Investigating 3′ Splice Site Recognition by the Pre-Spliceosome

by

Jeffrey Dean McDonald

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

Department of Biochemistry
University of Alberta

Abstract:

Eukaryotes feature a split gene structure in which pre-mRNAs are spliced into mature RNAs prior to the translation of proteins. The splicing process is catalyzed by a dynamic multi-subunit molecular machine known as the spliceosome that assembles in a stepwise fashion. An early complex formed in the spliceosome assembly pathway is the pre-spliceosome, otherwise referred to as the A complex.

The A complex factors SF3B1, SF3B6 and U2AF play a role in the recognition of important consensus sequences at the 3′ splice sites of pre-mRNA introns. SF3B1 and SF3B6 interact about the branch point and the U2AF heterodimer binds the polypyrimidine tract and 3′ splice site. The molecular mechanism of branch point and splice site selection by these factors remains unclear. SF3B6 has been biochemically suggested to directly bind the branch point adenosine, yet structural characterizations of spliceosomes have placed it in a position peripheral to this adenosine. These observed positional inconsistencies may reflect the dynamic nature of splicing and spliceosome assembly. And although essential in humans, SF3B6 may be deleted from the model organism *Schizosaccharomyces pombe* to yield a phenotypically distinct strain. To carry out splicing in SF3B6 knockout cells, spliceosomes must be assembling in the absence of SF3B6.

This thesis aimed to investigate the molecular mechanism of branch point and 3′ splice site selection by SF3B1, SF3B6 and U2AF in the pre-spliceosome. A series of SF3B1-SF3B6 and U2AF expression constructs were analyzed for the *in vitro* reconstitution of a ternary complex suitable for structural characterization and pre-mRNA binding analysis. Furthermore, biochemical characterization of the A complex in this thesis has demonstrated the formation of an SF3B6 deficient A complex. Preliminary work has also been conducted towards an affinity

purification of the A complex and toward the site-specific derivatization of SF3B6 to chemically map pre-spliceosome RNA and protein contacts.

Table of Contents

***List of Figures*:**

Chapter 1.

Chapter 2.

Chapter 3.

Chapter 5.

***List of Tables*:**

Chapter 3.

***List of Abbreviations*:**

Δ	delta, depleted
A complex	pre-spliceosome
ATP	adenosine triphosphate
B complex	pre-catalytic spliceosome
B^{act} complex	activated spliceosome
BBP	branch point binding protein
BME	β-mercaptoethanol
BPS	branch point sequence
BSL	branch point-interacting stem-loop
C complex	catalytic spliceosome
C. albicans	*Candida albicans*
CIAP	calf intestinal alkaline phosphatase
Cryo-EM	cryo-electron microscopy
DEPC	diethylpyrocarbonate
DMSO	dimethyl sulfoxide
DTT	dithiothreitol
E complex	early spliceosome
EDTA	ethylenediaminetetraacetic acid
EMSA	electrophoretic mobility shift assay
GST	glutathione S-transferase
His_6	polyhistidine tag
IPTG	isopropyl-β-d-thiogalactopyranoside
K_d	dissociation constant
LB	Luria-Bertani broth
MPA	mercaptopropionic acid
N	any nucleotide

Ni-NTA	nickel-nitrilotriacetic acid
NTP	nucleotide triphosphate
OD	optical density
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PMSF	phenylmethylsulfonyl fluoride
PPT	polypyrimidine tract
pre-mRNA	precursor messenger RNA
R	purine
RRM	RNA recognition motif
RS	serine-arginine-rich
RT	reverse transcriptase
S. cerevisiae	*Saccharomyces cerevisiae*
S. pombe	*Schizosaccharomyces pombe*
S-200	Superdex S-200
S-500	Sephacryl S-500
SAXS	small angle X-ray scattering
SDS-PAGE	Sodium dodecyl-sulfate polyacrylamide electrophoresis
SELEX	systematic evolution of ligands by exponential enrichment
SF1	splicing factor 1
SF3A	splicing factor 3A
SF3B	splicing factor 3B
snRNA	small nuclear RNA
snRNP	small nuclear ribonucleoprotein
SS	splice site
TBE	Tris-borate-ethylenediaminetetraacetic acid
TEV	Tobacco Etch Virus

tRNA	transfer RNA
Trp	tryptophan
U2AF	U2 auxiliary factor
UHM	U2AF homology motif
ULM	U2AF ligand motif
W	tryptophan
WT	wild type
Y	pyrimidine
ZF	zinc finger

Chapter 1

Introduction

Chapter 1. Introduction.

Pre-mRNA Splicing and Intronic Consensus Sequences.

RNA messages encoded by DNA are translated for the synthesis of cellular proteins. Eukaryotes, including humans, have genes structured in which protein coding sequences (exons) of pre-messenger RNA (pre-mRNA) are punctuated by non-coding sequences (introns). This split gene structure was first realized when electron micrographs of adenovirus mRNA-genomic DNA hybrids revealed portions of single-stranded looped out DNA (Berget *et al.* 1977; Chow *et al.* 1977) corresponding to excised introns. Consequently, the term "splicing" (Berget *et al.* 1977) was coined to describe the process of pre-mRNA intron removal and exon ligation to yield a mature mRNA molecule.

Specifically defining the points at which introns are removed and exons are ligated, introns are marked with 5′ and 3′ splice sites (SSs). The 5′ SS is defined by a conserved GU dinucleotide (Wahl *et al.* 2009) within organism-specific consensus sequences such as GURAGU (metazoans), or GUAAGU (budding yeast) (where the GU dinucleotide is underlined; R=purine). The 3′ half of an intron can be defined by three sequence elements known as the branch point sequence (BPS), the polypyrimidine tract (PPT) and the 3′ SS (Figure 1-1). The BPS in humans, YUNAY (where the branch point adenosine is underlined, Y=pyrimidine, N=any nucleotide), is highly degenerate and its recognition remains to be extensively characterized (Gao *et al.* 2008). Alternatively, *Saccharomyces cerevisiae* (*S.* cerevisiae) features a strictly conserved branch point sequence of UACUAAC (Gao *et al.* 2008). The 3′ SS is defined by an AG dinucleotide, which is conserved from yeast to metazoans (Wahl *et al.* 2009). Intervening the BPS and the 3′ SS lies the PPT, a pyrimidine-rich region, which serves to modulate BPS and 3′ SS recognition and utilization via splicing factor binding (House and Lynch, 2008; Matlin *et al.* 2005).

Stepwise Assembly of the Spliceosome.

Pre-mRNA splicing, which can be chemically described as two sequential transesterification reactions (Figure 1-2), is catalyzed by a dynamic multi-subunit molecular machine known as the spliceosome (reviewed by Will and Lührmann, 2011). The spliceosome is composed of the U1,

U2, U4, U5 and U6 small nuclear ribonucleoprotein (snRNP) particles, as well as numerous additional factors (Zhou *et al.* 2002). In a well conserved process from yeast to humans (Burge *et al.* 1999), the splicing reaction involves an assembly pathway dependent on the consecutive formation of several distinct complexes (Figure 1-3) followed by the two chemical steps.

Generally, the first complex to form during the spliceosome assembly pathway, known as the E complex, is responsible for the recognition of the 5′ SS and 3′ SS of an intron to be spliced from a pre-mRNA transcript. Spliceosome assembly begins with the association of the U1 snRNP and a pre-mRNA via base pairing interactions between the 5′ end of the RNA component of the snRNP, the U1 snRNA, and the 5′ SS (reviewed by Wahl *et al.* 2009). This RNA-mediated interaction of the U1 snRNP and the 5′ SS is stabilized by non-snRNP factors in higher eukaryotes, namely numerous arginine-serine-rich (RS) proteins. Next, E complex assembly finishes with splicing factor 1 (SF1)/branch point binding protein (BBP) binding to the BPS, which includes the branch point adenosine, as well as U2 auxiliary factor (U2AF) binding to the PPT and 3′ SS.

Following the formation of the spliceosomal E complex, the next discrete complex to be formed is the A complex. Displacing SF1/BBP, the U2 snRNP binds to the BPS mediated by an snRNA base pairing in an adenosine triphosphate (ATP)-dependent manner (reviewed by Wahl *et al.* 2009) which is initially stabilized by the U2 snRNP sub-complexes splicing factor 3A (SF3A) and splicing factor 3B (SF3B) as well as by the RS domain of U2AF65 (Valcarcel *et al.* 1996). In higher eukaryotes, U2 association with the BPS is also stabilized by an interaction with the SF3B subunit SF3B6 (also p14, SF3B14; hereafter SF3B6), which also binds at the branch point adenosine. Subsequent to SF1 dissociation, the N-terminal half of SF3B1 (also SF3B155, sap155; hereafter SF3B1) binds the third and most C-terminal RNA recognition motif (RRM) of U2AF65 (Gozani *et al.* 1998), which has since been classified as a U2AF homology motif (UHM) (reviewed by Kielkopf *et al.* 2004). Given the BPS degeneracy found in metazoans, this SF3B1-U2AF65 interaction may play a role in the recruitment and/or positioning of the U2 snRNP about the branch point (Gozani *et al.* 1998). Furthermore, SF3B6 forms a tight interaction with SF3B1 (Will *et al.* 2001), for which a minimal complex consisting of SF3B6 and residues 373-415 of SF3B1 has been structurally characterized by X-ray crystallography (Schellenberg *et al.* 2006). Thus, the possibly concomitant SF3B6-SF3B1, SF3B1-U2AF65 and

U2AF heterodimer interactions (Figure 1-4) may have important functional implications in SS selection, A complex formation and the proper processing of a pre-mRNA to eventually yield a functional gene product. Although the splicing process is fundamentally relevant to health, with mutations affecting splicing responsible for ~50% of all known disease-causing mutations (Hammond and Wood 2011), the molecular mechanism of SS selection remains poorly understood.

Prior to the first transesterification reaction, but after the formation of the A complex, the spliceosomal B complex assembles by the association of the U4/U6·U5 tri-snRNP with the pre-mRNA (reviewed by Wahl *et al.* 2009). The U1 and U4 snRNPs are then released, yielding the activated spliceosome (B^{act}), which can then carry out the first transesterification reaction. The first transesterification reaction yields the C complex, which undergoes conformational changes (Konarska *et al.* 2006) before catalyzing the second transesterification and ultimately spliceosome dissociation.

Spliceosome Dynamics.

Advances in cryo-electron microscopy (cryo-EM) have enabled structural investigations of large heteromeric spliceosome assemblies, giving rise to snapshots of spliceosomal complexes, such as that of the B^{act} complex (Zhang *et al.* 2018), the 17S U2 snRNP (Zhang *et al*. 2020) and the *S. cerevisiae* A complex (Plaschka *et al.* 2018). Interestingly, these cryo-EM studies have modelled splicing factors including SF3B1 and SF3B6 to be in various conformations and positions throughout the splicing pathway (Figure 1-5). For instance, the super-helical HEAT repeats of SF3B1 adopt an open conformation in an isolated core SF3B complex as well as in the 17S U2 snRNP as evidenced by a crystal structure and cryo-EM model respectively (Cretu *et al.* 2016; Zhang *et al*. 2020). However, in the B^{act} complex (Zhang *et al.* 2018) SF3B1 adopts a closed conformation enclosing and stabilizing the branch point-U2 snRNA interaction. Presumably, as a consequence of maintaining an interaction with residues 373-415 of SF3B1 (Schellenberg *et al.* 2006) throughout this conformational change, SF3B6 is also modelled in various positions peripheral to the branch point in the structures of each complex (Figure 1-5). Consistent with this implication of a non-fixed SF3B6 position throughout the splicing process, SF3B1-SF3B6 crosslinks were found to be maintained throughout spliceosome assembly (Zhang *et al*. 2020).

Prior to the cryo-EM models now available of spliceosomal complexes, SF3B6 has been modelled to interact with the branch point adenosine. SF3B6 was initially identified as a spliceosomal component that formed ATP-dependent crosslinks to the branch point adenosine modified with a benzophenone moiety tethered 15 Å from the N^6 position (MacMillan *et al.* 1994). Crosslinks to this photoprobe, which could occur at any position in a 15 Å radius about a bulged-out branch point adenosine, were found to persist from the A through to C complexes (MacMillan *et al.* 1994). Subsequently, SF3B6 was also shown to crosslink within 2 Å of an 8-azido modified branch point adenosine within minimal A complexes formed on pre-mRNAs composed only of the BPS and PPT (Query *et al.* 1996). Moreover, an X-ray crystallographic characterization of SF3B6 complexed with residues 373-415 of SF3B1 (Schellenberg *et al.* 2006) also implicated SF3B6 to directly associate with the branch point adenosine. Solved to a resolution of 2.5 Å, this structure interestingly found the β-sheet of the SF3B6 RRM, including the highly conserved RNP1 and RNP2 motifs, to be occluded by a C-terminal α helix and a portion of the stabilizing SF3B1 peptide (Figure 1-6). Canonically, the β-sheet of an RRM represents the RNA binding surface of RRM-containing proteins (Kenan *et al.* 1991). In the X-ray structure of human SF3B6, a portion of the RNP2 motif of SF3B6 was shown to form a conserved tyrosine (Y22) containing pocket (Figure 1-7) for which a crosslink to the branch point adenosine could be mapped (Schellenberg *et al.* 2006). Crosslinking and subsequent cyanogen bromide cleavage of a Y22M SF3B6 mutant further suggested the direct interaction between the branch point adenosine and the conserved aromatic in this pocket. Moreover, a subsequent analysis of cysteine (Cys)-less SF3B6-SF3B1 peptide crystals soaked with adenine demonstrated the ability of Y22 to interact with adenine via aromatic stacking in this pocket (Figure 1-8; Schellenberg *et al.* 2011). Altogether, these *in vitro* crosslinking and structural studies have suggested that the bulged-out branch point adenosine fits inside a conserved SF3B6 binding pocket in an early step of spliceosome assembly prior to the first transesterification of splicing.

The previously described cryo-EM models of spliceosomal complexes are inconsistent with the implication of direct SF3B6 contact with the branch point adenosine. It remains possible that this binding pocket of SF3B6 may not interact with the branch point adenosine, but rather accommodates some aromatic of an interacting protein *in vivo*. However, given that these cryo-EM models represent a snapshot image lacking temporal resolution, the different models may not

be mutually exclusive. Instead, the positional inconsistencies observed amongst various models might indicate the dynamic rearrangements of spliceosomes throughout splicing.

U2AF Interactions and 3′ SS Recognition.

Discovered as a nuclease-resistant factor required for U2 snRNP binding to the metazoan branch point (Ruskin *et al.* 1988), U2AF plays a key role in the formation of the A complex and the commitment of a pre-mRNA transcript to the first ATP-dependent transesterification reaction of splicing. An essential heterodimeric complex consisting of 65 and 35 kDa subunits (Zamore and Green, 1989), U2AF binds to the PPT and 3′ SS prior to U2 snRNP association with the branch point making protein and RNA contacts critical for 3′ SS recognition and spliceosome assembly.

Initial investigations of the U2AF65 domain structure revealed several functionally important regions, including an N-terminal RS motif, as well as three consecutive RNP domains (Zamore *et al.* 1992). Subsequent splicing and electrophoretic mobility shift assays (EMSAs) demonstrated U2AF65 to bind the PPT via these RNP domains and that the RS motif is required for *in vitro* splicing, but not for RNA binding (Zamore *et al.* 1992). Surprisingly, U2AF35 was found not to be required for splicing *in vitro* (Zamore *et al.* 1992), despite being essential for viability (Rudner *et al.* 1996) and being highly conserved amongst eukaryotes including models such as fission yeast (Wentz-Hunter and Potashkin, 1996), Drosophila (Rudner *et al.* 1996) and *C*aenorhabditis *elegans* (Zorio and Blumenthal, 1999).

As demonstrated by site-specific radiolabeling and crosslinking, U2AF35 recognizes the 3′ SS, directly contacting the consensus AG dinucleotide (Wu *et al.* 1999). Moreover, systematic evolution of ligands by exponential enrichment (SELEX), and splicing experiments have suggested the AG dinucleotide binding of U2AF35 to be sequence specific and required for the efficient splicing of AG-dependent introns with weak PPTs (Wu *et al.* 1999). In this model, U2AF65 is suggested to be sufficient for the efficient splicing of introns with strong PPTs while U2AF35 is further required for the splicing of introns with weak PPTs as mediated by specific binding to the 3′ SS AG dinucleotide. Supporting evidence of this model for U2AF35 activity can be found genetically upon comparison of fission and baker's yeast. Introns in *Schizosaccharomyces pombe* (*S. pombe*), which have a conserved U2AF35 orthologue (Wentz-

Hunter and Potashkin, 1996), are suggested to be AG-dependent (Romfo and Wise, 1997). Alternatively, introns in *S. cerevisiae,* which lacks U2AF35 (Abovich and Rosbash, 1997), are AG-independent. In conjunction with the linear scanning mechanism of 3′ SS selection in which the spliceosome scans downstream of the branchpoint adenosine and selects the first AG as the 3′ SS (Chen *et al.* 2000), these findings implicate U2AF35 as a fundamental factor in SS selection and therefore gene expression. Notably however, higher eukaryotes exhibit high degrees of SS degeneracy and alternative splicing (Modrek and Lee, 2002), which are also in part controlled by regulatory sequences such as intronic and exonic enhancers and silencers (reviewed by Chen and Manley, 2009).

Special case RRMs, UHMs interact with U2AF ligand motifs (ULMs) and facilitate the interactions between U2AF and numerous other splicing factors (Loerch and Kielkopf, 2016) including within the U2AF (Kielkopf *et al.* 2001) and the U2AF65-SF3B1 heterodimers (Figure 1-4) (Thickman *et al.* 2006). Canonical RRM domains are composed of two conserved RNP motifs, RNP1 and RNP2, which are separated by ~30 residues, in addition to a structurally conserved ~80 residue RNA-binding region (reviewed by Kielkopf *et al.* 2004). Although differing from canonical RRMs by sequence insertions between RNP1 and RNP2 as well as surface electrostatics, UHMs are also characterized by canonical RRM βαββαβ topology as evidenced by the crystal structure of the U2AF65-U2AF35 interaction (Kielkopf *et al.* 2001) and the NMR structure of the U2AF65-SF1 interaction (Selenko *et al.* 2003).

