GTP, or CTP.

Finally, compared to GST-UAP56, GST-SF3A1-UBL demonstrated lack of

significant binding to U1-SL3 (Figure 3.15, lanes 2-7), but bound U1-SL4 with a K_D of

124 ± 17 nM, as previously reported [232]. Overall, these results demonstrate that distinct

spliceosomal proteins interact with SL3 and SL4. UAP56 interacts with U1-SL3 but not

with U1-SL4, while SF3A1 interacts with U1-SL4 but not with U1-SL3.

91
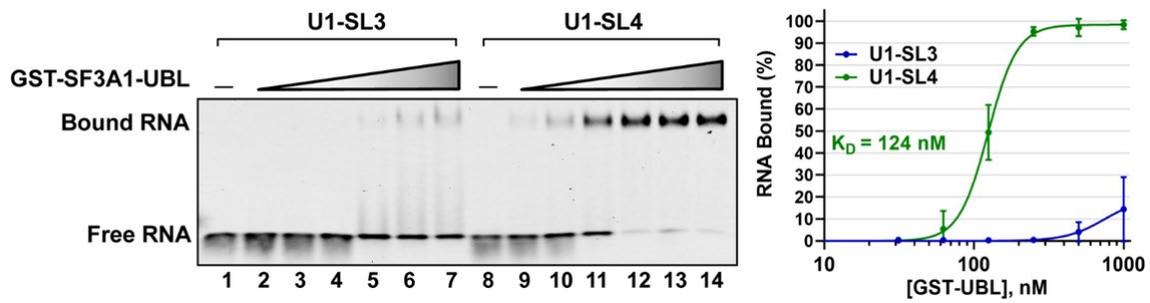
Figure 3.15 – U1-SL3 and SL4 bind to distinct spliceosomal proteins. EMSAs monitoring binding of Cy5-labeled U1-SL3 (lanes 1-7) or U1-SL4 (lanes 8-14) RNAs (10 nM) to GST-UBL (0.03125, 0.0625, 0.125 0.25, 0.5, 1.0 μM). Displayed dose-response curves were generated by plotting the average percent of bound U1-SL3 and U1-SL4 RNA (± s.d., n = 3) versus GST-UBL protein concentration and the apparent affinity constant values (K_D) are reported.

UAP56 association with the U1 snRNP requires ATP

To examine if UAP56 associates with U1 and U2 snRNPs, we performed affinity purification (AP) using 3′-biotinylated, 2′-O-methyl antisense-oligonucleotides (ASO). Nuclear extracts were incubated with the $U1_{1-13}$-ASO or $U2_{1-21}$-ASO in the absence of ATP or in the presence of ATP or ATP-γ-S. RNA and protein analysis of the AP complexes showed presence of the U1 snRNA and the U1 protein U1-70k in U1 complexes (Figure 3.16 and 3.17 lanes 3-5) and of the U2 snRNA and the U2 protein SF3A3 in U2 complexes (Figure 3.18 and 3.19 lanes 3-5) in all three conditions.
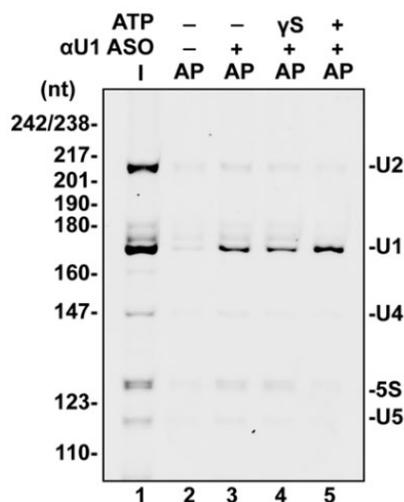
Figure 3.16 – To study the interaction of UAP56 with the U1 snRNP, U1 was affinity purified using anti-sense oligonucleotides from HeLa nuclear extract under varying conditions. Urea-PAGE analysis of RNA present in input (I) and U1 affinity purified (AP) complexes in the absence and presence of ATP-γ-S or ATP. RNA in the purified complexes were detected using GelRed total nucleic acid staining of urea-PAGE gels.
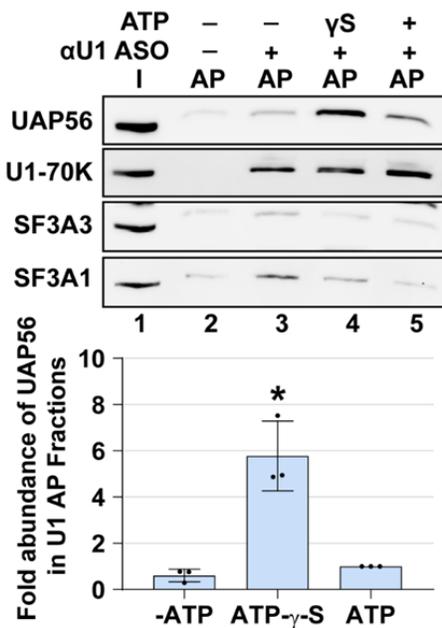
Figure 3.17 – UAP56 co-purifies with the U1 snRNP in the presence of ATP-γ-S specifically. Western analysis of proteins present in input (I) and U1 affinity purified (AP) complexes in the absence and presence of ATP-γ-S, and ATP. The intensity of the UAP56 band was normalized to that of U1-70K protein in the U1 complexes and then fold change was calculated relative to the plus ATP condition (± s.d., n = 3, * = $p < 0.05$).

93
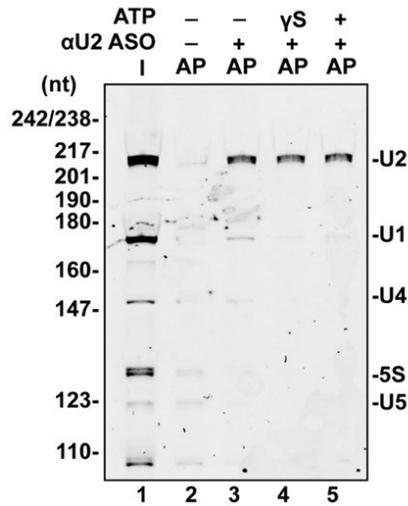
Figure 3.18 – To study the interaction of UAP56 with the U2 snRNP, U2 was affinity purified using anti-sense oligonucleotides from HeLa nuclear extract under varying conditions. Urea-PAGE analysis of RNA present in input (I) and U2 AP complexes in the absence and presence of ATP-γ-S or ATP. RNA in the purified complexes were detected using GelRed total nucleic acid staining of urea-PAGE gels.
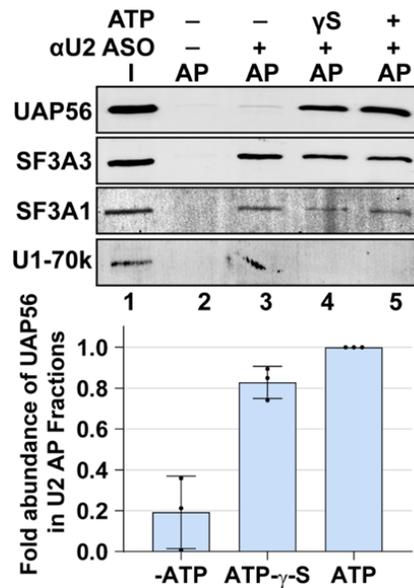
Figure 3.19 – Unlike U1, UAP56 co-purifies with the U2 snRNP in the presence of ATP-γ-S and ATP. Western analysis of proteins present in input (I) and U2 AP complexes in the absence and presence of ATP-γ-S, and ATP. The intensity of the UAP56 band was normalized to that of SF3A3 protein in the U2 complexes and then fold change was calculated relative to the plus ATP condition ($\pm$ s.d., n = 3, * = $p < 0.05$).

Association of UAP56 with U1 and U2 was ATP-dependent. UAP56 was observed in U2 complexes in the presence of both ATP-γ-S and ATP (Figure 3.19, lane 4 and 5), but in the case of U1, UAP56 was present in complexes assembled in the presence of ATP-γ-S but not ATP (Figure 3.17 lane 4 and 5). Thus, UAP56 has the capacity to interact with the U1 and U2 snRNPs. UAP56 has been shown to interact with the U2 snRNP associated protein U2AF65 and also to be recruited to the 3′-ss complex via this interaction [79, 234, 235]. Our results show that utilizing the energy from ATP hydrolysis, UAP56 can dissociate from U1 but not U2. Notably, ATP requirement is a common feature of U1-SL3 interaction with the ~50 kDa protein in HeLa nuclear extracts and purified GST-UAP56, and for the association of UAP56 with the U1 snRNP.

Discussion

Previously, we reported that an interaction between SL4 of the U1 snRNA and SF3A1 of the U2 snRNP occurred during the transition of the E to A complex [168]. In this study, mutation analysis by the U1 complementation assay revealed that SL3 mutations affect pre-mRNA splicing. Binding analysis demonstrated selective interactions of SL3 with UAP56 and SL4 with SF3A1. Selective binding of UAP56 to U1-SL3, but not to U1-SL4 is intriguing. UAP56 belongs to helicase super family 2 and has been shown to exhibit single and double stranded (ss and ds) RNA-dependent ATPase activity, and to unwind dsRNA, without sequence specificity, in an ATP-dependent manner [79, 236].

There are few examples of DEAD-box helicases that exhibit selective binding to RNA substrates. The *E. coli* DEAD-box protein A (DbpA) and its *Bacillus subtilis*

95

ortholog (YxiN) are RNA-dependent ATPases with an exceptional specificity for a short hairpin (H92) in the bacterial 23S rRNA [237, 238]. Recently, a human DEAD-box helicase DDX55 was shown to interact with domain IV of 28S rRNA in nuclear pre-ribosomal complexes with some specificity and was also reported to exhibited higher affinity for dsRNA than ssRNA [239]. Interestingly, ATPase activity of the yeast RNA helicase Prp5, the other helicase required for A complex formation, was demonstrated to be stimulated by U2 snRNA [240, 241]. To understand how UAP56 discriminates between free SL3 and SL4 and the mechanism underlying the action of the SL3-UAP56 complex in the context of early spliceosome assembly, a systematic analysis of the binding of wildtype and mutant SL3 RNAs by UAP56 and their impact on its ATPase and helicase activities is required.

The U1 snRNP and its accessible SL3 and SL4 structures have been reported as targets of hnRNP proteins in normal regulation of alternative splicing, and in pathogenesis of diseases associated with RNA-binding proteins. The SL4 of the U1 snRNA is targeted by the PTB protein during repression and enhancement of cassette exon inclusion. In the *Src* pre-mRNA, PTB bound to intronic sequences upstream and downstream of exon N1 interacts with SL4 of the pre-mRNA associated U1 snRNP [125, 126, 167]. This precludes U1 snRNA interactions with U2 snRNP components present in the downstream 3′-ss complex, thereby blocking formation of an active spliceosome. Studies on enhancement of exon inclusion by binding of PTB to sites that are present only in the intron downstream of the regulated exon also imply a role for the pre-mRNA bound U1 snRNA [242]. Repression of exon 4 of *CD45* pre-mRNA by the combined actions of hnRNPs L and A1 induces extended basepairing of 5′-ss bound U1 snRNA

[243], thereby stabilizing the U1/pre-mRNA association, which prevents the displacement of U1 by U6 and subsequent spliceosome assembly. It is likely that the non-canonical basepairing of U1 with exon 4 of the *CD45* pre-mRNA, and the interaction between pre-mRNA bound U1 snRNA and PTB in the *Src* pre-mRNA sterically prevent the optimal orientation of the U1 snRNA for an interaction with UAP56 and/or SF3A1. Recently, SL3 of the U1 snRNA was also identified as a contact of the fused in sarcoma (FUS) protein [165]. Aberrant cytoplasmic interactions between FUS and the U1 snRNA were shown to disrupt U1 biogenesis and was suggested as an underlying mechanism in FUS-induced amyotophic lateral sclerosis. Our work shows that the SL3 and SL4 of the human U1 snRNA make unique contacts with splicing factors associated with the early steps of spliceosome assembly and are crucial for formation of a functional spliceosome. Their inhibition by competing splicing regulators like FUS or PTB may be potential mechanisms of alternative splicing regulation.

Experimental

Plasmid constructs and transfections

The three-exon/two-intron reporter pDUP51p and the U1 snRNA expression plasmid pNS6U1 have been described previously [168]. The constructs expressing U1-5a snRNAs carrying SL3 mutations were generated by PCR mutagenesis using oligonucleotides and were verified by DNA sequencing. The sequences of all oligonucleotides used for U1-5a mutagenesis performed in this study are provided in Appendix D.

HeLa cells, originally purchased from ATCC, were a gift from Kurt Gustin (University of Arizona, College of Medicine-Phoenix). They were cultured in DMEM containing 10% FBS and antibiotics (100 U/mL penicillin, 100 μg/mL streptomycin). Culture supernatants were tested for mycoplasma by PCR using a pool of six mycoplasma specific primers, and found to be negative [219]. For transfection, 0.5×10^5 cells per well of a six-well plate were transfected with 0.4 μg of Dup51p reporter plasmid and 3.6 μg of control plasmid (pcDNA3.1) or U1 expression plasmid (pNS6U1) using Lipofectamine 2000 according to the manufacturer's instructions (Thermo Fisher Scientific). Cells were harvested 48 hours post-transfection and total RNA was extracted using TRIzol reagent or prepared using the Direct-zol RNA kit (Zymo Research).

Antibodies and Western blotting

For protein analysis, samples were boiled in SDS-PAGE sample buffer, separated on 10% SDS-PAGE gels and analyzed by Western blotting using PVDF membrane. Antibodies against SF3A1, SF3A3, and U1-70K have been described previously [168]. The anti-UAP56 rabbit polyclonal antibody was a gift from Robin Reed (Harvard Medical School). This antibody was raised against GST-UAP56 and cross-reacts with URH49 [244]. Other commercial antibodies used in this study included anti-α-Tubulin mouse monoclonal antibody (EMD Millipore; CP06-DM1A), and anti-Prp19 rabbit polyclonal antibody (Bethyl Laboratories; A300-102A). Secondary anti-mouse and anti-rabbit antibodies conjugated to Cy3 and Cy5 fluorophores were purchased from Thermo Fisher Scientific. Proteins were quantified by densitometric scanning of Western blots using ImageQuant.

Primer extension and RT-qPCR

Primer extensions to monitor splicing of the Dup51p reporter and for determining U1-5a snRNA expression were performed using ^{32}P-Dup3r and U1$_{7-26}$R oligonucleotides, respectively, as described previously [168]. Oligonucleotide sequences are provided in the Appendix E. Spliced products were quantified by densitometric scanning of urea-PAGE images using ImageQuant. For RT-qPCRs, reverse transcription was performed using SuperScript III Reverse Transcriptase kit and 1 ng of resulting cDNA was used as a template for qPCR amplification using SYBR Green reagent and StepOnePlus Real-Time PCR Machine; all according to manufacturer specifications (Thermo Fisher Scientific). Primer pairs used for amplification of U1 and U2 snRNA are reported in the Appendix E. U1 expression across all samples was normalized to U2 snRNA and fold-increase in expression was calculated relative to the pcDNA control.

UV crosslinking

Nuclear extract from HeLa S3 cells was prepared as described previously [245, 246]. ^{32}P-labeled U1-SL3 RNAs were *in vitro* transcribed from annealed DNA templates, gel purified, and ethanol precipitated. In Figure 3.5, ^{32}P-U1-SL3 RNAs were incubated at indicated final concentrations in a splicing reaction containing 2.2 mM MgCl$_2$, 0.4 mM ATP, 20 mM creatine phosphate, 10U RNaseOUT, and 60% nuclear extract in buffer DG (20 mM HEPES pH 7.9, 80 mM K-glutamate, 0.1 mM EDTA, 1 mM DTT, 20% glycerol). UV crosslinking was performed in a GS Gene Linker (Bio-Rad Laboratories)

for a total energy of 1800 mJ and separated on 10% SDS-PAGE gels and visualized using the Typhoon FLA 9500 imager.

RNA affinity purification and MS analysis

For RAP, biotinylated wildtype and mutant U1-SL3 RNAs were custom-synthesized (Integrated DNA Technologies). HeLa nuclear extracts were preincubated in the absence or presence of NaCl at 4°C for 20-30 min. in splicing conditions described above. The reaction mix was then added to 20 μl of Neutravidin beads that were pre-bound with 2 nmoles of biotinylated wildtype or mutant SL3 RNA and incubation was continued for 30 min. at RT. Beads were washed four times with 200 μL of buffer DG. Total RNA was extracted from the bound complexes using phenol:chloroform (5:1; pH 4.8), precipitated with ethanol, separated on 8% urea-PAGE gels, and visualized by ethidium bromide staining. For protein analysis, the bound complexes were eluted by treatment with RNase A/T1 cocktail (Life Technology) in 10 mM Tris-HCl (pH 7.2), 1 mM $MgCl_2$, and 40 mM NaCl. In Figure 3.8, nuclear extracts were preincubated with 0, 250, 500, and 1000 mM NaCl prior to addition of biotinylated RNAs. The eluted proteins were analyzed either by MS or separated on 10% SDS-PAGE gels and analyzed by Western blotting.

For MS, protein samples were reduced, alkylated, and digested using Lys-C and trypsin proteases as previously described [247]. Peptide mixtures were fractionated online using reversed phase chromatography and then analyzed by tandem mass spectrometry on a Q-Exactive mass spectrometer (Thermo Fisher Scientific) [248]. Data analysis was performed using the IP2 platform (Integrated Proteomics Applications)

using the ProLuCID and DTASelect algorithms and filtered at 5% false discovery rate for peptide spectrum matches as calculated using a decoy database approach [249-251]. NSAF values were calculated from the total number of spectrum-matching peptides from the protein (spectrum counts) that were then normalized for protein length [233].

For affinity purification of the U1 and U2 snRNPs, 100 μl splicing reactions containing HeLa S3 nuclear extract and 10 μM U1 or U2 hybridizing, 3′ biotinylated, 2′-*O*-methyl anti-sense oligonucleotides, U1$_{1-13}$ and U2$_{1-21}$ (Integrated DNA Technologies; Appendix E) were prepared as described above. These reactions, either lacking ATP or containing 0.5 mM ATP or ATP-γ-S, were incubated at 30°C for 30 min. and then added to 20 μl pre-blocked NeutrAvidin beads and kept at 4°C for 1 hour with end-over-end rotation [252]. Beads were washed four times with 200 μl of buffer DG and protein was eluted by boiling beads in 1X SDS-PAGE sample buffer. For analysis of RNA in ASO AP complexes, RNA was eluted and purified by standard TRIzol extraction.

GST-UAP56 purification and electrophoretic mobility shift assays

GST-UAP56 construct in the plasmid pGEX-5x was a gift from Robin Reed (Harvard Medical School) and the GST-SF3A1-UBL fusion construct was created by cloning cDNA of SF3A1-UBL domain (SF3A1 aa 704-793) into BamHI and XhoI restriction sites of pGEX-5x. GST, GST-UAP56, and GST-UBL proteins were expressed in *Escherichia coli* (BL21-DE3) by overnight induction with 0.2 mM IPTG at 18°C. Induced proteins were purified from bacterial lysates using glutathione agarose beads (Thermo Fisher Scientific) according to manufacturer's protocol, dialyzed against two liters of buffer DG, and stored at -20°C. EMSAs were performed as described previously

[232]. Briefly, binding reactions contained 10-50 nM 5′-Cy5-labeled U1-SL3 or U1-SL4 RNAs (Integrated DNA Technologies; see Appendix E for sequences), 2.2 mM $MgCl_2$, 60% buffer DG, and varying concentrations of purified GST, GST-UBL, and GST-UAP56 protein. ATP-γ-S and other NTPs were added at a final concentration of 0.5 mM. After incubation for 30 min. at RT, binding reactions were loaded onto horizontal 6% native-PAGE gels run at 100V for 45 minutes at 4°C [186, 213]. Native gels were visualized using the Typhoon FLA 9500 Imager.

CHAPTER 4

ELUCIDATING THE FUNCTIONS OF U1-SL3 AND U1-SL4 DURING
PRESPLICEOSOME ASSEMBLY

Publication Note

The research reported in this chapter has been prepared in a manuscript that has

been submitted for peer review at the journal *RNA Biology*. William Martelly, Bernice

Fellows, Paul Kang, Ajay Vashisht, James A. Wohlschlegel and Shalini Sharma. All co-

authors have granted permission for this work to be included in this dissertation.

Overview

Previously, we reported that the U2 snRNP specific protein SF3A1 interacts with

the protein-free SL4 structure of the U1 snRNA, via a noncanonical Ubiquitin-like

domain that generates a cross-intron interaction that is crucial for splicing [168, 232].

Investigation into the role of the other protein-free structure of the U1 snRNA identified

the RNA helicase UAP56 as an SL3 interacting protein (Chapter 3). UAP56 is known to

be required for the E→A complex transition in a manner that requires ATP hydrolysis,

although its precise role in pre-spliceosome formation in humans is not well characterized

[253]. The A complex is also the stage of spliceosome assembly at which SL4 interacts

with SF3A1 to form a bridge across the intron. Therefore, both SL3 and SL4 of the U1

snRNA may have overlapping functions involved in generating the pre-spliceosome A

complex.

Interestingly, mutations in the SL4 region of the U1 snRNA never casued

baseline levels of U1-dependent splicing activity [168]. Similarly, experiments applying

the U1-dependant minigene reporter Dup51p determined that mutations in SL3 reduce

the capacity for U1 to rescue full-length splicing, but do not entirely inhibit U1 activity

(Figure 3.3). Therefore, we hypothesized that these structures can individually support

splicing via interactions with their respective binding partners (SL3 with UAP56 and SL4

with SF3A1). To investigate this hypothesis, double mutations in both SL3 and SL4 were

introduced into U1-5a snRNA variants and their capacity to rescue the minigene reporter

was assessed. Additionally, to validate the functional significance of the SL3- and SL4-

specific binding proteins, an siRNA-based approach was used to silence UAP56 and

SF3A1 in combination with the minigene reporter assay. In previous analyses, excess U1-

SL4 RNA in splicing reactions was found to potently interfere with A complex assembly

by competing out the cross-intron bridging interaction between the 5′- and 3′-ss bound

U1 and U2 snRNPs [168]. To assess the impact of the SL3-UAP56 interaction, similar *in

vitro* splicing assays in HeLa cell nuclear extracts were performed in the presence of

excess U1-SL3 RNA and its impacts on splicing efficiency and pre-spliceosome

formation were determined.

Results

Stem-loops 3 and 4 have synergistic roles in U1 function

In the U1 complementation assays, sequence changes to SL3 caused a decrease in

U1 activity to ~60%, however, they did not reduce it to the baseline observed for the

pcDNA control (Figure 3.3 compare lanes 7-9 with 1), thereby indicating that U1 activity

was not completely abolished. This observation was similar to our previous analysis of

mutations in SL4 of the U1 snRNA, where SL4 mutations were found to compromise, but

not completely abrogate U1 activity [168]. So, we next examined if combining mutations

of SL3 and SL4 within the same U1-5a snRNA has larger effects than those observed for

a single mutation (Figure 4.1).
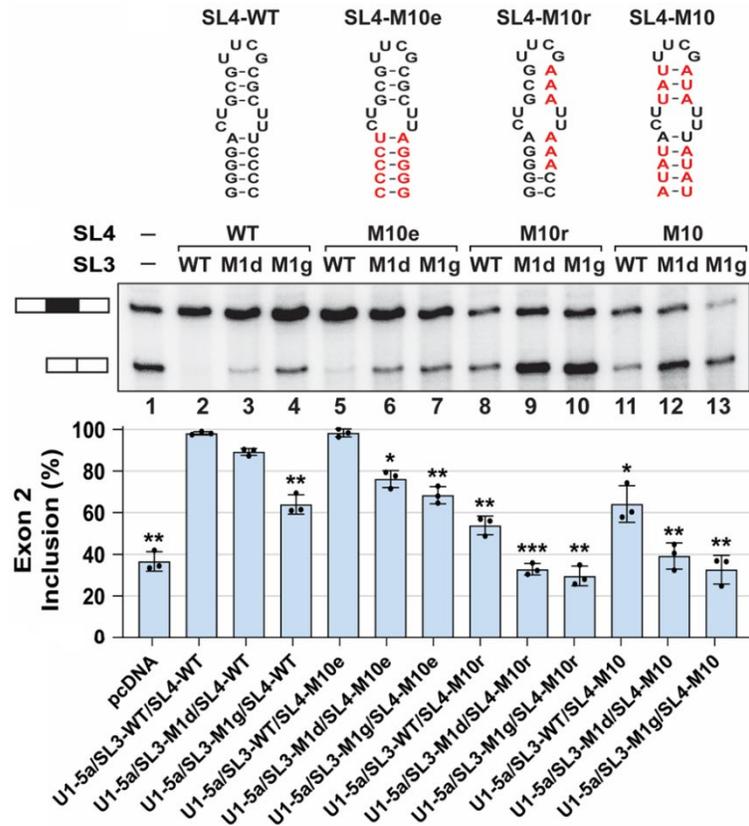
Figure 4.1 – Combined SL3 and SL4 mutations have synergistic effects on U1 function. Above the gel image a schematic of the SL4 secondary structures from wildtype and mutant U1 snRNAs used to create U1-SL3/SL4 double mutants is displayed. (B) Primer extension analysis to monitor splicing of the minigene reporter Dup51p after co-transfections with control or U1-5a plasmids for expression of wildtype and mutant U1 snRNA. The full-length and exon 2 skipped Dup51p mRNA products are depicted. The percentage of the full-length product ($\pm$ s.d.) is graphed below and statistical significance was determined by comparisons to the wildtype control (lane 2) (n = 3; * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$). The analysis for synergistic effects is shown in Table 4.1.

As reported earlier, single SL4 mutants U1-5a/SL4-M10r and U1-5a/SL4-M10

reduced exon 2 inclusion, whereas U1-5a/SL4-M10e did not have a significant effect

(Figure 4.1 lanes 5, 8, and 11) [168]. Relative to U1-5a carrying SL4-wildtype (WT) (lane 2), SL4-M10r and SL4-M10 reduced exon 2 inclusion by ~45% (lane 8; from ~97% to ~52%) and ~33% (lane 11; from ~97% to ~64%), respectively. A double mutant that carries both, SL3-M1d and SL4-M10r, mutations reduced exon 2 inclusion by ~64% (lane 9; from ~97% to ~33%). Thus, the reduction of exon 2 inclusion by the double mutant (~64%) is much larger than the effects seen with the single mutants SL3-M1d (~8%) and SL4-M10r (~45%). Similarly, the double mutant U1-5a/SL3-M1d/SL4-M10 (lane 12) has a much greater influence on exon 2 inclusion (reduced by ~58%) than the effects seen with the single mutants SL3-M1d (~8%) and SL4-M10 (~33%), thereby suggesting that the effects of double mutations may be synergistic.

To examine if the synergistic effects of combining SL3 and SL4 mutations on U1 activity were statistically significant, we applied the linear mixed model. For this, U1 activity (A) was defined as the fraction of exon 2 inclusion, and the predicted activity for a particular combination of single SL3 and SL4 mutations ($A_{pred} = A_{SL3}*A_{SL4}$) was compared with the observed activity (A_{obs}) of the U1 snRNAs carrying double mutations (Table 4.1). The effects of double mutations were considered synergistic if $A_{obs} < A_{pred}$ and the difference was ≥0.1 with $p \leq 0.01$. In these assays, the magnitude of effect is limited by the exon 2 inclusion baseline, which is ~35% (lane 1). As a result, the synergistic effects on U1 activity are more apparent for the double mutants carrying the milder SL3-M1d mutation - U1-5a/SL3-M1d/SL4-M10r, U1-5a/SL3-M1d/SL4-M10, and U1-5a/SL3-M1d/SL4-M10e. Notably, the differences between A_{obs} and A_{pred} for double mutants harboring the more severe mutation, SL3-M1g, suggested that the effects may be synergistic, but did not meet the stringent statistical criteria (Table 4.1). Although the U1-

5a variant carrying the single SL4-M10e mutation did not affect exon 2 inclusion, the effect of double mutant U1-5a/SL3-M1d/SL4-M10e was significantly synergistic. Overall, the results demonstrated that combined mutations of SL3 and SL4 exert synergistic effects, indicating that the roles of the two stem-loops are likely interconnected.

Combination (SL3 and SL4)	Exp #	A_{pred}	A_{obs}	Coefficient*	95% Conf. Interval	P-value	Synergy
SL3-M1d/SL4-M10e	1	0.867	0.793	0.113	0.074, 0.152	<0.001	Yes
	2	0.869	0.716				
	3	0.888	0.776				
SL3-M1g/SL4-M10e	1	0.589	0.727	-0.064	-0.124, -0.003	0.04	No
	2	0.591	0.646				
	3	0.682	0.679				
SL3-M1d/SL4-M10r	1	0.439	0.322	0.196	0.146, 0.247	<0.001	Yes
	2	0.507	0.244				
	3	0.514	0.304				
SL3-M1g/SL4-M10r	1	0.298	0.281	0.049	-0.008, 0.108	0.090	No
	2	0.344	0.349				
	3	0.395	0.258				
SL3-M1d/SL4-M10	1	0.668	0.448	0.186	0.111, 0.262	<0.001	Yes
	2	0.523	0.324				
	3	0.545	0.405				
SL3-M1g/SL4-M10	1	0.454	0.368	0.082	0.005, 0.160	0.037	No
	2	0.356	0.247				
	3	0.419	0.365				

Table 4.1 – Synergy analysis of the impacts of stem-loops 3 and 4 double mutations on activity of U1 snRNAs. Combinations of SL3-M1d and all SL4 mutations tested produce synergistic reductions in U1 activity when combined. The combinations with SL3-M1g overall produced positive coefficients but did not meet our stringent criteria. *Average values calculated from three independent experiments were considered synergistic if $A_{obs} < A_{pred}$ with a difference of ≥ 0.1 and $p < 0.01$. A positive coefficient value indicates synergy.