Although ULMs are highly degenerate in sequence and difficult to predict, these motifs are generally characterized by consensus sequences of basic residues preceding a central tryptophan (Trp) that inserts into a hydrophobic pocket of a UHM binding partner (Loerch and Kielkopf, 2016). UHM-ULM interactions are commonly characterized by these basic ULM residues interacting with an acidic UHM α-helix, as well as aromatic stacking interactions between the inserting ULM tryptophan and a conserved UHM phenylalanine residue (Loerch and Kielkopf, 2016). As elucidated by an X-ray structural determination of a minimal U2AF complex (Kielkopf *et al.* 2001), reciprocal “tongue-in-groove” tryptophan residues form the basis of the U2AF heterodimer interaction (Figure 1-9). That is, Trp92 of the U2AF65 ULM and Trp134 of the U2AF35 UHM fit into complementary binding sites stabilized by numerous aromatic stacking, hydrogen bonding and salt bridge interactions.

During the spliceosomal transition from the E to the A complex, the transient interaction of SF1 with U2AF65 (Rutz and Séraphin, 1999) is replaced with a U2 snRNP-U2AF65 interaction as a result of RNP helicase mediated RNA and protein rearrangements (Fleckner *et al.* 1997). The U2 snRNP associated factor SF3B1, which crosslinks to pre-mRNA about the BPS and adjacent to the PPT, interacts with a UHM of U2AF65 (Gozani *et al.* 1998). As revealed by circular dichroism spectroscopy, the N-terminal U2AF65 binding region of SF3B1 is largely unstructured (Thickman *et al*. 2006). Importantly, this U2AF binding region of SF3B1 contains seven tryptophan residues, five of which are preceded by basic residues and were found by mutational and intrinsic tryptophan fluorescence analysis to interact with the UHM of U2AF65 (Figure 1-4) (Thickman *et al*. 2006). Although not structurally demonstrated, the requirement of these SF3B1 tryptophan residues preceded by basic residues to mediate binding with the U2AF65 UHM suggests consistency with the interfaces observed previously in the U2AF heterodimer and SF1-U2AF65 structures. Furthermore, the presence of up to five biologically relevant binding sites for U2AF65 on SF3B1 may reflect the dynamic nature of the spliceosome and possible conformational rearrangements (Thickman *et al*. 2006). The model organism *S. pombe*, in which the interaction between SF3B1 and U2AF65 is conserved, also shares two of these putative tryptophan-containing binding motifs (Thickman *et al*. 2006). Additional isothermal calorimetry experiments found the SF3B1-U2AF65 Kd to be ~3 μM and the SF1-U2AF65 Kd to be ~11 nM revealing a much tighter association for the SF1-U2AF65 interaction, consistent with the requirement of ATP hydrolysis to drive SF1 dissociation and A complex formation (Thickman *et al*. 2006).

***S. pombe*: A Model Organism for Studying Splicing.**

S. cerevisiae has been used as an effective model organism for the characterization of the chemical mechanism of splicing (Fair and Pleiss, 2017) as well as for the identification of numerous conserved core spliceosomal components such as the pre-mRNA processing (PRP) genes (Lustig *et al.* 1986; Vijayraghavan *et al.* 1989). Although the *S. cerevisiae* model has proven invaluable in understanding the fundamentals of splicing, *S. cerevisiae* lacks key spliceosomal factors implicated in SS selection such as SF3B6 (Sun, 2020) and U2AF35 (Abovich and Rosbash, 1997). Furthermore, *S. cerevisiae* introns feature highly conserved

consensus sequences at the SSs and BPS (Figure 1-10), and alternative splicing events are rare (Fair and Pleiss, 2017). In contrast, alternative splicing is predominant in humans where >95% of intron-containing genes are estimated to undergo alternative splicing (Fair and Pleiss, 2017; Barbosa-Morais *et al.* 2012). Due to the simplicity of *S. cerevisiae* splicing relative to humans, *S. pombe* has become an important model organism for elucidating the mechanisms of splice site recognition and alternative splicing. *S. pombe* is more intron dense than *S. cerevisiae* (Fair and Pleiss, 2017), features BPS and SS degeneracy with greater similarity to that of humans (Figure 1-10) and contains SF3B6 and U2AF35 (Wentz-Hunter and Potashkin, 1996) orthologues.

Although essential in humans, as evidenced by a loss of viability upon SF3B6 knockout in HeLa cells (Schellenberg and MacMillan, unpublished), the role of SF3B6 in splicing remains poorly understood. Alternatively, SF3B6 can be deleted from *S. pombe* to yield a phenotypically distinct strain in which the cells tend to cluster and were found to exhibit resistance to oxidative stress and susceptibility to DNA replicative stress relative to their wild type (WT) counterparts (Madi, 2021). An *S. pombe* strain where SF3B6 has been knocked out (ΔSF3B6) is viable and therefore must be assembling spliceosomes and splicing pre-mRNAs at some level to yield functional proteins sufficient for survival. Differential protein expression levels observed between the WT and ΔSF3B6 strains (Madi, 2021) can presumably be attributed to differential splicing levels in the presence or absence of SF3B6, which is a key component of the spliceosomal A complex. WT *S. pombe* A complex formation has been previously characterized with a series of model pre-mRNAs (Huang *et al.* 2002) using established biochemical methods that can be feasibly applied to examine complex formation in the absence of SF3B6. Accordingly, this ΔSF3B6 *S. pombe* strain presents the opportunity to investigate the role of SF3B6 in splicing and A complex formation in a tractable model that shares some of the splicing complexities of humans.

Thesis Overview.

This thesis aims to investigate the molecular mechanism of branch point and 3′ SS recognition by SF3B6, SF3B1 and U2AF in the pre-spliceosome. Chapter 2 describes the development of an *in vitro* reconstitution of an *S. pombe* SF3B6-SF3B1-U2AF complex suitable for structural and pre-mRNA binding characterization. Chapter 3 examines the molecular role of SF3B6 in splicing

by characterizing *S. pombe* A complex formation on pre-mRNA substrates in the presence and absence of SF3B6.

Figure 1-1. Conserved intronic sequence elements. Schematic representation of a pre-mRNA including the conserved dinucleotides at the 5´ and 3´ SS, an optimized BPS (yeast consensus sequence), and PPT. The branch point adenosine (*) is indicated (adapted from Ritchie *et al.* 2009).

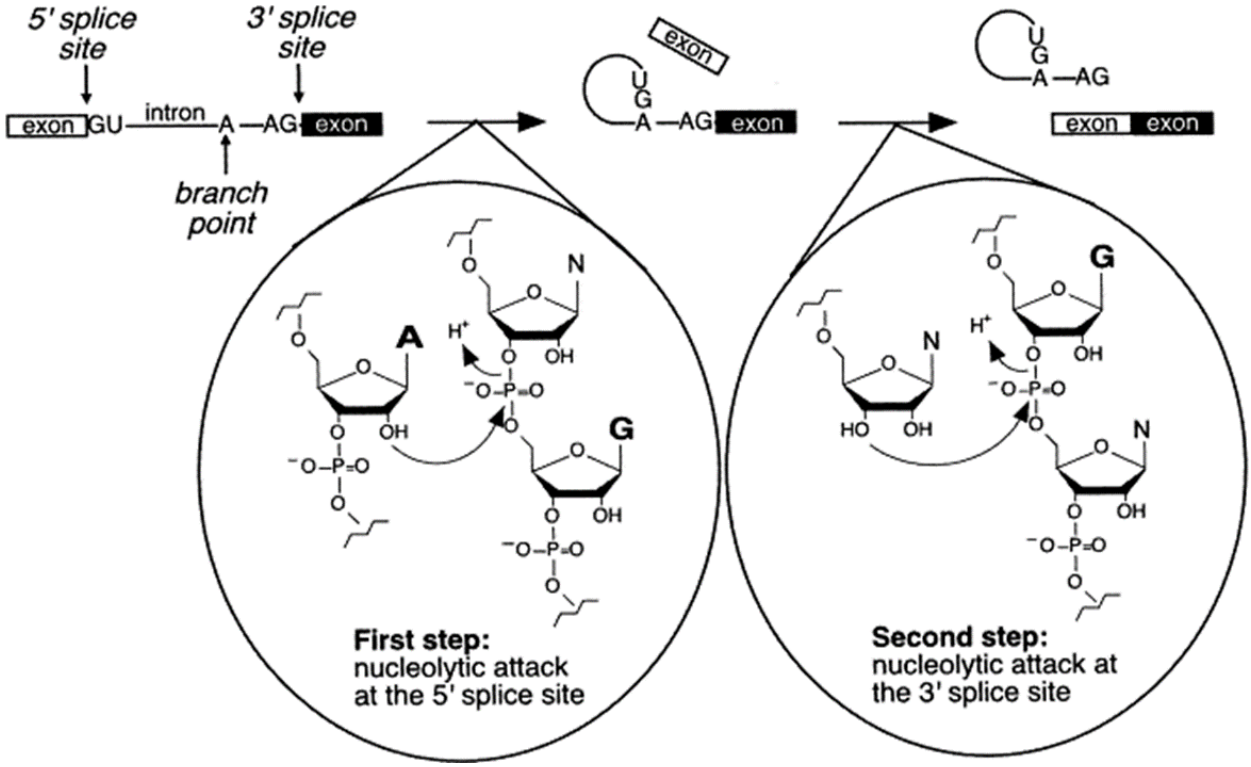

Figure 1-2. The chemistry of pre-mRNA splicing. Splicing occurs via two sequential transesterification reactions. In the first step the 2´ OH of the branch point adenosine performs a nucleophilic attack at the 5´ SS, producing the lariat intermediate and a free 5´ exon. In the second step, the 3´ OH of the 5´ exon attacks the 3´ SS joining the exons together and releasing a lariat product (adapted from Jurica and Moore, 2002).

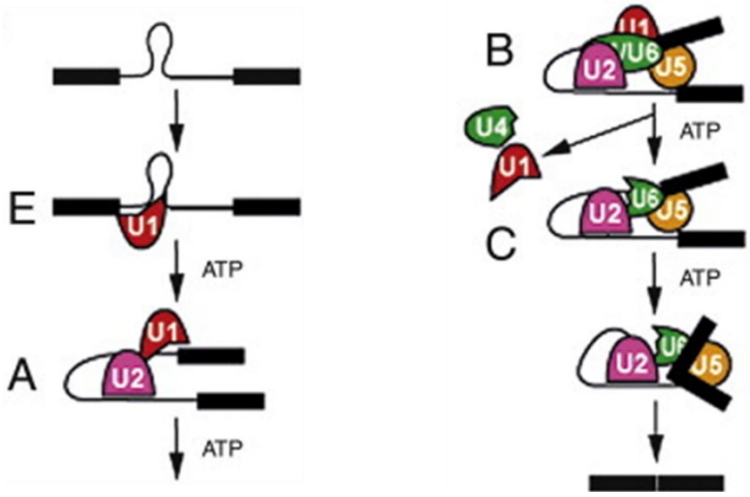

Figure 1-3. The sequential assembly of spliceosomal complexes. Schematic representation of the stepwise assembly of the spliceosomal E, A, B and C complexes including snRNP associations and dissociations (adapted from Ritchie *et al.* 2009).

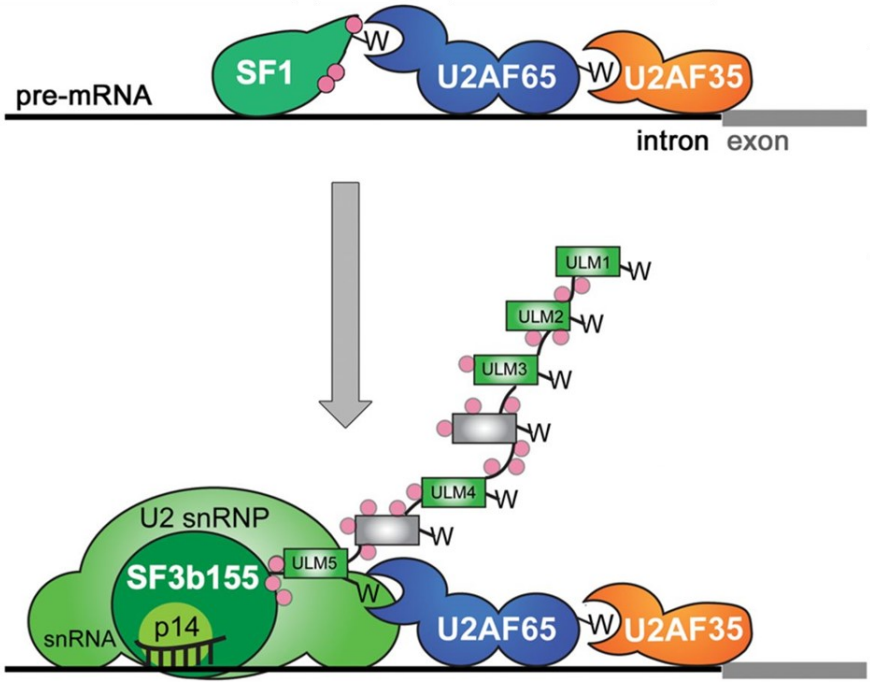

Figure 1-4. UHM-ULM interactions in the early stages of spliceosome assembly. Schematic representation of the SF1-U2AF65-U2AF35 and SF3B6-SF3B1-U2AF65-U2AF35 interactions in the early stages of spliceosome assembly. SF3B6 is labelled as p14 and SF3B1 is labelled as SF3b155. ULM tryptophans are represented by W. Pink circles indicate phosphorylation sites (adapted from Loerch and Kielkopf, 2016).

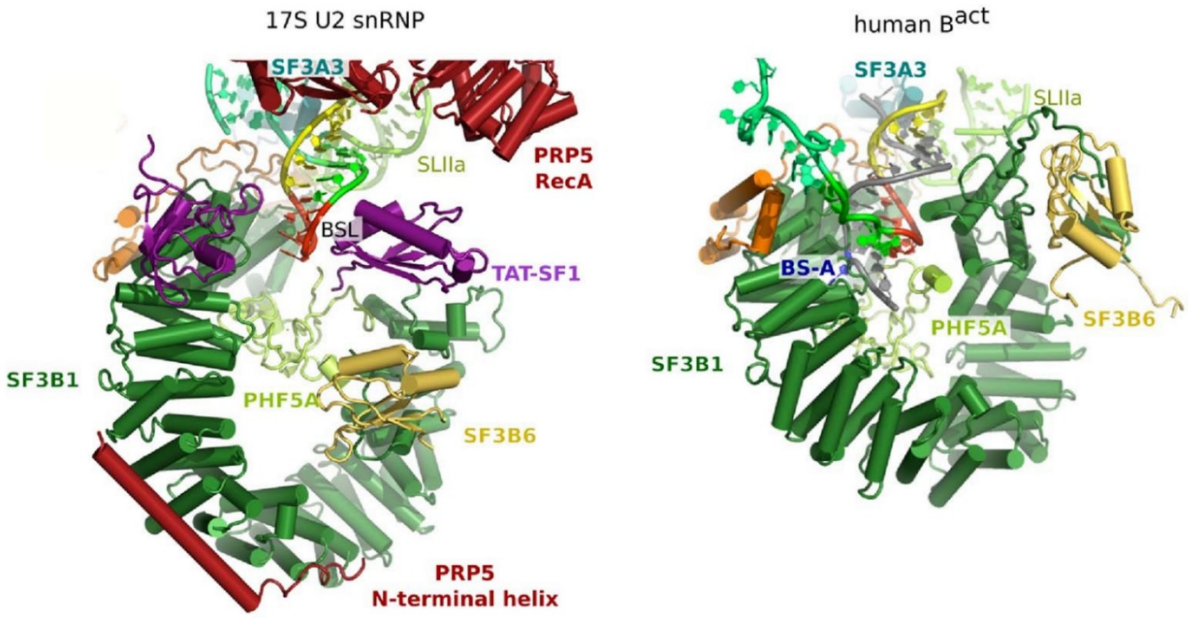

Figure 1-5. Cryo-EM models of the 17S U2 snRNP and B^{act} complex. A comparison of the spatial orientation of the branch point region within the human 17S U2 snRNP (left; PDB: 6Y5Q) and B^{act} (right; PDB: 6FF4) structures. Prior to pre-mRNA binding the BPS interacting region of the U2 snRNA is sequestered in a branch point-interacting stem-loop (BSL). The spatial organization of factors between each complex is similar, but not fixed. SF3B6 interacts with residues 373-415 of SF3B1 and is peripheral to the branch point (adapted from Zhang *et al.* 2020).

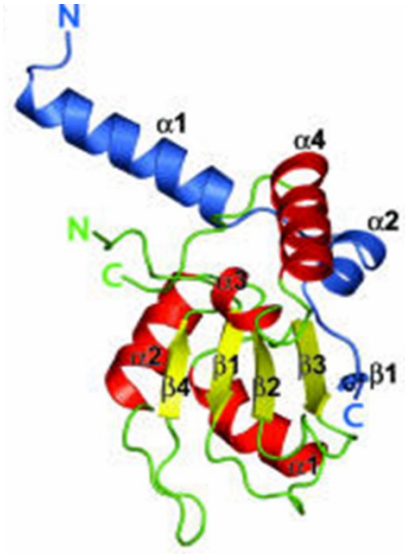

Figure 1-6. Structure of the SF3B6-SF3B1 peptide complex. Ribbon diagram of the SF3B6-SF3B1 peptide crystal structure. SF3B6 α-helices are colored red, the β-strands of SF3B6 are colored yellow, and SF3B1 is colored blue. The β-sheet of the SF3B6 RRM is largely occluded by the C-terminal α4 helix and a portion of the stabilizing SF3B1 peptide (adapted from Schellenberg *et al.* 2006).

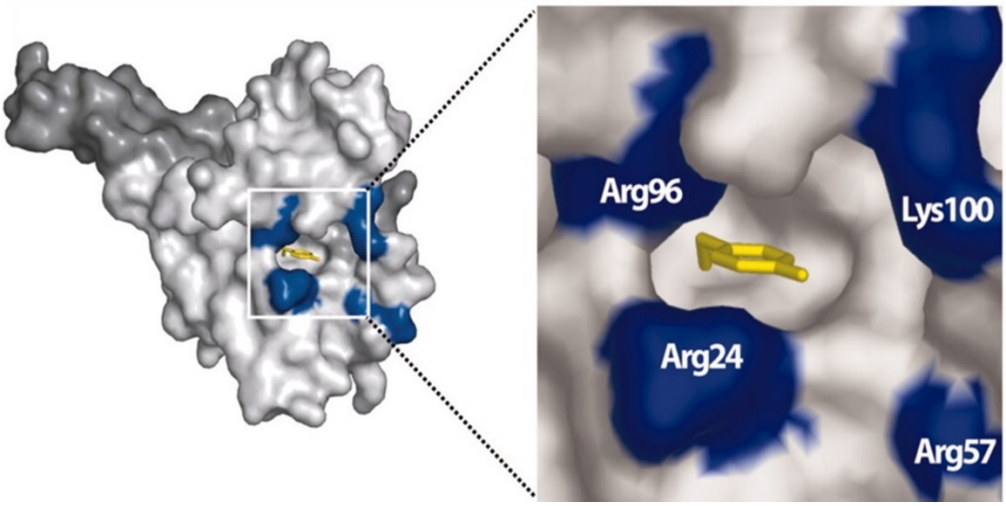

Figure 1-7. The exposed surface pocket of SF3B6 RNP2. Surface representation of the SF3B6-SF3B1 peptide crystal structure. The conserved Y22 (yellow) of RNP2 is exposed within a surface pocket surrounded by basic residues (blue) R24, R57, R96 and K100. SF3B6 is colored light grey and SF3B1 is colored dark grey (adapted from Schellenberg *et al.* 2006).

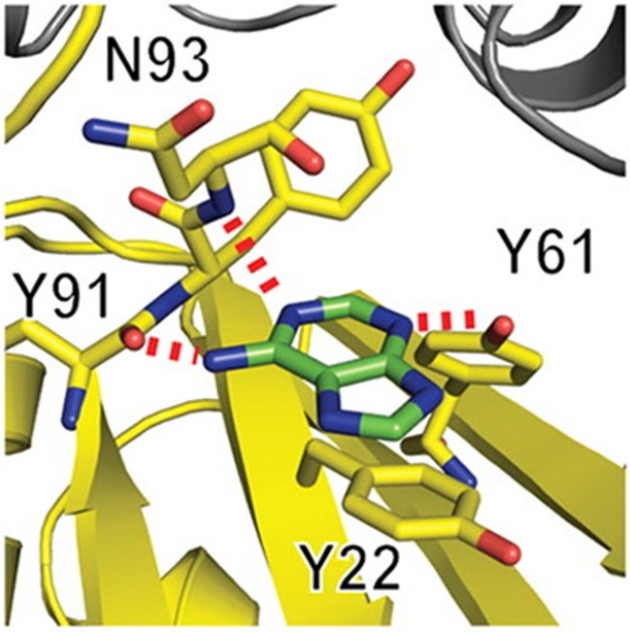

Figure 1-8. Adenine recognition by the exposed surface pocket of SF3B6 RNP2. Crystal structure of adenine bound within the surface pocket of an SF3B6-SF3B1 peptide complex. The adenine stacks on the conserved Y22 and forms hydrogen bonds (dashed red lines) with Y22, Y61, Y91 and N93. SF3B6 is colored yellow and SF3B1 is colored grey (adapted from Schellenberg *et al.* 2011).

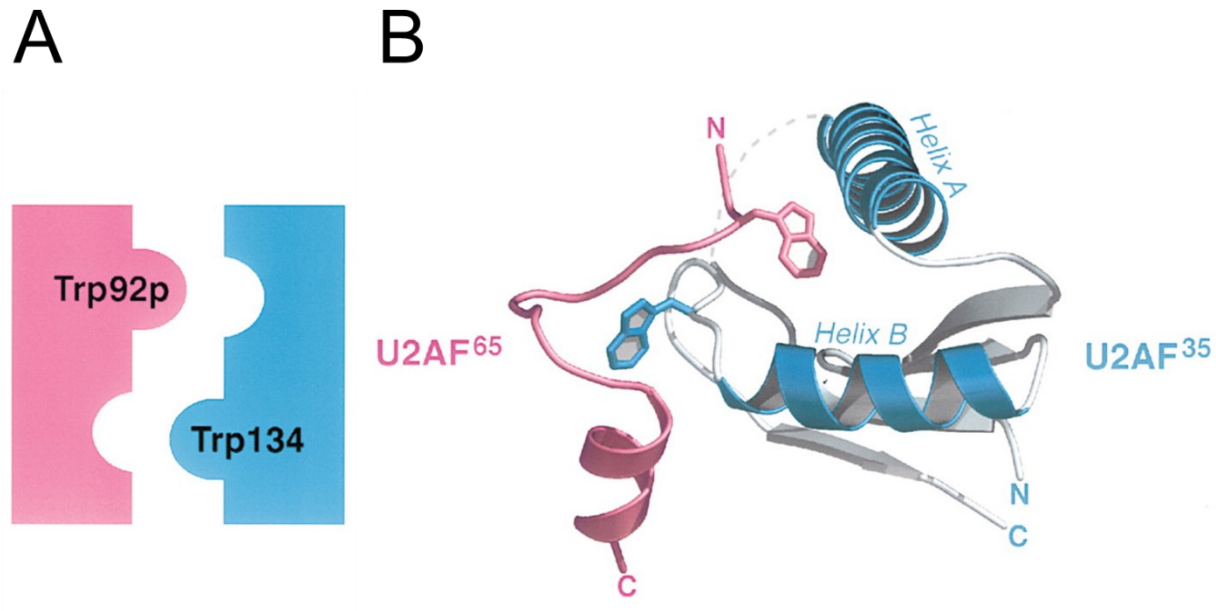

Figure 1-9. The U2AF35 UHM-U2AF65 ULM interaction. (A) Schematic representation of the reciprocal "tongue-in-groove" tryptophan residues at the interface of the U2AF heterodimer interaction. (B) Structural view of the U2AF heterodimer interaction. Trp92 of the U2AF65 ULM fits into a hydrophobic pocket of the U2AF35 UHM and Trp134 of U2AF35 fits into a proline-rich loop of the U2AF65 ULM (adapted from Kielkopf *et al.* 2001).

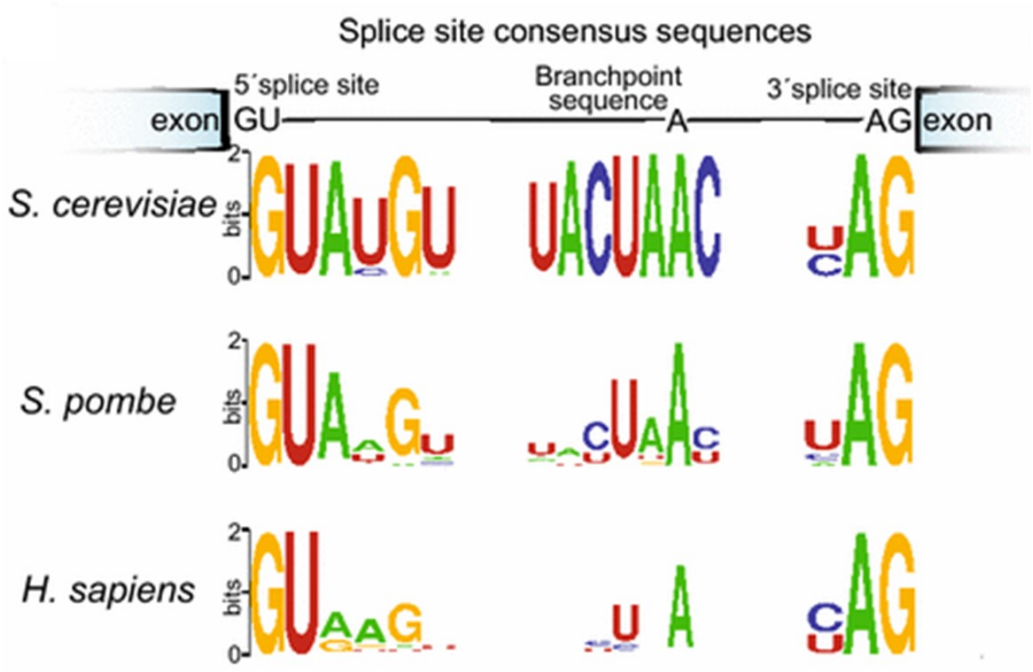

Figure 1-10. SS and BPS sequence conservation in yeasts and humans. Sequence logos of the consensus SS and BPS regions of introns in budding yeast, fission yeast and humans. The height at each position is proportional to nucleotide conservation at that position (adapted from Fair and Pleiss, 2017).

Chapter 2

Towards the Reconstitution of an SF3B6-SF3B1-U2AF Complex

Chapter 2. Towards the Reconstitution of an SF3B6-SF3B1-U2AF Complex.

Results and Discussion.

SF3B6-SF3B1 175-291-U2AF59 93-517-U2AF23 Reconstitution Analysis.

Contacting the branch point adenosine, PPT and the 3′ SS respectively, a complex containing SF3B6, U2AF65 (U2AF59 in *S. pombe*), and U2AF35 (U2AF23 in *S. pombe*) would represent a key assembly involved in the precise definition of the 3′ SS. To examine SF3B6-U2AF protein-protein interactions, as well as SF3B6 and U2AF-RNA interactions in the context of this assembly at a molecular level, a series of SF3B6 and U2AF constructs were purified, then combined in an attempt to reconstitute an *in vitro* complex suitable for structural analysis.

Because of the tendency that isolated SF3B6 has previously shown towards aggregation and precipitation, *S. pombe* SF3B6 was recombinantly co-expressed with residues 175-291 of SF3B1 (Figure 2-1) which, based on homology with the previously characterized human complex (Schellenberg *et al.* 2006), represents the SF3B6 binding interface. Full-length recombinant human U2AF heterodimer has been reported to have low protein yield and poor solubility upon expression, making it unsuitable for structural characterization (Kielkopf *et al.* 2001). Accordingly, residues 93-517 of *S. pombe* U2AF59 were recombinantly co-expressed with full-length U2AF23 to yield a U2AF heterodimer lacking the N-terminal SR domain of its large subunit. These SF3B6-SF3B1 175-291 and U2AF59 93-517-U2AF23 co-expressions were independently purified, then combined at an equimolar ratio and subjected to size exclusion chromatography in an attempt to isolate a distinct tetrameric complex. Unexpectedly however, the SF3B6-SF3B1 175-291 and U2AF59 93-517-U2AF23 dimers were observed to elute as two partially overlapped peaks (Figure 2-2 A), which were not found to correspond to the elution position of a stoichiometric tetramer. Consistent with the observed peak overlap, the U2AF59 93-517-U2AF23 dimer was detected to elute earlier in the elutions of peak 1 while the SF3B6-SF3B1 175-291 dimer was detected to elute in the elutions of peak 2 (Figure 2-2 B) suggesting that these dimers are not stably interacting but are rather independently eluted via gel filtration.

In the human system, primary sequence analysis of SF3B1 revealed 7 possible U2AF interaction motifs which are characterized by a tryptophan residue preceded by basic or lysine residues and followed by acidic residues (Thickman *et al.* 2006). Intrinsic tryptophan fluorescence experiments confirmed the interaction of 5 of these motifs with U2AF, two of which are

conserved within *S. pombe* SF3B1 (Thickman *et al.* 2006). These conserved U2AF-interacting tryptophan residues in *S. pombe* SF3B1 are W197 and W216 and are included in the previously discussed SF3B6 binding interface and the SF3B6-U2AF59 175-291 construct that was found to not stably interact with the U2AF59 93-517-U2AF23 dimer by gel filtration (Figure 2-2).

SF3B6-SF3B1 208-284-U2AF59 93-517-U2AF23 Reconstitution Analysis.

As only a single SF3B1 ULM is expected to interact with U2AF at a time (Thickman *et al.* 2006), SF3B1 208-284 (Figure 2-1) was recombinantly co-expressed with full length SF3B6 to generate a truncated construct containing only the W216-containing motif. This co-expression construct, SF3B6-SF3B1 208-284, was found to contain sequence sufficient to facilitate SF3B6 binding upon purification thereby making it useful for investigating U2AF binding in a simpler system. Accordingly, SF3B6-SF3B1 208-284 and U2AF59 93-517-U2AF23 co-expressions were independently purified, combined, and subjected to size exclusion chromatography similarly to the previous set of constructs. The resultant elution profile was observed to contain 3 clearly distinct peaks (Figure 2-3 A). SDS-PAGE analysis revealed peak 1 to contain both U2AF subunits, as well as a faintly observable SF3B6 band (Figure 2-3 B). Peak 2 was found to contain the U2AF subunits, and peak 3 was found to contain the majority of the injected SF3B6 sample (Figure 2-3 B). The protein contents of peak 1, as well as its earlier elution position suggest the isolation of an SF3B6-SF3B1 208-284-U2AF59 93-517-U2AF23 complex comprised of a fraction of the injected U2AF and SF3B6-SF3B1 peptide starting material. Given that peaks 2 and 3 are narrow and symmetrical, characteristic of monodisperse proteins which should be suitable for binding, the relatively small amount of fully complexed protein may indicate a relatively weak and/or transient SF3B1-U2AF interaction between these constructs.

SF3B6-SF3B1 175-291-U2AF59 93-517 Reconstitution Analysis.

In a further attempt to simplify the reconstitution of an SF3B6-SF3B1-U2AF complex, U2AF59 93-517 was independently purified, mixed with SF3B6-SF3B1 175-291, and subjected to gel filtration as previously described. A single peak was observed in the resultant elution profile (Figure 2-4 A). However, SDS-PAGE analysis revealed decoupled elution peaks, with a U2AF

peak elution at fraction 18, and an SF3B6-SF3B1 peptide peak elution at fraction 20 (Figure 2-4 B). Notably, U2AF59 93-517 was also found to have poor solubility and low expression yields relative to U2AF co-expressions. Consequently, U2AF59 93-517 often could not be sufficiently concentrated for the injection of a large quantity onto a size exclusion column which explains the relatively low quantity observed via SDS-PAGE (Figure 2-4 B). Due to the poor U2AF59 93-517 solubility, attempts to reconstitute a complex of SF3B6-SF3B1 208-284-U2AF59 93-517 were unsuccessful as a result of heavy precipitation. Altogether, the elution profile observed in this experiment (Figure 2-4 A) can be interpreted as the likely overlap of a small U2AF59 peak with a relatively large SF3B6-SF3B1 peptide peak.

SF3B6-SF3B1 175-291 and U2AF59 93-517-U2AF23 pre-mRNA binding.

Isothermal calorimetry experiments have revealed human SF3B1 and U2AF to bind with a K_d of ~2.8 μM (Thickman *et al.* 2006). The SF3B6-SF3B1 peptide and U2AF59-23 or U2AF59 constructs described above could typically only be concentrated and combined to micromolar quantities of protein before risking precipitation. Assuming that the human and *S. pombe* systems exhibit a similar affinity, when combining preparations for the reconstitution of a trimeric or tetrameric complex, the protein concentrations reached prior to size exclusion chromatography would likely be relatively near the K_d of the SF3B1-U2AF interaction. Not only would it have been desirable to use protein concentrations in excess of the assumed K_d to ensure a high proportion of complex was formed at equilibrium, but there may also be an effect of dilution further lowering protein concentrations as they migrate over the ~180 mL size exclusion column. Therefore, it is possible that the lack of interaction observed (Figure 2-2 and Figure 2-4), and the seemingly low proportion of complexed SF3B6-SF3B1-U2AF observed (Figure 2-3) may be explained by protein concentrations that are too low to accommodate stable SF3B1-U2AF binding. An alternative approach to evaluating complex formation by the previously described constructs is to visualize binding events of these proteins to a radiolabelled RNA by electrophoretic mobility shift. Accordingly, three model RNAs were transcribed and end-labelled with ^{32}P. Two of these RNAs, here referred to as the "U9 hairpin" and "U12 hairpin," are half-introns that include the 3′ SS and form hairpins designed to mimic U2 snRNA-BPS base pairing with a bulged branch point adenosine (Figure 2-5 A and B). The U9 and U12 hairpin RNAs

feature PPTs composed of nine and twelve uridines respectively. The third RNA transcribed, the "A3′ intron," is a half intron including the BPS, PPT, and 3′ SS but not designed to form an intramolecular hairpin and thus lacks a bulged adenosine (Figure 2-5 C). Native-PAGE of the U9 and U12 hairpins with a mixture of SF3B6-SF3B1 175-291 and U2AF59 93-517-U2AF23 recovered from the attempted reconstitution (Figure 2-2) revealed what appear to be two distinct shifted species (Figure 2-6). Under the limitation of not titrating the SF3B6 and U2AF constructs separately, it remains unclear if these shifts correspond to independent binding by SF3B6 and U2AF or if the uppermost band corresponds to a ternary complex. Only a single shifted band was observed in the case of the A3′ intron (Figure 2-6).

SF3B6-SF3B1 175-291-U2AF59 93-161-U2AF23 Reconstitution Analysis.