RT-qPCR quantification showed that U1-5a variant expression was ~2 to 4-fold higher than the endogenous U1 snRNA (Figure 4.2). Primer extension with $U1_{7-26}R$ oligonucleotide confirmed the presence of all mutant U1-5a snRNAs in these samples (Figure 4.3) and thus, the loss of exon 2 inclusion is not due to inefficient expression but due to loss of function of the U1 snRNP induced by single or double mutations in SL3 and SL4.
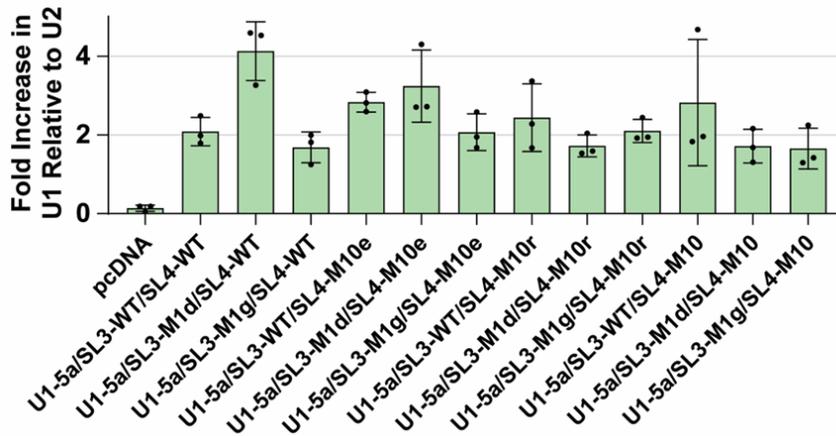
Figure 4.2 – Total U1 snRNA is overexpressed in HeLa cells transfected with U1-5a snRNA expression plasmids harboring single and double SL3/SL4 mutations. RT-qPCR analysis of U1 snRNA expression in HeLa cells co-transfected with Dup51p reporter and U1-5a variant plasmids. Fold change in U1 snRNA expression was calculated relative to the pcDNA control after normalization to U2; fold increase in U1 is graphed (± s.d., n = 3).
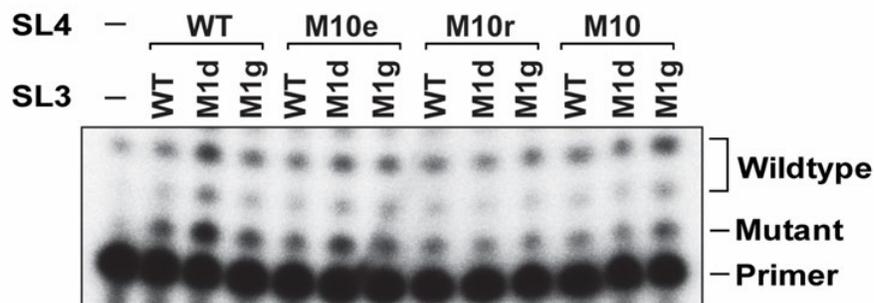
Figure 4.3 – U1-5a snRNA harboring single and double mutations in SL3/SL4 are also detectable by primer extension in transfected HeLa cells. Primer extension analysis with oligonucleotide ^{32}P-$U1_{7-26}R$ (Appendix E), showing expression of the endogenous U1 and variant U1-5a snRNAs.

Mutant U1 snRNAs exhibit nuclear localization and normal processing

The current model for biogenesis of the U1 snRNP posits existence of a 213 nt precursor snRNA that transiently traffics from the nucleus to the cytoplasm where 3′-end processing forms the 164 nt snRNA, and assembly of the Sm core takes place [254, 255]. Subsequent maturation occurs after import of the snRNP intermediate back into the nucleus and involves loading of the U1 specific proteins. To confirm processing of the U1 snRNA variants to mature length, we carried out Northern blotting using an oligonucleotide complementary to nts 27-46 of the U1 snRNA (U1$_{27\text{-}46}$R) to detect both endogenous U1 and U1-5a snRNAs. Results demonstrated that in cells expressing single and double mutants, the U1-5a snRNAs were processed to a length identical to the endogenous U1 snRNA (Figure 4.4).
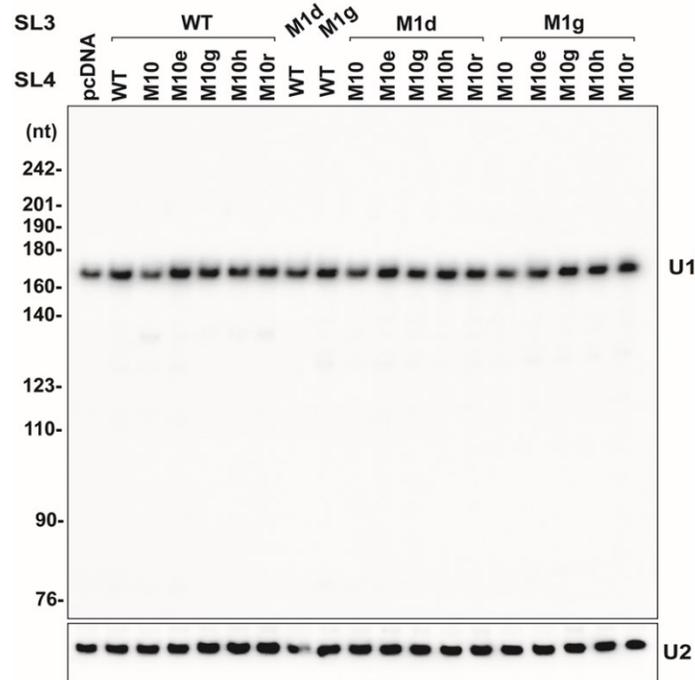
Figure 4.4 – Analysis of variant U1 snRNA expression by Northern blotting. To confirm processing of the U1 snRNA to mature length, total RNA from HeLa cells expressing U1-5a variants carrying single and double SL3/4 mutations was probed with ^{32}P-U1$_{26\text{-}46}$R oligonucleotide to detect both endogenous and transiently expressed U1 snRNAs. U2 snRNA was detected using ^{32}P-U2$_{114\text{-}135}$R and serves as a loading control.

To determine subcellular localization, we performed nuclear-cytoplasmic

fractionation of HeLa cells expressing U1-5a variants. Efficiency of the applied protocol

was assessed by RNA and protein analysis of the nuclear and cytoplasmic fractions

obtained from untransfected cells. Northern blotting confirmed enrichment of the U1 and

U2 snRNAs in the nuclear fraction and demonstrated relatively equal distribution of the

5S rRNA in both compartments (Figure 4.5). Western analysis demonstrated the presence

of α-Tubulin only in the cytoplasmic fraction, and predominantly nuclear localization of

the U1 protein U1-70k, and the U2 proteins SF3A1 and SF3B1 (Figure 4.6).
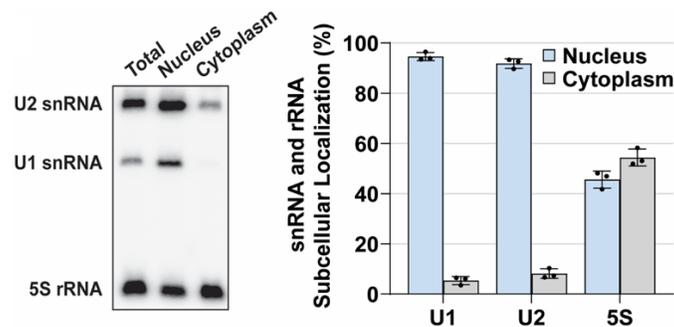
Figure 4.5 – Detergent based hypotonic lysis buffer efficiently fractionates HeLa cells for isolation of nuclear and cytosolic RNA. Northern blot of RNA isolated from nuclear and cytoplasmic fractions was probed with U1, U2, and 5S rRNA specific oligonucleotides (Appendix E). The average RNA levels ($\pm$ s.d., n = 3) in subcellular fractions calculated as a percentage of total RNA is graphed.
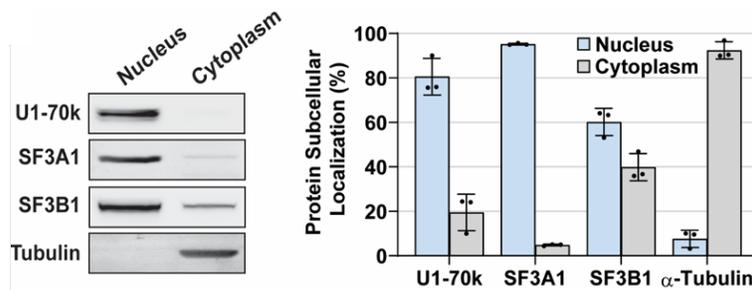
Figure 4.6 – Detergent based hypotonic lysis buffer efficiently fractionates HeLa cells for isolation of nuclear and cytosolic protein. Western analysis of nuclear and cytoplasmic fractions using antibodies against U1-70K, SF3A1, SF3B1, and α-Tubulin. The average protein level ($\pm$ s.d., n = 3) in subcellular fractions expressed as a percentage of total protein is graphed.

To specifically detect the SL4-M10r mutation, we designed a locked nucleic acid

(LNA) containing oligonucleotide (U1-M10r-LNA) (Appendix E). Northern blotting of

fractions obtained from transfected cells revealed nuclear localization of the U1 and U2

snRNAs in cells expressing U1-5a variants carrying wildtype stem-loops (Figure 4.7

lanes 4-6), single SL4-M10r mutation (lanes 7-9), and double mutations SL3-M1d/SL4-

M10r (lanes 10-12) and SL3-M1g/SL4-M10r (lanes 13-15). Probing with U1-M10r-LNA

indicated predominantly nuclear localization of the mutant U1 snRNAs (lanes 7-15). In

fractions from cells expressing U1 variants, Western analysis confirmed nuclear and

cytoplasmic localization of SF3A1 and α-Tubulin, respectively (Figure 4.8).
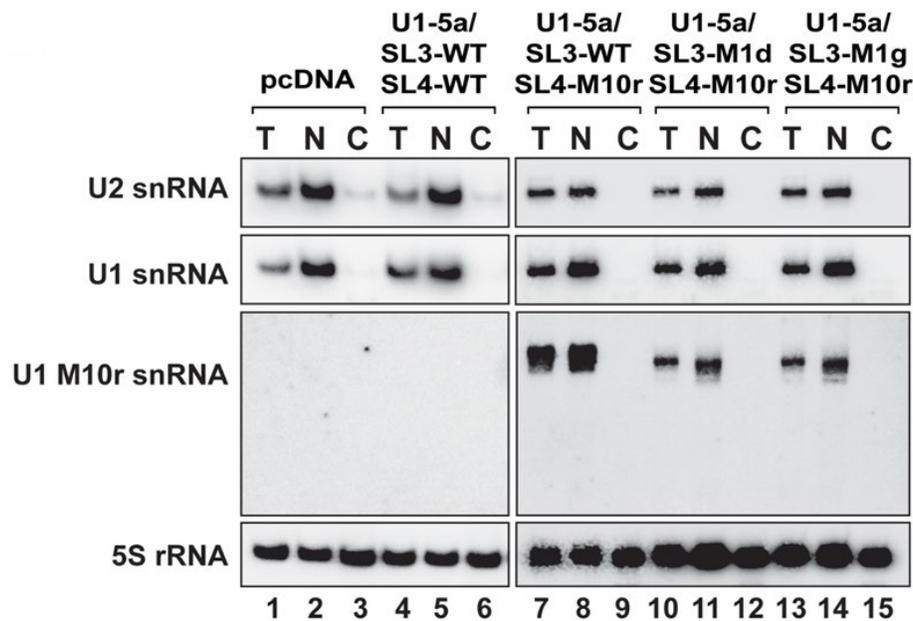
Figure 4.7 – Detection of SL4-M10r harboring mutant U1 snRNA in HeLa cell nuclear fractions confirm nuclear localization of U1 snRNAs harboring single and double SL3/SL4 mutations. Northern blotting of RNA from total cell lysate (T), and nuclear (N) and cytoplasmic (C) fractionations from HeLa cells expressing U1-5a variants carrying wildtype SL3/SL4, single, or double mutations. Northern blots were probed with ^{32}P-labeled oligonucleotides specific to U1 and U2 snRNAs, 5S rRNA, and the SL4-M10r mutation.
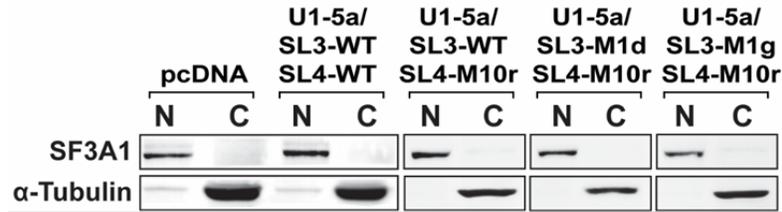
Figure 4.8 – Presence of SF3A1 and α-Tubulin proteins in nuclear and cytoplasmic fractions confirms efficiency of fractionation in U1 SL3/SL4 mutant expressing HeLa cells. Immunoblot of nuclear and cytoplasmic fractions using anti-SF3A1 and anti-α-Tubulin antibodies.

To assess maturation of the U1-5a variant particles, we used an anti-U1-70k antibody that efficiently immunoprecipitates (IPs) the U1 snRNP (Figure 4.9). Northern blotting of IP complexes from cells expressing U1-5a/SL4-M10r, U1-5a/SL3-M1d/SL4-M10r, and U1-5a/SL3-M1g/SL4-M10r demonstrated the presence of U1 but not U2 snRNA (Figure 4.10 compare lanes 3, 6, and 9 with 12, 15, and 18). Probing with the U1-M10r confirmed the presence of U1-5a variants carrying single and double mutations in the IP complexes (lanes 21, 24, and 27).
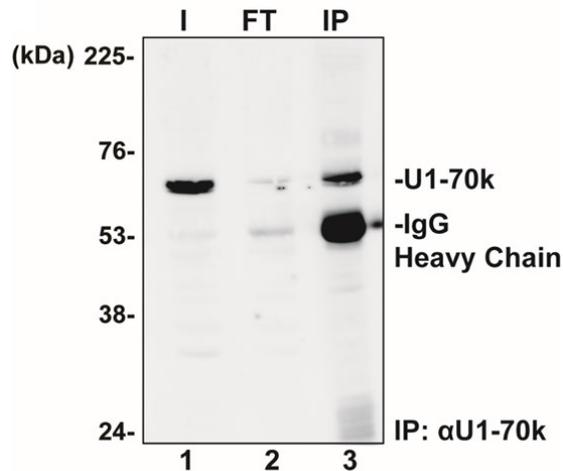
Figure 4.9 – U1 snRNP can be efficiently immunoprecipitated using U1-70k specific antibody. Anti-U1-70k antibody efficiently immunoprecipitates U1 snRNP from HeLa nuclear extracts. Protein from input (I), flow through (FT), and immunoprecipitated (IP) complexes were collected during IP of HeLa nuclear extracts with anti-U1-70k antibody and analyzed by Western blotting using the same U1-70k antibody. The positions of U1-70k and the IgG heavy chain are indicated.
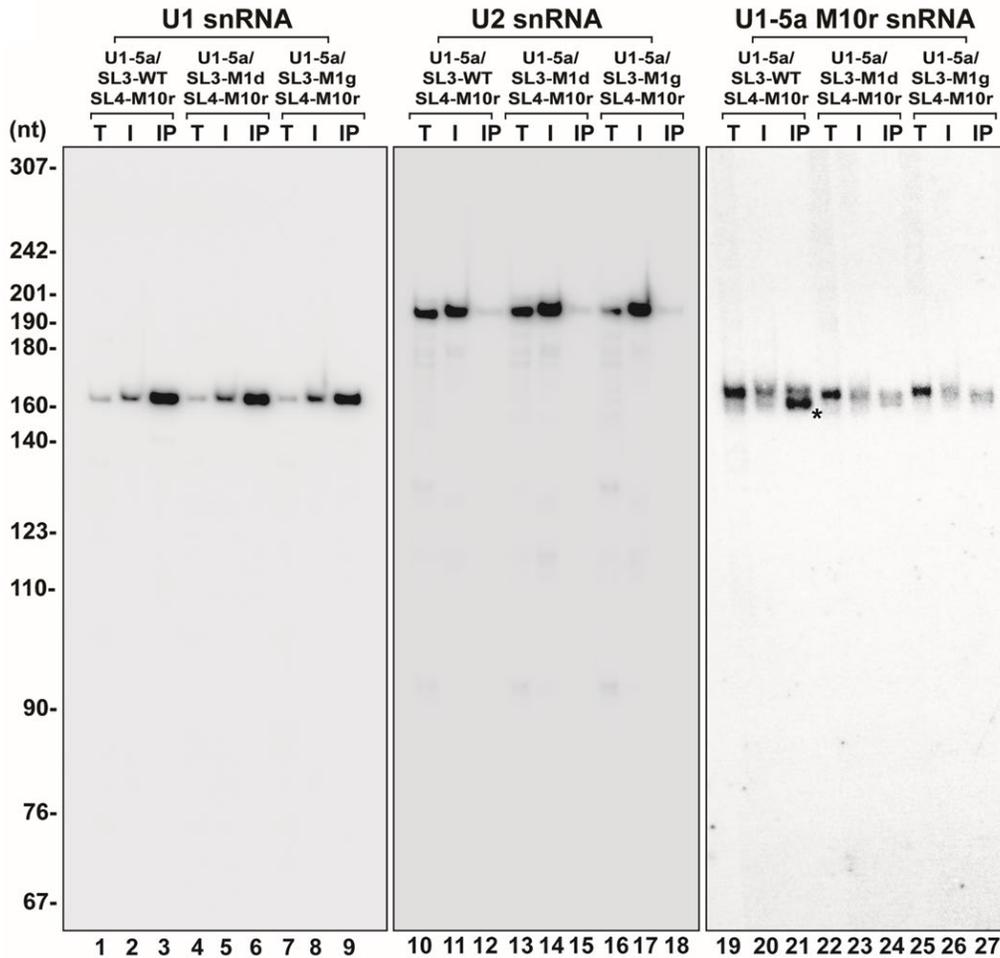
Figure 4.10 – U1 snRNAs carrying single and double mutations are present in U1-70k immunoprecipitated complexes. Northern blotting analysis of RNA extracted from total cell lysate (T), nuclear fractions used as input (I), and anti-U1-70k antibody IP complexes (IP). Lysates and nuclear fractions were derived from HeLa cells expressing single SL4-M10r or double SL3/4 mutations. A slight change in the mobility of the U1-5a M10r mutant snRNA was observed in the immunoprecipitated RNA fraction (Lane 21; *). This reduced mobility band was not seen in total or input RNA from the same lysate, and may, therefore, be occurring as a result of processing and handling of lysate during immunoprecipitation.

Overall, the results demonstrated that the U1-5a snRNA variants were expressed

efficiently and did not affect integrity of the endogenous U1 snRNA. Although others

have reported aberrant cleavage of U1 snRNAs carrying the SL4-M10 mutation [256],

we found that variants carrying single and double mutations were the same length as the

endogenous U1 snRNA. Therefore, the variant snRNAs were processed to mature length,

localized within the nucleus, and associated with U1-70K, indicating that they were likely exported to cytoplasm and then imported back into the nucleus for maturation.

U1 snRNA acts through UAP56 and SF3A1

To establish that SL3 of the U1 snRNA acts through UAP56, we tested the impact of siRNA mediated UAP56 knockdown on the activity of the U1-5a variants carrying either SL3 or SL4 mutation. Our rationale was that if SL3 was acting through UAP56, UAP56 knockdown would phenocopy SL3 mutations in the U1 complementation assay. Mammalian cells also express a UAP56 paralog called UAP56-related helicase, 49 kDa (URH49; also known as DDX39A). The two proteins are 90% identical, and have redundant functions in pre-mRNA splicing and nuclear export of mature mRNA [257-259]. Treatment of HeLa cells with siUAP56 or siURH49 caused a ~60% decrease in levels of these proteins, but not of SF3A1 or SF3B1 (Figure 4.11). Simultaneous knockdown of UAP56 and URH49 drastically reduced cell viability and the yield of total RNA from siRNA treated cells. Therefore, in determining whether SL3 action involves UAP56 and URH49, we resorted to performing the U1 complementation assay after individual knockdowns.
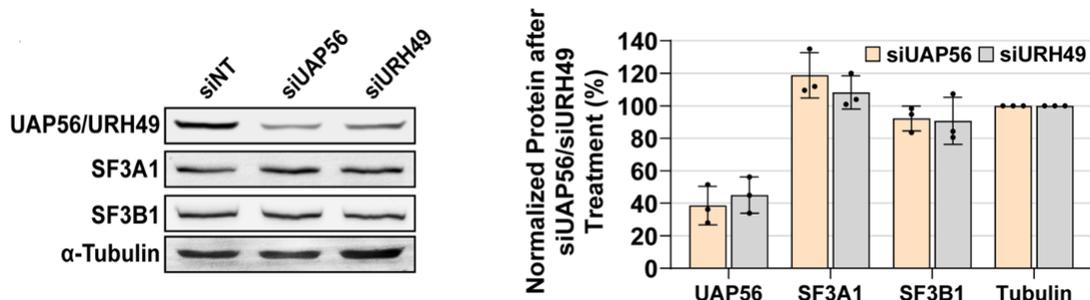
Figure 4.11 – UAP56 and URH49 targeting siRNAs efficiently knockdown protein expression. Western analysis of HeLa cell lysates after treatment with non-targeting control (siNT), UAP56 targeting (siUAP56), and URH49 targeting (siURH49) siRNAs. Average protein expression (± s.d., n = 3) normalized to α-Tubulin is shown.

Treatment with non-targeting siRNA (siNT) did not have any effect on rescue of exon 2 inclusion by the U1-5a snRNA carrying wildtype SL3 and SL4 (Figure 4.12 Left and Right Panels lane 1). As before, in comparison to U1-5a/WT, mutants U1-5a/SL3-M1d and U1-5a/SL4-M10 reduced exon 2 inclusion (Figure 4.12 Left and Right Panels lanes 2 and 3). In co-transfections with U1-5a/SL3-M1d, knockdown of UAP56 or URH49 did not significantly exacerbate the effects of the SL3-M1d mutation on exon 2 inclusion (Figure 4.12 Left and Right Panels compare lane 5 to lane 2). However, in the case of co-transfections with U1-5a/SL4-M10, knockdown of UAP56 or URH49 elicited a much greater decrease in exon 2 inclusion (Figure 4.12 Left and Right Panels compare lane 6 to lane 3).
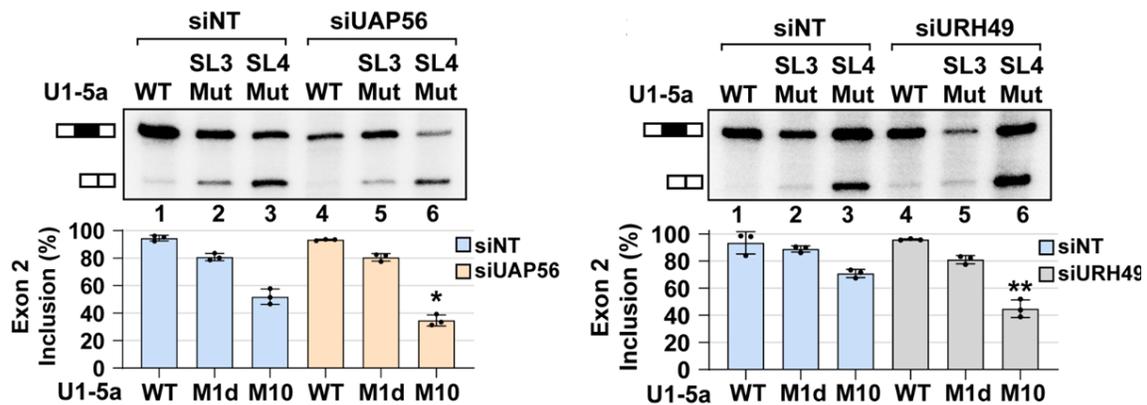
Figure 4.12 – UAP56 and URH49 knockdowns phenocopy SL3 mutation. Left panel – Primer extension analysis of Dup51p reporter transcripts after complementation with U1-5a variants and treatment with control siNT or siUAP56. Right panel – Primer extension analysis of Dup51p reporter transcripts after complementation with U1-5a variants and treatment with control siNT or siURH49. The average percentage of the full-length product (± s.d.) is graphed below (n = 3; * = $p < 0.05$, ** = $p < 0.01$). Statistical comparisons were performed for each U1-5a snRNA tested under the siNT versus siRNA treatment conditions. Analysis for synergistic effects is shown in Table 4.2.

To examine if the effects of the combination of UAP56 or URH49 knockdown and SL3 or SL4 mutations were synergistic, we compared the predicted U1 activity for a combination of siUAP56/URH49 treatment and a particular stem-loop mutation (A_{pred} =

$A_{siUAP56/URH49}$*$A_{SL3/SL4}$) with the observed U1 activity (A_{obs}) when stem-loop mutations were expressed after siUAP56/URH49 treatment. This comparison revealed that the effects were synergistic ($A_{obs} < A_{pred}$; a difference in exon 2 inclusion of ≥ 0.1 with $p \leq 0.01$) for a combination of UAP56 or URH49 knockdown with U1-5a/SL4-M10, but not with U1-5a/SL3-M1d (Table 4.2). The effects of UAP56 or URH49 knockdown on U1 activity were synergistic only when combined with the U1-SL4 mutation. Therefore, UAP56 or URH49 knockdown and U1-SL3 mutation phenocopy one another, indicating that SL3 of the U1 snRNA likely acts through UAP56 or URH49.

Combination (SL3 or SL4 mutation and siRNA treatment)	Exp #	A_{pred}	A_{obs}	Coefficient*	95% Conf. Interval	P-value	Synergy
WT/siSF3A1	1	0.529	0.557	-0.027	-0.078, 0.024	0.297	No
	2	0.531	0.567				
	3	0.681	0.698				
SL3-M1d/siSF3A1	1	0.447	0.380	0.120	0.051, 0.189	0.001	Yes
	2	0.450	0.380				
	3	0.629	0.407				
SL4-M10r/siSF3A1	1	0.308	0.323	0.036	-0.054, 0.127	0.432	No
	2	0.321	0.385				
	3	0.417	0.229				
WT/siUAP56	1	0.862	0.935	-0.051	-0.064, -0.039	<0.001	No
	2	0.880	0.926				
	3	0.900	0.936				
SL3-M1d/siUAP56	1	0.781	0.823	-0.053	-0.066, -0.039	<0.001	No
	2	0.742	0.819				
	3	0.735	0.775				
SL4-M10/siUAP56	1	0.540	0.392	0.137	0.087, 0.187	<0.001	Yes
	2	0.431	0.314				
	3	0.481	0.335				
WT/siURH49	1	0.802	0.956	-0.063	-0.118, -0.007	0.028	No
	2	0.939	0.954				
	3	0.944	0.964				
SL3-M1d/siURH49	1	0.858	0.832	0.041	0.005, 0.078	0.027	No
	2	0.822	0.775				
	3	0.871	0.821				
SL4-M10/siURH49	1	0.683	0.389	0.229	0.173, 0.286	<0.001	Yes
	2	0.701	0.438				
	3	0.650	0.518				

Table 4.2 – Synergy analysis of the impacts of SF3A1 or UAP56/URH49 knockdown on activity of U1 snRNAs carrying WT or single stem-loop 3 or 4 mutations. *Average values calculated from three independent experiments were considered synergistic if $A_{obs} < A_{pred}$ with a difference of ≥ 0.1 and p < 0.01. A positive coefficient value indicates synergy.

The SL4 of the U1 snRNA was previously reported to interact with the U2 protein SF3A1 and also with the polypyrimidine tract binding protein 1 (PTBP1) [167, 168]. So, we also examined the impact of knockdown of these two proteins on the activity of U1 variants. siSF3A1 treatment of HeLa cells caused a ~75% decrease in levels of SF3A1, without affecting UAP56 or SF3B1 (Figure 4.13 Left Panel). SF3A1 knockdown caused a decrease in exon 2 inclusion in co-transfections with U1-5a carrying wildtype and mutant stem-loops (Figure 4.13 Right Panel compare lanes 1-3 to 4-6), thereby indicating that loss of SF3A1 caused a general decrease in splicing, which was not observed upon UAP56 or URH49 knockdown. This, in our opinion, is due to the comparatively lower knockdown efficiencies of UAP56 and URH49, and the residual protein levels may be sufficient to support the function of U1-5a snRNAs carrying wildtype stem-loops. Importantly, upon SF3A1 knockdown, effect on exon 2 inclusion in co-transfections with U1-5a/SL3-M1d or U1-5a/SL4-M10r was larger than that for U1-5a/WT (Figure 4.13 Right Panel, compare lanes 5 and 6 to 4).
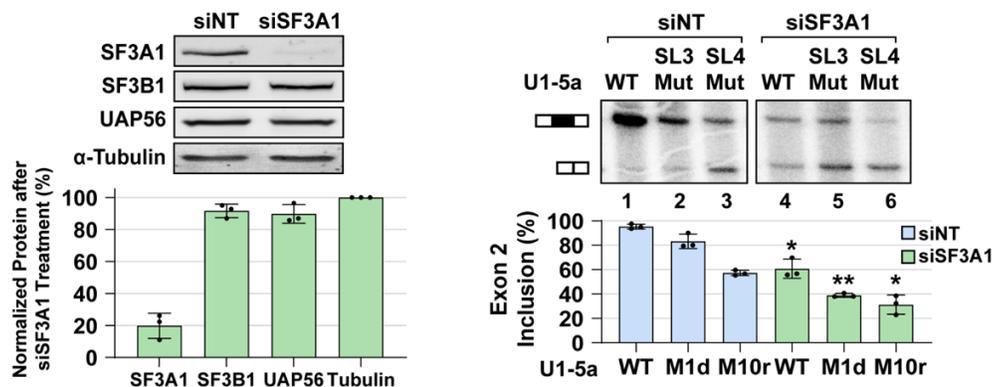
Figure 4.13 – SF3A1 knockdown phenocopy SL4 mutations. Left panel – Western analysis of HeLa cell lysates after treatment with control siNT or SF3A1 targeting (siSF3A1) siRNAs. Average protein expression (± s.d., n = 3) normalized to α-Tubulin is shown. Right panel – Primer extension analysis of Dup51p reporter transcripts after complementation with U1-5a variants and treatment with control siNT or siSF3A1. The average percentage of the full-length product (± s.d.) is graphed below (n = 3; * = p < 0.05, ** = p < 0.01). Statistical comparisons were performed for each U1-5a snRNA tested under the siNT versus siRNA treatment conditions. Analysis for synergistic effects is shown in Table 4.2.