RNAs designed to bind SF3B6, as well as U2AF could potentially be used as a scaffold to improve SF3B1-U2AF binding stability, or to generate a complex independent of SF3B1-U2AF binding. A minimal complex of U2AF59 93-161-U2AF23 (Figure 2-7), where a short U2AF59 peptide was used to improve U2AF23 stability, has been structurally characterized for the elucidation of 3′ SS recognition by U2AF23 (Yoshida *et al.* 2015). The U2AF59 peptide of this minimal complex does not include its SF3B1-interacting UHM (Loerch and Kielkopf, 2016). However, this may be useful in reconstituting an RNA-mediated ternary SF3B6-SF3B1-U2AF complex. For instance, it may be possible to assemble SF3B6-SF3B1 peptide and U2AF59 93-161-U2AF23 complexes on the BPS and 3′SS regions of a pre-mRNA respectively. Preliminary investigations of human SF3B6 and U2AF65 pre-mRNA binding in our lab revealed SF3B6 RNA binding to be stimulated by U2AF65. Subsequent pull-down experiments using a maltose binding protein (MBP)-tagged SF3B6 and a series of glutathione S-transferase (GST)-tagged U2AF65 deletion constructs identified residues 94-151 and RRM3 of U2AF65 to interact with SF3B6 (Schellenberg and MacMillan, unpublished). The SF3B6-interacting residues 94-151 of U2AF65 align with residues 137-203 of U2AF59 (Figure 2-7). That is, the U2AF59 peptide of this previously characterized U2AF59 93-161-U2AF23 construct (Yoshida *et al.* 2015) contains a portion of an SF3B6-interacting region identified the human system. As such, this U2AF59 93-161-U2AF23 complex which has been previously utilized for investigation of 3′SS recognition may also be capable of SF3B6 interaction for ternary complex assembly. U2AF59 93-161-

U2AF23 was recombinantly co-expressed, combined with SF3B6-SF3B1 175-291 and subjected to gel filtration as previously described. These proteins were not observed to elute as a single monodisperse peak at a position appropriate for an assembled complex (Figure 2-8 A). Interestingly, U2AF23 seemingly eluted as two peaks straddling a centered SF3B6 elution peak (Figure 2-8 B). Electrophoretic mobility shift of the radiolabelled A3′ intron upon titration of this minimal U2AF complex, or with SF3B6-SF3B1-U2AF each resulted in a single shifted band (Figure 2-9). Simultaneous titration of both U2AF and SF3B6 dimers resulted in greater proportion of shifted signal than that of each dimer individually. Although the signal of the highest combined protein concentration lanes did not migrate particularly far into the gel, the band is distinct from the individual shifts and potentially represents a super-shifted ternary complex (Figure 2-9). Consistent with the concern that the binding affinity of the SF3B1-U2AF59 interaction may be too weak for stable reconstitution by gel filtration, this potential ternary complex was only observed to form at protein concentrations ≥50 μM which was well above the ~2.8 μM K_d of the human interaction.

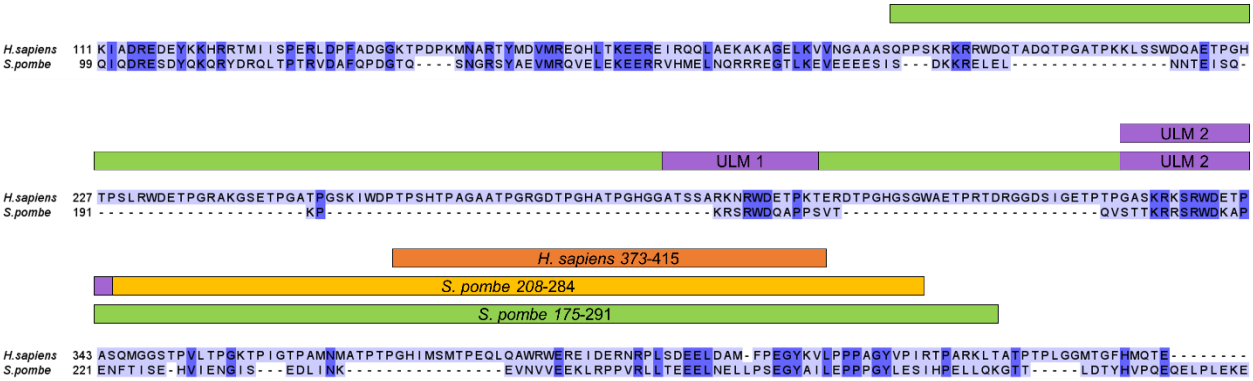

Figure 2-1. SF3B6 and U2AF59 interacting regions of SF3B1. Sequence alignment of human and *S. pombe* SF3B1 (human residues 111-449, *S. pombe* residues 99-311). Identical residues are blocked in dark blue and similar residues are blocked in light blue. Human SF3B6-interacting residues (Schellenberg *et al.* 2006) are indicated by the orange box. *S. pombe* constructs including one and two putative U2AF-interacting motifs (purple boxes) are indicated by the yellow and green boxes respectively.

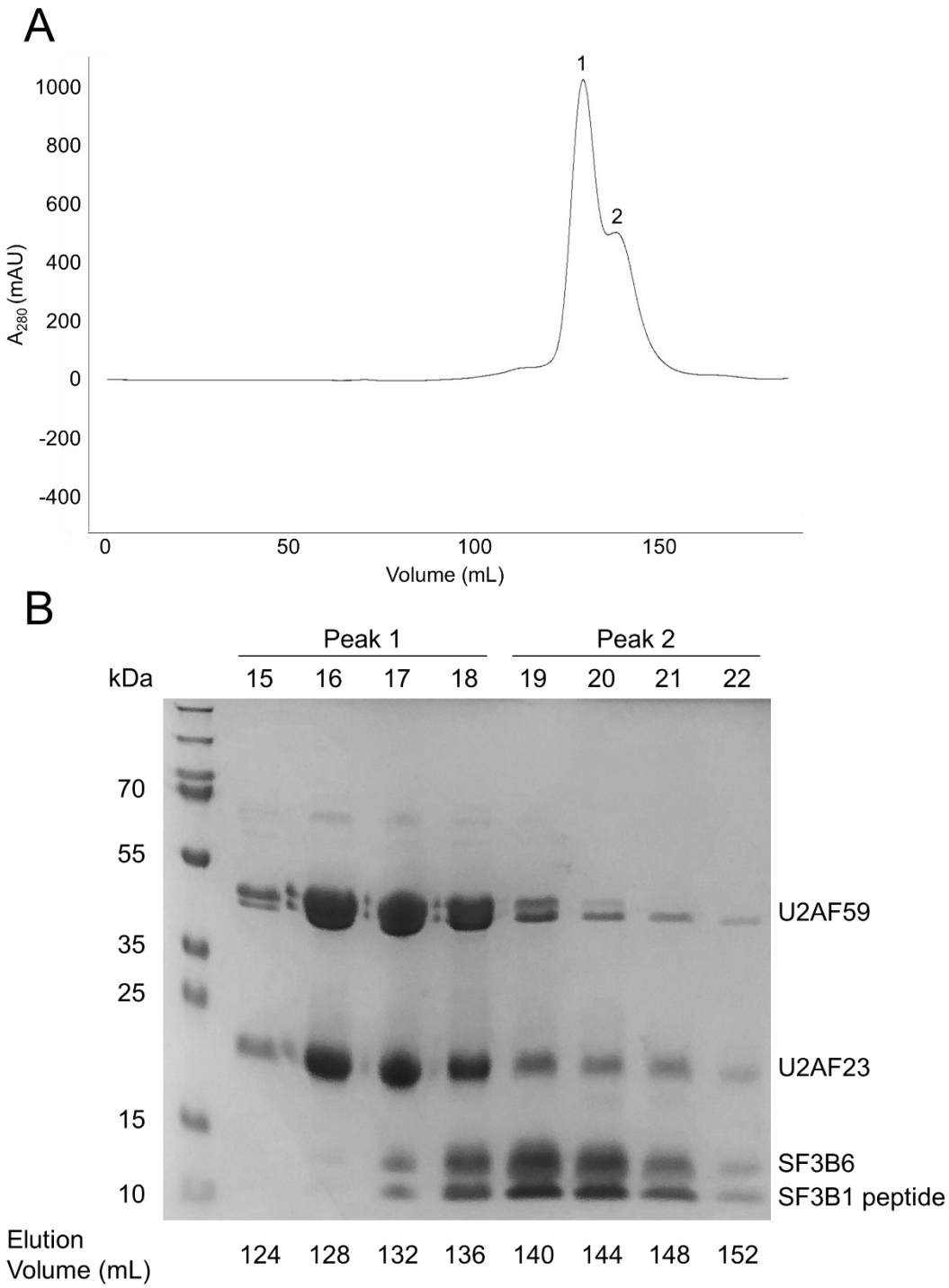

Figure 2-2. Attempted reconstitution of an SF3B6-SF3B1 + U2AF59-U2AF23 complex with both putative SF3B1 ULMs. (**A**) Elution profile of independently purified SF3B6-SF3B1 175-291 and U2AF59 93-517-U2AF23 mixed at an equimolar ratio and purified via size exclusion chromatography. Elution peaks are numbered. (**B**) Sodium dodecyl-sulfate polyacrylamide electrophoresis (SDS-PAGE) of peak elution fractions.

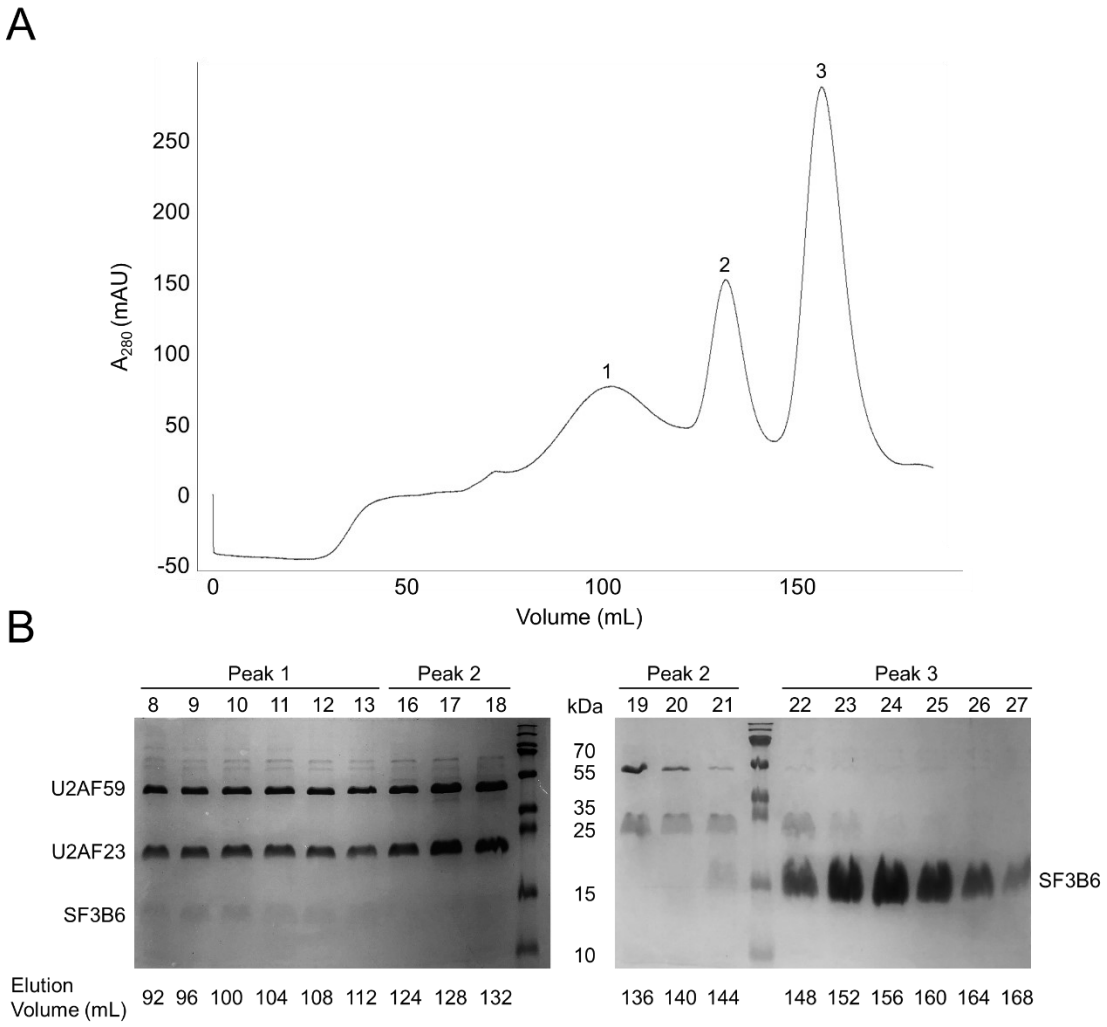

Figure 2-3. Attempted reconstitution of an SF3B6-SF3B1 + U2AF59-U2AF23 complex with a single putative SF3B1 ULM. (**A**) Elution profile of independently purified SF3B6-SF3B1 208-284 and U2AF59 93-517-U2AF23 mixed at an equimolar ratio and purified via size exclusion chromatography. Elution peaks are numbered. (**B**) SDS-PAGE of peak elution fractions. Note that the SF3B1 208-284 peptide is approximately 8.8 kDa and often difficult to visualize, presumably due to it migrating through the gel or by poor stain accommodation.

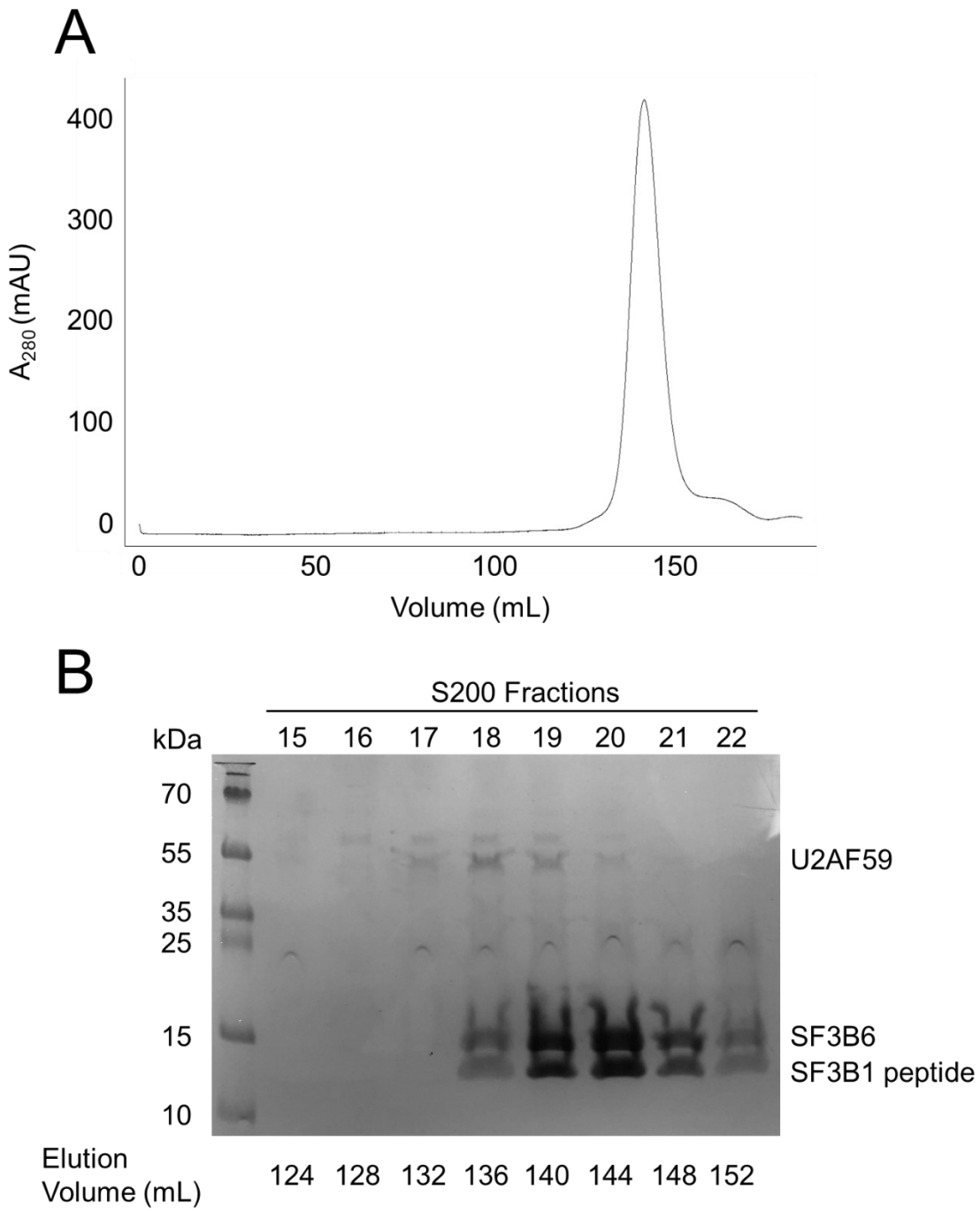

Figure 2-4. Attempted reconstitution of an SF3B6-SF3B1 + U2AF59 complex with both putative SF3B1 ULMs. (**A**) Elution profile of independently purified SF3B6-SF3B1 175-291 and U2AF59 93-517 mixed and purified via size exclusion chromatography. (**B**) SDS-PAGE of peak elution fractions.

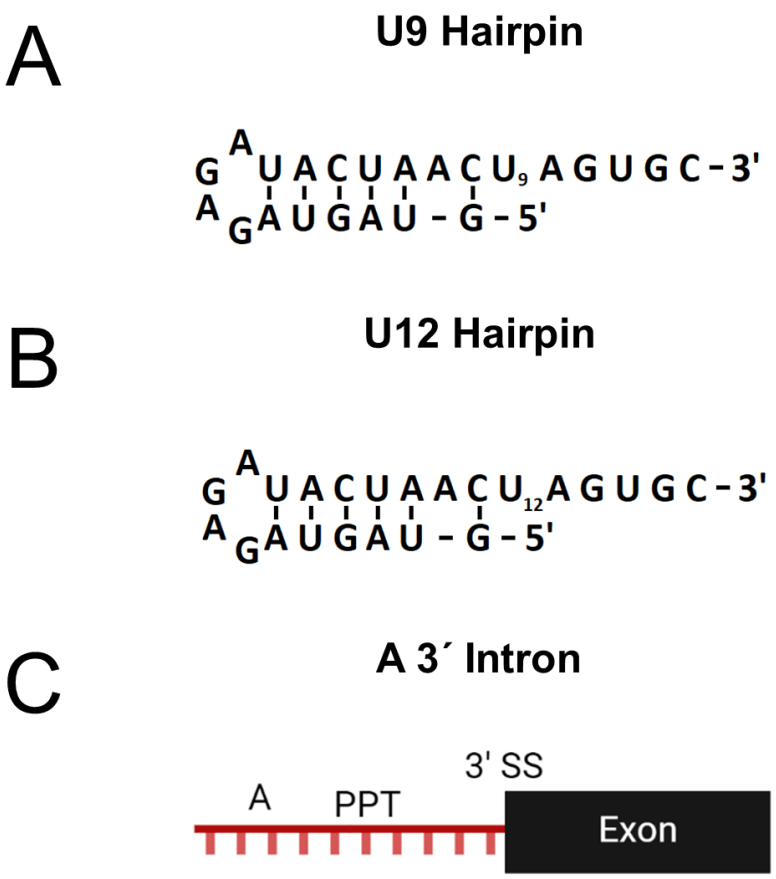

Figure 2-5. Schematic representation of model intron RNAs. Sequence and secondary structure of the U9 hairpin (**A**) and U12 hairpin (**B**) RNAs. Hairpin stem mimics an optimized base pairing of the BPS (UACUAAC) with U2 snRNA, with an unpaired, bulged branch point adenosine. U_9 and U_{12} represent 9 and 12 nucleotide polyuridine PPTs respectively [3´ SSs (AG) are included]. (**C**) Schematic diagram of the 64 nucleotide A3´ intron RNA. A indicates branch point adenosine.

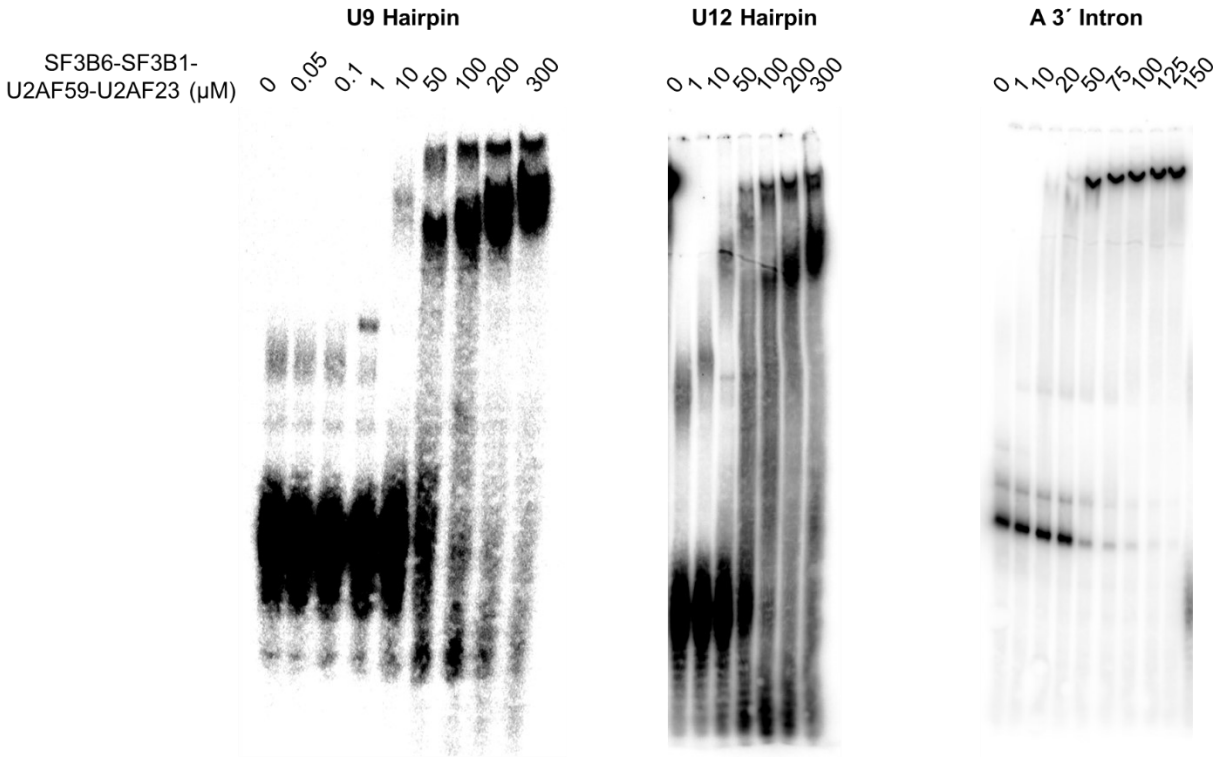

Figure 2-6. Mobility shift of model introns by pooled SF3B6-SF3B1-U2AF59-U2AF23 elutions. Peak 1 and 2 elutions from the attempted reconstitution of SF3B6-SF3B1 175-291 + U2AF59 93-517-U2AF23 (Figure 1-2) were pooled and titrated against radiolabelled U9 hairpin, U12 hairpin and A3´ intron RNAs. Pooled protein concentration was approximately quantified assuming a 1:1 ratio of SF3B6-SF3B1:U2AF59-U2AF23.

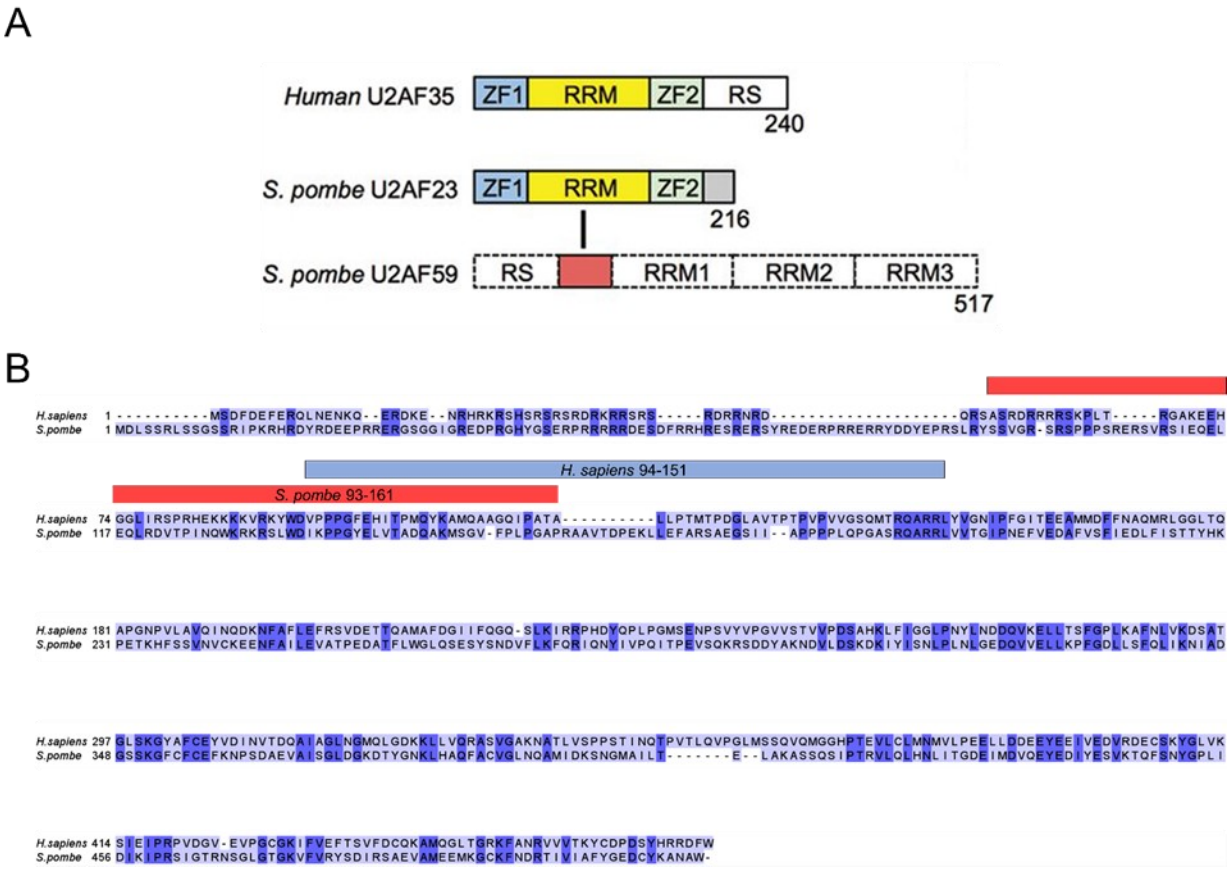

Figure 2-7. U2AF23 and SF3B6 interacting regions of U2AF59. (A) Schematic representation of *S. pombe* U2AF domain structure. Zinc finger 1 (ZF1), the RRM and ZF2 are colored blue, yellow and green respectively. The U2AF23-interacting region of U2AF59 is colored red (adapted from Yoshida *et al.* 2015). (B) Sequence alignment of human U2AF65 and *S. pombe* U2AF59. Identical residues are blocked in dark blue and similar residues are blocked in light blue. Human SF3B6-interacting residues are indicated by the blue box and the U2AF59 peptide used in the minimal U2AF heterodimer is indicated by the red box.

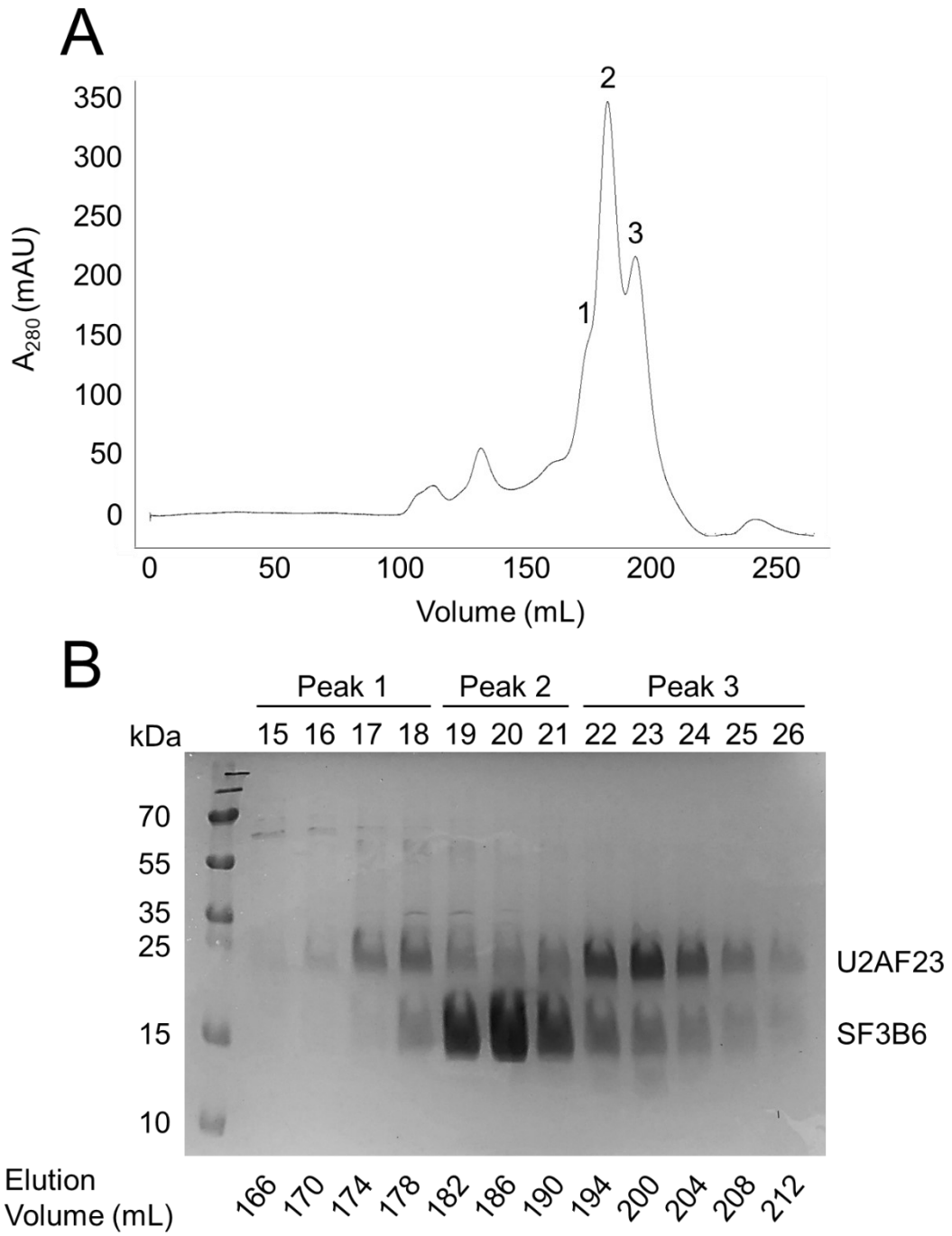

Figure 2-8. Attempted reconstitution of an SF3B6-SF3B1 + U2AF59-U2AF23 complex containing both putative SF3B1 ULMs and a minimal U2AF heterodimer. (A) Elution profile of independently purified SF3B6-SF3B1 175-291 and U2AF59 93-161-U2AF23 mixed at an equimolar ratio and purified via size exclusion chromatography. Elution peaks are numbered. (B) SDS-PAGE of peak elution fractions. Note that the SF3B1 175-291 peptide is approximately 13.5 kDa, and resolves as a band close to SF3B6, which has been difficult to distinctly visualize on occasion, especially in the case of a relatively high protein load.

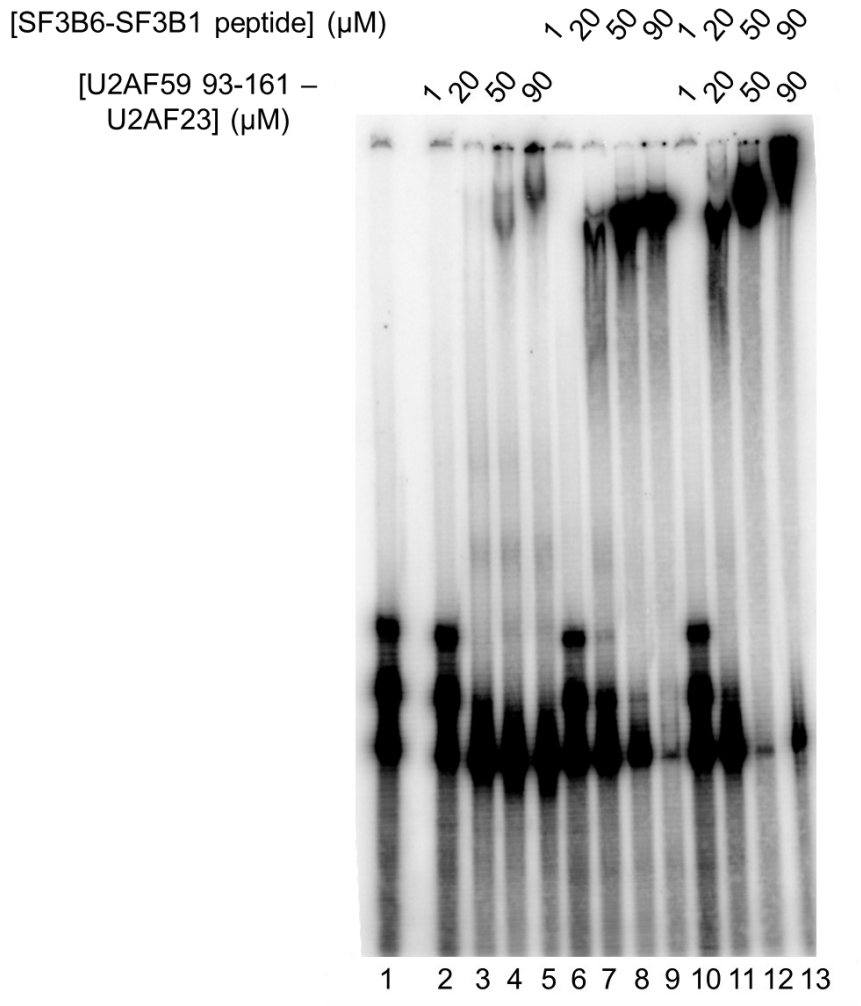

Figure 2-9. Analyzing RNA-mediated SF3B6-U2AF complex formation and intron affinity. EMSA of SF3B6-SF3B1 175-291 and U2AF59 93-161-U2AF23 titrated both independently and together against radiolabelled A3′ intron RNA. RNA migration in the absence of protein is shown in lane 1.

Materials and Methods

Recombinant Protein Expression. An overnight culture of BL21 Gold cells transformed with the appropriate expression vector was used to inoculate 2 L Luria-Bertani broth (LB) cultures containing antibiotic and grown at 37°C until reaching an OD_{600} of 0.5. Cultures were then cooled at 4°C for 1 hour without shaking. Subsequently, protein expression was induced by the addition of isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM, followed by incubation with shaking at 15°C overnight. For U2AF construct expression 100 μM $ZnCl_2$ was also supplemented upon induction.