Comparison of the predicted U1 activity ($A_{pred} = A_{siSF3A1} * A_{SL3/SL4}$) with the observed activity (A_{obs}) revealed that the effects on U1 activity were synergistic ($A_{obs} < A_{pred}$; a difference in exon 2 inclusion ≥ 0.1 with $p \leq 0.01$) for a combination of SF3A1 knockdown with U1-5a/SL3-M1d, but not with U1-5a/SL4-M10r (Table 4.2). Thus, the effects of SF3A1 knockdown on U1 activity phenocopy loss-of-function SL4 mutations in the U1 snRNA. siPTBP1 treatment of HeLa cells caused a ~80% reduction in PTBP1 levels, however, it did not exacerbate the effects of either SL3 or SL4 mutations, and synergistic effects were not observed for a combination of PTBP1 knockdown and SL3 or SL4 mutations (Figure 4.14 and Table 4.2).
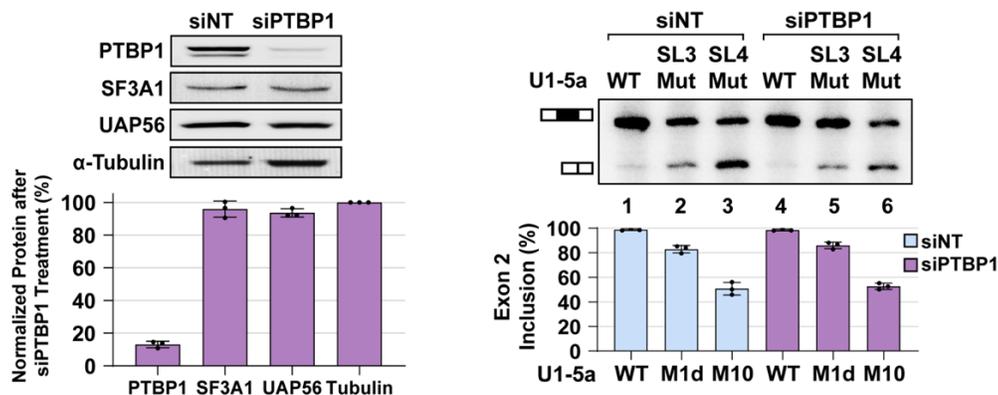
Figure 4.14 – PTBP1 knockdown does not induce synergistic impacts on U1 splicing activity when combined with either mutations in SL3 or SL4. Left panel – Western analysis of HeLa cell lysates after treatment with control siNT or PTBP1 targeting (siPTBP1) siRNA. Average protein expression (± s.d., n = 3) normalized to α-Tubulin is shown. Right panel – Primer extension analysis of Dup51p reporter transcripts after complementation with U1-5a variants and treatment with control siNT or siPTBP1. The average percentage of the full-length product (± s.d.) is graphed below (n = 3; * = p < 0.05, ** = p < 0.01). Statistical comparisons were performed for each U1-5a snRNA tested under the siNT versus siRNA treatment conditions. Analysis for synergistic effects is shown in Table 4.2.

These results emphasize the importance context has on regulation of splice-site choice and U1 snRNP activity. Although PTBP1 can interact with the U1 snRNA, PTBP1 binding to ESS is a prerequisite for regulation of cassette exon splicing. Previous

work has demonstrated that the Dup51 pre-mRNA lacks PTBP1 binding sites and its splicing is not regulated by this protein [260], therefore its silencing did not lead to synergistic reduction of U1 activity when combined with either SL3 or SL4 mutations (Figure 4.14). Overall, the results show that the impact of introducing double mutations in SL3 and SL4 (Figure 4.1) can be replicated by combining protein silencing with individual stem-loop mutations since the knockdown of SF3A1 phenocopies SL4 mutations and UAP56 knockdown phenocopies SL3 mutations. The lack of synergy, when SL3 mutations were combined with UAP56/URH49 knockdown or when SL4 mutations were combined with SF3A1 knockdown, indicated that each member of the stem-loop/protein pairs contributes to the same phenotype i.e. an epistatic relationship exists between SL3 and UAP56, and between SL4 and SF3A1. This epistatic relationship enables the synergistic effects of SL3 and SL4 double mutations on U1 activity to be replicated by combining SL3 mutations with SF3A1 knockdown or SL4 mutations with UAP56 knockdown and indicates that the stem-loop/protein pairs likely act together during constitutive pre-mRNA splicing.

Spatial orientation of SL3 and SL4 is important for U1 function

In the U1 snRNP, SL3 and SL4 are located diagonally opposite to each other and do not interact with any of the U1 specific proteins [77, 223]. To determine, if the relative orientation of these stem-loops within the U1 snRNP was important for activity, we created U1-5a snRNA constructs carrying tandem SL3 (SL3/SL3), tandem SL4 (SL4/SL4), and also swapped the positions of the stem-loops (SL4/SL3), and tested their activity in the Dup51p reporter assay (Figure 4.15).
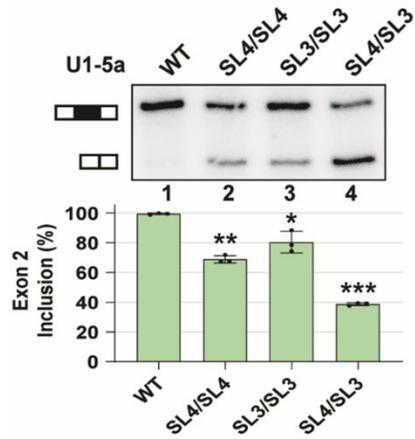
Figure 4.15 – Orientation and position of the SL3 and SL4 structures are vital for U1 splicing activity in transfected HeLa cells. Primer extension analysis to monitor splicing of Dup51p after co-transfections with U1-5a plasmids with wildtype U1 or U1 snRNA harboring tandem SL4 (SL4/SL4), tandem SL3 (SL3/SL3), or swapped SL3 and SL4 (SL4/SL3) structures. The full-length and exon 2 skipped Dup51p mRNA products are depicted. The percentage of the full-length product (± s.d.) is represented in the graph below and statistical significance was determined by comparisons to the wildtype control (lane 1) (n = 3; * = p < 0.05, ** = p < 0.01, *** = p < 0.001).
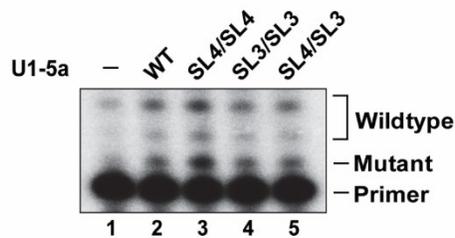
Figure 4.16 – U1-5a snRNA harboring tandem or swapped SL3/SL4 mutations are detectable by primer extension in transfected HeLa cells. Primer extension analysis with oligonucleotide ^{32}P-U1$_{7\text{-}26}$R showing expression of U1-5a snRNA harboring wildtype stem-loops, tandem SL4 (SL4/SL4), tandem SL3 (SL3/SL3), or swapped SL3 and SL4 (SL4/SL3) structures in the minigene Dup51p reporter assay.

Primer extension with U1$_{7\text{-}26}$R oligonucleotide confirmed the expression of U1-5a

snRNAs carrying tandem and swapped stem-loops in HeLa cells (Figure 4.16). U1-5a

constructs carrying tandem SL3 or SL4 decreased exon 2 inclusion to levels observed for

single SL3 or SL4 mutations (Figure 4.15, lanes 2 and 3). The swap-construct, on the

other hand, caused a much larger decrease in exon 2 inclusion (lane 4), suggesting that

exchanging the positions of SL3 and SL4 leads to complete loss of U1 activity as observed for SL3/SL4 double mutants. Thus, the inability of the swap-construct to support U1 activity suggests that the three-dimensional orientation of the stem-loops relative to other U1 components may be important for their recognition by spliceosomal proteins and critical for optimal U1 activity.

U1-SL3 promotes the SL4-SF3A1 interaction and A complex assembly

We next examined if free U1-SL3 could directly influence the interaction between U1-SL4 and its interacting U2 protein SF3A1 [168]. For this, uniformly ^{32}P-labeled U1-SL4-WT RNA was incubated in HeLa nuclear extract under splicing conditions, UV crosslinked, and then analyzed by SDS-PAGE. In HeLa nuclear extracts, U1-SL4 crosslinks to two proteins of ~120 kDa and ~50 kDa (Figure 4.17 lanes 1 and 2). In previous work, by a combination of crosslinking and immunoprecipitation, the ~120 kDa protein was identified as SF3A1 [168]. To test the effect of U1-SL3 on the U1-SL4-SF3A1 interaction, the reactions were preincubated with free cold U1-SL3-WT, U1-SL4-WT, U1-SL3-M1g or U1-SL4-M10 RNAs (see Figures 3.2 and 4.1 for sequences). Preincubation with U1-SL4-WT competed out crosslinking of the ~120 kDa SF3A1 band and the ~50 kDa protein, indicating the specificity of crosslinking (Figure 4.17, lanes 7-10). In contrast, preincubation with U1-SL3-WT enhanced U1-SL4-SF3A1 crosslinking but did not affect the ~50 kDa band (Figure 4.17, compare lane 2 to lanes 3-6). U1-SL3-M1g, which was found to reduce U1 activity in the complementation assay, did not enhance SF3A1 crosslinking (lanes 11-14). The U1-SL4-M10 mutant carries changes to the upper and lower stems but retains the CU-rich bulge and loop regions (Figure 4.1).

121

We have previously shown that these changes lead to loss of SF3A1 binding [168]. Preincubation with U1-SL4-M10 competed out the ~50 kDa product but did not affect binding to SF3A1 (Figure 4.17, lanes 15-18). U1-SL4-M10 slightly enhanced U1-SL4-SF3A1 crosslinking. This could be due to depletion of the ~50 kDa protein(s) by the excess SL4 mutant, leading to increased availability of the ^{32}P-U1-SL4 for interaction with SF3A1. Previously, analysis by RNA affinity purification and immunoblotting indicated that PTBP1 may be the ~50 kDa protein that binds to U1-SL4-WT [168].
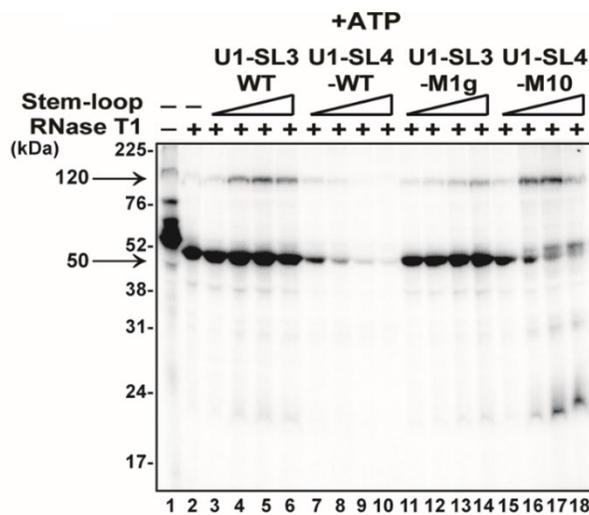
Figure 4.17 – U1-SL3 can promote the U1-SL4/SF3A1 interaction in vitro. UV crosslinking of ^{32}P-labeled U1-SL4 RNA in HeLa nuclear extracts in the presence of ATP. To determine the effect of free U1-SL3 and U1-SL4, the reactions were preincubated with 0.625, 1.25, 2.5, and 5.0 µM of the indicated cold stem-loop RNAs prior to addition of ^{32}P-U1-SL4.

Because the U1 snRNA-SF3A1 interaction during spliceosome assembly was found to be ATP-dependent [168], we examined if the enhancement of the U1-SL4-SF3A1 interaction by U1-SL3 was ATP-dependent. UV crosslinking in the absence of ATP or in the presence of ATP-γ-S revealed that under both conditions, wildtype U1-SL4 was able to compete out crosslinking (Figure 4.18 and 4.19 lanes 7-10), but wildtype U1-SL3 did not enhance U1-SL4-SF3A1 crosslinking (Figure 4.18 and 4.19 lanes 3-6),

thereby indicating that ATP-hydrolysis is required for the U1-SL3 mediated effect. These results suggest that the wildtype U1-SL3, but not the M1g mutant, may be promoting the interaction between U1-SL4 and SF3A1 in an ATP-dependent manner.
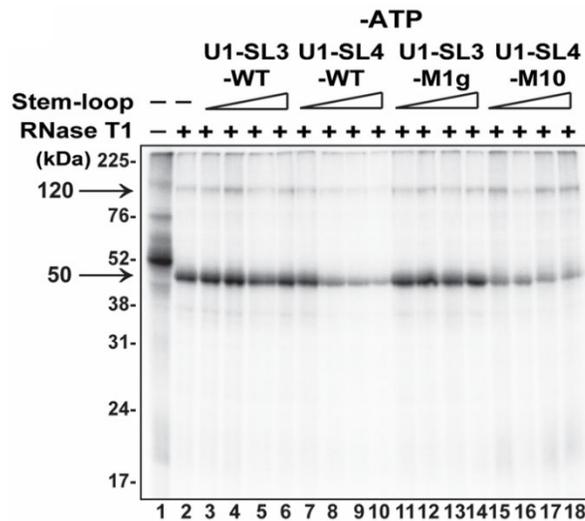
Figure 4.18 – U1-SL3 promotes U1-SL4/SF3A1 interaction in an ATP-dependent manner. HeLa nuclear extracts containing ^{32}P-U1-SL4 were incubated in the absence of ATP. The complexes were UV crosslinked, treated with RNase T1, and then analyzed by SDS-PAGE. To determine the effect of free U1-SL3 and U1-SL4, the reactions were preincubated with 0.625, 1.25, 2.5, and 5.0 μM of the indicated cold competitor stem-loop RNAs prior to addition of ^{32}P-U1-SL4.
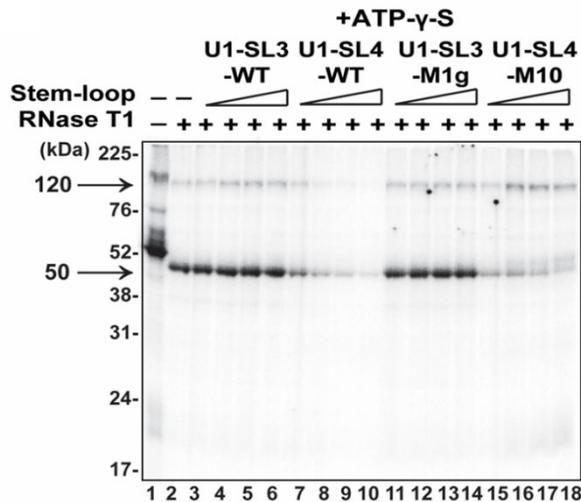
Figure 4.19 – U1-SL3 requires ATP hydrolysis to promote the U1-SL4/SF3A1 interaction. HeLa nuclear extracts containing ^{32}P-U1-SL4 were incubated in the presence of ATP-γ-S. The complexes were UV crosslinked, treated with RNase T1, and then analyzed by SDS-PAGE. To determine the effect of free U1-SL3 and U1-SL4, the reactions were preincubated with 0.625, 1.25, 2.5, and 5.0 μM of the indicated cold competitor stem-loop RNAs prior to addition of ^{32}P-U1-SL4.

Previously, we demonstrated that addition of free U1-SL4 to HeLa nuclear extracts inhibits pre-mRNA splicing by blocking formation of the A complex and not affecting the E complex [168]. We rationalized that if SL3 had a role in promoting the interaction between SL4 of pre-mRNA bound U1 snRNA and SF3A1, the effect of free U1-SL3 RNA addition would be to enhance *in vitro* splicing. To test this, HeLa nuclear extract was preincubated with increasing concentrations of free wildtype U1-SL3 and U1-SL4 RNAs prior to addition of ^{32}P-labeled adenovirus major late (AdML) pre-mRNA. Analysis of the spliced product showed that in contrast to U1-SL4-WT, U1-SL3-WT enhanced splicing by ~1.4 fold (Figure 4.20 compare lanes 7-10 with 3-6; U1-SL4-WT IC_{50} = ~2.0 μM).
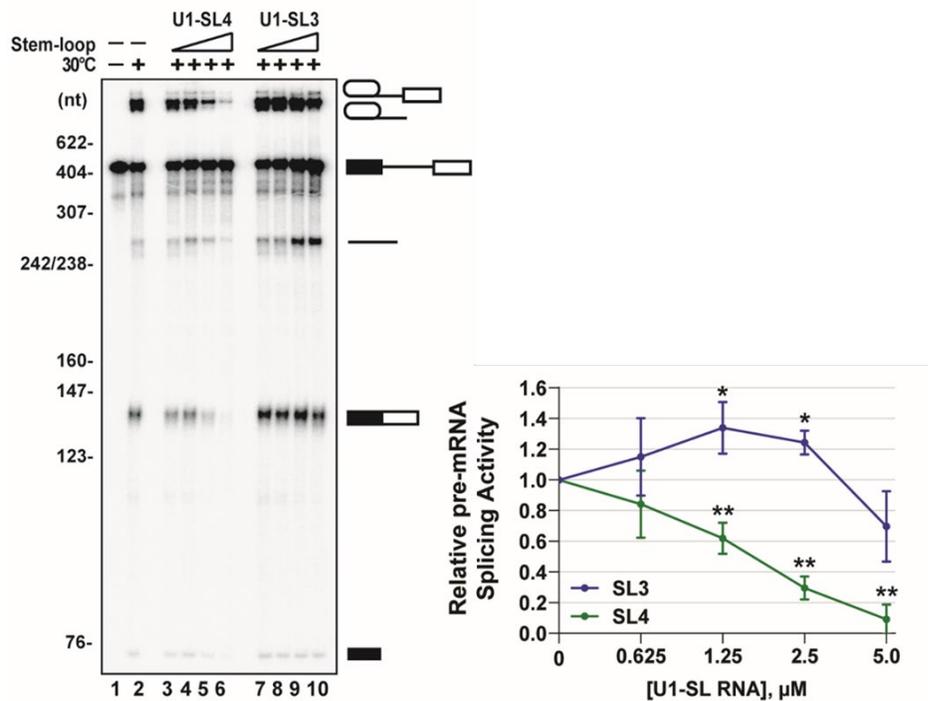
Figure 4.20 – U1-SL3 titration into *in vitro* splicing reactions can enhance splicing activity. *In vitro* splicing of uniformly ^{32}P-labeled AdML pre-mRNA in the absence of stem-loop RNA or in the presence of 0.625, 1.25, 2.5, and 5.0 μM wildtype U1-SL4 or U1-SL3. Splicing intermediates and products are depicted. Fold change in splicing activity is the mRNA/pre-mRNA ratio calculated relative to the no stem-loop control. Statistical analysis compared activity in the presence of U1-SL3 or U1-SL4 to the no stem-loop control ($\pm$ s.d., n = 4; * = $p < 0.05$, ** = $p < 0.01$).

124

Analysis of ATP-dependent spliceosomal complexes demonstrated that in contrast with U1-SL4-WT, U1-SL3-WT caused an increase in A complex formation (Figure 4.21, compare lanes 5-8 to 1-4). The reduction in splicing activity and A complex formation at concentrations of U1-SL3 above 2.5 μM was notable (Figure 4.20 lane 10, and Figure 4.21 lane 4). Although, the cause of this reduction is not clear, the trend of enhanced splicing activity and A complex formation in the presence of lower U1-SL3 concentrations is consistent.
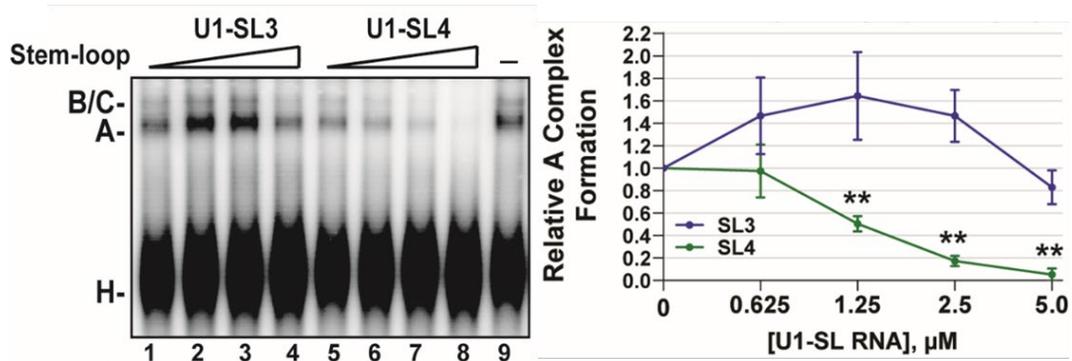
Figure 4.21 – U1-SL3 enhances splicing by stimulating A complex formation. Native agarose gel analysis of ATP-dependent spliceosomal complexes assembled on uniformly ^{32}P-labeled AdML pre-mRNA in the absence of stem-loop RNA or in the presence of 0.625, 1.25, 2.5, and 5.0 μM wildtype U1-SL4 or U1-SL3. Fold change in A complex formation is the A complex/H complex ratio calculated relative to the no stem-loop control ($\pm$ s.d., n = 3, ** = p < 0.01).

Analysis of the ATP-independent complexes showed that like U1-SL4 [168], U1-SL3 did not affect E complex formation (Figure 4.22), thereby suggesting that U1-SL3 likely promotes the E to A transition. Overall, these results underscore the ATP-dependence of U1-SL3 functions. Similar to its ability to bind UAP56, the ability of free wildtype U1-SL3 to promote the interaction between the U1-SL4 RNA and the U2 protein SF3A1 is also ATP-dependent. These observations, along with the fact that

association of UAP56 with the U1 snRNP also requires ATP, strongly suggest that the ability of U1-SL3 to promote the E to A complex transition and enhance pre-mRNA splicing may be due to its association with UAP56.
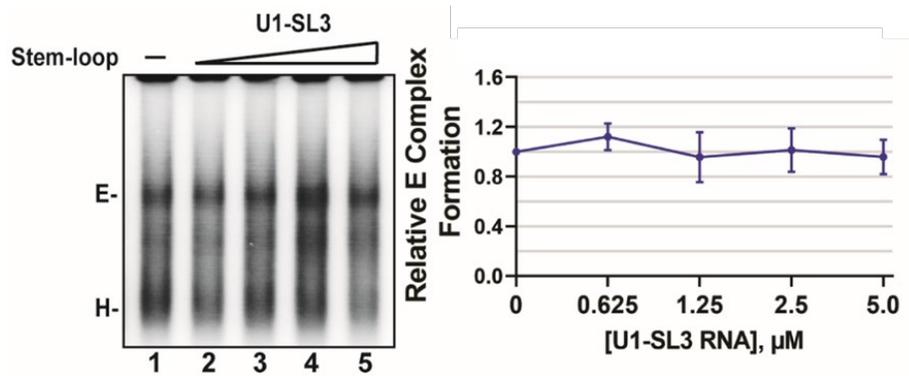
Figure 4.22 – U1-SL3 has comparatively no effect on E complex formation. Native agarose gel analysis of ATP-independent E complex assembled on uniformly ^{32}P-labeled AdML pre-mRNA in the absence or in the presence of 0.625, 1.25, 2.5, and 5.0 µM wildtype U1-SL3. Fold change in E complex formation is the E complex/H complex ratio calculated relative to the no stem-loop control (± s.d., n = 3).

Discussion

Our studies demonstrate that in addition to recognizing 5′-ss, the human U1 snRNA has the key functional role of bringing intron/exon ends to proximity during pre-mRNA splicing. This U1 function is enabled by the stem-loops 3 and 4 of the U1 snRNA that are not bound by any of the U1 specific proteins, and are available to interact with other factors during the early steps of spliceosome assembly. Previously, we reported that an interaction between SL4 of the U1 snRNA and SF3A1 of the U2 snRNP occurred during the transition of the E to A complex [168]. In this study, mutation analysis by the U1 complementation assay revealed that SL3 mutations affect pre-mRNA splicing and combining mutations of SL3 and SL4 synergistically compromises U1 snRNP activity,

indicating that the roles of SL3 and SL4 are interdependent. Binding analysis demonstrated selective interactions of SL3 with UAP56 and SL4 with SF3A1 (Chapter 3). Additionally, U1 complementation assays in combination with siRNA mediated knockdown confirmed that SL3 and SL4 of the U1 snRNA act through UAP56 and SF3A1, respectively. Finally, addition of free U1-SL3 to nuclear extracts was found to promote the U1-SL4-SF3A1 interaction in an ATP-dependent manner, and enhance pre-mRNA splicing *in vitro* by promoting the E to A complex transition.

Based on these observations, we propose that interactions of SL3 and SL4 of the pre-mRNA bound U1 with UAP56 and SF3A1 bridge the 5′- and 3′-ss complexes during the early steps of spliceosome assembly (Figure 4.23). The SL3-UAP56 complex stabilizes the interaction between SL4 and SF3A1 in an ATP-dependent manner, thereby enhancing pre-mRNA splicing.
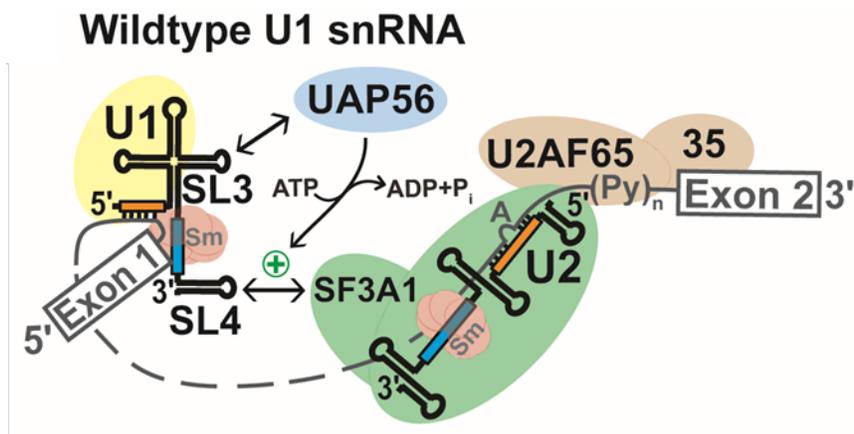
Figure 4.23 – Model for the role of the U1 snRNA during early spliceosome formation. During the early steps of spliceosome assembly, SL3 and SL4 of the U1 snRNA interact with the RNA helicase UAP56 and the U2 snRNP specific protein SF3A1, respectively (double headed black arrows). The SL4-SF3A1 contact bridges the 5′- and 3′-ss complexes. The SL3-UAP56 complex, directly or indirectly, promotes the SL4-SF3A1 interaction (green plus symbol) in an ATP-dependent manner, leading to enhancement of pre-mRNA splicing.

In the absence of the U1 snRNA-UAP56 contact, due to SL3 mutations or UAP56 knockdown, the SL4-SF3A1 interaction occurs, but is weaker or less efficient, overall leading to reduced splicing (Figure 4.24 Top Panel). Similarly, in the absence of the U1 snRNA-SF3A1 contact, due to SL4 mutations or SF3A1 knockdown, splicing is reduced (Figure 2.24 Bottom Panel). In this scenario, the 5′- and 3′-ss complexes are likely bridged by interactions of UAP56 with SL3 of the U1 snRNA and with U2AF65. UAP56 was shown to be essential for A complex formation and was found to interact with U2AF65 in an ATP-dependent manner in yeast and humans [79, 234, 235].
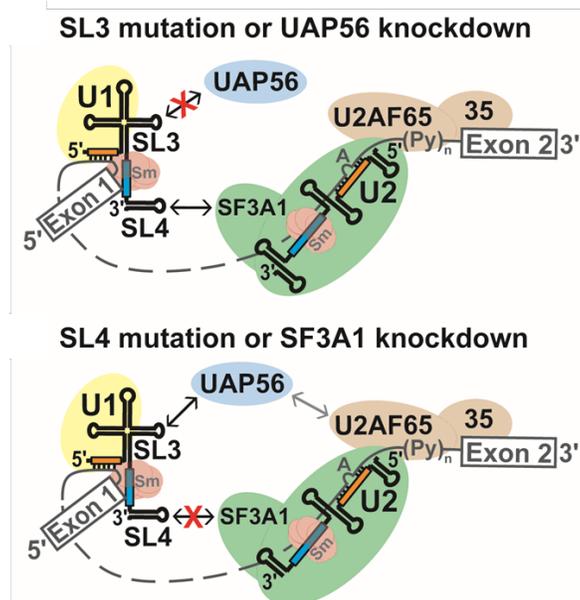
Figure 4.24 – Model for the impacts of stem-loop 3 or 4 mutations or respective binding protein knockdown. Top Panel – Disruption of the SL3-UAP56 contact by either SL3 mutations or UAP56 knockdown prevents stabilization of the SL4-SF3A1 interaction, resulting in reduced splicing. Bottom Panel – Disruption of the SL4-SF3A1 interaction by either SL4 mutations or SF3A1 knockdown reduces but does not completely abrogate splicing as the SL3-UAP56 interaction can occur. In the absence of the SL4-SF3A1 contact, interaction of UAP56 with U2AF65 likely bridges the 5′- and 3′-ss complexes (gray double headed arrow).

The primary role of the SL4-SF3A1 contact in cross-intron bridging is underscored by the observation that addition of free U1-SL4 prevents the E to A complex transition and is sufficient for inhibition of pre-mRNA splicing *in vitro* (Figure 4.25 Top

Panel) [168]. Addition of free U1-SL3, on the other hand, has the profoundly different effect of enhancing splicing. This suggests that the SL3-UAP56 complex may be stabilizing the cross-intron contact between SL4 of the U1 snRNA and the U2 protein SF3A1 during A complex formation (Figure 4.25 Bottom Panel). This idea is supported by the observations that the addition of U1-SL3 in trans promotes the U1-SL4-SF3A1 contact in an ATP-dependent manner and that free U1-SL3 enhances pre-mRNA splicing by promoting the E to A complex transition. It is very likely that these U1 snRNA contacts occur in addition to the other reported contacts between U1 and 3′-ss complex components including interactions between U1-70k and U2AF65, Prp40 and SF1, and an indirect contact between U1A and SF3B1 [230, 261-264].
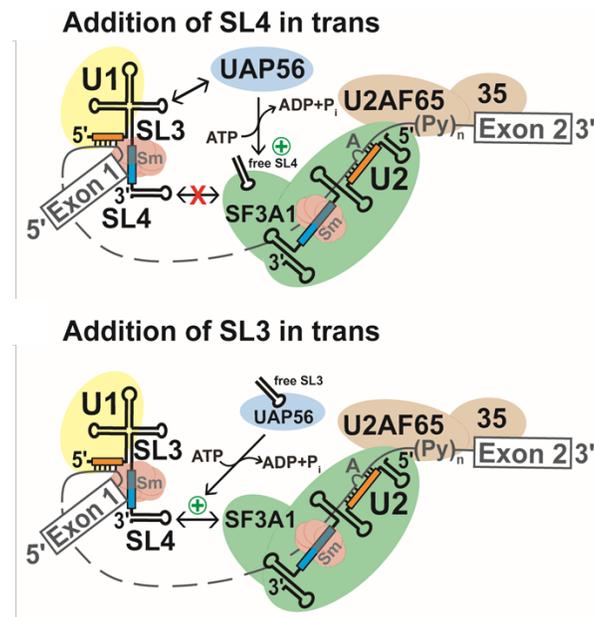
Figure 4.25 – Model for the impacts of stem-loops 3 and 4 on splicing when present in excess and added in trans. Top Panel – The addition of excess U1-SL4 in trans competes out the interaction of SF3A1 with endogenous U1 snRNA, reducing A complex formation and inhibiting splicing in vitro (Sharma et al., 2014). Bottom Panel – By contrast, addition of excess U1-SL3 in trans enhances pre-mRNA splicing by binding to endogenous UAP56. The U1-SL3-UAP56 complex promotes the SL4-SF3A1 interaction in an ATP-dependent manner, enhancing A complex formation and splicing.