SF3B6-SF3B1 and U2AF59-U2AF23 Individual Purifications. Overexpressed cells were harvested by centrifugation for 30 minutes at 3000 x g and 4°C. Cell pellets were resuspended in 30 mL of lysis buffer [20 mM Tris pH 8.0, 100 mM NaCl, 20 mM imidazole, 1 mM phenylmethylsulfonyl fluoride (PMSF), 1 mg/mL lysozyme, 5 mM β-mercaptoethanol (BME)] per liter of cells and stirred for 1 hour at 4°C. Subsequently, the lysate was sonicated on ice for 6 cycles of 30 seconds with 0.5 second pulses at 80% amplitude, with 1 minute of rest between each cycle. The lysate was cleared by centrifugation at 14 500 x g at 4°C, then the supernatant was incubated with 2 mL of Ni-nitrilotriacetic acid (Ni-NTA) resin previously equilibrated with wash buffer (20 mM Tris pH 8.0, 100 mM NaCl, 20 mM imidazole, 5 mM BME) for 1 hour at 4°C with gentle mixing. After allowing binding to occur, the resin-lysate mixture was transferred to a gravity flow column and a flowthrough fraction was collected. The resin was washed with 5 x 10 mL of wash buffer, then eluted with 3 x 10 mL of elution buffer (20 mM Tris pH 8.0, 100 mM NaCl, 250 mM imidazole, 5 mM BME). Elutions were concentrated over a 3 to 30 kDa molecular weight cut off ultrafiltration device (Amicon) depending on construct size. Concentrates were purified by gel filtration on a Superdex-75 26/60 column (GE Healthcare) equilibrated with wash buffer lacking imidazole. Peak fractions were analyzed by SDS-PAGE and concentrated. The polyhistidine (His_6) tags were cleaved via overnight incubation at 4°C with Tobacco Etch Virus (TEV) protease. Subsequently, SF3B6-SF3B1 constructs were purified by anion exchange with a MonoQ column (GE Healthcare) using a 0-300 mM NaCl gradient over 40 mL. U2AF constructs were purified using a 0-350 mM NaCl gradient over 70 mL. Peak fractions were analyzed by SDS-PAGE, concentrated, and quantified via absorbance at 280 nm.

SF3B6-SF3B1-U2AF59-U2AF23 Reconstitution Analysis. Individually purified SF3B6-SF3B1 and U2AF concentrates (see above) were combined in equimolar ratio at 4℃, and subjected to size exclusion chromatography over a Superdex-200 16/60 column (GE Healthcare) equilibrated with wash buffer lacking imidazole. Chromatogram fractions were analyzed by 16-20% (200:1 acrylamide:bisacrylamide) SDS-PAGE.

RNA Transcription. Transcription reactions were prepared *in vitro* to contain 40 mM Tris pH 8.0, 10 mM $MgCl_2$, 4 mM of each nucleotide triphosphate (NTP), 4 mM spermidine, 4 mM dithiothreitol (DTT), 25 mM NaCl, 2.5% dimethyl sulfoxide (DMSO), 0.3 μM template DNA [purified polymerase chain reaction (PCR) products], and 0.5 U/μL T7 RNA polymerase. Reactions were incubated at room temperature overnight. RNAs were resolved by electrophoresis through 6 to 10% (19:1 acrylamide:bisacrylamide) gels with 8M urea and 1X Tris-borate-ethylenediaminetetraacetic acid (TBE) depending on transcript length. RNA bands were visualized by UV shadowing, excised, and crushed. Crushed gel bands were resuspended in 360 μL diethylpyrocarbonate (DEPC) treated water, 40 μL phenol, and 40 μL of 3M NaOAc, then incubated at 37℃ overnight with shaking to extract RNA. Gel pieces were separated from extracted RNAs via spin column, then purified by phenol/chloroform extraction using 1 volume of phenol, followed by 1 volume of chloroform/isoamyl alcohol (24:1), and finally precipitated with 1 mL of chilled ethanol. Following precipitation, RNAs were resuspended in DEPC treated water and quantified by absorbance at 260 nm.

5′-End Labelling. 200 pmol of each transcript was 5′ dephosphorylated by incubation with 4 U/μL of calf intestinal alkaline phosphatase (CIAP, New England Biolabs) in 1X NEB3 buffer at 37℃ for 1 hour. NaOAc was added to each dephosphorylated RNA to a concentration of 0.3 M, then RNAs were phenol/chloroform extracted and ethanol precipitated as previously described. Precipitated RNAs were resuspended in DEPC treated water, then 200-2000 ng of each were end-labelled by incubation at 37℃ for 1 hour with 1 μL of [γ-^{32}P]-ATP using 2 U/μL T4 polynucleotide kinase and 1X kinase buffer (Thermo). RNAs were phenol/chloroform extracted, ethanol precipitated, resuspended in DEPC treated water and stored at -20℃.

EMSAs. Samples containing pre-mRNAs (5-10 $\times$ 10^3 cpm per gel lane), 10 mM $MgCl_2$, 0.05 mg/mL yeast transfer RNA (tRNA), 0.5 U/μL RNaseOut (Thermo), and 0-300 μM protein were prepared and incubated on ice for 30 minutes to allow protein-RNA binding. Following equilibration, samples were resolved by electrophoresis (200 V, 5 hours, 4°C) through a 6% 50 mM Tris-glycine gel. A phosphor screen was exposed to gels dried under vacuum at 80°C and scanned using a Typhoon 9400 Variable Mode Imager (GE Healthcare).

Chapter 3

Biochemical Characterization of the A3′ Complex in the Presence and Absence of SF3B6

Chapter 3. Biochemical Characterization of the A3′ Complex in the Presence and Absence of SF3B6.

Results and Discussion.

WT and ΔSF3B6 A3′ Complex Formation.

Splicing extracts of WT and ΔSF3B6 *S. pombe* cultures were prepared to investigate the mechanistic role of SF3B6 in splicing by biochemical characterization of the A complex in the presence and absence of SF3B6. To form and study homogenous A complex, free of subsequently formed splicing complexes, a series of model 3′ half pre-mRNAs were transcribed (Figure 3-1). These half pre-mRNAs contained an optimized branch point (UACUAAC) and 3′ SS but lacked the 5′ SS which is required for the first transesterification reaction of splicing. Consequently, spliceosomes accumulate and are stalled in an A complex mimic (Lamond *et al.* 1987) here denoted as the "A3′ complex." With the aim of purifying A3′ complex utilizing an MS2 aptamer based pull-down system (Jurica and Moore, 2002), an MS2-tagged A3′ half intron (A3′-MS2 intron) was transcribed. Time-course splicing reactions with radiolabelled A3′-MS2 intron were prepared in the presence and absence of ATP with WT and ΔSF3B6 splicing extracts to assess A3′ complex formation via electrophoretic mobility shift. A3′ complex was observed to form upon incubation at 30℃ in the case of both WT and ΔSF3B6 extracts when analyzed by native agarose gel (Figure 3-2 A) which was consistent with a characterization of the *S. pombe* A complex (Huang *et al.* 2002). Interestingly, A3′ complex was found to accumulate to a maximum signal intensity by 15-30 minutes followed by diminishing signal strength with further incubation at 30℃ (Figure 3-2 A). This signal loss could not be detected as unbound RNA nor as an otherwise shifted complex. This potentially suggests RNA degradation over time at 30℃ due to RNase presence in splicing extracts despite the addition of RNase inhibitor in the reactions and may not necessarily reflect intrinsic complex stability. Surprisingly, the A3′ complex was found to form on extract previously depleted of ATP (Figure 3-2 A) inconsistent with a previous characterization indicating the requirement of an ATP-dependent dissociation of the U2/5/6 tri-snRNP prior to A complex formation (Huang *et al.* 2002). An ATP-independent formation of the A3′ complex was also observed in ΔSF3B6 extract when resolved by native-PAGE (Figure 3-2 B). Time-course samples had to be equilibrated with heparin, a polyanion disrupting non-specific RNA-protein interactions, in order to resolve complexes by native-PAGE. With heparin

treatment, WT A3′ complex was not found to form while ΔSF3B6 complex persisted (Figure 3-2 B) possibly suggesting greater ΔSF3B6 complex stability.

WT and ΔSF3B6 A3′ Complex Stability.

A series of A3′-MS2 intron chase experiments were conducted to further investigate A3′ complex stability in the presence and absence of SF3B6. Splicing reactions pre-incubated with unlabelled competitor A3′-MS2 RNA followed by the addition of radiolabelled A3′-MS2 RNA and further incubation, "hot chases", can be used to estimate the concentration of A3′ complex available in a reaction. The reverse experiment, reactions incubated with radiolabelled A3′-MS2 RNA followed by the addition of unlabelled competitor A3′-MS2 RNA, "cold chases", can be used as an indication of complex stability and turnover. With a hot chase, WT *S. pombe* exhibited a loss of A3′ complex in favor of H complex or unbound RNA beginning at 400-2000 nM competitor (Figure 3-3 A). In a cold chase, the WT A3′ complex was observed to largely persist upon titration of competitor intron up to 2000 nM (Figure 3-3 B) in excess of the estimated available complex. A hot chase of ΔSF3B6 splicing reactions revealed a distinct loss of A3′ complex at 10 nM (Figure 3-4 A). A cold chase of ΔSF3B6 reactions exhibited largely stable A3′ complex up to 400 nM (Figure 3-4 B), also in excess of the estimated available complex. With the out-competition of radiolabelled A3′-MS2 RNA observed only in considerable excess of the available complex (Figures 3-3 and 3-4), these chase experiments suggest the A3′ complex to be relatively stable outside of prolonged 30℃ incubation (Figure 3-2) in the presence and absence of SF3B6. The low amount of complex turnover was consistent with previously characterized *S. pombe* A complex (Huang *et al.* 2002).

snRNA Dependence of ΔSF3B6 A3′ Complex Formation.

To investigate the snRNA dependence of the ΔSF3B6 A3′ complex, extract depleted of U1, U2, U3 or U6 snRNA via RNase H and antisense oligonucleotide treatment was analyzed for complex formation. A shifted A3′ complex was observed to form upon incubation at 30℃ with comparable intensity to a no-treatment control following targeted degradation of the U1, U3 and U6 snRNAs (Figure 3-5). Consistent with previously characterized *S. pombe* A complex (Huang

et al. 2002), targeted degradation of the branch point binding region of the U2 snRNA eliminated the formation of the A3′ complex. This suggested that U2 snRNP recruitment is also required for A3′ complex formation in ΔSF3B6 extract, and that the U2 snRNP assembles on pre-mRNA in the absence of SF3B6.

To further confirm that the observed A3′ complex was U2 snRNA dependent, WT and ΔSF3B6 H and A3′ complexes were resolved by native agarose, then visualized, excised and subjected to reverse transcriptase (RT)-PCR analysis using U1 and U2 snRNA specific oligonucleotides. Consistent with the previous analysis, U2 snRNA was detected in both the WT and ΔSF3B6 A3′ complexes (Figure 3-6). Surprisingly, U1 snRNA was also detected in both the WT and ΔSF3B6 A3′ complexes, despite the lack of a U1 snRNA binding region and 5′ SS in the model A3′ introns. U1 and U2 snRNAs were also detected in the WT and ΔSF3B6 H complexes. The presence of the U1 and U2 snRNAs in these shifted complexes reinforces the suggestion that they correspond to spliceosomal complexes. The unexpected detection of the U1 snRNA in the A3′ complexes may be explained by comigration of endogenous *S. pombe* A complex with the labelled A3′-MS2 RNA.

MBP-Affinity A3′ Complex Purification for Component Analysis by Mass Spectrometry. Aiming to validate that the U2 snRNP forms and is recruited to pre-mRNA in the absence of SF3B6, a pull-down procedure for A3′ complex spliceosomes was developed for the identification of A complex components (Figure 3-7). Complex forming reactions were prepared with WT and ΔSF3B6 extract using A3′-MS2 RNA pre-bound to MS2-MBP fusion protein, then subjected to size exclusion and amylose affinity chromatography. Doping reactions with radiolabelled A3′-MS2 RNA, the tagged-intron was traced throughout the purification procedure. As the size-exclusion step of the purification was designed only to crudely separate complexes excluded in the void volume from monomeric proteins and unbound or degraded RNA retained by the resin, a small column packed with Sephacryl S-500 (S-500) was chosen. However, analysis of the elution profile from the small S-500 column revealed only a single RNA-containing peak (Figure 3-8). This single elution peak was similarly observed for both WT and ΔSF3B6 complexes. To improve resolution of complexes from monomeric proteins and unbound RNA, Superdex S-200 (S-200) was selected as a substitute gel filtration resin. Over the

S-200 resin, complex forming reactions exhibited two distinct RNA-containing elution peaks (Figure 3-9 B) which were shifted slightly relative to the peak generated by the independent elution of long A3′-MS2 RNA (Figure 3-9 A). Native agarose analysis of the S-200 fractions confirmed that peak 1 of the complex forming reactions corresponded to a mobility shift relative to the RNA only peak (Figure 3-10 A). Peak fractions were then pooled and concentrated over a 100 kDa molecular weight cut-off ultrafiltration unit, which should retain only A3′ RNA complexed with additional factors. However, the shifts observed for peak 1 prior to concentration (Figure 3-10 A), as well as after concentration (Figure 3-10 B) did not appear to clearly correspond to the A3′ complex loading control. Rather, the shifts observed appeared smeared and largely corresponding to the H complex, likely on account of non-specific RNA-protein binding. Despite the lack of a clearly defined A3′ complex, this pull-down experiment was attempted a second time, this attempt with the inclusion of MS2-MBP to assess subsequent amylose affinity purification of the complex peak. Complexes concentrated following amylose affinity elution with 10 mM maltose were found to exhibit a tight shift that appears distinct from both MS2-MBP bound RNA as well as the A3′ complex (Figure 3-11). SDS-PAGE analysis of the A3′ complex amylose elution revealed a band corresponding to MS2-MBP, further supporting the effectiveness of the aptamer-based pull-down system (Figure 3-12). Several protein bands were also found to be enriched throughout the affinity pull-down (Figures 3-12 and 3-13). Coupled with the observed binding of radiolabelled long A3′ RNA on the amylose resin until elution with maltose (Table 1), which must be mediated by MS2-MBP, this distinct shift likely represents a complex of the RNA with MS2-MBP and additional *S. pombe* factors. Furthermore, a similar enrichment of proteins was achieved upon complex elution with RNase A in the absence of maltose (Figure 3-13), indicating that complex-resin interactions were RNA dependent. However, based on the native gel shift position alone (Figure 3-11), it remained unclear if this unique shift represented the binding of A complex factors, or other nonspecific binding partners.

Several pull-downs of both WT and ΔSF3B6 complexes were analyzed with the adjustment of multiple variables, including reaction scales ranging from 0.2-20 mL, as well as MS2-MBP and pre-mRNA concentrations ranging from 1-10 μM and 10-100 nM respectively. Buffer conditions were also varied, testing the inclusion of 0.05% Nonidet P-40 or 2 mM magnesium acetate, or the exclusion of ethylenediaminetetraacetic acid (EDTA). Pull-downs were also analyzed with

and without equilibration to 0.5 mg/mL heparin following complex formation. A one-step amylose affinity pull-down was also analyzed. Each of the examined conditions yielded similar protein enrichment results via SDS-PAGE. The presence of the MS2-MBP fusion protein in the amylose affinity elutions with 10 mM maltose was confirmed by mass spectrometry. However, mass spectrometry analyses of these A3′ complex pull-downs were unable to detect any SF3B subunits, nor many A complex factors expected to be present (data not shown).

Generating a Cys-SF3B6 Construct for N-terminal Derivatization.

As an alternative approach to confirm the formation of the A3′ complex in the absence of SF3B6, a recombinant SF3B6 construct was generated for site-specific N-terminal derivatization. Peptides with an N-terminal Cys residue can be site-specifically derivatized using previously established chemistry (Erlanson *et al.* 1996; Kent *et al.* 2003) with a variety of thioester reagents useful for biochemical characterization. One such derivatization reagent that could be utilized is biotin-3-mercaptopropionic acid (MPA). SF3B6 with an N-terminal Cys residue (Cys-SF3B6) could be site-specifically derivatized with biotin-3-MPA (Figure 3-14 A) and added back to ΔSF3B6 A3′ complex forming reactions. ΔSF3B6 splicing complexes supplemented with biotinylated SF3B6 could be resolved by native-PAGE for the extraction of the A3′ complex. Subsequently, the extracted A3′ complex could be resolved by SDS-PAGE for Western analysis using a streptavidin conjugated dye for the detection of biotinylated SF3B6. If the A3′ complex is assembling on a pre-mRNA within ΔSF3B6 extract, then the biotinylated SF3B6 supplemented to the reaction should be incorporated as a component of SF3B. An alternative to spliceosome purification and protein identification, this proposed detection of exogenously added SF3B6 in ΔSF3B6 A3′ complex would implicate SF3B assembly in the absence of the SF3B6 subunit.

Cleavage of an affinity tag with TEV protease leaves a protein of interest with one non-native N-terminal residue, the P1′ residue of its recognition sequence. TEV recognition sequences typically contain a glycine or serine residue in the P1′ position for optimal cleavage efficiency but other residues including Cys may be substituted (Kapust *et al.* 2002). An SF3B6 construct was designed to feature an N-terminal His_6 tag, which can be cleaved following nickel affinity purification with TEV protease to yield full-length SF3B6 with an introduced N-terminal Cys

residue (Figure 3-14 B). This SF3B6 precursor construct was successfully purified and efficiently cleaved with TEV protease (Figure 3-14 C). Therefore, a Cys-SF3B6 construct suitable for derivatization has been prepared and may be used to demonstrate SF3B assembly in ΔSF3B6 *S. pombe* extract.

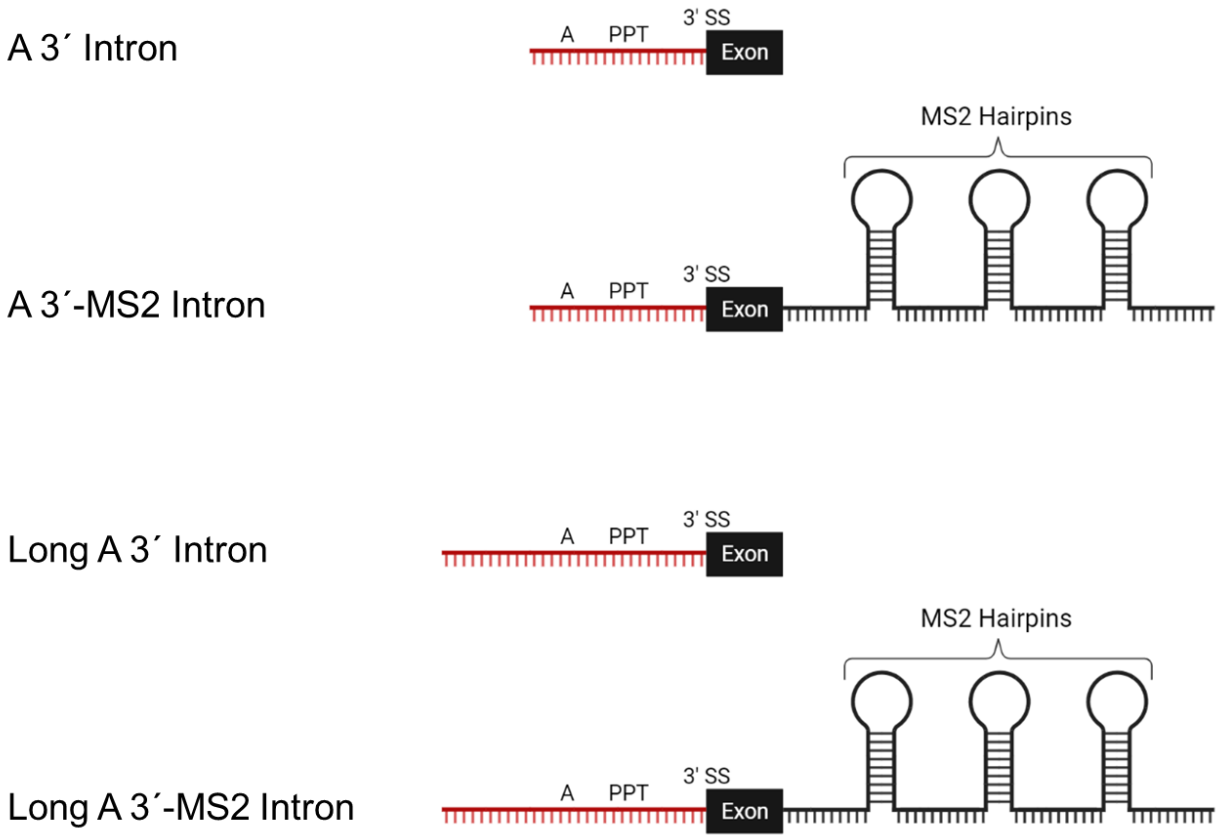

Figure 3-1. Schematic representation of A3´ half introns. A3´half introns used for the biochemical characterization of the *S. pombe* A3´ complex. The 3´ MS2 aptamer tags of the tagged introns are shown. The long A3´ intron sequence matches the "3´ half RNA" used in a previous characterization of the *S. pombe* A complex (Huang *et al.* 2002).

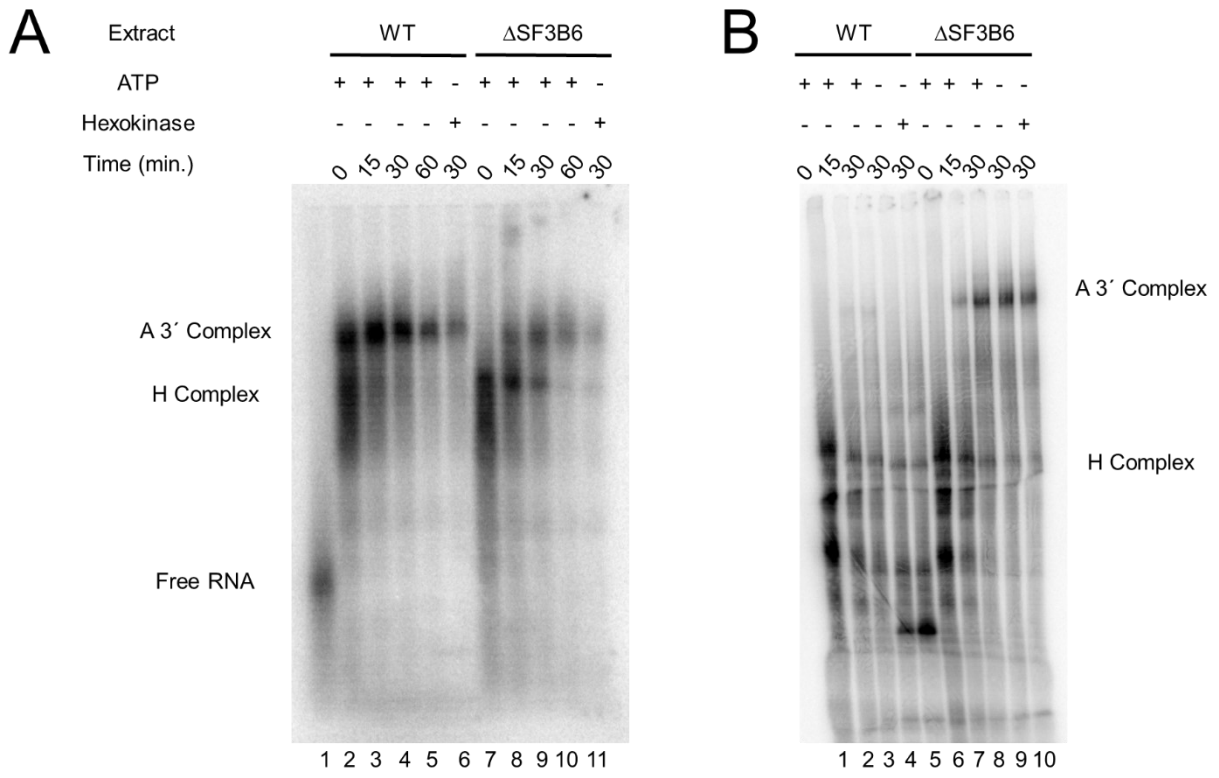

Figure 3-2: A3´ complex formation on MS2-tagged pre-mRNAs in the presence and absence of SF3B6. EMSA of WT and ΔSF3B6 *S. pombe* splicing time-course reactions with radiolabelled A3´intron – MS2 RNA resolved by (A) 1.2% agarose in the absence of heparin and (B) 4% acrylamide following equilibration with 0.5 mg/mL heparin. RNA migration in the absence of *S. pombe* extract is shown in lane 1.

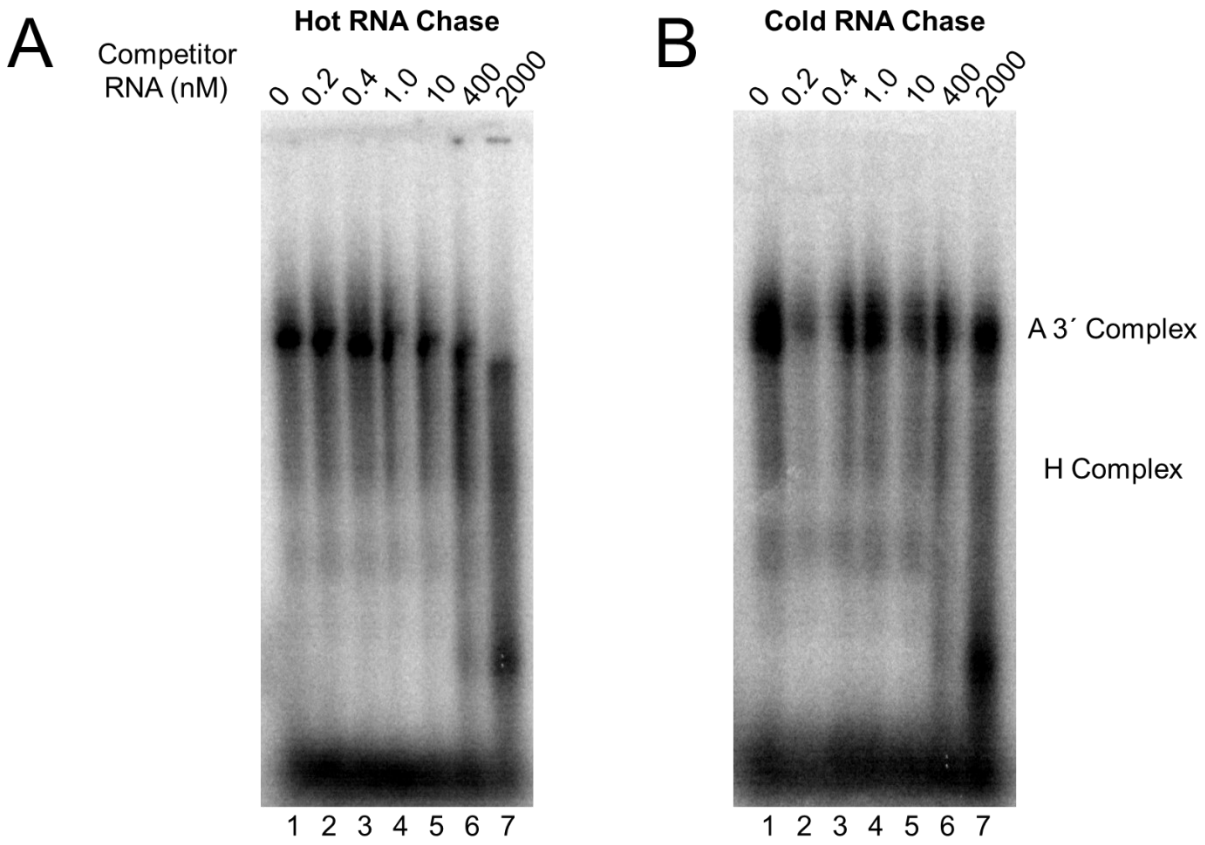

Figure 3-3: A3´ complex stability in the presence of SF3B6. Unlabelled competitor A3´ intron – MS2 RNA was incubated for 30 minutes at 30℃ either before (A) or after (B) complex formation on radiolabelled A3´ intron – MS2 RNA via splicing reactions at 30℃ for 30 minutes with WT *S. pombe* extract.

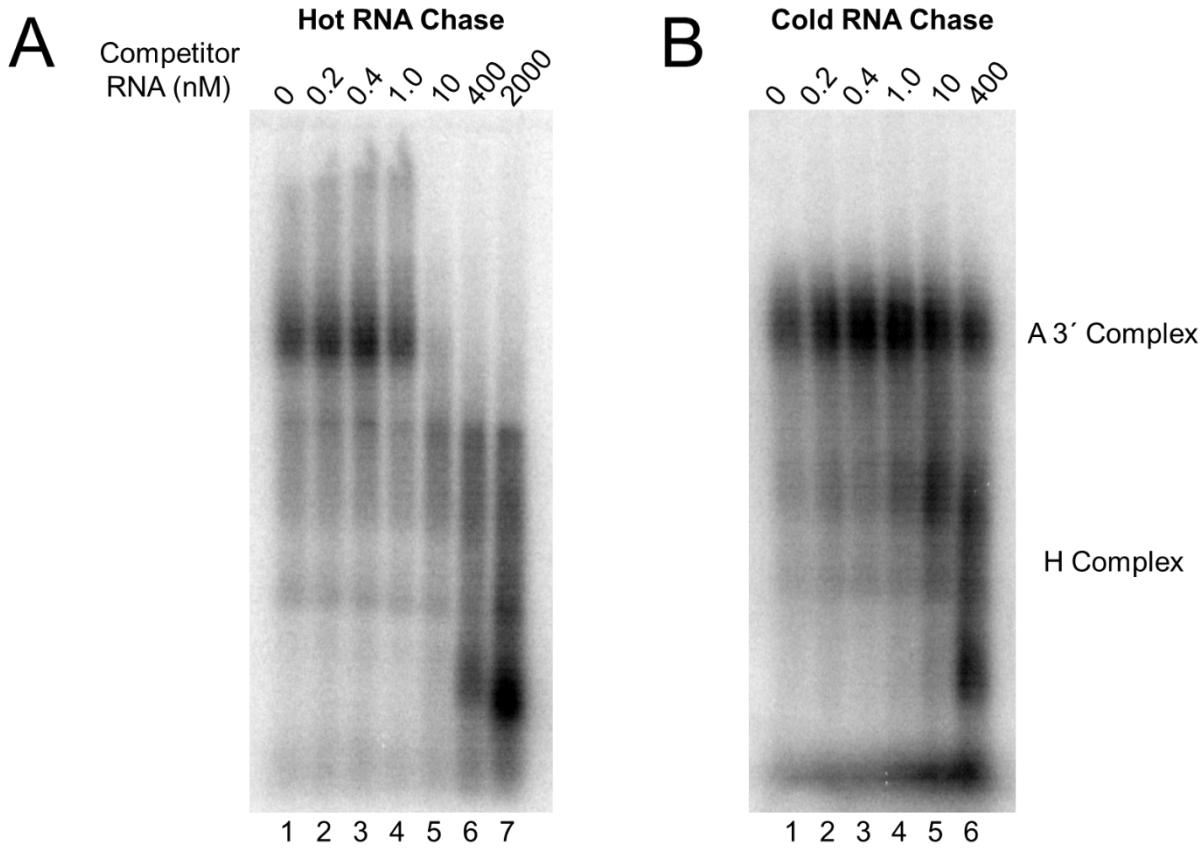

Figure 3-4: A3´ complex stability in the absence of SF3B6. Unlabelled competitor A3´ intron – MS2 RNA was incubated for 30 minutes at 30℃ either before (A), or after (B) complex formation on radiolabelled A3´ intron – MS2 RNA via splicing reactions at 30℃ for 30 minutes with ΔSF3B6 *S. pombe* extract.

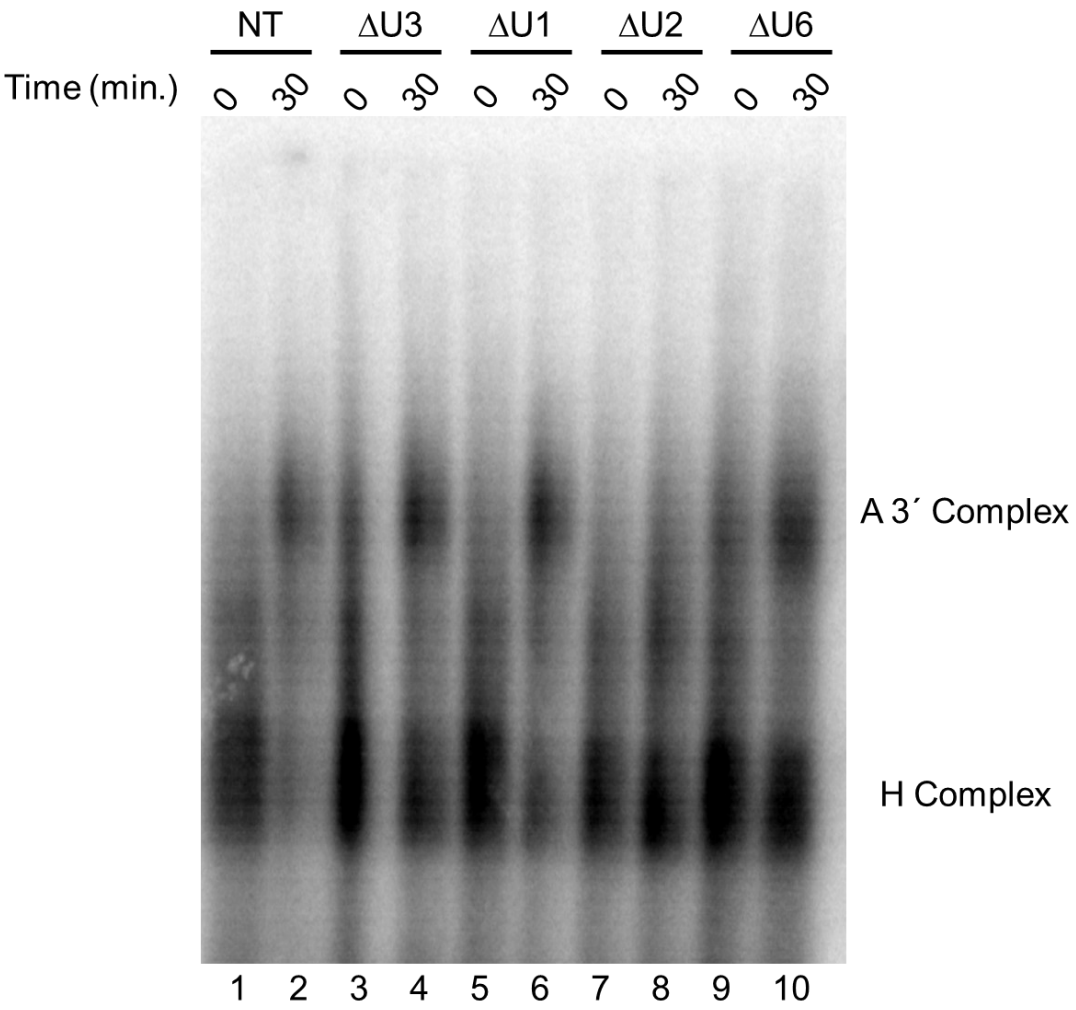

Figure 3-5: A3´ complex snRNA dependence in the absence of SF3B6. A3´ Complex formation on A3´ intron – MS2 RNA upon individual snRNA degradation by targeted antisense DNA oligonucleotide and RNase H treatment of ΔSF3B6 *S. pombe* extract. NT indicates no RNase H treatment, ΔU3 indicates extract treated with oligonucleotide complementary to the U3 snRNA, ΔU1 indicates extract treated with oligonucleotide complementary to the 5´ SS pairing region of the U1 snRNA, ΔU2 indicates extract treated with oligonucleotide complementary to the BPS pairing region of the U2 snRNA, and ΔU6 indicates extract treated with oligonucleotide complementary to the U6 snRNA.