Experimental

Plasmid constructs and transfections

The three-exon/two-intron reporter pDUP51p and the U1 snRNA expression plasmid pNS6U1 have been described previously [168]. The constructs expressing U1-5a snRNAs carrying SL3 mutations, SL3/SL4 double mutations, and SL3/SL4 tandem and swap mutations were generated by PCR mutagenesis using oligonucleotides and were verified by DNA sequencing. The sequences of all oligonucleotides used for U1-5a mutagenesis performed in this study are provided in Appendix D.

HeLa cells, originally purchased from ATCC, were a gift from Kurt Gustin (University of Arizona, College of Medicine-Phoenix). They were cultured in DMEM containing 10% FBS and antibiotics (100 U/mL penicillin, 100 μg/mL streptomycin). Culture supernatants were tested for mycoplasma by PCR using a pool of six mycoplasma specific primers, and found to be negative [219]. For transfection, 0.5×10^5 cells per well of a six-well plate were transfected with 0.4 μg of Dup51p reporter plasmid and 3.6 μg of control plasmid (pcDNA3.1) or U1 expression plasmid (pNS6U1) using Lipofectamine 2000 according to the manufacturer's instructions (Thermo Fisher Scientific). Cells were harvested 48 hours post-transfection and total RNA was extracted using TRIzol reagent or prepared using the Direct-zol RNA kit (Zymo Research).

For siRNA and reporter double transfection experiments, 0.24×10^5 cells per well of a six-well plate were transfected with 50 nM synthetic siRNA using lipofectamine RNAiMAX transfection reagent (Thermo Fisher Scientific). After 24 hours, cells were transfected with Dup51p reporter and U1 plasmids as described above. After incubation for another 24 hours, cells were harvested, and total RNA was extracted also as described

above. siRNAs targeting SF3A1, UAP56, URH49, and PTBP1 have been described previously [181, 265-267]. All siRNAs, including the non-targeting control (siNT; siGENOME Non-Targeting Pool #1) were purchased from Horizon Discovery; sequences are provided in the S Appendix E.

For SL3/SL4 tandem and swap mutants, transfections were performed in a 12-well plate format with a pNS6U1-5a to Dup51p ratio of 7.5:1 (1.5 μg pNS6U1-5a and 0.2 μg Dup51p) with the addition of pNS6U1-WT (0.3 μg) to maintain the total levels of U1 expression plasmid the same as in previous experiments (final ratio of pNS6U1:Dup51p = 9:1). 2.0 x10^5 cells per well of a 12-well plate were transfected using Lipofectamine 2000 reagent as before and total RNA was harvested by TRIzol extraction 48 hours post-transfection prior to primer extension analysis of Dup51p reporter transcripts.

Nucleo-cytoplasmic fractionation

Sub-cellular fractionation of HeLa cells was performed using the protocol by Gagnom *et al.*, with some modifications [268]. Briefly, 300 μl of ice-cold IGEPAL hypotonic lysis buffer (IHLB; 10 mM Tris-HCl (pH 7.5), 10 mM NaCl, 3 mM MgCl$_2$, 0.3% IGEPAL CA-630) was added to each well of a 6-well plate. Cells were lifted by scraping and the lysates were kept on ice for 10 min., vortexed, and centrifuged at 5,000 x *g* at 4°C for 10 min. Supernatants containing the cytoplasmic fractions were treated with SDS/Proteinase K and total cytoplasmic RNA was extracted using phenol:chloroform. For cytoplasmic protein, the fractions were treated with 500 U Nuclease A for 15 min. at room temperature (RT) prior to addition of SDS-PAGE sample buffer. The nuclear pellets were washed by re-suspension in 50 μl ice-cold IHLB for 10

min. and centrifuged as above. From the nuclear pellet, total RNA was extracted using TRIzol. For nuclear protein, the pellet was re-suspended in 300 μl of PBS and treated with 500 U Nuclease A for 15 min. at RT.

Antibodies, Western blotting, and immunoprecipitation

For protein analysis, samples were boiled in SDS-PAGE sample buffer, separated on 10% SDS-PAGE gels and analyzed by Western blotting using PVDF membrane. Antibodies against SF3A1, SF3A3, SF3B1, and U1-70K have been described previously [168]. The anti-UAP56 rabbit polyclonal antibody was a gift from Robin Reed (Harvard Medical School). This antibody was raised against GST-UAP56 and cross-reacts with URH49 [244]. Other commercial antibodies used in this study included anti-α-Tubulin mouse monoclonal antibody (EMD Millipore; CP06-DM1A) and secondary anti-mouse and anti-rabbit antibodies conjugated to Cy3 and Cy5 fluorophores were purchased from Thermo Fisher Scientific. Proteins were quantified by densitometric scanning of Western blots using ImageQuant.

For U1-70k IP, HeLa cells transiently expressing WT and mutant U1-5a snRNA from two wells of a 6-well plate were pooled after trypsinization and pelleted by centrifugation at 600 x g for 5 min. at RT. HeLa cell nuclei were then purified as described in the fractionation protocol above. After washing in IHLB, the nuclei were resuspended in 250 μl of Buffer C (20 mM HEPES-KOH pH 7.9, 1.5 mM $MgCl_2$, 0.42 M NaCl, 0.2 mM EDTA, 1 mM DTT, 0.1 mM PMSF, and 20% glycerol). To extract nuclear components, the Buffer C nuclear suspension was incubated with rotation at 4°C for 30 min. and then centrifuged at 5,000 x g for 10 min. at 4°C. The supernatant was added to

20 μl packed-volume of GammaBind Sepharose beads (Cytiva) that were pre-bound with 5 μg of anti-U1-70k antibody and incubated for 1 hour at RT with rotation. Beads were washed four times in 1x PBS. Bound protein was extracted by boiling in SDS-PAGE sample buffer and bound RNA was extracted using TRIzol.

Primer extension, RT-qPCR, and Northern blotting

Primer extensions to monitor splicing of the Dup51p reporter and for determining U1-5a snRNA expression were performed using ^{32}P-Dup3r and U1$_{7-26}$R oligonucleotides, respectively, as described previously [168]. Oligonucleotide sequences are provided in Appendix E. Spliced products were quantified by densitometric scanning of urea-PAGE images using ImageQuant. For RT-qPCRs, reverse transcription was performed using SuperScript III Reverse Transcriptase kit and 1 ng of resulting cDNA was used as a template for qPCR amplification using SYBR Green reagent and StepOnePlus Real-Time PCR Machine; all according to manufacturer specifications (Thermo Fisher Scientific). Primer pairs used for amplification of U1 and U2 snRNA are reported in the Appendix E. U1 expression across all samples was normalized to U2 snRNA and fold-increase in expression was calculated relative to the pcDNA control.

For Northern blotting, RNA samples were separated on 10% urea-PAGE gels and transferred onto Amersham Hybond nylon membrane (Cytiva) for 1 hour at 15V, 400 mA using the Trans-blot Turbo semi-dry transfer system (Bio-Rad). Transferred RNA was UV crosslinked to nylon membranes for 10 min. and pre-hybridized in 15 ml of ULTRAhyb hybridization buffer. The membranes were probed with ^{32}P-labeled oligo probes (1.0 x 10^6 cpm/ml) in hybridization buffer overnight at 42°C. Membranes were

washed at the temperature of hybridization once with 2X saline-sodium citrate buffer (2X SSC; 300 mM NaCl, and 30 mM sodium citrate) containing 0.1% SDS for 10 min. and twice with 2X SSC for 10 min. For locked-nucleic acid (LNA)-modified oligonucleotide probes (Exiqon), membranes were incubated at 37°C overnight and washed at RT once with 2X SSC containing 0.1% SDS and twice with 2X SSC alone. Northern blots were visualized using the Typhoon FLA 9500 imager and RNA bands were quantified by densitometric scanning using ImageQuant.

UV crosslinking and pre-mRNA splicing

Nuclear extract from HeLa S3 cells was prepared as described previously [245, 246]. The ^{32}P-labeled U1-SL3 and U1-SL4 RNAs were *in vitro* transcribed from annealed DNA templates, gel purified, and ethanol precipitated. The RNAs were incubated at a final concentration of 20 nM in a splicing reaction containing 2.2 mM MgCl$_2$, 0.4 mM ATP, 20 mM creatine phosphate, 10U RNaseOUT, and 60% nuclear extract in buffer DG (20 mM HEPES pH 7.9, 80 mM K-glutamate, 0.1 mM EDTA, 1 mM DTT, 20% glycerol). In Figures 4.17, 4.18, and 4.19, the reactions were preincubated with cold 5′-biotinylated U1-SL3-WT, U1-SL4-WT, U1-SL3-M1g or U1-SL4-M10 RNAs (Integrated DNA Technologies) at the indicated concentrations for 20 min. on ice. After preincubation, ^{32}P-U1-SL4 was added and incubation was continued at 30°C for 30 min. UV crosslinking was performed in a GS Gene Linker (Bio-Rad Laboratories) for a total energy of 1800 mJ. Next, reactions were treated with 100U of RNase T1 at RT for 5 min. and crosslinked proteins were separated on 10% SDS-PAGE gels and visualized using the Typhoon FLA 9500 imager.

For *in vitro* splicing, uniformly ^{32}P-labeled pre-mRNA substrate was transcribed from the pSPAd plasmid, gel purified, and added to splicing reactions containing HeLa nuclear extract and all other components as described above. To examine effects of stem-loop RNAs on AdML splicing, reactions were preincubated with cold wildtype U1-SL3 or U1-SL4 at the indicated concentrations for 20 minutes at RT prior to addition of the pre-mRNA substrate and further incubation at 30°C for 1 hour. Analysis of spliceosomal complexes was performed by native agarose gels as described previously [125].

Statistical analysis

All statistical comparisons were performed using the two-tailed Student t-test in Microsoft Excel. For Figure 4.1, a difference in exon 2 inclusion of $\geq 10\%$ with $p < 0.05$ was considered statistically significant. The analyses for synergy in U1 activity (A; fraction of exon 2 inclusion in the Dup51p reporter mRNA) were performed by STATA version 14 using the linear mixed model. For U1 snRNA mutations, these analyses compared the predicted activity for a particular combination of single SL3 or SL4 mutations ($A_{pred} = A_{SL3}*A_{SL4}$) to the observed activity (A_{obs}) of U1 snRNAs carrying double mutations (Table 4.1). The effects of double mutations were considered synergistic if $A_{obs}<A_{pred}$ with a difference of ≥ 0.1 and $p \leq 0.01$. For combinations of protein knockdown and U1 snRNA mutations, the predicted U1 activity for a combination of siRNA treatment and a particular stem-loop mutation ($A_{pred} = A_{siRNA}*A_{SL3/SL4}$) was compared with the observed U1 activity (A_{obs}) when stem-loop mutations were expressed after siRNA treatment (Table 4.2). All measurements approximated a normal distribution following a log transformation.

135

CHAPTER 5

CONCLUSION

Intron architecture dictates modes of splice-site recognition in yeast and humans

Although the fundamental mechanism of splicing is highly conserved and can be traced back to ancient mobile genetic elements like self-splicing group II introns, key differences in intron architectures have likely required unique processes of spliceosome assembly to evolve. In the unicellular budding yeast, *S. cerevisiae*, the consensus sequences are nearly invariant and rarely engage in alternative splicing (Figure 1.5). Yeast also lack SR proteins that play a large role in metazoans with more degenerate splice-site sequences where these auxiliary proteins facilitate recruitment of snRNPs for constitutive splicing and during regulation of alternative splicing [269]. In budding yeast, only ~4% of genes contain introns these are primarily short, ranging from ~100-400 nucleotides in length [270, 271]. However, the opposite is found in higher eukaryotes where large introns separating short exons is the rule and correlated with increased phenotypic complexity of the organism [272].

Communication between the 5′- and 3′-ss complexes reinforce each other, enhancing splice-site specificity and rates of spliceosome assembly. Applying model pre-mRNA substrates with single introns, prespliceosome interactions across short introns have been observed between the U1 snRNP and components of the 3′-ss (like U2AF65 and the U2 snRNP) in both yeast and humans [86, 230]. However, due to the prevalence of short introns in yeast, it is thought that spliceosome assembly via intron definition (ID) prevails in that organisms, building upon cross-intron interactions formed in the prespliceosome (Figure 5.1) [273, 274]. In humans however, it is thought that short exons

are recognized first through exon definition (ED) prior to spliceosome formation across the intron [275]. Once introns exceed ~250 nts, splice-site recognition is thought to occur through ED, with exon-bound SR proteins primarily mediating communication across the exon (Figure 5.1) [13, 276]. For example, U1 bound to a 5′-ss can promote binding of U2AF65 at an upstream 3′-ss, and the presence of an upstream and downstream 5′-ss can synergistically enhance U2 snRNP recruitment to the branchpoint via both cross-intron and cross-exon interactions [277, 278]. Consistent with these models of spliceosome assembly, alternative splicing in budding yeast is exclusively in the form of intron retention, whereas in metazoans like humans, exon skipping is most prevalent [279].
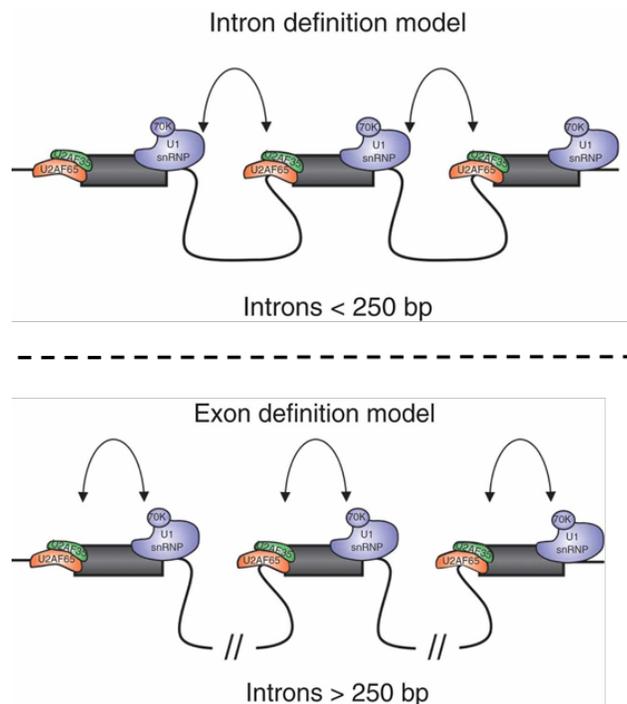
Figure 5.1 – Spliceosome assembly can occur after intron definition or exon definition. Top Panel – Prespliceosome assembly on pre-mRNA substrates with short introns, like those that predominate in budding yeast, is thought to occur via cross-intron interactions between the 5′- and 3′-ss (intron definition). Bottom Panel – When introns are long and vastly outsize exons as in humans, prespliceosome assembly is thought to occur via cross-exon interactions first (exon definition). To juxtapose exons for splicing and orient the 5′- and 3′-ss into catalytic centers, the exon-defined prespliceosome would need to transition to interactions across the intron during spliceosome assembly. Figure adapted from De Conti et al. 2013.

Once exons in humans are identified by ED, mechanisms to convert the exon-defined spliceosome to the intron-defined spliceosome must occur through processes likely not present in budding yeast. How this transition occurs however, remains unclear [280]. This conversion to an intron-defined spliceosome is vital for splicing and can be inhibited by hnRNPs like PTB and RBM5 to promote exon skipping [126, 127]. Therefore, further characterizing the 5′- and 3′-ss interactions is necessary to further elucidate mechanisms of prespliceosome formation and alternative splicing in higher eukaryotes. Findings from the studies presented in this dissertation demonstrate that spliceosomes in higher eukaryotes leverage unique interactions during prespliceosome formation. Additionally, the observations made in these experiments expand on our fundamental understanding of RNA-protein interactions performed by non-canonical RNA binding domains and RNA helicases. Investigating the cross-intron contact between U2 protein SF3A1 and the U1 stem-loop 4 structure revealed that Ubiquitin-like domains have the capacity to bind RNA. Additionally, the RNA-dependent ATPase UAP56 was identified as a U1 snRNP interacting splicing factor found to bind U1 stem-loop 3 with remarkable specificity; indicating that other helicases may also have specific RNA substrates. The interplay among the interactions made by U1 stem-loops 3 and 4 provide further insight into the contacts between splice-site complexes formed in the prespliceosome, and may contribute to an understanding of the conversion from an exon-defined to an intron-defined spliceosome.

The SL4-SF3A1 interaction is unique to higher eukaryotes

The U2 snRNP is largely a composed of two sub-complexes formed by SF3A and SF3B that are conserved from budding yeast to humans and are involved in formation of the prespliceosome (Figure 1.6). The SF3B complex contains SF3B1, a protein that plays a critical role stabilizing the U2-branchpoint duplex and binds the branchpoint (BP) adenosine [281]. The SF3A complex is a heterotrimer composed of the SF3A1, SF3A2, and SF3A3 proteins that are required for mature 17S U2 snRNP biogenesis [211]. The SF3A2 and SF3A3 components interact directly with the U2 snRNA via zinc-finger RNA binding domains, while SF3A1 associates with the U2 snRNP via protein-protein interactions with SF3A2 and SF3A3 (Figure 2.1) [182, 211].

SF3A1 and SF3B1 may help loosely associate U2 to the E complex; SF3A1 by interactions between its N-terminal SURP1 domain and SF1, and SF3B1 by binding to U2AF65 [175, 180]. Once the A complex is formed, all components of the SF3A complex, along with proteins in SF3B, interact with a region upstream of the BP called the anchoring site that stabilized the association of the U2-snRNP at the BP [82, 282]. The organization of the major U2-snRNP subunits are highly conserved from yeast to humans and have essential roles in prespliceosome formation in both species [179, 283, 284].

The SF3A1 component of human SF3A however is highly divergent from its budding yeast homolog (Prp21). Both proteins share a conserved N-terminal region harboring tandem SURP domains (Figure 2.1) with SURP2 involved in binding SF3A3 in both systems [178]. However, human SF3A1 is ~2.8-times the size of Prp21 and contains a large unstructured region that spans the central portion of the protein and is followed by

a C-terminal nuclear localization signal and Ubiquitin-like domain. [178, 184].

Unfortunately, due to the flexibility of the C-terminal half of the protein, the full-length

structure of SF3A1 has not been captured by cryo-EM studies of the purified U2 snRNP

[281] or in any of the early-spliceosomal complexes in humans [88, 90, 94]. Recently

however, crosslinking of the central residues in SF3A1 (aa 458-473) to components of

the tri-snRNP have been identified in a cryo-EM structure of the human pre-B complex,

implicating SF3A1 in tri-snRNP recruitment [90].

Previous work identified a new role for SF3A1 in prespliceosome formation in

contacting the stem-loop 4 structure of the U1 snRNA across the intron [168]. In the

research reported in Chapter 2, I demonstrate further that the C-terminal Ubiquitin-like

(UBL) domain of SF3A1 is necessary and sufficient for binding free U1-SL4 and the

intact U1 snRNP [232]. The SF3A1-UBL represents a novel non-canonical RNA binding

domain and this cross-intron bridging interaction is specific to human prespliceosomes,

since budding yeast lack this region in their SF3A1 homolog Prp21 and also contain no

SL4-equivalent structure in their U1 snRNA [177]. Thus, these findings expand the

family of known RNA binding domains and support a role for a metazoan-specific cross-

intron interaction that may inform future studies into the mechanisms of exon and intron

definition in complex eukaryotes.

Expanding the role for UAP56 in prespliceosome formation

Helicases have many diverse functions in cells, apart from canonical RNA

unwinding activities, and can leverage ATP binding and hydrolysis to clamp down and

remain bound to substrates or to displace protein from RNA [285]. For example, the other

DExD/H box RNA helicase required for prespliceosome formation, Prp5, does not require helicase activity to promote A complex assembly [286, 287]. Rather, through ATP hydrolysis, Prp5 displaces proteins from the U2 snRNA branchpoint-interacting stem-loop (BSL) causing this structure to denature and form the thermodynamically favored duplex with the BP sequence in pre-mRNA [281]. Displacement of splicing factors to facilitate new interactions is a common mechanism employed by helicases throughout the splicing cycle and may be how UAP56 promotes the SL4-SF3A1 interaction [170, 288]. Perhaps the UBL of SF3A1 is sequestered in the U2 snRNP and UAP56 via ATP hydrolysis, either directly or in-directly, displaces and frees the UBL domain for interaction with U1-SL4 during prespliceosome formation. More experiments are required to gain a mechanistic understanding for how UAP56, and its interaction with U1-SL3, promotes prespliceosome formation.

The capacity for UAP56 to discriminate between the U1-SL3 and U1-SL4 structures *in vitro* was striking (Figure 3.10). Generally, helicases are considered to be non-specific in their interactions with RNA, requiring the use of N- and C-terminal extensions/domains, or interactions with binding partners, to recruit them to target substrates [238, 239, 289-291]. Consistent with these observations, UAP56 *in vitro* has been found to unwind short 13-nucleotide RNA duplexes in a sequence independent manner, while its ATPase activity is enhanced the most in the presence of single-stranded RNA [253]. Surprisingly, we observed that UAP56 can specifically bind U1-SL3 *in vitro* by EMSA, while showing comparably no binding activity for a similar stem-loop RNA structure, U1-SL4. This is interesting because UAP56 has no obvious N- or C-terminal extension to mediate RNA binding specificity as observed in many other helicases [292].

Therefore, the UAP56-SL3 interaction may be an indication that minimal RNA helicases with little more than core helicase motifs can indeed recognize specific RNA substrates.

UAP56 was initially identified as a U2AF65-interacting spliceosomal helicase in humans that was required for stable incorporation of the U2 snRNP at the BP [168]. Experiments in *S. cerevisiae* identified the yeast homolog of UAP56 (Sub2) as a helicase that interacts with yeast U2AF65 (Mud2p) [235, 293, 294]. In budding yeast, the 3′-ss is initially recognized by the branch point binding protein (BBP) which also interacts with Mud2p to form a heterodimer [295, 296]. This early 3′-ss complex is analogous to that in humans where the SF1-U2AF65-U2AF35 heterotrimer binds the BP, polypyrimidine tract, and 3'ss, respectively [297]. There are differences however, in that in budding yeast there is no homolog of U2AF35 [298], and the BBP-Mud2p heterodimer (yeast SF1-U2AF65) is completely displaced during the addition of the U2 snRNP at the BP [230, 299]. An initial model for the role of UAP56 function in prespliceosome formation in yeast proposed that Sub2 binds Mud2p and, using the energy from ATP hydrolysis, displaces the BBP-Mud2p heterodimer thereby promoting annealing of the U2 snRNA to the now exposed BP [235].

Like yeast, displacement of the branch point binding protein SF1 is required for stable integration of the U2 snRNP in humans, however, there is evidence that unlike the yeast system U2AF65 is not displaced during A complex formation. Unlike Mud2p, U2AF65 may stabilize the U2 snRNA at the branchpoint via its RS domain [83, 299, 300]. Once hybridized, SF3B1 of the U2 snRNP may directly bind U2AF65 though U2AF-homology motif interactions [176, 301, 302]. Although not modeled in current prespliceosomal structures and bound less-stably than in the E complex, U2AF65 is

142

detectable as late as the A and pre-B complexes by mass spectrometry [67, 71, 303, 304]. The current model for UAP56 function in human prespliceosome assembly posits that, unlike yeast where Sub2 displaces the entire BBP-Mud2p heterodimer, UAP56 instead is involved in the displacement of SF1 only, although there is little direct evidence for this model.

The functions of UAP56 in human spliceosome assembly may be more complex than those observed in budding yeast. In humans, the interaction of UAP56 with U2AF65 may be involved in later stages of spliceosome assembly, as UAP56 was found to unwind the U4/U6 duplex *in vitro* in a U2AF65 dependent fashion [79]. Accordingly, this function was dependent on UAP56 helicase activity and mutations that disrupted this function interfered with progression past the A complex. In this study by Shen and colleagues, no interaction between U1 and UAP56 was detected by UV-crosslinking and immunoprecipitation. However, these experiments were performed in the presence of ATP and I observed that the interaction of UAP56 with the U1 snRNP was primarily detected in the presence of ATP-γ-S (Figure 3.17) and may explain why the interaction between UAP56 and U1 had not been detected previously.

Shen and colleagues also observed that mutations which disrupted the ATP binding or ATP hydrolysis activities of UAP56 inhibited the conversion from the E→A complexes [79]. The requirement for the ATP binding and hydrolysis activities of UAP56 in promoting A complex formation is consistent with observations outlined in Chapters 3 and 4 where the interaction of UAP56 with U1-SL3 (Figure 3.14), with the U1 snRNP (Figure 3.17), and the ability for U1-SL3 to promote crosslinking of U1-SL4 with SF3A1, presumably through the action of UAP56 (Figure 4.17), all had an ATP

requirement. Our findings support a role for UAP56 in promoting the SL4-SF3A1 interaction across introns that enhances the efficiency of A complex formation and might be involved in the conversion from an exon- to an intron-defined spliceosome in humans.

Future perspectives

Moving forward, experiments aimed towards further characterizing the interaction of UAP56 with SL3 and the intact U1 snRNP, in addition to elucidating the mechanisms underlying the capacity for U1-SL3 to enhance the SL4-SF3A1 interaction, are needed to provide further evidence for the model of A complex assembly presented in this dissertation. Additionally, to determine if the SL4-SF3A1 interaction is a component of constitutive splicing or perhaps only required for splicing of particular genes across the human transcriptome, identifying and quantifying SF3A1-dependent splicing events is required.

Identifying the features of UAP56 which may be responsible for conferring its specificity for the U1-SL3 RNA would advance our understanding of DExD/H box helicase RNA binding activity. A series of *in vitro* binding experiments as performed in Chapters 2 and 3 such as UV crosslinking and gel shift assays to assess the binding activity of UAP56 mutants might reveal structures unique to UAP56 which confer affinity to U1-SL3. Deletion of the short N- and C-terminal extensions outside of the core RNA helicase domain are obvious regions which can be truncated in recombinant proteins to determine if they are necessary to bind SL3. The N-terminal extension of UAP56 has been found to bind the influenza A nucleoprotein [305] and therefore these extensions may be physiologically relevant. Additionally, it would be interesting to

144

determine if U1-SL3 may stimulate UAP56 ATPase and helicase activities in a particularly distinct fashion compared to other RNA substrates. These experiments can be performed *in vitro* using a variety of assays that monitor rates of ATP hydrolysis or duplex unwinding with either radioactive or fluorescent probes [306]. Determining the RNA binding activity of UAP56 harboring mutations that can differentially uncouple helicase activity from ATPase activity (like the D199A helicase mutation) will also be important to confirm which activities of UAP56 are required for interacting with U1-SL3 [253]. The binding of UAP56 to the U1 snRNP may not be entirely mediated by its association with the U1-SL3 structure as a recent yeast two-hybrid screen identified U1A as a UAP56 interacting protein (https://thebiogrid.org/interaction/2700332) [307, 308]. Validating the capacity for UAP56 to interact with U1A would be particularly interesting as it could be mediated by the RRM2 of U1A which currently has no known function [309], and would shed light on the mechanism underlying UAP56 recruitment to the prespliceosome (Chapter 3).

The capacity for U1-SL3 to enhance the SL4-SF3A1 is likely mediated by UAP56 as this phenomenon was observed to be ATP-dependent and did not occur with a mutant SL3 which does not interact with UAP56 (Table 3.1 and Figures 4.17, 4.18, and 4.19). Therefore, a direct effect of UAP56 on SF3A1 or other components of the SF3A complex may be involved. It might be revealing to determine if UAP56 can bind any components of the SF3A complex directly. I observed that UAP56 can co-purify with the U2 snRNP by anti-sense oligo affinity purification in the presence of both ATP-γ-S and ATP (Figure 3.19). UAP56 was not found to crosslink to U2 snRNA in the presence of ATP [79] and therefore the U2-UAP56 interaction may be mediated by binding of UAP56 directly to

U2 associated protein, like those of the SF3A complex. Co-purification experiments with recombinant proteins could determine which SF3A-complex proteins, or other components of the U2 snRNP, may be involved in binding UAP56 and may shed light on its capacity to promote the SL4-SF3A1 interaction.

To determine the number of SF3A1-dependent introns across the human genome, whole-transcriptome RNA sequencing (RNA-seq) would be required. Although complete SF3A1 knock-out cell lines cannot be generated due to the essential nature of this gene [181], the transient knock-out of protein expression by the siRNA strategy employed in Chapter 4 or the use of more precise strategies such as CRISPR interference [310] followed by RNA-seq could be leveraged to monitor SF3A1-dependent splicing events.

A more precise method to monitor SF3A1-UBL-dependent splicing changes could be to generate a cell line expressing an SF3A1 truncation with the UBL domain deleted leaving the rest of the protein intact. A cell line harboring this mutation may be more viable than a complete SF3A1 knockout. SF3A1 protein lacking the UBL domain has been shown to localize properly in the nucleus and to be distributed in nuclear speckles in a manner comparable to that of full-length SF3A1[182]. The domains necessary to interact with other component of the SF3A complex would be untouched and therefore the UBL deletion might not affect assembly of the mature 17S U2 snRNP. This could be achieved by applying another CRISPR/Cas-based approach which makes use of a Cas9 protein fused to deaminase base editors that allows for single nucleotide changes to be introduced at specific sites, as opposed the creation of double stranded breaks [311]. Combining this technique with editing reporters and flow cytometry which have greatly improved the efficiency of generating edited cell-lines [312], a premature

146

stop codon upstream of the UBL domain in the endogenous SF3A1 gene could be introduced to create a cell-line expressing truncated SF3A1. Next-generation sequencing of RNA harvested from this cell-line compared to an un-edited control could be very powerful for precisely identifying SF3A1-UBL dependent splicing events in the human transcriptome.

Although the requirement of UAP56 in prespliceosome assembly is clear, identifying the mechanisms underlying its role in prespliceosome formation in humans has been challenging as this helicase has not been captured by any cryo-EM structures or mass-spectrometry analyses of prespliceosomal complexes to date. The role of UAP56 in human prespliceosome assembly may be advanced by studying its interaction with the U1 and U2 snRNPs. Additionally, to elucidate the contribution and influence of the SL4-SF3A1 interaction towards global splicing in humans, transcriptome-level analysis of SF3A1-dependent splicing events is needed. Studying these interactions will further our understanding of constitutive and alternative splicing regulation at homeostasis, and when dysregulated in disease.

REFERENCES

1. Thieffry, D. and S. Sarkar, *Forty years under the central dogma.* Trends in Biochemical Sciences, 1998. **23**(8): p. 312-316.

2. Crick, F., *Central dogma of molecular biology.* Nature, 1970. **227**(5258): p. 561-563.

3. Monod, J. and F. Jacob, *Teleonomic mechanisms in cellular metabolism, growth, and differentiation.* Cold Spring Harbor Symposia on Quantitative Biology, 1961. **26**(1): p. 389-401.

4. Sharp, P.A., *The discovery of split genes and RNA splicing.* Trends in Biochemical Sciences, 2005. **30**(6): p. 279-281.

5. O'Malley, M.A., et al., *How do microbial populations and communities function as model systems?* The Quarterly Review of Biology, 2015. **90**(3): p. 269-293.

6. Friedmann, H.C., *From Butyribacterium to E. coli : An essay on unity in biochemistry.* Perspectives in Biology and Medicine, 2004. **47**(1): p. 47-66.

7. Berka, A.J., *Discovery of RNA splicing and genes in pieces.* Proceedings of the National Academy of Sciences of the United States of America, 2016. **113**(4): p. 801-805.

8. Thomas, M., R.L. White, and R.W. Davis, *Hybridization of RNA to double-stranded DNA: formation of R-loops.* Proceedings of the National Academy of Sciences, 1976. **73**(7): p. 2294-2298.

9. Chow, L.T., et al., *A map of cytoplasmic RNA transcripts from lytic adenovirus type 2, determined by electron microscopy of RNA:DNA hybrids.* Cell, 1977. **11**(4): p. 819-836.

10. Meyer, J., et al., *Electron microscopy of late adenovirus type 2 mRNA hybridized to double-stranded viral DNA.* Journal of Virology, 1977. **21**(3): p. 101001018.

11. Chow, L.T., et al., *An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA.* Cell, 1977. **12**(1): p. 1-8.

12. Berget, S.M., C. Moore, and P.A. Sharp, *Spliced segments at the 5' terminus of adenovirus 2 late mRNA.* Proceedings of the National Academy of Sciences, 1977. **74**(8): p. 3171-3175.

13. Maniatis, T. and B. Tasic, *Alternative pre-mRNA splicing and proteome expansion in metazoans.* Nature, 2002. **418**(6894): p. 236-243.

14. Graveley, B.R., *Alternative splicing: increasing diversity in the proteomic world.* Trends in Genetics, 2001. **17**(2): p. 100-107.

15. Kim, E., A. Magen, and G. Ast, *Different levels of alternative splicing among eukaryotes.* Nucleic Acids Research, 2007. **35**(1): p. 125-131.

16. Louhichi, A., A. Fourati, and A. Rebaï, *IGD: A resource for intronless genes in the human genome.* Gene, 2011. **488**(1-2): p. 35-40.

17. Grzybowska, E.A., *Human intronless genes: Functional groups, associated diseases, evolution, and mRNA processing in absence of splicing.* Biochemical and Biophysical Research Communications, 2012. **424**(1): p. 1-6.

18. Sakharkar, M.K., V.T.K. Chow, and P. Kangueane, *Distributions of exons and introns in the human genome.* In Silico Biology, 2004. **4**(4): p. 387-393.

19. Koonin, E.V., *The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate?* Biology Direct, 2006. **1**(1): p. 22.

20. Modrek, B. and C.J. Lee, *Alternative splicing in the human, mouse and rat genomes is associated with an increased frequency of exon creation and/or loss.* Nature Genetics, 2003. **34**(2): p. 177-180.

21. Brett, D., et al., *EST comparison indicates 38% of human mRNAs contain possible alternative splice forms.* FEBS Letters, 2000. **474**(1): p. 83-86.

22. Pan, Q., et al., *Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing.* Nature Genetics, 2008. **40**: p. 1413-1415.

23. Wang, E.T., et al., *Alternative isoform regulation in human tissue transcriptomes.* Nature, 2008. **456**(7221): p. 470-476.

24. Black, D.L., *Protein Diversity from Alternative Splicing: A Challenge forBioinformatics and Post-Genome Biology.* Cell, 2000. **103**(3): p. 367-370.

25. Rowen, L., et al., *Analysis of the Human Neurexin Genes: Alternative Splicing and the Generation of Protein Diversity.* Genomics, 2002. **79**(4): p. 587-597.

26. Schad, E., P. Tompa, and H. Hegyi, *The relationship between proteome size, structural disorder and organism complexity.* Genome Biology, 2011. **12**(12): p. 1-13.

27. Hahn, M.W. and G.A. Wray, *The g-value paradox.* Evolution and Development, 2002. **4**(2): p. 73-75.

28.   Lander, E.S., et al., *Initial sequencing and analysis of the human genome.* Nature, 2001. **409**(6822): p. 860-921.

29.   Hatje, K., et al., *The protein-coding human genome: Annotating high-hanging fruits.* BioEssays, 2019. **41**(11): p. 1900066.

30.   Pertea, M. and S.L. Salzberg, *Between a chicken and a grape: estimating the number of human genes.* Genome Biology, 2010. **11**(5): p. 206.

31.   Gan, X., et al., *Multiple reference genomes and transcriptomes for Arabidopsis thaliana.* Nature, 2011. **477**(7365): p. 419-423.

32.   Hillier, L.W., *Genomics in C. elegans: So many genes, such a little worm.* Genome Research, 2005. **15**(12): p. 1651-1660.

33.   Ramani, A.K., et al., *Genome-wide analysis of alternative splicing in Caenorhabditis elegans.* Genome Research, 2011. **21**(2): p. 342-348.

34.   Kurokawa, K., et al., *Brief naturalistic stress induces an alternative splice variant of SMG-1 lacking exon 63 in peripheral leukocytes.* Neuroscience Letters, 2010. **484**(2): p. 128-132.

35.   Zhang, H., C. Lin, and L. Gu, *Light Regulation of Alternative Pre-mRNA Splicing in Plants.* Photochemistry and Photobiology, 2017. **93**(1): p. 159-165.

36.   Weyn-Vanhentenryck, S.M., et al., *Precise temporal regulation of alternative splicing during neural development.* Nature Communications, 2018. **9**(1): p. 1-17.

37.   Bland, C.S., et al., *Global regulation of alternative splicing during myogenic differentiation.* Nucleic Acids Research, 2010. **38**(21): p. 7651-7664.

38.   Castle, J.C., et al., *Expression of 24,426 human alternative splicing events and predicted cis regulation in 48 tissues and cell lines.* Nature Genetics, 2008. **40**(12): p. 1416-1425.

39.   Minovitsky, S., *The splicing regulatory element, UGCAUG, is phylogenetically and spatially conserved in introns that flank tissue-specific alternative exons.* Nucleic Acids Research, 2005. **33**(2): p. 714-724.

40.   Mount, S.M., *A catalogue of splice junction sequences.* Nucleic Acids Research, 1982. **10**(2): p. 459-472.

41.   Lim, L.P. and C.B. Burge, *A computational analysis of sequence features involved in recognition of short introns.* Proceedings of the National Academy of Sciences, 2001. **98**(20): p. 11193-11198.

42. Reed, R., *The organization of 3' splice-site sequences in mammalian introns.* Genes & Development, 1989. **3**(12b): p. 2113-2123.

43. Rogozin, I.B. and L. Milanesi, *Analysis of donor splice sites in different eukaryotic organisms.* Journal of Molecular Evolution, 1997. **45**(1): p. 50-59.

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

45. Sharp, P.A., *On the origin of RNA splicing and introns.* Cell, 1985. **42**(2): p. 397-400.

46. Konarska, M.M., et al., *Characterization of the branch site in lariat RNAs produced by splicing of mRNA precursors.* Nature, 1985. **313**(6003): p. 552-557.

47. Jacquier, A., *Self-splicing group II and nuclear pre-mRNA introns: how similar are they?* Trends in Biochemical Sciences, 1990. **15**(9): p. 351-354.

48. Qin, P.Z. and A.M. Pyle, *The architectural organization and mechanistic function of group II intron structural elements.* Current Opinion in Structural Biology, 1998. **8**(3): p. 301-308.

49. Costa, M., et al., *Multiple tertiary interactions involving domain II of group II self-splicing introns 1 1Edited by M. Yaniv.* Journal of Molecular Biology, 1997. **267**(3): p. 520-536.

50. Irimia, M. and S.W. Roy, *Origin of spliceosomal introns and alternative splicing.* Cold Spring Harbor Perspectives in Biology, 2014. **6**(6): p. a016071.

51. Cech, T.R., *The generality of self-splicing RNA: Relationship to nuclear mRNA splicing.* Cell, 1986. **44**(2): p. 207-210.

52. Suchy, M. and C. Schmelzer, *Restoration of the self-splicing activity of a defective group II intron by a small trans-acting RNA.* Journal of Molecular Biology, 1991. **222**(2): p. 179-187.

53. Roger, A.J. and W.F. Doolittle, *Why introns-in-pieces?* Nature, 1993. **364**(6435): p. 289-290.

54. Kim, V.N., *MicroRNA biogenesis: coordinated cropping and dicing.* Nature Reviews Molecular Cell Biology, 2005. **6**(5): p. 376-385.

55. Weeks, K.M. and T.R. Cech, *Assembly of a ribonucleoprotein catalyst by tertiary structure capture.* Science, 1996. **271**(5247): p. 345-348.

56. Piñol-Roma, S. and G. Dreyfuss, *Shuttling of pre-mRNA binding proteins between nucleus and cytoplasm.* Nature, 1992. **355**(6362): p. 730-732.

57. Nissen, P., *The structural basis of ribosome activity in peptide bond synthesis.* Science, 2000. **289**(5481): p. 920-930.

58. Collins, L.J., et al., *The Modern RNP World of Eukaryotes.* Journal of Heredity, 2009. **100**(5): p. 597-604.

59. Dreyfuss, G., V.N. Kim, and N. Kataoka, *Messenger-RNA-binding proteins and the messages they carry.* Nature Reviews Molecular Cell Biology, 2002. **3**(3): p. 195-205.

60. Sheth, N., et al., *Comprehensive splice-site analysis using comparative genomics.* Nucleic Acids Res, 2006. **34**(14): p. 3955-67.

61. Yu, Y.-T., et al., *The growing world of small nuclear ribonucleoproteins.* The RNA World, 1999. **2**(1): p. 487-524.

62. Moore, M.J., C.C. Query, and P.A. Sharp, *Splicing of precursors to mRNA by the spliceosome.* The RNA World, 1999. **2**(1): p. 525-560.

63. Liu, S., *The network of protein-protein interactions within the human U4/U6.U5 tri-snRNP.* RNA, 2006. **12**(7): p. 1418-1430.

64. Migliorini, P., et al., *Anti-Sm and anti-RNP antibodies.* Autoimmunity, 2005. **38**(1): p. 47-54.

65. Reeves, W.H., et al., *Autoimmune sera reactive with Sm antigen contain high levels of RNP-like antibodies.* Journal of Clinical Investigation, 1985. **75**(2): p. 580-587.

66. Achsel, T., et al., *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, 1999. **18**(20): p. 5789-5802.

67. Wahl, M.C., C.L. Will, and R. Lührmann, *The Spliceosome: Design Principles of a Dynamic RNP Machine.* Cell, 2009. **136**(4): p. 701-718.

68. Staley, J.P. and C. Guthrie, *Mechanical devices of the spliceosome: motors, clocks, springs, and things.* Cell, 1998. **92**: p. 315-326.

69. Das, R. and R. Reed, *Resolution of the mammalian E complex and the ATP-dependent spliceosomal complexes on native agarose mini-gels.* RNA, 1999. **5**(11): p. 1504-1508.

70.   Jurica, M.S. and M.J. Moore, *Capturing splicing complexes to study structure and mechanism.* Methods, 2002. **28**(3): p. 336-345.

71.   Agafonov, D.E., et al., *Semiquantitative Proteomic Analysis of the Human Spliceosome via a Novel Two-Dimensional Gel Electrophoresis Method.* Molecular and Cellular Biology, 2011. **31**(13): p. 2667-2682.

72.   Cvitkovic, I. and M.S. Jurica, *Spliceosome Database: a tool for tracking components of the spliceosome.* Nucleic Acids Research, 2013. **41**(D1): p. D132-D141.

73.   Yan, C., R. Wan, and Y. Shi, *Molecular mechanisms of pre-mRNA splicing through structural biology of the spliceosome.* Cold Spring Harbor Perspectives in Biology, 2019. **11**(1): p. a032409.

74.   Wilkinson, M.E., C. Charenton, and K. Nagai, *RNA splicing by the spliceosome.* Annual Review of Biochemistry, 2020. **89**(1): p. 359-388.

75.   Cate, J.H.D., *A big bang in spliceosome structural biology.* Science, 2016. **351**(6280): p. 1390-1392.

76.   Michaud, S. and R. Reed, *A functional association between the 5' and 3' splice site is established in the earliest prespliceosome complex (E) in mammals.* Genes & Development, 1993. **7**(6): p. 1008-1020.

77.   Kondo, Y., et al., *Crystal structure of human U1 snRNP, a small nuclear ribonucleoprotein particle, reveals the mechanism of 5' splice site recognition.* eLife, 2015. **4**: p. 1-19.

78.   Zhuang, Y. and A.M. Weiner, *A compensatory base change in U1 snRNA suppresses a 5' splice site mutation.* Cell, 1986. **46**(6): p. 827-835.

79.   Shen, H., et al., *Distinct activities of the DExD/H-box splicing factor hUAP56 facilitate stepwise assembly of the spliceosome.* Genes and Development, 2008. **22**(13): p. 1796-1803.

80.   Liang, W.W. and S.C. Cheng, *A novel mechanism for Prp5 function in prespliceosome formation and proofreading the branch site sequence.* Genes and Development, 2015. **29**: p. 81-93.

81.   Zhuang, Y. and A.M. Weiner, *A compensatory base change in human U2 snRNA can suppress a branch site mutation.* Genes & Development, 1989. **3**(10): p. 1545-1552.

82.   Gozani, O., R. Feld, and R. Reed, *Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for*

assembly of spliceosomal complex A. Genes and Development, 1996. **10**(2): p. 233-243.

83. Valcárcel, J., et al., *Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA.* Science, 1996. **273**(5282): p. 1706-1709.

84. Dönmez, G., et al., *The 5' end of U2 snRNA is in close proximity to U1 and functional sites of the pre-mRNA in early spliceosomal complexes.* Molecular Cell, 2007. **25**(3): p. 399-411.

85. Xu, Y.Z., et al., *Prp5 bridges U1 and U2 snRNPs and enables stable U2 snRNP association with intron RNA.* EMBO Journal, 2004. **23**(2): p. 376-385.

86. Plaschka, C., et al., *Prespliceosome structure provides insights into spliceosome assembly and regulation.* Nature, 2018. **559**(7714): p. 419-422.

87. Boesler, C., et al., *A spliceosome intermediate with loosely associated tri-snRNP accumulates in the absence of Prp28 ATPase activity.* Nature Communications, 2016. **7**(1): p. 11997.

88. Zhan, X., et al., *Structures of the human pre-catalytic spliceosome and Its precursor spliceosome.* Cell Research, 2018. **28**(12): p. 1129-1140.

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

90. Charenton, C., M.E. Wilkinson, and K. Nagai, *Mechanism of 5' splice site transfer for human spliceosome activation.* Science, 2019. **364**(6438): p. 362-367.

91. Plaschka, C., A.J. Newman, and K. Nagai, *Structural basis of nuclear pre-mRNA splicing: Lessons from Yeast.* Cold Spring Harbor Perspectives in Biology, 2019. **11**(5): p. a032391.

92. Wassarman, D. and J. Steitz, *Interactions of small nuclear RNA's with precursor messenger RNA during in vitro splicing.* Science, 1992. **257**(5078): p. 1918-1925.

93. Raghunathan, P.L. and C. Guthrie, *RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2.* Current Biology, 1998. **8**(15): p. 847-855.

94. Zhang, X., et al., *Structure of the human activated spliceosome in three conformational states.* Cell Research, 2018. **28**(3): p. 307-322.

95. Huppler, A., et al., *Metal binding and base ionization in the U6 RNA intramolecular stem-loop structure.* Nature Structural Biology, 2002. **9**(6): p. 431-435.

96.   Haselbach, D., et al., *Structure and conformational dynamics of the human spliceosomal Bact complex.* Cell, 2018. **172**(3): p. 454-464.

97.   Kim, S.H. and R.J. Lin, *Spliceosome activation by PRP2 ATPase prior to the first transesterification reaction of pre-mRNA splicing.* Molecular and Cellular Biology, 1996. **16**(12): p. 6810-6819.

98.   Lardelli, R.M., et al., *Release of SF3 from the intron branchpoint activates the first step of pre-mRNA splicing.* RNA, 2010. **16**(3): p. 516-528.

99.   Liu, H.L. and S.C. Cheng, *The Interaction of Prp2 with a Defined Region of the Intron Is Required for the First Splicing Reaction.* Molecular and Cellular Biology, 2012. **32**(24): p. 5056-5066.

100.   Yean, S.-L., et al., *Metal-ion coordination by U6 small nuclear RNA contributes to catalysis in the spliceosome.* Nature, 2000. **408**(6814): p. 881-884.

101.   Lee, C., et al., *Metal binding and substrate positioning by evolutionarily invariant U6 sequences in catalytically active protein-free snRNAs.* RNA, 2010. **16**(11): p. 2226-38.

102.   O'Keefe, R.T., C. Norman, and A.J. Newman, *The invariant U5 snRNA loop 1 sequence Is dispensable for the first catalytic step of pre-mRNA splicing in yeast.* Cell, 1996. **86**(4): p. 679-689.

103.   Zhan, X., et al., *Structure of a human catalytic step I spliceosome.* Science, 2018. **359**(6375): p. 537-545.

104.   Sontheimer, E. and J. Steitz, *The U5 and U6 small nuclear RNAs as active site components of the spliceosome.* Science, 1993. **262**(5142): p. 1989-1996.

105.   Schwer, B. and C. Guthrie, *A conformational rearrangement in the spliceosome is dependent on PRP16 and ATP hydrolysis.* The EMBO Journal, 1992. **11**(13): p. 5033-5039.

106.   Fica, S.M., et al., *A human postcatalytic spliceosome structure reveals essential roles of metazoan factors for exon ligation.* Science, 2019. **363**(6428): p. 710-714.

107.   Company, M., J. Arenas, and J. Abelson, *Requirement of the RNA helicase-like protein PRP22 for release of messenger RNA from spliceosomes.* Nature, 1991. **349**(6309): p. 487-493.

108.   Arenas, J.E. and J.N. Abelson, *Prp43: An RNA helicase-like factor involved in spliceosome disassembly.* Proceedings of the National Academy of Sciences, 1997. **94**(22): p. 11798-11802.

109. Tsai, R.T., *Spliceosome disassembly catalyzed by Prp43 and its associated components Ntr1 and Ntr2.* Genes & Development, 2005. **19**(24): p. 2991-3003.

110. Fourmann, J.B., et al., *Dissection of the factor requirements for spliceosome disassembly and the elucidation of its dissociation products using a purified splicing system.* Genes & Development, 2013. **27**(4): p. 413-428.

111. Lim, S.R. and K.J. Hertel, *Commitment to splice site pairing coincides with a complex formation.* Molecular Cell, 2004. **15**(3): p. 477-483.

112. Kent, O.A. and A.M. MacMillan, *Early organization of pre-mRNA during spliceosome assembly.* Nature Structural Biology, 2002. **9**(8): p. 576-581.

113. Kotlajich, M.V., T.L. Crabb, and K.J. Hertel, *Spliceosome assembly pathways for different types of alternative splicing converge during commitment to splice site pairing in the A complex.* Molecular and cellular biology, 2009. **29**(4): p. 1072-82.

114. Ule, J. and B.J. Blencowe, *Alternative splicing regulatory networks: functions, mechanisms, and evolution.* Molecular Cell, 2019. **76**(2): p. 329-345.

115. Kotlajich, M.V. and K.J. Hertel, *Death by splicing: tumor suppressor RBM5 freezes splice-site pairing.* Molecular Cell, 2008. **32**(2): p. 162-164.

116. Schaal, T.D. and T. Maniatis, *Selection and characterization of pre-mRNA splicing enhancers: identification of novel SR protein-specific enhancer sequences.* Molecular and Cellular Biology, 1999. **19**(3): p. 1705-1719.

117. Zhu, J., *Pre-mRNA splicing in the absence of an SR protein RS domain.* Genes & Development, 2000. **14**(24): p. 3166-3178.

118. Zhou, Z. and X.-D. Fu, *Regulation of splicing by SR proteins and SR protein-specific kinases.* Chromosoma, 2013. **122**(3): p. 191-207.

119. Macdougall, C., D. Harbison, and M. Bownes, *The developmental consequences of alternate splicing in sex determination and differentiation in Drosophila.* Developmental Biology, 1995. **172**(2): p. 353-376.

120. Spingola, M. and M. Ares, *A yeast intronic splicing enhancer and Nam8p are required for Mer1p-activated splicing.* Molecular Cell, 2000. **6**(2): p. 329-338.

121. Lynch, K.W. and T. Maniatis, *Assembly of specific SR protein complexes on distinct regulatory elements of the Drosophila doublesex splicing enhancer.* Genes & Development, 1996. **10**(16): p. 2089-2101.

122. Santoro, B., et al., *Interactive cloning with the SH3 domain of N-src identifies a new brain specific ion channel protein, with homology to Eag and cyclic*

nucleotide-gated channels. Proceedings of the National Academy of Sciences, 1997. **94**(26): p. 14815-14820.

123.   Modafferi, E.F. and D.L. Black, *A complex intronic splicing enhancer from the c-src pre-mRNA activates inclusion of a heterologous exon.* Molecular and cellular biology, 1997. **17**(11): p. 6537-45.

124.   Chan, R.C. and D.L. Black, *The polypyrimidine tract binding protein binds upstream of neural cell-specific c-src exon N1 to repress the splicing of the intron downstream.* Molecular and Cellular Biology, 1997. **17**(8): p. 4667-4676.

125.   Sharma, S., A.M. Falick, and D.L. Black, *Polypyrimidine tract binding protein blocks the 5' splice site-dependent assembly of U2AF and the prespliceosomal e complex.* Molecular Cell, 2005. **19**(4): p. 485-496.

126.   Sharma, S., et al., *Polypyrimidine tract binding protein controls the transition from exon definition to an intron defined spliceosome.* Nature structural & molecular biology, 2008. **15**(2): p. 183-191.

127.   Bonnal, S., et al., *RBM5/Luca-15/H37 regulates fas alternative splice site pairing after exon definition.* Molecular Cell, 2008. **32**(1): p. 81-95.

128.   House, A.E. and K.W. Lynch, *An exonic splicing silencer represses spliceosome assembly after ATP-dependent exon recognition.* Nature Structural & Molecular Biology, 2006. **13**(10): p. 937-944.

129.   Krawczak, M., J. Reiss, and D. Cooper, *The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: Causes and consequences.* Human genetics., 1992. **90**(1-2): p. 41.

130.   Ward, A.J. and T.A. Cooper, *The pathobiology of splicing.* The Journal of Pathology, 2010. **220**(2): p. 152-163.

131.   Kountouris, P., et al., *IthaGenes: an interactive database for haemoglobin variations and epidemiology.* PLoS ONE, 2014. **9**(7): p. e103020.

132.   Cheng, T.C., et al., *beta-Thalassemia in Chinese: use of in vivo RNA analysis and oligonucleotide hybridization in systematic characterization of molecular defects.* Proceedings of the National Academy of Sciences, 1984. **81**(9): p. 2821-2825.

133.   Spritz, R.A., et al., *Base substitution in an intervening sequence of a beta+-thalassemic human globin gene.* Proceedings of the National Academy of Sciences, 1981. **78**(4): p. 2455-2459.

134.   Xu, S., et al., *Editing aberrant splice sites efficiently restores β-globin expression in β-thalassemia.* Blood, 2019. **133**(21): p. 2255-2262.

135.   Popplewell, L.L. and S.J. Forman, *Is there an upper age limit for bone marrow transplantation?* Bone Marrow Transplantation, 2002. **29**(4): p. 277-284.

136.   Ma, X., *Epidemiology of myelodysplastic syndromes.* American Journal of Medicine, 2012. **125**(7): p. S2-S5.

137.   Bejar, R. and D.P. Steensma, *Recent developments in myelodysplastic syndromes.* Blood, 2014. **124**(18): p. 2793-2803.

138.   Bejar, R., R. Levine, and B.L. Ebert, *Unraveling the molecular pathophysiology of myelodysplastic syndromes.* Journal of Clinical Oncology, 2011. **29**(5): p. 504-515.

139.   Papaemmanuil, E., et al., *Clinical and biological implications of driver mutations in myelodysplastic syndromes.* Blood, 2013. **122**(22): p. 3616-3627.

140.   Haferlach, T., et al., *Landscape of genetic lesions in 944 patients with myelodysplastic syndromes.* Leukemia, 2014. **28**(2): p. 241-247.

141.   Yoshida, K., et al., *Frequent pathway mutations of splicing machinery in myelodysplasia.* Nature, 2011. **478**: p. 64-9.

142.   Maguire, S.L., et al., *SF3B1 mutations constitute a novel therapeutic target in breast cancer.* Journal of Pathology, 2015. **235**(4): p. 571-580.

143.   Ciriello, G., et al., *Comprehensive molecular portraits of invasive lobular breast cancer.* Cell, 2015. **163**(2): p. 506-519.

144.   Bailey, P., et al., *Genomic analyses identify molecular subtypes of pancreatic cancer.* Nature, 2016. **531**(7592): p. 47-52.

145.   Wang, L., et al., *SF3B1 and other novel cancer genes in chronic lymphocytic leukemia.* New England Journal of Medicine, 2011. **365**(26): p. 2497-2506.

146.   Lee, S.C.-W., et al., *Synthetic lethal and convergent biological effects of cancer-associated spliceosomal gene mutations.* Cancer Cell, 2018. **34**(2): p. 225-241.

147.   Malcovati, L., et al., *SF3B1 mutation identifies a distinct subset of myelodysplastic syndrome with ring sideroblasts.* Blood, 2015. **126**(2): p. 233-241.

148.   Nagai, K., et al., *Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A.* Nature, 1990. **348**(6301): p. 515-520.

149.   Maris, C., C. Dominguez, and F.H.T. Allain, *The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression.* FEBS Journal, 2005. **272**(9): p. 2118-2131.

150.   Oubridge, C., et al., *Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin.* Nature, 1994. **372**(6505): p. 432-438.

151.   Stark, H., et al., *Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle.* Nature, 2001. **409**(6819): p. 539-542.

152.   Krummel, D.A.P., et al., *Crystal structure of human spliceosomal U1 snRNP at 5.5 Å resolution.* Nature, 2009. **458**(7237): p. 475-480.

153.   Nelissen, R.L.H., et al., *The association of the U1-specific 70K and C proteins with U1 snRNPs is mediated in part by common U snRNP proteins.* EMBO Journal, 1994. **13**(17): p. 4113-4125.

154.   Heinrichs, V., et al., *U1-specific protein C needed for efficient complex formation of U1 snRNP with a 5' splice site.* Science, 1990. **247**(4938): p. 69-72.

155.   Han, J., et al., *Pre-mRNA splicing: where and when in the nucleus.* Trends in Cell Biology, 2011. **21**(6): p. 336-343.

156.   Hicks, M.J., et al., *Linking splicing to pol II transcription stabilizes pre-mRNAs and influences splicing patterns.* PLoS Biology, 2006. **4**(6): p. e147.

157.   Das, R., et al., *Functional coupling of RNAP II transcription to spliceosome assembly.* Genes & Development, 2006. **20**(9): p. 1100-1109.

158.   Das, R., et al., *SR proteins function in coupling RNAP II transcription to pre-mRNA splicing.* Molecular Cell, 2007. **26**(6): p. 867-881.

159.   Zhang, S., et al., *Structure of a transcribing RNA polymerase II–U1 snRNP complex.* Science, 2021. **371**(6526): p. 305-309.

160.   Di, C., et al., *U1 snRNP telescripting roles in transcription and its mechanism.* Cold Spring Harbor Symposia on Quantitative Biology, 2019. **84**(1): p. 115-122.

161.   Kaida, D., et al., *U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation.* Nature, 2010. **468**(7324): p. 664-668.

162.   Michael, et al., *U1 snRNP determines mRNA length and regulates isoform expression.* Cell, 2012. **150**(1): p. 53-64.

163.   Gunderson, S.I., M. Polycarpou-Schwarz, and I.W. Mattaj, *U1 snRNP inhibits pre-mRNA polyadenylation through a direct interaction between U1-70K and poly(A) polymerase.* Molecular Cell, 1998. **1**(2): p. 255-264.

164. So, B.R., et al., *A complex of U1 snRNP with cleavage and polyadenylation factors controls telescripting, regulating mRNA transcription in human cells.* Molecular Cell, 2019. **76**(4): p. 590-599.

165. Jutzi, D., et al., *Aberrant interaction of FUS with the U1 snRNA provides a molecular mechanism of FUS induced amyotrophic lateral sclerosis.* Nature Communications, 2020. **11**(1): p. 1-14.

166. Yu, Y. and R. Reed, *FUS functions in coupling transcription to splicing by mediating an interaction between RNAP II and U1 snRNP.* Proceedings of the National Academy of Sciences. **112**(28): p. 8608-8613.

167. Sharma, S., et al., *U1 snRNA directly interacts with polypyrimidine tract-binding protein during splicing repression.* Molecular cell, 2011. **41**(5): p. 579-88.

168. Sharma, S., et al., *Stem–loop 4 of U1 snRNA is essential for splicing and interacts with the U2 snRNP-specific SF3A1 protein during spliceosome assembly.* Genes and Development, 2014. **28**(22): p. 2518-2531.

169. Côté, J., et al., *The U1 small nuclear ribonucleoprotein/5' splice site interaction affects U2AF65 binding to the downstream 3' Splice Site.* Journal of Biological Chemistry, 1995. **270**(8): p. 4031-4036.

170. Will, C.L. and R. Lührmann, *Spliceosome structure and function.* Cold Spring Harbor perspectives in biology, 2011. **3**(7): p. a003707.

171. Bertram, K., et al., *Cryo-EM Structure of a pre-catalytic human spliceosome primed for activation.* Cell, 2017. **170**(4): p. 701-713.

172. Plaschka, C., P.C. Lin, and K. Nagai, *Structure of a pre-catalytic spliceosome.* Nature, 2017. **546**(7660): p. 617-621.

173. Fica, S.M., et al., *Structure of a spliceosome remodelled for exon ligation.* Nature, 2017. **542**(7641): p. 377-380.

174. Galej, W.P., et al., *Cryo-EM structure of the spliceosome immediately after branching.* Nature, 2016. **537**(7619): p. 197-201.

175. Das, R., Z. Zhou, and R. Reed, *Functional association of U2 snRNP with the ATP-independent splicecosomal complex E.* Molecular Cell, 2000. **5**(5): p. 779-787.

176. Gozani, O., J. Potashkin, and R. Reed, *A potential role for U2AF-SAP 155 interactions in recruiting U2 snRNP to the branch site.* Molecular and cellular biology, 1998. **18**(8): p. 4752-60.

177.   Li, X., et al., *CryoEM structure of saccharomyces cerevisiae U1 snRNP offers insight into alternative splicing.* Nature Communications, 2017. **8**(1): p. 1-13.

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

179.   Brosi, R., et al., *Interaction of mammalian splicing factor SF3a with U2 snRNP and relation of its 60-kD subunit to yeast PRP9.* Science, 2006. **262**(5130): p. 102-105.

180.   Crisci, A., et al., *Mammalian splicing factor SF1 interacts with SURP domains of U2 snRNP-associated proteins.* Nucleic Acids Research, 2015. **43**(21): p. 10456-10473.

181.   Tanackovic, G. and A. Kramer, *Human splicing factor SF3a, but not SF1, is essential for pre-mRNA splicing in vivo.* Mol Biol Cell, 2005. **16**(3): p. 1366-77.

182.   Huang, C.J., et al., *Interaction domains and nuclear targeting signals in subunits of the U2 small nuclear ribonucleoprotein particle-associated splicing factor SF3a.* Journal of Biological Chemistry, 2011. **286**(15): p. 13106-13114.

183.   Vijay-kumar, S., C.E. Bugg, and W.J. Cook, *Structure of ubiquitin refined at 1.8 Å resolution.* Journal of Molecular Biology, 1987. **194**(3): p. 531-544.

184.   A Krämer, F.M.C.W.K.G. and G. Bilbe, *Mammalian splicing factor SF3a120 represents a new member of the SURP family of proteins and is homologous to the essential splicing factor PRP21p of Saccharomyces cerevisiae.* 1995. **1**(1): p. 260-272.

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

186.   Ryder, S.P., M.I. Recht, and J.R. Williamson, *Quantitative analysis of protein-RNA interactions by gel mobility shift.* RNA-Protein Interaction Protocols, 2008. **488**(1): p. 99-115.

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

188.   Masuda, S., et al., *Recruitment of the human TREX complex to mRNA during splicing.* Genes and Development, 2005. **19**(13): p. 1512-1517.

189.   Chang, C.k., et al., *The N-terminus of TDP-43 promotes its oligomerization and enhances DNA binding affinity.* Biochemical and Biophysical Research Communications, 2012. **425**(2): p. 219-224.

190. Afroz, T., et al., *Functional and dynamic polymerization of the ALS-linked protein TDP-43 antagonizes its pathologic aggregation.* Nature Communications, 2017. **8**(1): p. 1-15.

191. Smet-Nocca, C., et al., *SUMO-1 regulates the conformational dynamics of Thymine-DNA Glycosylase regulatory domain and competes with its DNA binding activity.* BMC Biochemistry, 2011. **12**(1): p. 1-15.

192. Zhou, B., et al., *Microtubule-associated protein 1 light chain 3 is a fibronectin mRNA- binding protein linked to mrna translation in lamb vascular smooth muscle cells.* Journal of Clinical Investigation, 1997. **100**(12): p. 3070-3082.

193. Ying, L., et al., *LC3-mediated fibronectin mRNA translation induces fibrosarcoma growth by increasing connective tissue growth factor.* Journal of Cell Science, 2009. **122**(9): p. 1441-1451.

194. Kirschner, L.S. and C.A. Stratakis, *Structure of the human ubiquitin fusion gene Uba80 (RPS27a) and one of its pseudogenes.* Biochemical and Biophysical Research Communications, 2000. **270**(3): p. 1106-1110.

195. Finley, D., B. Bartel, and A. Varshavsky, *The tails of ubiquitin precursors are ribosomal proteins whose fusion to ubiquitin facilitates ribosome biogenesis.* Nature, 1989. **338**(6214): p. 394-401.

196. Cindy L. Will, et al., *The human 18S U11/U12 snRNP contains a set of novel proteins not found in the U2-dependent spliceosome.* RNA, 2004. **10**(6): p. 929-941.

197. Qin, H., et al., *TDP-43 N terminus encodes a novel ubiquitin-like fold and its unfolded form in equilibrium that can be shifted by binding to ssDNA.* Proceedings of the National Academy of Sciences, 2014. **111**(52): p. 18619-18624.

198. Eilebrecht, S., et al., *SUMO-1 possesses DNA binding activity.* BMC Research Notes, 2010. **3**(1): p. 1-7.

199. Kranz, J.K., J. Lu, and K.B. Hall, *Contribution of the tyrosines to the structure and function of the human U1A N-terminal RNA binding domain.* Protein Science, 1996. **5**(8): p. 1567-1583.

200. Auweter, S.D., F.C. Oberstrass, and F.H.T. Allain, *Solving the structure of PTB in complex with pyrimidine tracts: an NMR study of protein-RNA complexes of weak affinities.* Journal of Molecular Biology, 2007. **367**(1): p. 174-186.

201. Katsamba, P.S., D.G. Myszka, and I.A. Laird-Offringa, *Two functionally distinct steps mediate high affinity binding of U1A protein to U1 hairpin II RNA.* Journal of Biological Chemistry, 2001. **276**(24): p. 21476-21481.

202.   Kim, E., et al., *SRSF2 mutations contribute to myelodysplasia by mutant-specific effects on exon recognition.* Cancer Cell, 2015. **27**(5): p. 617-630.

203.   Zhang, J., et al., *Disease-associated mutation in SRSF2 misregulates splicing by altering RNA-binding affinities.* Proceedings of the National Academy of Sciences, 2015. **112**(34): p. E4726-E4734.

204.   Okeyo-Owuor, T., et al., *U2AF1 mutations alter sequence specificity of pre-mRNA binding and splicing.* Leukemia, 2015. **29**(4): p. 909-917.

205.   Fei, D.L., et al., *Wild-type U2AF1 antagonizes the splicing program characteristic of U2AF1-mutant tumors and is required for cell survival.* PLoS Genetics, 2016. **12**(10): p. e1006384.

206.   Loughlin, F.E., et al., *The solution structure of FUS bound to RNA reveals a bipartite mode of RNA recognition with both sequence and shape specificity.* Molecular Cell, 2019. **73**(3): p. 490-504.

207.   Vasilyev, N., et al., *Crystal structure reveals specific recognition of a G-quadruplex RNA by a β-turn in the RGG motif of FMRP.* Proceedings of the National Academy of Sciences, 2015. **112**(39): p. E5391-E5400.

208.   Corbin-Lickfett, K.A., et al., *The HSV-1 ICP27 RGG box specifically binds flexible, GC-rich sequences but not G-quartet structures.* Nucleic Acids Research, 2009. **37**(21): p. 7290-7301.

209.   Ozdilek, B.A., et al., *Intrinsically disordered RGG/RG domains mediate degenerate specificity in RNA binding.* Nucleic Acids Research, 2017. **45**(13): p. 7984-7996.

210.   Chong, P.A., R.M. Vernon, and J.D. Forman-Kay, *RGG/RG motif regions in RNA binding and phase separation.* Journal of Molecular Biology, 2018. **430**(23): p. 4650-4665.

211.   Nesic, D. and A. Kramer, *Domains in human splicing factors SF3a60 and SF3a66 required for binding to SF3a120, assembly of the 17S U2 snRNP, and prespliceosome formation.* Molecular and Cellular Biology, 2001. **21**(19): p. 6406-6417.

212.   Bessonov, S., et al., *Characterization of purified human Bact spliceosomal complexes reveals compositional and morphological changes during spliceosome activation and first step catalysis.* RNA, 2010. **16**(12): p. 2384-2403.

213.   Dowdle, M.E., et al., *Horizontal gel electrophoresis for enhanced detection of protein-RNA complexes.* Journal of visualized experiments : JoVE, 2017(125).

214. Katsamba, P.S., S. Park, and I.a. Laird-Offringa, *Kinetic studies of RNA-protein interactions using surface plasmon resonance.* Methods (San Diego, Calif.), 2002. **26**(2): p. 95-104.

215. Zhou, J., et al., *iUUCD 2.0: An update with rich annotations for ubiquitin and ubiquitin-like conjugations.* Nucleic Acids Research, 2018. **46**(D1): p. D447-D453.

216. Mitchell, A.L., et al., *InterPro in 2019: Improving coverage, classification and access to protein sequence annotations.* Nucleic Acids Research, 2019. **47**(D1): p. D351-D360.

217. Thompson, J.D., D.G. Higgins, and T.J. Gibson, *CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.* Nucleic Acids Research, 1994. **22**(22): p. 4673-4680.

218. Schrödinger, L., *The PyMOL Molecular Graphics System, Version 2.3.2.* 2015.

219. Pisal, R.V., et al., *Detection of Mycoplasma Contamination Directly from Culture Supernatant Using Polymerase Chain Reaction.* Folia Biol (Praha), 2016. **62**(5): p. 203-206.

220. Perriman, R. and M. Ares, Jr., *Invariant U2 snRNA nucleotides form a stem loop to recognize the intron early in splicing.* Mol Cell, 2010. **38**(3): p. 416-27.

221. Xu, Y.Z. and C.C. Query, *Competition between the ATPase Prp5 and branch region-U2 snRNA pairing modulates the fidelity of spliceosome assembly.* Mol Cell, 2007. **28**(5): p. 838-49.

222. Yan, C., et al., *Structure of a yeast step II catalytically activated spliceosome.* Science, 2017. **355**(6321): p. 149-155.

223. Weber, G., et al., *Functional organization of the Sm core in the crystal structure of human U1 snRNP.* EMBO J, 2010. **29**(24): p. 4172-84.

224. Kretzner, L., B.C. Rymond, and M. Rosbash, *S. cerevisiae U1 RNA is large and has limited primary sequence homology to metazoan U1 snRNA.* Cell, 1987. **50**(4): p. 593-602.

225. Siliciano, P.G., W.J. Kivens, and C. Guthrie, *More than half of yeast U1 snRNA is dispensable for growth.* Nucleic Acids Res, 1991. **19**(23): p. 6367-72.

226. Carlson, S.M., et al., *RBM25 is a global splicing factor promoting inclusion of alternatively spliced exons and is itself regulated by lysine mono-methylation.* J Biol Chem, 2017. **292**(32): p. 13381-13390.

227. Izquierdo, J.M., et al., *Regulation of Fas alternative splicing by antagonistic effects of TIA-1 and PTB on exon definition.* Mol Cell, 2005. **19**(4): p. 475-84.

228. Becerra, S., et al., *Prp40 pre-mRNA processing factor 40 homolog B (PRPF40B) associates with SF1 and U2AF65 and modulates alternative pre-mRNA splicing in vivo.* RNA, 2015. **21**(3): p. 438-57.

229. Li, X., et al., *A unified mechanism for intron and exon definition and back-splicing.* Nature, 2019. **573**(7774): p. 375-380.

230. Abovich, N. and M. Rosbash, *Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals.* Cell, 1997. **89**(3): p. 403-412.

231. Bai, R., et al., *Structures of the fully assembled Saccharomyces cerevisiae spliceosome before activation.* Science, 2018. **360**(6396): p. 1423-1429

232. Martelly, W., et al., *Identification of a noncanonical RNA binding domain in the U2 snRNP protein SF3A1.* RNA, 2019. **25**(11): p. 1509-1521.

233. Florens, L., et al., *Analyzing chromatin remodeling complexes using shotgun proteomics and normalized spectral abundance factors.* Methods, 2006. **40**(4): p. 303-11.

234. Fleckner, J., et al., *U2AF65 recruits a novel human DEAD box protein required for the U2 snRNP-branchpoint interaction.* Genes Dev, 1997. **11**(14): p. 1864-72.

235. Kistler, A.L. and C. Guthrie, *Deletion of MUD2, the yeast homolog of U2AF65, can bypass the requirement for Sub2, an essential spliceosomal ATPase.* Genes and Development, 2001. **15**(1): p. 42-49.

236. Fairman-Williams, M.E., U.P. Guenther, and E. Jankowsky, *SF1 and SF2 helicases: family matters.* Curr Opin Struct Biol, 2010. **20**(3): p. 313-24.

237. Diges, C.M. and O.C. Uhlenbeck, *Escherichia coli DbpA is an RNA helicase that requires hairpin 92 of 23S rRNA.* EMBO J, 2001. **20**(19): p. 5503-12.

238. Kossen, K., F.V. Karginov, and O.C. Uhlenbeck, *The carboxy-terminal domain of the DExD/H protein YxiN is sufficient to confer specificity for 23S rRNA.* Journal of Molecular Biology, 2002. **324**(4): p. 625-636.

239. Choudhury, P., et al., *The DExD box ATPase DDX55 is recruited to domain IV of the 28S ribosomal RNA by its C-terminal region.* RNA Biology, 2020: p. 1-12.

240. Wu, G., et al., *Pseudouridines in U2 snRNA stimulate the ATPase activity of Prp5 during spliceosome assembly.* EMBO J, 2016. **35**(6): p. 654-67.

241. O'Day, C.L., G. Dalbadie-McFarland, and J. Abelson, *The Saccharomyces cerevisiae Prp5 protein has RNA-dependent ATPase activity with specificity for U2 small nuclear RNA.* J Biol Chem, 1996. **271**(52): p. 33261-7.

242. Hamid, F.M. and E.V. Makeyev, *A mechanism underlying position-specific regulation of alternative splicing.* Nucleic Acids Res, 2017. **45**(21): p. 12455-12468.

243. Chiou, N.T., G. Shankarling, and K.W. Lynch, *hnRNP L and hnRNP A1 induce extended U1 snRNA interactions with an exon to repress spliceosome assembly.* Mol Cell, 2013. **49**(5): p. 972-82.

244. Dias, A.P., et al., *A role for TREX components in the release of spliced mRNA from nuclear speckle domains.* Nat Commun, 2010. **1**(1): p. 1-10.

245. Chan, R.C. and D.L. Black, *Conserved intron elements repress splicing of a neuron-specific c-src exon in vitro.* Mol Cell Biol, 1995. **15**(11): p. 6377-85.

246. Dignam, J.D., *Preparation of extracts from higher eukaryotes.* Methods Enzymol, 1990. **182**: p. 194-203.

247. Kaiser, P. and J. Wohlschlegel, *Identification of ubiquitination sites and determination of ubiquitin-chain architectures by mass spectrometry.* Methods Enzymol, 2005. **399**: p. 266-77.

248. Kelstrup, C.D., et al., *Optimized fast and sensitive acquisition methods for shotgun proteomics on a quadrupole orbitrap mass spectrometer.* J Proteome Res, 2012. **11**(6): p. 3487-97.

249. Xu, T., et al., *ProLuCID, a fast and sensitive tandem mass spectra-based protein identification program.* Molecular and Cellular Proteomics, 2006. **5**: p. S174-S174.

250. Tabb, D.L., W.H. McDonald, and J.R. Yates, 3rd, *DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics.* J Proteome Res, 2002. **1**(1): p. 21-6.

251. Elias, J.E. and S.P. Gygi, *Target-decoy search strategy for mass spectrometry-based proteomics.* Methods Mol Biol, 2010. **604**(1): p. 55-71.

252. Blencowe, B.J. and A.I. Lamond, *Purification and depletion of RNP particles by antisense affinity chromatography.* Methods Mol Biol, 1999. **118**(1): p. 275-87.

253. Shen, J., L. Zhang, and R. Zhao, *Biochemical characterization of the ATPase and helicase activity of UAP56, an essential pre-mRNA splicing and mRNA export factor.* Journal of Biological Chemistry, 2007. **282**(31): p. 22544-22550.

254. Yong, J., et al., *Gemin5 delivers snRNA precursors to the SMN complex for snRNP biogenesis.* Mol Cell, 2010. **38**(4): p. 551-62.

255. Fischer, U., C. Englbrecht, and A. Chari, *Biogenesis of spliceosomal small nuclear ribonucleoproteins.* Wiley Interdiscip Rev RNA, 2011. **2**(5): p. 718-731.

256. Rogalska, M.E., et al., *Therapeutic activity of modified U1 core spliceosomal particles.* Nat Commun, 2016. **7**(1): p. 1-13.

257. Luo, M.L., et al., *Pre-mRNA splicing and mRNA export linked by direct interactions between UAP56 and Aly.* Nature, 2001. **413**(6856): p. 644-647.

258. Zhang, M. and M.R. Green, *Identification and characterization of yUAP/Sub2p, a yeast homolog of the essential human pre-mRNA splicing factor hUAP56.* Genes Dev, 2001. **15**(1): p. 30-35.

259. Pryor, A., et al., *Growth-regulated expression and G0-specific turnover of the mRNA that encodes URH49, a mammalian DExH/D box protein that is highly related to the mRNA export protein UAP56.* Nucleic Acids Res, 2004. **32**(6): p. 1857-65.

260. Amir-Ahmady, B., et al., *Exon repression by polypyrimidine tract binding protein.* RNA, 2005. **11**(5): p. 699-716.

261. Wu, J.Y. and T. Maniatis, *Specific interactions between proteins implicated in splice site selection and regulated alternative splicing.* Cell, 1993. **75**(6): p. 1061-70.

262. Boukis, L.A., et al., *Ser/Arg-rich protein-mediated communication between U1 and U2 small nuclear ribonucleoprotein particles.* J Biol Chem, 2004. **279**(28): p. 29647-53.

263. Staknis, D. and R. Reed, *SR proteins promote the first specific recognition of Pre-mRNA and are present together with the U1 small nuclear ribonucleoprotein particle in a general splicing enhancer complex.* Mol Cell Biol, 1994. **14**(11): p. 7670-82.

264. Cho, S., et al., *Interaction between the RNA binding domains of Ser-Arg splicing factor 1 and U1-70K snRNP protein determines early spliceosome assembly.* Proc Natl Acad Sci U S A, 2011. **108**(20): p. 8233-8.

265. Kapadia, F., et al., *Nuclear localization of poly(A)+ mRNA following siRNA reduction of expression of the mammalian RNA helicases UAP56 and URH49.* Gene, 2006. **384**: p. 37-44.

266. Yamazaki, T., et al., *The closely related RNA helicases, UAP56 and URH49, preferentially form distinct mRNA export machineries and coordinately regulate mitotic progression.* Mol Biol Cell, 2010. **21**(16): p. 2953-65.

267. Wang, C., et al., *Polypyrimidine tract-binding protein (PTB) differentially affects malignancy in a cell line-dependent manner.* J Biol Chem, 2008. **283**(29): p. 20277-87.

268. Gagnon, K.T., et al., *Analysis of nuclear RNA interference in human cells by subcellular fractionation and Argonaute loading.* Nat Protoc, 2014. **9**(9): p. 2045-60.

269. Ram, O. and G. Ast, *SR proteins: a foot on the exon before the transition from intron to exon definition.* Trends in Genetics, 2007. **23**(1): p. 5-7.

270. Spingola, M., et al., *Genome-wide bioinformatic and molecular analysis of introns in Saccharomyces cerevisiae.* RNA, 1999. **5**(2): p. 221-234.

271. Ares, M., L. Grate, and M.H. Pauling, *A handful of intron-containing genes produces the lion's share of yeast mRNA.* RNA, 1999. **5**(9): p. 1138-1139.

272. Ast, G., *How did alternative splicing evolve?* Nature Reviews Genetics, 2004. **5**(10): p. 773-782.

273. Shao, W., et al., *A U1-U2 snRNP Interaction Network during Intron Definition.* Molecular and Cellular Biology, 2012. **32**(2): p. 470-478.

274. Romfo, C.M., et al., *Evidence for splice site pairing via intron definition in Schizosaccharomyces pombe.* Molecular and Cellular Biology, 2000. **20**(21): p. 7955-7970.

275. De Conti, L., M. Baralle, and E. Buratti, *Exon and intron definition in pre-mRNA splicing.* Wiley Interdisciplinary Reviews: RNA, 2013. **4**(1): p. 49-60.

276. Fox-Walsh, K.L., et al., *The architecture of pre-mRNAs affects mechanisms of splice-site pairing.* Proceedings of the National Academy of Sciences, 2005. **102**(45): p. 16176-16181.

277. Braun, J.E., et al., *Synergistic assembly of human pre-spliceosomes across introns and exons.* eLife, 2018. **7**(1): p. e37751.

278. Hoffman, B.E. and P.J. Grabowski, *U1 snRNP targets an essential splicing factor, U2AF65, to the 3' splice site by a network of interactions spanning the exon.* Genes & Development, 1992. **6**(12b): p. 2554-2568.

279. Berget, S.M., *Exon recognition in vertebrate splicing.* Journal of Biological Chemistry, 1995. **270**(6): p. 2411-2414.

280. Moldón, A. and C. Query, *Crossing the exon.* Molecular Cell, 2010. **38**(2): p. 159-161.

281. Zhang, Z., et al., *Molecular architecture of the human 17S U2 snRNP.* Nature, 2020. **583**(7815): p. 310-313.

282. Champion-Arnaud, P. and R. Reed, *The prespliceosome components SAP 49 and SAP 145 interact in a complex implicated in tethering U2 snRNP to the branch site.* Genes & Development, 1994. **8**(16): p. 1974-1983.

283. Bennett, M. and R. Reed, *Correspondence between a mammalian spliceosome component and an essential yeast splicing factor.* Science, 1993. **262**(5130): p. 105-108.

284. Igel, H., et al., *Conservation of structure and subunit interactions in yeast homologues of splicing factor 3b (SF3b) subunits.* RNA, 1998. **4**(1): p. 1-10.

285. Linder, P. and E. Jankowsky, *From unwinding to clamping - the DEAD box RNA helicase family.* Nature Reviews Molecular Cell Biology, 2011. **12**(8): p. 505-516.

286. Wu, G., et al., *Pseudouridines in U2 snRNA stimulate the ATPase activity of Prp5 during spliceosome assembly.* The EMBO Journal, 2016. **35**(6): p. 654-667.

287. Beier, D.H., et al., *Dynamics of the DEAD-box ATPase Prp5 RecA-like domains provide a conformational switch during spliceosome assembly.* Nucleic Acids Research, 2019. **47**(20): p. 10842-10851.

288. Papasaikas, P. and J. Valcárcel, *The spliceosome: the ultimate RNA chaperone and sculptor.* 2016. **41**(1): p. 33-45.

289. Ngo, T.D., A.C. Partin, and Y. Nam, *RNA specificity and autoregulation of DDX17, a modulator of microRNA Bbiogenesis.* Cell Reports, 2019. **29**(12): p. 4024-4035.

290. Tsu, C.A., K. Kossen, and O.C. Uhlenbeck, *The Escherichia coli DEAD protein DbpA recognizes a small RNA hairpin in 23S rRNA.* RNA, 2001. **7**(5): p. 702-709.

291. Silverman, E., G. Edwalds-Gilbert, and R.J. Lin, *DExD/H-box proteins and their partners: Helping RNA helicases unwind.* Gene, 2003. **312**(1-2): p. 1-16.

292. Shi, H., et al., *Crystal structure of the human ATP-dependent splicing and export factor UAP56.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(51): p. 17628-17633.

293. Zhang, M. and M.R. Green, *Identification and characterization of yUAP/Sub2p, a yeast homolog of the essential human pre-mRNA splicing factor hUAP56.* Genes and Development, 2001. **15**(1): p. 30-35.

294. Libri, D., et al., *Multiple roles for the yeast SUB2/yUAP56 gene in splicing.* Genes and Development, 2001. **15**(1): p. 36-41.

295. Selenko, P., et al., *Structural Basis for the Molecular Recognition between Human Splicing Factors U2AF 65 and SF1 / mBBP.* Molecular Cell, 2003. **11**(4): p. 965-976.

296. Berglund, J.A., et al., *The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC.* Cell, 1997. **89**(5): p. 781-787.

297. Chatrikhi, R., et al., *SF1 Phosphorylation Enhances Specific Binding to U2AF65 and Reduces Binding to 3'-Splice-Site RNA.* Biophysical Journal, 2016. **111**(12): p. 2570-2586.

298. Wang, Q., et al., *A BBP-Mud2p heterodimer mediates branchpoint recognition and influences splicing substrate abundance in budding yeast.* Nucleic Acids Research, 2008. **36**(8): p. 2787-2798.

299. Rutz, B. and B. Séraphin, *Transient interaction of BBP/ScSF1 and Mud2 with the splicing machinery affects the kinetics of spliceosome assembly.* RNA, 1999. **5**(6): p. 819-831.

300. Shen, H. and M.R. Green, *A pathway of sequential arginine-serine-rich domain-splicing signal interactions during mammalian spliceosome assembly.* Molecular Cell, 2004. **16**(3): p. 363-373.

301. Corsini, L., et al., *U2AF-homology motif interactions are required for alternative splicing regulation by SPF45.* Nature Structural and Molecular Biology, 2007. **14**(7): p. 620-629.

302. Loerch, S. and C.L. Kielkopf, *Unmasking the U2AF homology motif family: A bona fide protein-protein interaction motif in disguise.* RNA, 2016. **22**(12): p. 1795-1807.

303. Behzadnia, N., et al., *Composition and three-dimensional EM structure of double affinity-purified, human prespliceosomal A complexes.* EMBO Journal, 2002. **26**(18): p. 1737-1748.

304. Deckert, J., et al., *Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions.* Molecular and Cellular Biology, 2006. **26**(14): p. 5528-5543.

305.   Morris, A.K., et al., *Cellular mRNA export factor UAP56 recognizes nucleic acid binding site of in fl uenza virus NP protein.* Biochemical and Biophysical Research Communications, 2020. **525**(2): p. 259-264.

306.   Mojumdar, A. and J. Deka, *Assaying the activity of helicases.* 2019, Elsevier. p. 235-246.

307.   Luck, K., et al., *A reference map of the human binary protein interactome.* Nature, 2020. **580**(7803): p. 402-408.

308.   Oughtred, R., et al., *TheBioGRIDdatabase: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions.* Protein Science, 2021. **30**(1): p. 187-200.

309.   Campagne, S., et al., *An in vitro reconstituted U1 snRNP allows the study of the disordered regions of the particle and the interactions with proteins and ligands.* Nucleic Acids Research, 2021: p. 1-13.

310.   Huebsch, N., et al., *CRISPR interference efficiently induces specific and reversible gene silencing in human iPSCs.* Stem Cell, 2016. **18**(4): p. 541-553.

311.   Komor, A.C., A.H. Badran, and D.R. Liu, *Editing the genome without double-stranded DNA breaks.* ACS Chemical Biology, 2018. **13**(2): p. 383-388.

312.   Standage-Beier, K., et al., *A transient reporter for editing enrichment (TREE) in human cells.* Nucleic Acids Research, 2019. **47**(19): p. e120.

APPENDIX A

SATURATION BINDING CURVES OF WT AND MUTANT UBL PROTEINS
FOR U1-SL4 FROM POINT STUDY ANALYSIS OF SPR DATA
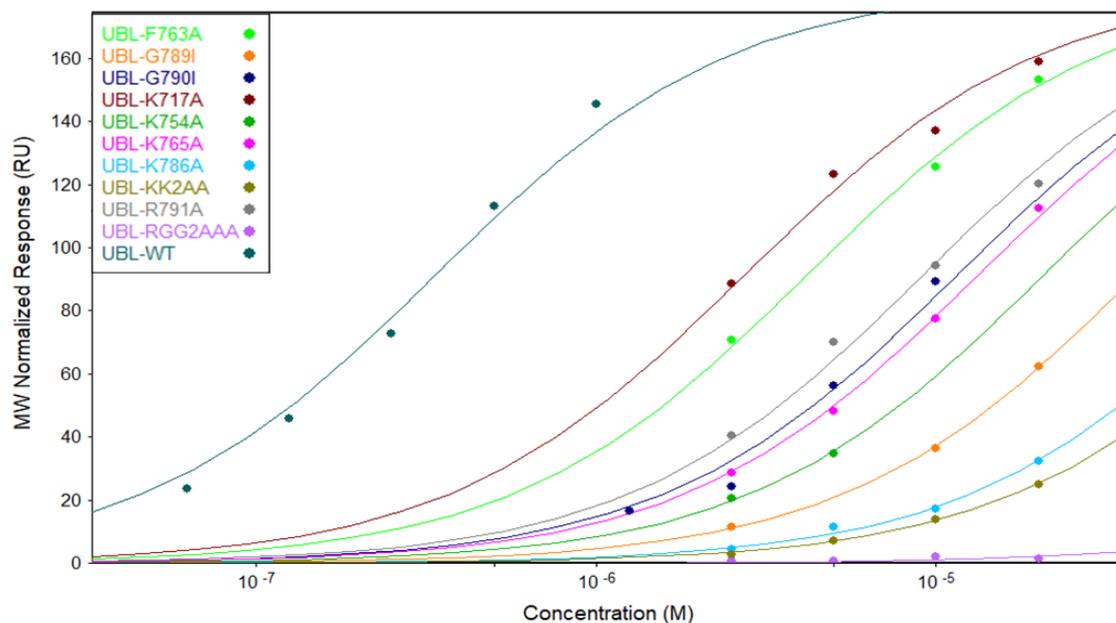
Titrated WT and mutant UBL proteins were injected onto a U1-SL4 coated biosensor and responses were recorded (RU). WT UBL protein (blue-green) binds with high affinity while all mutants experience rightward shifts in their saturation binding curves indicating reduction in U1-SL4 binding activity. These curves were used to derive the affinity constants reported in Table 2.3.

APPENDIX B

ALIGNMENT OF FULL-LENGTH SF3A1 PROTEIN FROM HUMAN AND YEAST

```
                10        20        30        40        50        60        70        80
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
hSF3A1  MPAGPVQAVPPPPPVPTEPKQPTEEEASSKEDSAPSKPVVGIIYPPPEVRNIVDKTASFVARNGPEFEARIRQNEINNPK
Prp21   ----------------M..-----.DTQL...--------------------IKT.VNYIKQH.V...NKLLED----ER
                                                                             SURP1

                90        100       110       120       130       140       150       160
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
hSF3A1  FNFLNPNDPYHAYYRHKVSEFKEGKAQEPSAAIPKVMQQQQQTTQQQLPQKVQAQVIQETIVPKEPPPEFEFIADPPSIS
Prp21   .S.IKKD..L.E..TKLMN.PTDTVSG.----------DNDRKSEREIAR----------------..D.L.SQYDTG..

                170       180       190       200       210       220       230       240
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
hSF3A1  AFDLDVVKLTAQFVARNGRQFLTQLMQKEQRNYQFDFLRPQHSLFNYFTKLVEQYTKILIPPKGLFSKLKKEAENPREVL
Prp21   RR.ME.I....RYY.KD-KSIVE.MIS.DG-EARLN.MNSS.P.HKT..DF.A..KRVYSFTG------QEIKKSK.TI.
                            SURP2

                250       260       270       280       290       300       310       320
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
hSF3A1  DQVCYRVEWAKFQERERKKEEEEKEKERVAYAQIDWHDFVVVETVDFQPNEQGNFPPPTTPEELGARILIQERYEKFGES
Prp21   .NCFE.TQYWE.EKDKDREHDKLV.LCKIQF.A.P.DK.TQ.AKFSI---------.ED.EIFE.SLD.E.M.LRRVQTG
                      Charged Region

                330       340       350       360       370       380       390       400
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
hSF3A1  EEVEMEVESDEEDDKQEKAEEPPSQLDQDTQVQDMDEGSDDEEEGQKVPPPPETPMPPPLPPTPDQVIVRKDYDPKASKP
Prp21   IKLFDSIKPTN.EE.IVSD--------------.GKQK.G.SKGKKR.IRAVG..RLKKSKK--------------------
                                                           Proline-rich Region 1

                410       420       430       440       450       460       470       480
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
hSF3A1  LPPAPAPDEYLVSPITGEKIPASKMQEHMRIGLLDPRWLEQRDRSIREKQSDDEVYAPGLDIESSLKQLAERRTDIFGVE
Prp21   ---------------------------------------------------------------------------------

                490       500       510       520       530       540       550       560
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
hSF3A1  ETAIGKKIGEEEIQKPEEKVTWDGHSGSMARTQQAAQANITLQEQIEAIHKAKGLVPEDDTKEKIGPSKPNEIPQQPPPP
Prp21   ------------------------------------------------------------------------------

                570       580       590       600       610       620       630       640
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
hSF3A1  SSATNIPSSAPPITSVPRPPTMPPPVRTTVVSAVPVMPRPPMASVVRLPPGSVIAPMPPIIHAPRINVVPMPPSAPPIMA
Prp21   ------------------------------------------------------------------------------
                      Proline-rich Region 2

                650       660       670       680       690       700       710       720
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
hSF3A1  PRPPPMIVPTAFVPAPPVAPVPAPAPMPPVHPPPPMEDEPTSKKLKTEDSLMPEEEFLRRNKGPVSIKVQVPNMQDKTEW
Prp21   ------------------------------------------------------------------------------
                                                          NLS

                730       740       750       760       770       780       790
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
hSF3A1  KLNGQVLVFTLPLTDQVSVIKVKIHEATGMPAGKQKLQYEGIFIKDSNSLAYYNMANGAVIHLALKERGGRKK
Prp21   ----------------------------------------------------------------------
                      Ubiquitin-like Domain
```

SF3A1 from Saccharomyces cerevisiae (Prp21; UniProt ID: P32524) was aligned to human SF3A1 (hSF3A1; UniProt ID: Q15459) using ClustalW. The N-terminal features of SF3A1 are conserved between yeast and humans; including the two SURP domains and the small region of charged residues. Prp21 lacks the large proline rich regions, nuclear localization signal, and the UBL domain. In the aligned sequences, a dot inZdicates the presence of a residue identical to reference (hSF3A1) and a dash indicates a gap.

175

APPENDIX C

ACCESSION NUMBERS OF UBL DOMAIN CONTAINING PROTEINS WITH
SF3A1-LIKE FEATURES

Protein Name	UBL Family	UniProt Accession
SF3A1	UBL	Q15459
MAP2K5	PB1	Q13163
PARD6B	PB1	Q9BYG5
UBIQUITIN	UB	P0CG47
RPS27A	NEDD8	P62979
NEDD8	NEDD8	Q15843
ZFAND4	UBL	Q86XD8
URM1	UBL	Q9BTM9
UBD UBL-2	UBL	O15205
OASL UBL-2	UBL	Q15646
DDI1	UBL	Q8WTU0
DDI2	UBL	Q5TDH0
SQSTM1	PB1	Q13501
SNRNP25	UBL	Q9BV90
UHRF1	UBL	Q96T88
RAD23A	PIM	P54725
RAD23B	PIM	P54727
UBL5	UBL	Q9BZL1
PRKCI	PB1	P41743
UBD UBL-1	UBL	O15205
OASL UBL-1	UBL	Q15646
UBL7	UBL	Q96S82
UHRF2	PIM	Q96PU4
UBL4B	PIM	Q8N7F7
BAG6	PIM	A0A0G2JK23
MAP1LC3A	ATG8	Q9H492
MAP1LC3B	ATG8	Q9GZQ8
MAP1LC3B2	ATG8	A6NCE7
MAP1LC3C	ATG8	Q9BXW4
GABARAP	ATG8	O95166
GABARAPL1	ATG8	Q9H0R8
GABARAPL2	ATG8	P60520
PCGF1	RAWUL	Q9BSM1
TBK1	UBL	Q9UHD2
MIDN	UBL	Q504T8

Table of genes and their accession numbers aligned in Figure 2.29.

APPENDIX D

SEQUENCE OF PRIMERS USED FOR SF3A1 AND U1-5A SNRNA MUTAGENESIS

Final Clone Name	Template Plasmid	Forward Primer	Reverse Primer
pcDNA3.1:FLAG-RNAiR-SF3A1-WT	pcDNA3.1:FLAG-SF3A1-WT	GCcCCgTCcAAGCCAGTTGTGGGGATTATTTAC	cGAgTCtTCCTTTGAAGATGCTTCTTCTTCTGTGG
pcDNA3.1:FLAG-RNAiR-SF3A1 K717A	pcDNA3.1:FLAG-RNAiR-SF3A1-WT	GCGACGGAATGGAAACTGAATGGGC	ATCCTGCATGTTGGGCACCTGGAC
pcDNA3.1:FLAG-RNAiR-SF3A1 K754A	pcDNA3.1:FLAG-RNAiR-SF3A1-WT	GCACAGAAGCTACAGTATGAGGGTATCTTCATC	CCCTGCAGGCATGCCTGTGGCTTC
pcDNA3.1:FLAG-RNAiR-SF3A1 F763A	pcDNA3.1:FLAG-RNAiR-SF3A1-WT	CATCAAAGATTCCAACTCACTGGCTTACTAC	GCGATACCCTCATACTGTAGCTTCTGTTTCC
pcDNA3.1:FLAG-RNAiR-SF3A1 K765A	pcDNA3.1:FLAG-RNAiR-SF3A1-WT	GCAGATTCCAACTCACTGGCTTACTACAAC	GATGAAGATACCCTCATACTGTAGCTTCTG
pcDNA3.1:FLAG-RNAiR-SF3A1 K786A	pcDNA3.1:FLAG-RNAiR-SF3A1-WT	GCGGAGAGAGGCGGGAGGAAGAAGTAGG	GAGGGCCAGGTGGATGACTGCGCCATTG
pcDNA3.1:FLAG-RNAiR-SF3A1 R788A	pcDNA3.1:FLAG-RNAiR-SF3A1-WT	GGCGGGAGGAAGAAGTAGGCGGCCG	TGCCTCCTTGAGGGCCAGGTGGATGACTGC
pcDNA3.1:FLAG-RNAiR-SF3A1 RGG2AAA	pcDNA3.1:FLAG-RNAiR-SF3A1-WT	GCTGCTAGGAAGAAGTAGGCGGCCGCTC	TGCCTCCTTGAGGGCCAGGTGGATGACTGC
pcDNA3.1:FLAG-RNAiR-SF3A1 G789I	pcDNA3.1:FLAG-RNAiR-SF3A1-WT	ATCGGGAGGAAGAAGTAGGCGGCCGCTC	TCTCTCCTTGAGGGCCAGGTGGATGAC
pcDNA3.1:FLAG-RNAiR-SF3A1 G790I	pcDNA3.1:FLAG-RNAiR-SF3A1-WT	GGCATCAGGAAGAAGTAGGCGGCCGCTC	TCTCTCCTTGAGGGCCAGGTGGATGAC
pcDNA3.1:FLAG-RNAiR-SF3A1 R791A	pcDNA3.1:FLAG-RNAiR-SF3A1-WT	GCCAAGAAGTAGGCGGCCGCTCGAGTCTAGAG	CCCGCCTCTCTCCTTGAGGGCCAG
pcDNA3.1:FLAG-RNAiR-SF3A1 KK2AA	pcDNA3.1:FLAG-RNAiR-SF3A1-WT	GCTGCTTAGGCGGCCGCTCGAGTCTAGAGG	CCTCCCGCCTCTCTCCTTGAGGG
pcDNA3.1:FLAG-RNAiR-SF3A1 RKK2AAA	pcDNA3.1:FLAG-RNAiR-SF3A1-WT	GCTGCTGCTTAGGCGGCCGCTCGAGTCTAGAG	CCCGCCTCTCTCCTTGAGGGCCAG
pcDNA3.1:FLAG-RNAiR- K754A+RGG2AAA	pcDNA3.1:FLAG-RNAiR-SF3A1-RGG2AAA	GCACAGAAGCTACAGTATGAGGGTATCTTCATC	CCCTGCAGGCATGCCTGTGGCTTC
pcDNA3.1:FLAG-RNAiR- K765A+RGG2AAA	pcDNA3.1:FLAG-RNAiR-SF3A1-RGG2AAA	GCAGATTCCAACTCACTGGCTTACTACAAC	GATGAAGATACCCTCATACTGTAGCTTCTG
pcDNA3.1:FLAG-RNAiR-SF3A1 Y772C	pcDNA3.1:FLAG-RNAiR-SF3A1-WT	GCTACAACATGGCCAATGG	AAGCCAGTGAGTTGGAATC
pcDNA3.1:FLAG-RNAiR-SF3A1 Y773C	pcDNA3.1:FLAG-RNAiR-SF3A1-WT	GCAACATGGCCAATGGCGC	AGTAAGCCAGTGAGTTGGAATCTTTG
pNS6:U1-5a/SL3-M1a	pNS6:U1-5a/WT	GTGGGAAATCGACTGCATAATTTGTGGTAGTG	ATTTGGGGAAATCGCAGGGGTCAGCACATCCGGAG
pNS6:U1-5a/SL3-M1b	pNS6:U1-5a/WT	GTAGGAAACTCGACTGCATAATTTGTGGTAGTG	ATTTGGGGAAATCGCAGGGGTCAGCACATCCGGAG
pNS6:U1-5a/SL3-M1c	pNS6:U1-5a/WT	GTAAGAAACTCGACTGCATAATTTGTGGTAGTG	ATTTGGGGAAATCGCAGGGGTCAGCACATCCGGAG
pNS6:U1-5a/SL3-M1d	pNS6:U1-5a/WT	GTAAAAAACTCGACTGCATAATTTGTGGTAGTG	ATTTGGGGAAATCGCAGGGGTCAGCACATCCGGAG
pNS6:U1-5a/SL3-M1e	pNS6:U1-5a/WT	GTAAATAACTCGACTGCATAATTTGTGGTAGTG	ATTTGGGGAAATCGCAGGGGTCAGCACATCCGGAG
pNS6:U1-5a/SL3-M1f	pNS6:U1-5a/WT	GTAAATTACTCGACTGCATAATTTGTGGTAGTG	ATTTGGGGAAATCGCAGGGGTCAGCACATCCGGAG
pNS6:U1-5a/SL3-M1g	pNS6:U1-5a/WT	GTAAATTTCTCGACTGCATAATTTGTGGTAGTG	ATTTGGGGAAATCGCAGGGGTCAGCACATCCGGAG
pNS6:U1-5a/SL3-M1h	pNS6:U1-5a/WT	GTCCCTTTCAGCACTGCATAATTTGTGGTAGTG	ATTTGCCCTTTAGCCAGGGGTCAGCACATCCGGAG
pNS6:U1-5a/SL3-M1i	pNS6:U1-5a/WT	GTCCCTTTAGCACTGCATAATTTGTGGTAGTG	ATTTGCCCTTTAGCCAGGGGTCAGCACATCCGGAG
pNS6:U1-5a/SL3-M1j	pNS6:U1-5a/WT	GTTTTAAACTCGACTGCATAATTTGTGGTAGTG	ATTTGTTTAAATCGCAGGGGTCAGCACATCCGGAG
pNS6:U1-5a/SL3-M1k	pNS6:U1-5a/WT	GTGGGGGGCTCGACTGCATAATTTGTGGTAGTG	ATTTGGGGGGGTCGCAGGGGTCAGCACATCCGGAG
pNS6:U1-5a/SL3-M1l	pNS6:U1-5a/WT	GTTTTGGGCTCGACTGCATAATTTGTGGTAGTG	ATTTGTTTGGGTCGCAGGGGTCAGCACATCCGGAG
pNS6:U1-5a/SL3-M2a	pNS6:U1-5a/WT	GTGGGAAATCGACTGCATAATTTGTGGTAGTG	TTTGGGGAAATCGCAGGGGTCAGCACATCCGGAG
pNS6:U1-5a/SL3-M2b	pNS6:U1-5a/WT	GTGGGAAATCGACTGCATAATTTGTGGTAGTG	TTGGGGAAATCGCAGGGGTCAGCACATCCGGAG
pNS6:U1-5a/SL3-M2c	pNS6:U1-5a/WT	GTGGGAAATCGACTGCATAATTTGTGGTAGTG	TGGGGAAATCGCAGGGGTCAGCACATCCGGAG
pNS6:U1-5a/SL3-M2d	pNS6:U1-5a/WT	GTGGGAAATCGACTGCATAATTTGTGGTAGTG	GGGGAAATCGCAGGGGTCAGCACATCCGGAG
pNS6:U1-5a/SL3-M1d/SL4-M10e	pNS6:U1-5a/SL3-M1d	CGCCTTAGGGGTGACTTTCTGGAGTTTCAAAAACAGACCG	CGAACGCAGAGGGGCACTACCACAAATTATGCAGTCGAG
pNS6:U1-5a/SL3-M1d/SL4-M10r	pNS6:U1-5a/SL3-M1d	GAAATTAAACCTGACTTTCTGGAGTTTCAAAAACAGACC	GAACGCAGTCCCCCACTACCACAAATTATGCAGTCGAG
pNS6:U1-5a/SL3-M1d/SL4-M10	pNS6:U1-5a/SL3-M1d	GATATTTATATTGACTTTCTGGAGTTTCAAAAACAGACCG	GAAATATGTATATCACTACCACAAATTATGCAGTCGAG
pNS6:U1-5a/SL3-M1g/SL4-M10e	pNS6:U1-5a/SL3-M1g	CGCCTTAGGGGTGACTTTCTGGAGTTTCAAAAACAGACCG	CGAACGCAGAGGGGCACTACCACAAATTATGCAGTCGAG
pNS6:U1-5a/SL3-M1g/SL4-M10r	pNS6:U1-5a/SL3-M1g	GAAATTAAACCTGACTTTCTGGAGTTTCAAAAACAGACC	GAACGCAGTCCCCCACTACCACAAATTATGCAGTCGAG
pNS6:U1-5a/SL3-M1g/SL4-M10	pNS6:U1-5a/SL3-M1g	GATATTTATATTGACTTTCTGGAGTTTCAAAAACAGACCG	GAAATATGTATATCACTACCACAAATTATGCAGTCGAG
pNS6:U1-5a/SL4/SL4	pNS6:U1-5a/WT	CGCGCTTTCCCCACTGCATAATTTGTGGTAGTG	AACGCAGTCCCCCAGGGGTCAGCACATCCGGAG
pNS6:U1-5a/SL3/SL3	pNS6:U1-5a/WT	GTGGGAAACTCGTGACTTTCTGGAGTTTCAAAAAC	ATTTGGGGAAATCGCACTACCACAAATTATGCAGT
pNS6:U1-5a/SL4/SL3	pNS6:U1-5a/SL3/SL3	CGCGCTTTCCCCACTGCATAATTTGTGGTAGTG	AACGCAGTCCCCCAGGGGTCAGCACATCCGGAG

APPENDIX E

SEQUENCE OF OLIGONUCLEOTIDES AND SMALL INTERFERING RNAS USED
IN THIS DISSERTATION

180

Oligo name	Sequence (5′→3′)	Technique
Dup3r	AACAGCATCAGGAGTGGACAGATCCC	Primer extension
U1$_{7-26}$R	TGGTATCTCCCCTGCCAGGT	Primer extension
U1$_{17-39}$F	GGAGATACCATGATCACGAAGG	RT-qPCR
U1$_{58-80}$R	CATCCGGAGTGCAATGGATAAG	RT-qPCR
U2$_{8-19}$F	CTCGGCCTTTTGGCTAAGATCA	RT-qPCR
U2$_{62-84}$R	TCCTCGGATAGAGGACGTATCA	RT-qPCR
U1$_{27-46}$R	GAAAACCACCTTCGTGATCA	Northern blotting
U2$_{114-135}$R	GGAGCAAGCTCCTATTCCATCT	Northern blotting
5S$_{83-103}$R	TATTCCCAGGCGGTCTCCCAT	Northern blotting
U1-M10r-LNA	A+T+T+TCGAACGCAGTC	Northern blotting
Cy5-U1-SL3	Cy5-CGAUUUCCCCAAAUGUGGGAAACUCG	EMSA
Cy5-U1-SL4	Cy5-GGGGACUGCGUUCGCGCUUUCCCC	EMSA
U1$_{1-13}$	mGmCmCmAmGmGmUmAmAmGmUmAmUTTTT-Biotin	ASO
U2$_{1-21}$	mGmCmCmAmAmAmAmGmGmCmCmGmAmGmAmAmGmCmGmAmUTTTT-Biotin	ASO
siSF3A1	GGAGGAUUCUGCACCUUCU	RNAi
siUAP56	AAGGGCUUGGCUAUCACAU	RNAi
siURH49	AAAGACAUCAAGGGAUCCUACGUUU	RNAi
siPTBP1	UGACAAGAGCCGUGACUAC	RNAi

+ indicates a locked nucleic acid, m indicates 2′-O-methyl modification.

APPENDIX F

NSAF ANALYSIS OF MS DATA

#	LocusID	L	MW	pI	SL3-WT SpectrumCount	SL3-WT NSAFe5	SL3-WT SequenceCount	SL3-WT Coverage	SL3-MUT SpectrumCount	SL3-MUT NSAFe5	SL3-MUT SequenceCount	SL3-MUT Coverage	Description
1	P84103	164	19330	11.6	12	367.717804	8	41.5	21	1397.215441	10	46.3	Splicing factor, arginine/serine-rich 3 OS=Homo sapiens GN=SFRS3 PE=1 SV=1
2	P09651	372	38747	9.1	27	364.752338	22	47	30	879.966714	21	47	Heterogeneous nuclear ribonucleoprotein A1 OS=Homo sapiens GN=HNRNPA1 PE=1 SV=5
3	P53999	127	14395	9.6	9	356.136141	9	48	11	945.0981086	9	48.8	Activated RNA polymerase II transcriptional coactivator p15 OS=Homo sapiens GN=SUB1 PE=1 SV=3
4	P22626	353	37430	8.9	21	298.966033	19	46.7	25	772.7753012	21	58.6	Heterogeneous nuclear ribonucleoproteins A2/B1 OS=Homo sapiens GN=HNRNPA2B1 PE=1 SV=2
5	Q14103	355	38434	7.8	20	283.125446	18	34.6	30	922.1059651	23	35.8	Heterogeneous nuclear ribonucleoprotein D0 OS=Homo sapiens GN=HNRNPD PE=1 SV=1
6	P67809	324	35924	9.9	15	232.660956	14	49.1	32	1077.68763	22	61.4	Nuclease-sensitive element-binding protein 1 OS=Homo sapiens GN=YBX1 PE=1 SV=3
7	P61978	463	50976	5.5	21	227.937386	18	40.4	60	1414.028586	26	48.4	Heterogeneous nuclear ribonucleoprotein K OS=Homo sapiens GN=HNRNPK PE=1 SV=1
8	Q9Y333	95	10835	6.5	4	211.599017	3	32.6	5	574.294066	3	32.6	U6 snRNA-associated Sm-like protein LSm2 OS=Homo sapiens GN=LSM2 PE=1 SV=1
9	Q15365	356	37498	7.1	14	197.631105	12	54.8	47	1440.574722	18	68.3	Poly(rC)-binding protein 1 OS=Homo sapiens GN=PCBP1 PE=1 SV=2
10	P62312	80	9128	9.6	3	188.455375	3	36.2	4	545.5793627	4	45	U6 snRNA-associated Sm-like protein LSm6 OS=Homo sapiens GN=LSM6 PE=1 SV=1
11	Q9BUJ2	856	95739	6.9	29	170.255634	26	30.4	40	509.8872548	30	33.9	Heterogeneous nuclear ribonucleoprotein U-like protein 1 OS=Homo sapiens GN=HNRNPUL1 PE=1 SV=2
12	O60506	623	69603	8.6	18	145.198363	17	32.3	34	595.4959336	24	42.5	Heterogeneous nuclear ribonucleoprotein Q OS=Homo sapiens GN=SYNCRIP PE=1 SV=2
13	P98179	157	17170	8.9	4	128.037622	4	28.7	10	695.0055576	9	40.8	Putative RNA-binding protein 3 OS=Homo sapiens GN=RBM3 PE=1 SV=1
14	P62316	118	13527	9.9	3	127.766356	3	33.1	2	184.9421568	2	16.9	Small nuclear ribonucleoprotein Sm D2 OS=Homo sapiens GN=SNRPD2 PE=1 SV=1
15	Q16629	238	27367	11.8	6	126.692689	5	19.7	8	366.7760421	7	30.7	Splicing factor, arginine/serine-rich 7 OS=Homo sapiens GN=SFRS7 PE=1 SV=1
16	P62306	86	9725	4.7	2	116.87155	2	24.4	2	253.7578431	2	24.4	Small nuclear ribonucleoprotein F OS=Homo sapiens GN=SNRPF PE=1 SV=1
17	Q14011	172	18648	9.5	4	116.87155	3	20.9	2	126.8789216	2	14.5	Cold-inducible RNA-binding protein OS=Homo sapiens GN=CIRBP PE=1 SV=1
18	Q00839	825	90585	6	19	115.73825	16	24.8	19	251.297161	17	30.3	Heterogeneous nuclear ribonucleoprotein U OS=Homo sapiens GN=HNRNPU PE=1 SV=6
19	Q15366	365	38580	6.8	8	110.147434	7	23.6	29	866.9480284	14	51	Poly(rC)-binding protein 2 OS=Homo sapiens GN=PCBP2 PE=1 SV=1
20	Q9H2U1	1008	114776	7.7	21	104.69743	21	24.1	5	54.12493677	5	7.9	Probable ATP-dependent RNA helicase DHX36 OS=Homo sapiens GN=DHX36 PE=1 SV=1
21	O43390	633	70943	8.1	13	103.208841	13	21.2	23	396.4715748	20	28.3	Heterogeneous nuclear ribonucleoprotein R OS=Homo sapiens GN=HNRNPR PE=1 SV=1
22	P11940	636	70671	9.5	13	102.722007	13	21.7	12	205.8790048	12	19.3	Polyadenylate-binding protein 1 OS=Homo sapiens GN=PABPC1 PE=1 SV=2
23	Q13247	344	39587	11.4	7	102.262606	7	13.4	8	253.7578431	8	17.4	Splicing factor, arginine/serine-rich 6 OS=Homo sapiens GN=SFRS6 PE=1 SV=2
24	Q8NC51	408	44965	8.6	8	94.538758	8	17.6	4	106.9763456	4	11.3	Plasminogen activator inhibitor 1 RNA-binding protein OS=Homo sapiens GN=SERBP1 PE=1 SV=2
25	P17844	614	69148	8.9	12	98.2177848	12	20.7	8	142.170518	8	13.4	Probable ATP-dependent RNA helicase DDX5 OS=Homo sapiens GN=DDX5 PE=1 SV=1
26	P51991	378	39595	9	7	93.0643825	7	22.8	5	144.3331647	5	20.6	Heterogeneous nuclear ribonucleoprotein A3 OS=Homo sapiens GN=HNRNPA3 PE=1 SV=2
27	Q99729	332	36225	8.2	6	90.8218673	6	12.3	24	788.78944	12	18.7	Heterogeneous nuclear ribonucleoprotein A/B OS=Homo sapiens GN=HNRNPAB PE=1 SV=2
28	P14866	589	64133	8.2	10	85.3221843	9	22.4	14	259.3586104	13	32.8	Heterogeneous nuclear ribonucleoprotein L OS=Homo sapiens GN=HNRNPL PE=1 SV=2
29	P78527	4128	469093	7.1	68	82.7840147	62	17.2	58	153.3120302	55	16.2	DNA-dependent protein kinase catalytic subunit OS=Homo sapiens GN=PRKDC PE=1 SV=3
30	Q92499	740	82432	7.2	12	81.4942161	11	19.9	11	162.19927	11	19.2	ATP-dependent RNA helicase DDX1 OS=Homo sapiens GN=DDX1 PE=1 SV=2
31	P35637	526	53426	9.4	8	76.4331051	8	16.5	6	124.4667748	5	14.6	RNA-binding protein FUS OS=Homo sapiens GN=FUS PE=1 SV=1
32	Q13243	272	31264	11.6	4	73.9040685	4	16.5	7	280.8129073	7	22.8	Splicing factor, arginine/serine-rich 5 OS=Homo sapiens GN=SFRS5 PE=1 SV=1
33	P52272	730	77516	8.7	10	68.842146	10	16.8	4	59.7895192	4	7	Heterogeneous nuclear ribonucleoprotein M OS=Homo sapiens GN=HNRNPM PE=1 SV=3
34	Q01130	221	25476	11.9	3	68.2191401	3	13.1	5	246.8684899	5	14.9	Splicing factor, arginine/serine-rich 2 OS=Homo sapiens GN=SFRS2 PE=1 SV=4
35	P52597	415	45672	5.6	5	60.5479115	5	12	3	78.878944	3	6.3	Heterogeneous nuclear ribonucleoprotein F OS=Homo sapiens GN=HNRNPF PE=1 SV=3
36	O14979	420	46437	9.6	5	59.8271031	5	12.9	16	415.6795144	12	13.1	Heterogeneous nuclear ribonucleoprotein D-like OS=Homo sapiens GN=HNRPDL PE=1 SV=3
37	P26368	475	53501	9.1	5	52.8997543	4	11.6	4	91.88705056	4	11.6	Splicing factor U2AF 65 kDa subunit OS=Homo sapiens GN=U2AF2 PE=1 SV=4
38	P46063	649	73457	7.9	6	46.460493	6	10.3	10	168.1292335	10	23.3	ATP-dependent DNA helicase Q1 OS=Homo sapiens GN=RECQL PE=1 SV=3
39	P08621	437	51557	9.9	4	45.9997863	4	8.7	2	49.93861443	2	6.2	U1 small nuclear ribonucleoprotein 70 kDa OS=Homo sapiens GN=SNRNP70 PE=1 SV=2
40	P31943	449	49229	6.3	4	44.7703934	4	11.6	3	72.9059282	3	5.8	Heterogeneous nuclear ribonucleoprotein H OS=Homo sapiens GN=HNRNPH1 PE=1 SV=4
41	P55795	449	49264	6.3	4	44.7703934	4	12	3	72.9059282	3	5.8	Heterogeneous nuclear ribonucleoprotein H2 OS=Homo sapiens GN=HNRNPH2 PE=1 SV=1
42	Q07955	248	27745	10.4	2	40.5280376	2	8.5	13	571.9783641	11	43.5	Splicing factor, arginine/serine-rich 1 OS=Homo sapiens GN=SFRS1 PE=1 SV=2
43	Q86V81	257	26888	11.2	2	39.1087678	2	11.3	5	212.2876898	4	23.3	THO complex subunit 4 OS=Homo sapiens GN=THOC4 PE=1 SV=3
44	Q96AE4	644	67560	7.6	5	39.0176759	5	8.7	57	965.7771327	32	51.6	Far upstream element-binding protein 1 OS=Homo sapiens GN=FUBP1 PE=1 SV=3
45	P38159	391	42332	10.1	3	38.5586444	3	9	2	55.81374554	2	5.4	Heterogeneous nuclear ribonucleoprotein G OS=Homo sapiens GN=RBMX PE=1 SV=3
46	Q1KMD3	747	85105	4.9	5	33.6377286	5	7.8	6	87.64327111	6	11.8	Heterogeneous nuclear ribonucleoprotein U-like protein 2 OS=Homo sapiens GN=HNRNPUL2 PE=1 SV=1
47	Q92945	710	73147	7.3	4	28.3125446	4	6.1	47	722.3163393	29	44.8	Far upstream element-binding protein 2 OS=Homo sapiens GN=KHSRP PE=1 SV=4
48	O43148	476	54844	6.6	2	21.1154481	2	5.9	6	137.5410158	6	17.9	mRNA cap guanine-N7 methyltransferase OS=Homo sapiens GN=RNMT PE=1 SV=1
49	Q05519	484	53542	10.5	2	20.7664325	2	5.2	2	45.08920353	2	6.2	Splicing factor, arginine/serine-rich 11 OS=Homo sapiens GN=SFRS11 PE=1 SV=1
50	Q09161	790	91839	6.4	3	19.0840886	3	3.9	7	96.68495035	7	9.2	Nuclear cap-binding protein subunit 1 OS=Homo sapiens GN=NCBP1 PE=1 SV=1
51	Q15459	793	88886	5.2	3	19.0118915	3	5.3	2	27.51976609	2	3.9	Splicing factor 3A subunit 1 OS=Homo sapiens GN=SF3A1 PE=1 SV=1
52	Q14498	530	59380	10.1	2	18.9640629	2	5.3	9	185.2911043	8	20.2	RNA-binding protein 39 OS=Homo sapiens GN=RBM39 PE=1 SV=2
53	Q01844	656	68478	9.3	2	15.3215752	2	3.5	2	33.26703431	2	4.3	RNA-binding protein EWS OS=Homo sapiens GN=EWSR1 PE=1 SV=1
54	Q13435	872	97657	5.7	2	11.5263226	2	2.3	3	37.5398644	3	6.1	Splicing factor 3B subunit 2 OS=Homo sapiens GN=SF3B2 PE=1 SV=1
55	Q14839	1912	217989	5.9	4	10.5135495	4	2.8	3	17.1206913	3	1.7	Chromodomain-helicase-DNA-binding protein 4 OS=Homo sapiens GN=CHD4 PE=1 SV=1
56	Q99878	128	13936	10.9	4	154.274034	2	12.5	0	0	0	0	Histone H2A type 1-J OS=Homo sapiens GN=HIST1H2AJ PE=1 SV=3
57	Q93077	130	14105	11.1	4	154.274034	2	12.3	0	0	0	0	Histone H2A type 1-C OS=Homo sapiens GN=HIST1H2AC PE=1 SV=3
58	Q71UI9	128	13509	10.6	4	154.274034	2	12.5	0	0	0	0	Histone H2A.V OS=Homo sapiens GN=H2AFV PE=1 SV=3
59	P0C0S8	130	14091	10.9	4	154.274034	2	12.3	0	0	0	0	Histone H2A type 1 OS=Homo sapiens GN=HIST1H2AG PE=1 SV=2
60	P0C0S5	128	13553	10.6	4	154.274034	2	12.5	0	0	0	0	Histone H2A.Z OS=Homo sapiens GN=H2AFZ PE=1 SV=2

#	Accession												Description
61	P16104	143	15145	10.7	4	154.274034	2	11.2	0	0	0	0	Histone H2A.x OS=Homo sapiens GN=H2AFX PE=1 SV=2
62	Q96QV6	131	14233	10.9	4	154.274034	2	12.2	0	0	0	0	Histone H2A type 1-A OS=Homo sapiens GN=HIST1H2AA PE=1 SV=3
63	Q6FI13	130	14095	10.9	4	154.274034	2	12.3	0	0	0	0	Histone H2A type 2-A OS=Homo sapiens GN=HIST2H2AA3 PE=1 SV=3
64	P20671	130	14107	10.9	4	154.274034	2	12.3	0	0	0	0	Histone H2A type 1-D OS=Homo sapiens GN=HIST1H2AD PE=1 SV=2
65	Q16777	129	13988	10.9	4	154.274034	2	12.4	0	0	0	0	Histone H2A type 2-C OS=Homo sapiens GN=HIST2H2AC PE=1 SV=4
66	Q7L7L0	130	14121	11.1	4	154.274034	2	12.3	0	0	0	0	Histone H2A type 3 OS=Homo sapiens GN=HIST3H2A PE=1 SV=3
67	Q9BTM1	129	14019	10.9	4	154.274034	2	12.4	0	0	0	0	Histone H2A.J OS=Homo sapiens GN=H2AFJ PE=1 SV=1
68	Q96KK5	128	13906	10.9	4	154.274034	2	12.5	0	0	0	0	Histone H2A type 1-H OS=Homo sapiens GN=HIST1H2AH PE=1 SV=3
69	P04908	130	14135	11.1	4	154.274034	2	12.3	0	0	0	0	Histone H2A type 1-B/E OS=Homo sapiens GN=HIST1H2AB PE=1 SV=2
70	Q16778	126	13920	10.3	3	119.654206	3	16.7	0	0	0	0	Histone H2B type 2-E OS=Homo sapiens GN=HIST2H2BE PE=1 SV=3
71	P33778	126	13950	10.3	3	119.654206	3	16.7	0	0	0	0	Histone H2B type 1-B OS=Homo sapiens GN=HIST1H2BB PE=1 SV=2
72	Q8N257	126	13908	10.3	3	119.654206	3	16.7	0	0	0	0	Histone H2B type 3-B OS=Homo sapiens GN=HIST3H2BB PE=1 SV=3
73	P06899	126	13904	10.3	3	119.654206	3	16.7	0	0	0	0	Histone H2B type 1-J OS=Homo sapiens GN=HIST1H2BJ PE=1 SV=3
74	P23527	126	13906	10.3	3	119.654206	3	16.7	0	0	0	0	Histone H2B type 1-O OS=Homo sapiens GN=HIST1H2BO PE=1 SV=3
75	P16401	226	22580	10.9	5	111.183112	3	14.2	0	0	0	0	Histone H1.5 OS=Homo sapiens GN=HIST1H1B PE=1 SV=3
76	P84243	136	15328	11.3	3	110.856103	2	13.2	0	0	0	0	Histone H3.3 OS=Homo sapiens GN=H3F3A PE=1 SV=2
77	Q71DI3	136	15388	11.3	3	110.856103	2	13.2	0	0	0	0	Histone H3.2 OS=Homo sapiens GN=HIST2H3A PE=1 SV=3
78	Q16695	136	15508	11.1	3	110.856103	2	13.2	0	0	0	0	Histone H3.1t OS=Homo sapiens GN=HIST3H3 PE=1 SV=3
79	P68431	136	15404	11.1	3	110.856103	2	13.2	0	0	0	0	Histone H3.1 OS=Homo sapiens GN=HIST1H3A PE=1 SV=2
80	Q99877	126	13922	10.3	2	79.7694708	2	15.9	0	0	0	0	Histone H2B type 1-N OS=Homo sapiens GN=HIST1H2BN PE=1 SV=3
81	Q93079	126	13892	10.3	2	79.7694708	2	15.9	0	0	0	0	Histone H2B type 1-H OS=Homo sapiens GN=HIST1H2BH PE=1 SV=3
82	Q99880	126	13952	10.3	2	79.7694708	2	15.9	0	0	0	0	Histone H2B type 1-L OS=Homo sapiens GN=HIST1H2BL PE=1 SV=3
83	O60814	126	13890	10.3	2	79.7694708	2	15.9	0	0	0	0	Histone H2B type 1-K OS=Homo sapiens GN=HIST1H2BK PE=1 SV=3
84	P58876	126	13936	10.3	2	79.7694708	2	15.9	0	0	0	0	Histone H2B type 1-D OS=Homo sapiens GN=HIST1H2BD PE=1 SV=2
85	P57053	126	13944	10.4	2	79.7694708	2	15.9	0	0	0	0	Histone H2B type F-S OS=Homo sapiens GN=H2BFS PE=1 SV=2
86	Q99879	126	13989	10.3	2	79.7694708	2	15.9	0	0	0	0	Histone H2B type 1-M OS=Homo sapiens GN=HIST1H2BM PE=1 SV=3
87	Q5QNW6	126	13920	10.3	2	79.7694708	2	15.9	0	0	0	0	Histone H2B type 2-F OS=Homo sapiens GN=HIST2H2BF PE=1 SV=3
88	P62807	126	13906	10.3	2	79.7694708	2	15.9	0	0	0	0	Histone H2B type 1-C/E/F/G/I OS=Homo sapiens GN=HIST1H2BC PE=1 SV=4
89	P07910	306	33670	5.1	4	65.6925053	4	11.4	0	0	0	0	Heterogeneous nuclear ribonucleoproteins C1/C2 OS=Homo sapiens GN=HNRNPC PE=1 SV=4
90	O00571	662	73244	7.2	8	60.7308357	8	14.5	0	0	0	0	ATP-dependent RNA helicase DDX3X OS=Homo sapiens GN=DDX3X PE=1 SV=3
91	O00148	427	49130	5.7	5	58.8463309	5	11.5	0	0	0	0	ATP-dependent RNA helicase DDX39 OS=Homo sapiens GN=DDX39 PE=1 SV=2
92	Q9NY12	217	22348	10.9	2	46.3177572	2	9.2	0	0	0	0	H/ACA ribonucleoprotein complex subunit 1 OS=Homo sapiens GN=GAR1 PE=1 SV=1
93	P23246	707	76150	9.4	6	42.649024	6	11.7	0	0	0	0	Splicing factor, proline- and glutamine-rich OS=Homo sapiens GN=SFPQ PE=1 SV=2
94	Q08211	1270	140958	6.8	10	39.5706823	10	9	0	0	0	0	ATP-dependent RNA helicase A OS=Homo sapiens GN=DHX9 PE=1 SV=4
95	Q13310	644	70783	9.3	5	39.0176759	5	8.5	0	0	0	0	Polyadenylate-binding protein 4 OS=Homo sapiens GN=PABPC4 PE=1 SV=1
96	Q92841	650	72372	8.6	5	38.6575128	5	9.2	0	0	0	0	Probable ATP-dependent RNA helicase DDX17 OS=Homo sapiens GN=DDX17 PE=1 SV=1
97	Q9NR30	783	87344	9.3	4	25.6729331	4	5	0	0	0	0	Nucleolar RNA helicase 2 OS=Homo sapiens GN=DDX21 PE=1 SV=5
98	Q15393	1217	135577	5.3	5	20.6469871	5	5.3	0	0	0	0	Splicing factor 3B subunit 3 OS=Homo sapiens GN=SF3B3 PE=1 SV=4
99	Q12874	501	58849	5.4	2	20.0617831	2	6	0	0	0	0	Splicing factor 3A subunit 3 OS=Homo sapiens GN=SF3A3 PE=1 SV=1
100	Q9UMS4	504	55181	6.6	2	19.9423677	2	4.6	0	0	0	0	Pre-mRNA-processing factor 19 OS=Homo sapiens GN=PRPF19 PE=1 SV=1
101	Q15029	972	109436	5	3	15.5107304	2	2.3	0	0	0	0	116 kDa U5 small nuclear ribonucleoprotein component OS=Homo sapiens GN=EFTUD2 PE=1 SV=1
102	O75533	1304	145830	7.1	4	15.4155726	4	4.1	0	0	0	0	Splicing factor 3B subunit 1 OS=Homo sapiens GN=SF3B1 PE=1 SV=3
103	Q6P2Q9	2335	273599	8.8	5	10.761192	5	2	0	0	0	0	Pre-mRNA-processing-splicing factor 8 OS=Homo sapiens GN=PRPF8 PE=1 SV=2
104	O75643	2136	244505	6.1	4	9.41100498	4	1.9	0	0	0	0	U5 small nuclear ribonucleoprotein 200 kDa helicase OS=Homo sapiens GN=SNRNP200 PE=1 SV=2
105	P57721	339	35938	8.1	0	0	0	0	19	611.5638874	6	21.5	Poly(rC)-binding protein 3 OS=Homo sapiens GN=PCBP3 PE=1 SV=1
106	Q9Y580	266	30503	9.6	0	0	0	0	11	451.2310518	9	35.7	RNA-binding protein 7 OS=Homo sapiens GN=RBM7 PE=1 SV=1
107	P62310	102	11845	4.7	0	0	0	0	4	427.9053825	3	22.5	U6 snRNA-associated Sm-like protein LSm3 OS=Homo sapiens GN=LSM3 PE=1 SV=2
108	Q9UK45	103	11602	5.3	0	0	0	0	4	423.7509613	2	15.5	U6 snRNA-associated Sm-like protein LSm7 OS=Homo sapiens GN=LSM7 PE=1 SV=1
109	Q7Z2W4	902	101431	8.4	0	0	0	0	35	423.3986185	32	42.6	Zinc finger CCCH-type antiviral protein 1 OS=Homo sapiens GN=ZC3HAV1 PE=1 SV=3
110	Q6NZY4	707	78577	4.9	0	0	0	0	26	401.2747788	23	36.5	Zinc finger CCHC domain-containing protein 8 OS=Homo sapiens GN=ZCCHC8 PE=1 SV=2
111	Q9Y4Z0	139	15350	10	0	0	0	0	5	392.5031386	5	38.1	U6 snRNA-associated Sm-like protein LSm4 OS=Homo sapiens GN=LSM4 PE=1 SV=1
112	P49458	86	10112	8	0	0	0	0	3	380.6367647	3	34.9	Signal recognition particle 9 kDa protein OS=Homo sapiens GN=SRP9 PE=1 SV=2
113	Q14671	1186	126473	6.8	0	0	0	0	38	349.6124078	37	34.8	Pumilio homolog 1 OS=Homo sapiens GN=PUM1 PE=1 SV=3
114	O75534	798	88885	6.2	0	0	0	0	21	287.147033	21	29.1	Cold shock domain-containing protein E1 OS=Homo sapiens GN=CSDE1 PE=1 SV=2
115	P83731	157	17779	11.3	0	0	0	0	4	278.002223	4	25.5	60S ribosomal protein L24 OS=Homo sapiens GN=RPL24 PE=1 SV=1
116	Q15370	118	13133	4.9	0	0	0	0	3	277.4132353	3	35.6	Transcription elongation factor B polypeptide 2 OS=Homo sapiens GN=TCEB2 PE=1 SV=1
117	Q15717	326	36092	9.2	0	0	0	0	8	267.7690124	8	35.9	ELAV-like protein 1 OS=Homo sapiens GN=ELAVL1 PE=1 SV=2
118	Q15056	248	27385	7.2	0	0	0	0	6	263.9900142	6	25.8	Eukaryotic translation initiation factor 4H OS=Homo sapiens GN=EIF4H PE=1 SV=5
119	Q01085	375	41591	7.8	0	0	0	0	9	261.8780941	9	32.8	Nucleolysin TIAR OS=Homo sapiens GN=TIAL1 PE=1 SV=1
120	O76071	339	37840	5	0	0	0	0	8	257.5005842	8	32.7	Probable cytosolic iron-sulfur protein assembly protein CIAO1 OS=Homo sapiens GN=CIAO1 PE=1 SV=1
121	Q9H3K6	86	10117	6.5	0	0	0	0	2	253.7578431	2	29.1	BolA-like protein 2 OS=Homo sapiens GN=BOLA2 PE=1 SV=1
122	O00425	579	63705	8.9	0	0	0	0	13	244.9924599	12	24.4	Insulin-like growth factor 2 mRNA-binding protein 3 OS=Homo sapiens GN=IGF2BP3 PE=1 SV=2
123	P63165	101	11557	5.5	0	0	0	0	2	216.0710347	2	20.8	Small ubiquitin-related modifier 1 OS=Homo sapiens GN=SUMO1 PE=1 SV=1
124	Q96I24	572	61640	8.4	0	0	0	0	11	209.8382164	10	17.7	Far upstream element-binding protein 3 OS=Homo sapiens GN=FUBP3 PE=1 SV=2
125	Q5RKV6	272	28235	6.3	0	0	0	0	5	200.580648	4	23.9	Exosome complex exonuclease MTR3 OS=Homo sapiens GN=EXOSC6 PE=1 SV=1
126	Q96EP5	407	43383	8.6	0	0	0	0	7	187.6685768	6	24.6	DAZ-associated protein 1 OS=Homo sapiens GN=DAZAP1 PE=1 SV=1
127	Q13868	293	32789	7.5	0	0	0	0	5	186.2045606	5	21.8	Exosome complex exonuclease RRP4 OS=Homo sapiens GN=EXOSC2 PE=1 SV=2
128	P34896	483	53083	7.7	0	0	0	0	8	180.7302237	8	20.1	Serine hydroxymethyltransferase, cytosolic OS=Homo sapiens GN=SHMT1 PE=1 SV=1

#	Accession	Length	MW	pI					Count	Score	Count2	Cov	Description
129	Q9UHX1	559	59876	5.3	0	0	0	0	9	175.6785068	9	24.5	Poly(U)-binding-splicing factor PUF60 OS=Homo sapiens GN=PUF60 PE=1 SV=1
130	Q5UI30	125	14199	5.3	0	0	0	0	2	174.5853961	2	20	tRNA methyltransferase 112 homolog OS=Homo sapiens GN=TRMT112 PE=1 SV=1
131	Q96PZ0	661	75035	6.4	0	0	0	0	10	165.076963	10	20.1	Pseudouridylate synthase 7 homolog OS=Homo sapiens GN=PUS7 PE=1 SV=2
132	O15116	133	15179	5.2	0	0	0	0	2	164.0840189	2	30.8	U6 snRNA-associated Sm-like protein LSm1 OS=Homo sapiens GN=LSM1 PE=1 SV=1
133	O43414	337	37238	8.1	0	0	0	0	5	161.892986	5	17.2	ERI1 exoribonuclease 3 OS=Homo sapiens GN=ERI3 PE=1 SV=2
134	O43399	206	22238	5.4	0	0	0	0	3	158.9066105	3	18.4	Tumor protein D54 OS=Homo sapiens GN=TPD52L2 PE=1 SV=2
135	Q16630	551	59210	7.2	0	0	0	0	8	158.4259492	7	16.3	Cleavage and polyadenylation specificity factor subunit 6 OS=Homo sapiens GN=CPSF6 PE=1 SV=2
136	P07951	284	32851	4.7	0	0	0	0	4	153.6843275	4	15.8	Tropomyosin beta chain OS=Homo sapiens GN=TPM2 PE=1 SV=1
137	P47813	144	16460	5.2	0	0	0	0	2	151.549823	2	14.6	Eukaryotic translation initiation factor 1A, X-chromosomal OS=Homo sapiens GN=EIF1AX PE=1 SV=2
138	Q15024	291	31835	5.2	0	0	0	0	4	149.9874537	4	18.6	Exosome complex exonuclease RRP42 OS=Homo sapiens GN=EXOSC7 PE=1 SV=2
139	P34096	147	16840	9	0	0	0	0	2	148.4569694	2	17	Ribonuclease 4 OS=Homo sapiens GN=RNASE4 PE=1 SV=3
140	Q9BXS6	441	49452	9.9	0	0	0	0	6	148.4569694	5	15	Nucleolar and spindle-associated protein 1 OS=Homo sapiens GN=NUSAP1 PE=1 SV=1
141	Q13242	221	25542	8.6	0	0	0	0	3	148.1210939	3	14	Splicing factor, arginine/serine-rich 9 OS=Homo sapiens GN=SFRS9 PE=1 SV=1
142	Q9Y6H1	151	15513	9.2	0	0	0	0	2	144.5243345	2	15.2	Coiled-coil-helix-coiled-coil-helix domain-containing protein 2, mitochondrial OS=Homo sapiens GN=CHCHD2 PE=1 SV=1
143	Q9UHV9	154	16648	6.6	0	0	0	0	2	141.7089254	2	12.3	Prefoldin subunit 2 OS=Homo sapiens GN=PFDN2 PE=1 SV=1
144	Q9Y2L1	958	109003	7.1	0	0	0	0	12	136.6795898	12	18.5	Exosome complex exonuclease RRP44 OS=Homo sapiens GN=DIS3 PE=1 SV=2
145	Q8IYB7	885	99210	6	0	0	0	0	11	135.6242483	11	12.7	DIS3-like exonuclease 2 OS=Homo sapiens GN=DIS3L2 PE=1 SV=3
146	Q9H2H8	161	18155	6.8	0	0	0	0	2	135.5476677	2	15.5	Peptidyl-prolyl cis-trans isomerase-like 3 OS=Homo sapiens GN=PPIL3 PE=1 SV=1
147	Q92879	486	52063	8.5	0	0	0	0	6	134.7109537	6	14.4	CUGBP Elav-like family member 1 OS=Homo sapiens GN=CELF1 PE=1 SV=2
148	Q9Y3C6	166	18237	8	0	0	0	0	2	131.4649067	2	14.5	Peptidyl-prolyl cis-trans isomerase-like 1 OS=Homo sapiens GN=PPIL1 PE=1 SV=1
149	Q8IV48	349	40064	6.7	0	0	0	0	4	125.061172	4	13.5	3'-5' exoribonuclease 1 OS=Homo sapiens GN=ERI1 PE=1 SV=3
150	P15927	270	29247	6.1	0	0	0	0	3	121.2398584	3	20.7	Replication protein A 32 kDa subunit OS=Homo sapiens GN=RPA2 PE=1 SV=1
151	Q96B26	276	30040	5.3	0	0	0	0	3	118.6042093	3	12	Exosome complex exonuclease RRP43 OS=Homo sapiens GN=EXOSC8 PE=1 SV=1
152	P31483	386	42963	7.8	0	0	0	0	4	113.073443	4	11.1	Nucleolysin TIA-1 isoform p40 OS=Homo sapiens GN=TIA1 PE=1 SV=3
153	P31350	389	44878	5.4	0	0	0	0	4	112.2014113	4	14.1	Ribonucleoside-diphosphate reductase subunit M2 OS=Homo sapiens GN=RRM2 PE=1 SV=1
154	Q08170	494	56678	11.5	0	0	0	0	5	110.4411665	5	8.1	Splicing factor, arginine/serine-rich 4 OS=Homo sapiens GN=SFRS4 PE=1 SV=2
155	Q9BUL9	199	20632	9.6	0	0	0	0	2	109.6641935	2	12.1	Ribonuclease P protein subunit p25 OS=Homo sapiens GN=RPP25 PE=1 SV=1
156	O15347	200	22980	8.4	0	0	0	0	2	109.1158725	2	14.5	High mobility group protein B3 OS=Homo sapiens GN=HMGB3 PE=1 SV=4
157	Q9GZS3	305	33581	5.5	0	0	0	0	3	107.3270877	3	14.4	WD repeat-containing protein 61 OS=Homo sapiens GN=WDR61 PE=1 SV=1
158	Q86TB9	770	86850	6.7	0	0	0	0	7	99.19624776	7	12.1	Protein PAT1 homolog 1 OS=Homo sapiens GN=PATL1 PE=1 SV=2
159	O95218	330	37404	10	0	0	0	0	3	99.19624776	3	11.8	Zinc finger Ran-binding domain-containing protein 2 OS=Homo sapiens GN=ZRANB2 PE=1 SV=2
160	Q9NSI2	230	25456	11.1	0	0	0	0	2	94.88336742	2	8.3	Uncharacterized protein C21orf70 OS=Homo sapiens GN=C21orf70 PE=1 SV=2
161	A0AV96	593	64067	7.7	0	0	0	0	5	92.00326521	5	10.1	RNA-binding protein 47 OS=Homo sapiens GN=RBM47 PE=1 SV=1
162	P28072	239	25358	4.9	0	0	0	0	2	91.31035359	2	8.4	Proteasome subunit beta type-6 OS=Homo sapiens GN=PSMB6 PE=1 SV=4
163	Q01081	240	27872	8.8	0	0	0	0	2	90.92989378	2	10.8	Splicing factor U2AF 35 kDa subunit OS=Homo sapiens GN=U2AF1 PE=1 SV=3
164	P28340	1107	123631	7	0	0	0	0	9	88.71209149	8	8.3	DNA polymerase delta catalytic subunit OS=Homo sapiens GN=POLD1 PE=1 SV=2
165	Q8NDH3	523	55861	6.9	0	0	0	0	4	83.45382221	4	10.3	Probable aminopeptidase NPEPL1 OS=Homo sapiens GN=NPEPL1 PE=1 SV=3
166	O43143	795	90933	7.5	0	0	0	0	6	82.35160191	6	8.2	Putative pre-mRNA-splicing factor ATP-dependent RNA helicase DHX15 OS=Homo sapiens GN=DHX15 PE=1 SV=2
167	P10155	538	60671	8	0	0	0	0	4	81.12704278	4	10.4	60 kDa SS-A/Ro ribonucleoprotein OS=Homo sapiens GN=TROVE2 PE=1 SV=1
168	Q07666	443	48227	8.7	0	0	0	0	3	73.89336741	3	7.2	KH domain-containing, RNA-binding, signal transduction-associated protein 1 OS=Homo sapiens GN=KHDRBS1 PE=1 SV=1
169	Q92804	592	61830	8	0	0	0	0	4	73.7269409	3	5.2	TATA-binding protein-associated factor 2N OS=Homo sapiens GN=TAF15 PE=1 SV=1
170	Q9UNP9	301	33431	5.6	0	0	0	0	2	72.50224089	2	12.3	Peptidyl-prolyl cis-trans isomerase E OS=Homo sapiens GN=PPIE PE=1 SV=1
171	Q86U42	306	32749	5.1	0	0	0	0	2	71.31756375	2	6.9	Polyadenylate-binding protein 2 OS=Homo sapiens GN=PABPN1 PE=1 SV=3
172	P60510	307	35080	5.1	0	0	0	0	2	71.08525898	2	7.5	Serine/threonine-protein phosphatase 4 catalytic subunit OS=Homo sapiens GN=PPP4C PE=1 SV=1
173	Q6UXN9	313	35079	7.7	0	0	0	0	2	69.72260226	2	6.7	WD repeat-containing protein 82 OS=Homo sapiens GN=WDR82 PE=1 SV=1
174	P49321	788	85238	4.3	0	0	0	0	5	69.23595973	5	10	Nuclear autoantigenic sperm protein OS=Homo sapiens GN=NASP PE=1 SV=2
175	Q15293	331	38890	5	0	0	0	0	2	65.93104081	2	6.9	Reticulocalbin-1 OS=Homo sapiens GN=RCN1 PE=1 SV=1
176	Q9ULR0	331	37566	5.8	0	0	0	0	2	65.93104081	2	8.2	Pre-mRNA-splicing factor ISY1 homolog OS=Homo sapiens GN=ISY1 PE=1 SV=2
177	Q8TB72	1066	114216	7.1	0	0	0	0	6	61.41606334	6	6	Pumilio homolog 2 OS=Homo sapiens GN=PUM2 PE=1 SV=2
178	O00303	357	37564	5.4	0	0	0	0	2	61.12934036	2	5	Eukaryotic translation initiation factor 3 subunit F OS=Homo sapiens GN=EIF3F PE=1 SV=1
179	O94992	359	40623	4.9	0	0	0	0	2	60.78878693	2	7.8	Protein HEXIM1 OS=Homo sapiens GN=HEXIM1 PE=1 SV=1
180	Q13619	759	87680	8.1	0	0	0	0	4	57.50507117	4	6.6	Cullin-4A OS=Homo sapiens GN=CUL4A PE=1 SV=3
181	Q9NZI8	577	63457	9.2	0	0	0	0	3	56.73268936	2	4.5	Insulin-like growth factor 2 mRNA-binding protein 1 OS=Homo sapiens GN=IGF2BP1 PE=1 SV=1
182	Q9BY77	421	46089	10	0	0	0	0	2	51.83651902	2	7.1	Polymerase delta-interacting protein 3 OS=Homo sapiens GN=POLDIP3 PE=1 SV=2
183	Q15637	639	68330	9	0	0	0	0	3	51.22810917	3	8.6	Splicing factor 1 OS=Homo sapiens GN=SF1 PE=1 SV=4
184	Q06265	439	48949	5.3	0	0	0	0	2	49.71110366	2	4.8	Exosome complex exonuclease RRP45 OS=Homo sapiens GN=EXOSC9 PE=1 SV=3
185	Q9UMR2	479	53927	6.3	0	0	0	0	2	45.60747023	2	4.8	ATP-dependent RNA helicase DDX19B OS=Homo sapiens GN=DDX19B PE=1 SV=1
186	Q9NUU7	478	53975	6.6	0	0	0	0	2	45.60747023	2	4.8	ATP-dependent RNA helicase DDX19A OS=Homo sapiens GN=DDX19A PE=1 SV=1
187	Q5VYS8	1495	171228	6.8	0	0	0	0	6	43.79232343	6	5.9	Terminal uridylyltransferase 7 OS=Homo sapiens GN=ZCCHC6 PE=1 SV=1
188	P08651	508	55675	8.4	0	0	0	0	2	42.95900494	2	4.7	Nuclear factor 1 C-type OS=Homo sapiens GN=NFIC PE=1 SV=2
189	Q8WVV9	542	60083	7.7	0	0	0	0	2	40.26415961	2	5	Heterogeneous nuclear ribonucleoprotein L-like OS=Homo sapiens GN=HNRPLL PE=1 SV=1
190	Q96JP5	570	63445	7.4	0	0	0	0	2	38.28627107	2	6.5	Zinc finger protein 91 homolog OS=Homo sapiens GN=ZFP91 PE=1 SV=1
191	Q8WXI9	593	65261	9.7	0	0	0	0	2	36.80130608	2	4.7	Transcriptional repressor p66-beta OS=Homo sapiens GN=GATAD2B PE=1 SV=1
192	P43246	934	104743	5.8	0	0	0	0	3	35.0479248	3	3.4	DNA mismatch repair protein Msh2 OS=Homo sapiens GN=MSH2 PE=1 SV=1
193	Q93009	1102	128302	5.6	0	0	0	0	3	29.70486548	3	3.4	Ubiquitin carboxyl-terminal hydrolase 7 OS=Homo sapiens GN=USP7 PE=1 SV=2
194	P49790	1475	153938	8.7	0	0	0	0	4	29.59074509	4	3.7	Nuclear pore complex protein Nup153 OS=Homo sapiens GN=NUP153 PE=1 SV=2
195	O60306	1485	171294	6.4	0	0	0	0	4	29.39148082	4	3.3	Intron-binding protein aquarius OS=Homo sapiens GN=AQR PE=1 SV=4
196	Q9NZB2	1118	121888	8.9	0	0	0	0	3	29.27975113	3	3.6	Constitutive coactivator of PPAR-gamma-like protein 1 OS=Homo sapiens GN=FAM120A PE=1 SV=2
197	P47712	749	85211	5.4	0	0	0	0	2	29.13641456	2	2.8	Cytosolic phospholipase A2 OS=Homo sapiens GN=PLA2G4A PE=1 SV=1
198	Q01780	885	100831	8.5	0	0	0	0	2	24.65895425	2	3.1	Exosome component 10 OS=Homo sapiens GN=EXOSC10 PE=1 SV=2
199	P51532	1647	184644	7.9	0	0	0	0	2	13.25025775	2	1.5	Transcription activator BRG1 OS=Homo sapiens GN=SMARCA4 PE=1 SV=2
200	P51610	2035	208730	7.5	0	0	0	0	2	10.72391868	2	1.4	Host cell factor 1 OS=Homo sapiens GN=HCFC1 PE=1 SV=2
201	Q9UQ35	2752	299616	12.1	0	0	0	0	2	7.929932597	2	1.2	Serine/arginine repetitive matrix protein 2 OS=Homo sapiens GN=SRRM2 PE=1 SV=2