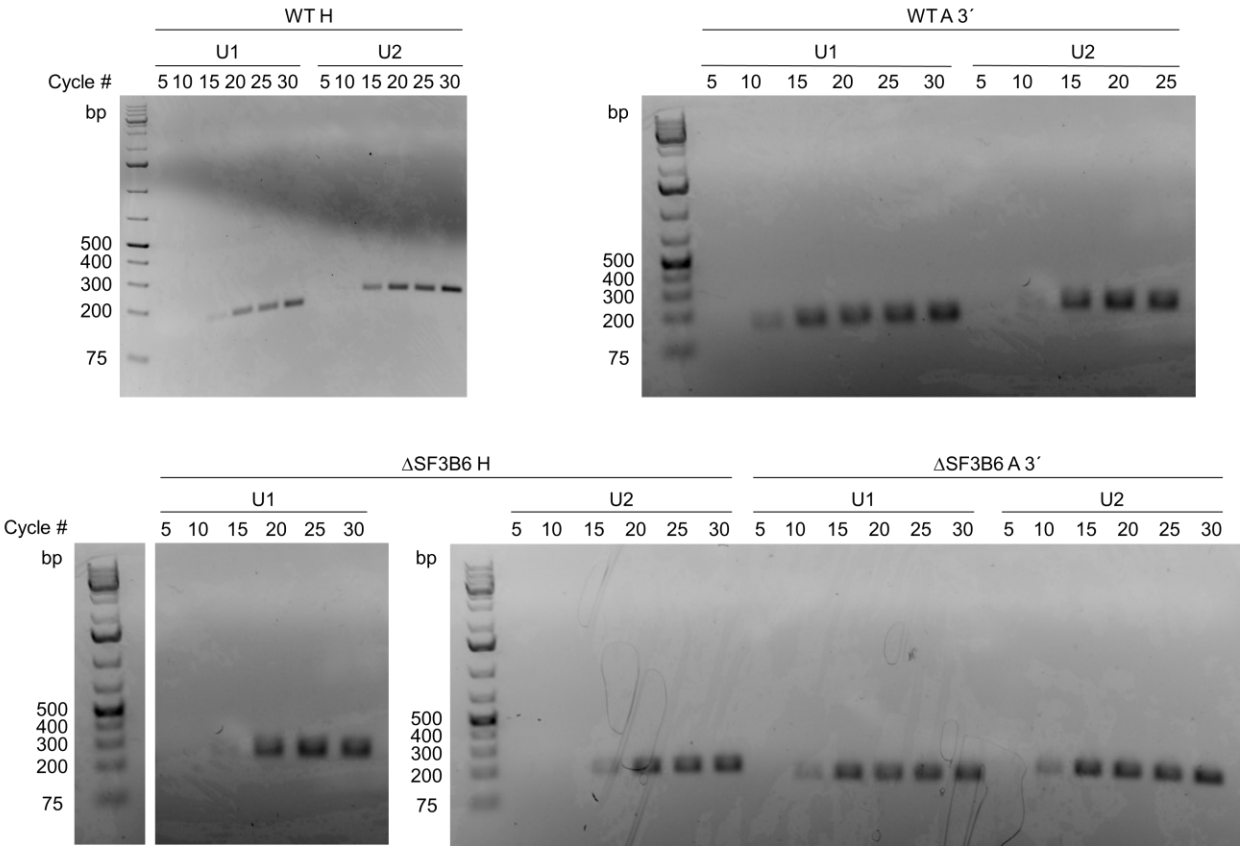

Figure 3-6. U1 and U2 snRNA detection in H and A3´ complexes. RT-PCR analysis of U1 and U2 snRNAs in WT and ΔSF3B6 H and A3´ complexes resolved by native agarose gel electrophoresis.

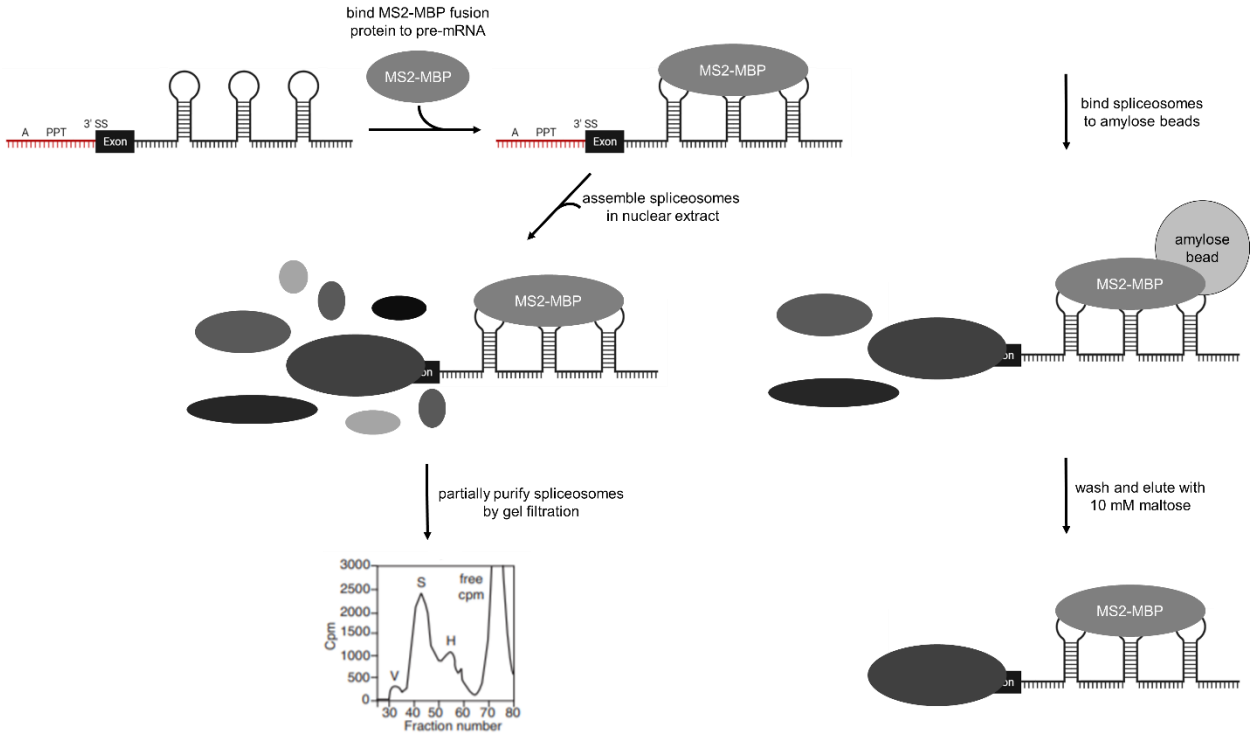

Figure 3-7. A3´ complex purification strategy. Schematic representation of a proposed MS2-MBP-mediated A3´ complex purification (adapted from Zhou and Reed, 2002).

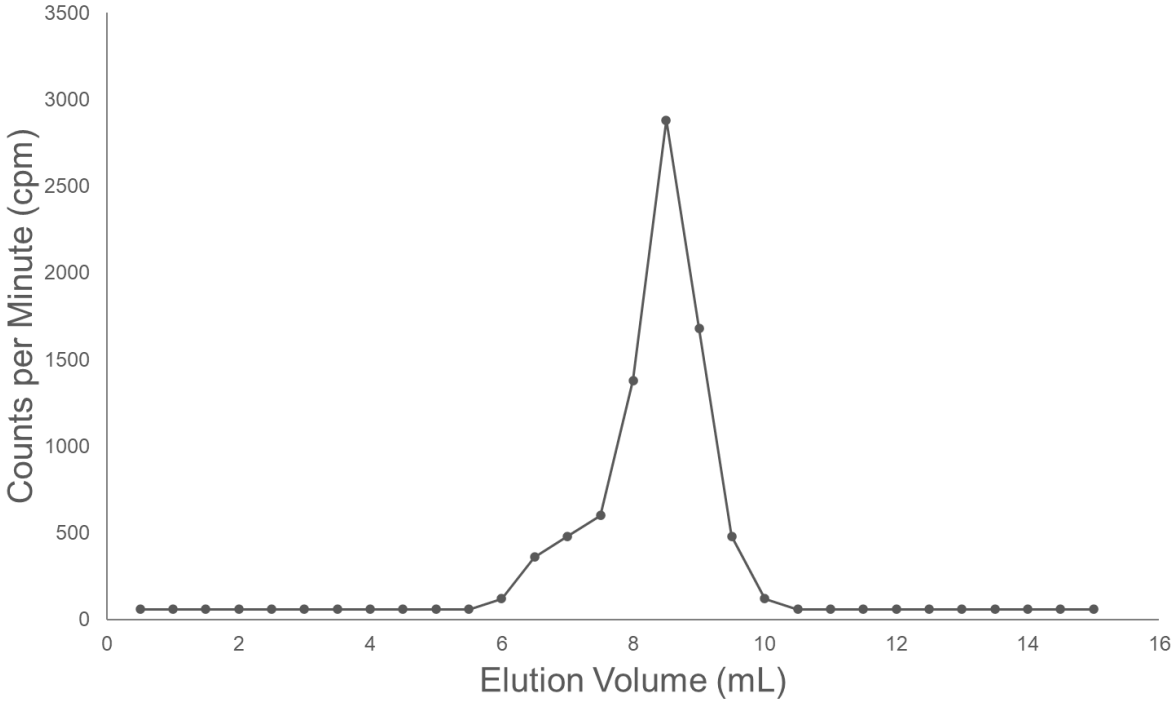

Figure 3-8. A3´ complex elution over S-500 column. Representative S-500 elution chromatogram of A3´ complexes formed on A3´-MS2 RNA pre-bound to MS2-MBP with WT *S. pombe* extract. Reactions were doped with radiolabelled RNA and fractions were quantified with a Geiger counter.

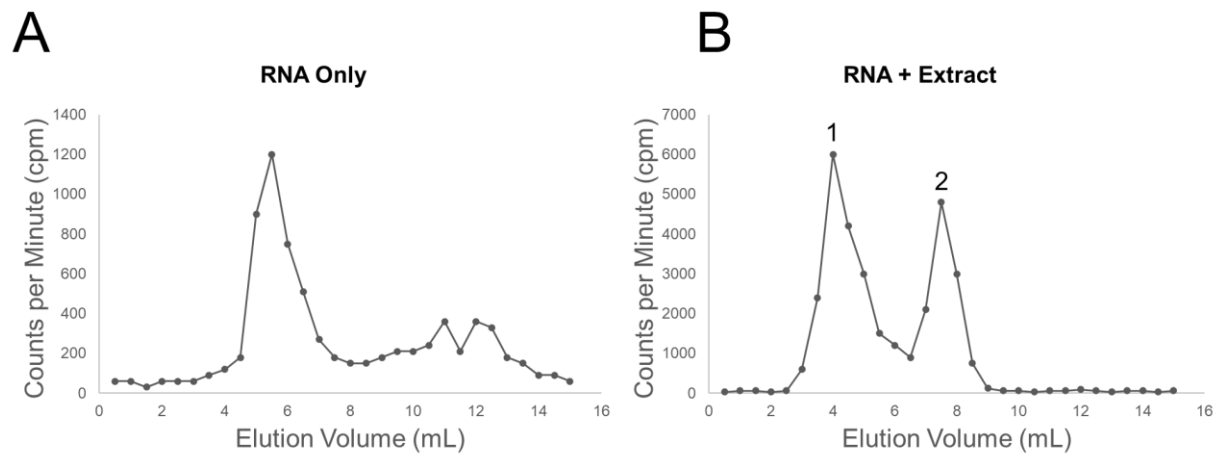

Figure 3-9. A3´ complex elution over S-200 column. Representative S-200 elution chromatogram of A3´ complexes formed on long A3´-MS2 RNA with WT *S. pombe* extract. Reactions were doped with radiolabelled RNA and fractions were quantified with a Geiger counter.

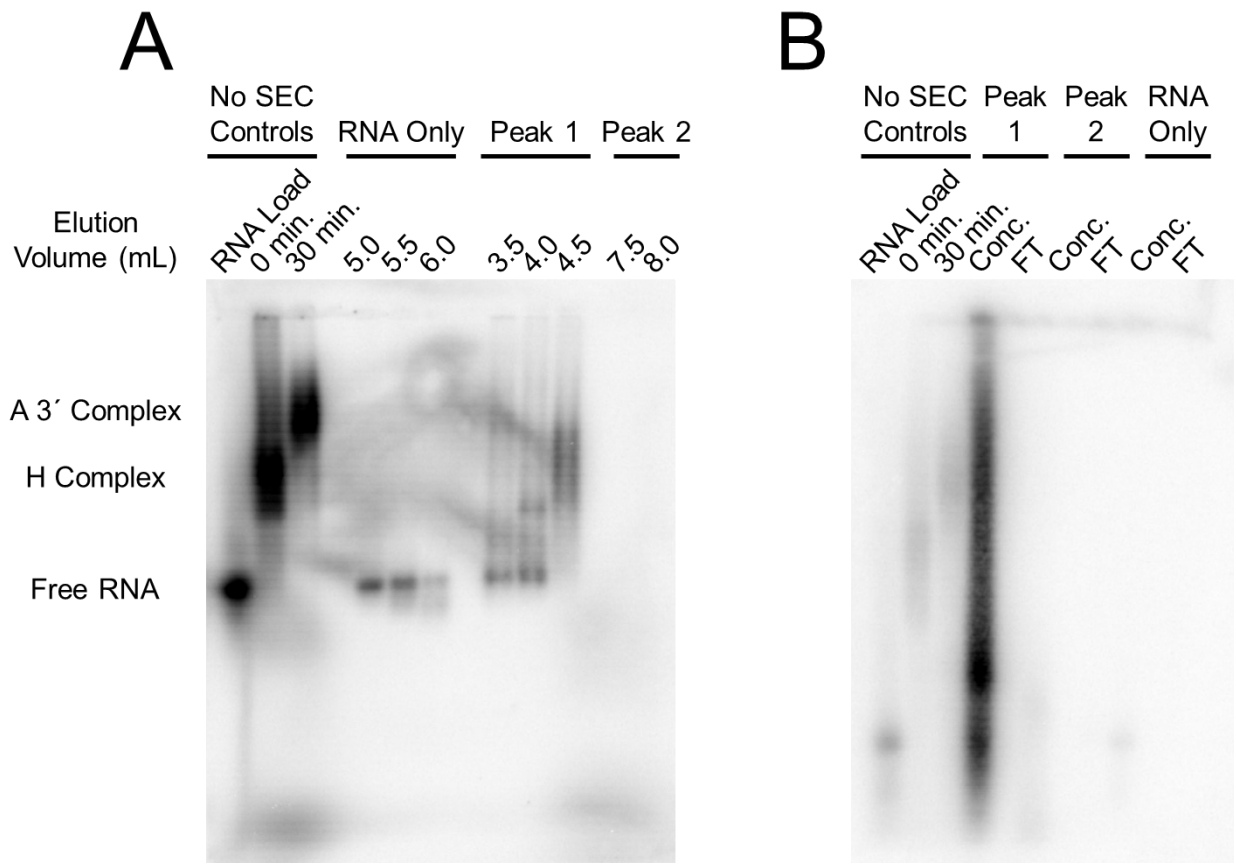

Figure 3-10. Mobility shift analysis of A3´ complex eluted over S-200. (A) EMSA of long A3´-MS2 RNA only and A3´ complex elution peaks (Figure 3-8 A and B). (B) EMSA of peak ultrafiltration concentrate (conc.) and flow through (FT) samples. SEC indicates size exclusion chromatography.

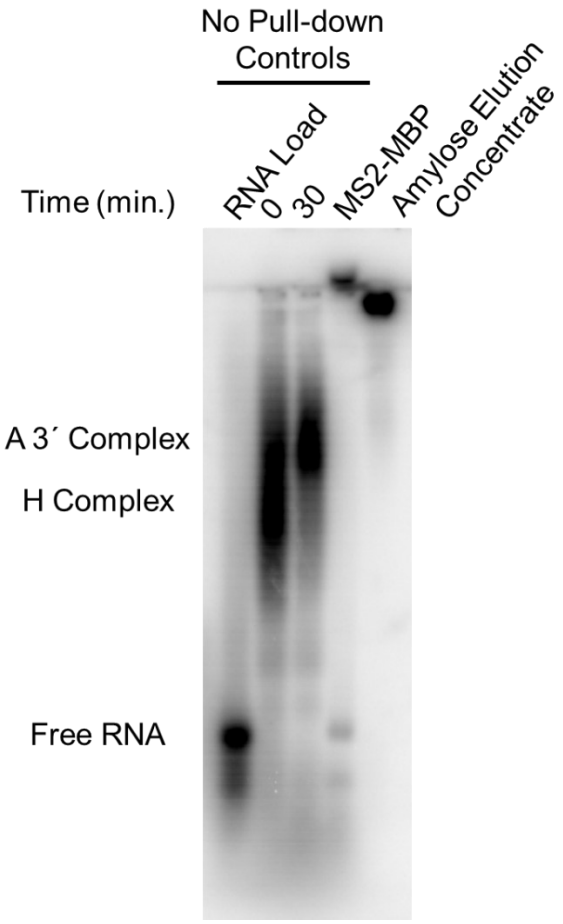

Figure 3-11. Mobility shift analysis of A3´ complex amylose affinity elution. EMSA of A3´ complex formed on long A3´-MS2 RNA pre-bound to MS2-MBP. Complexes were purified by gel filtration over S-200 resin followed by amylose affinity chromatography. Affinity elutions were concentrated.

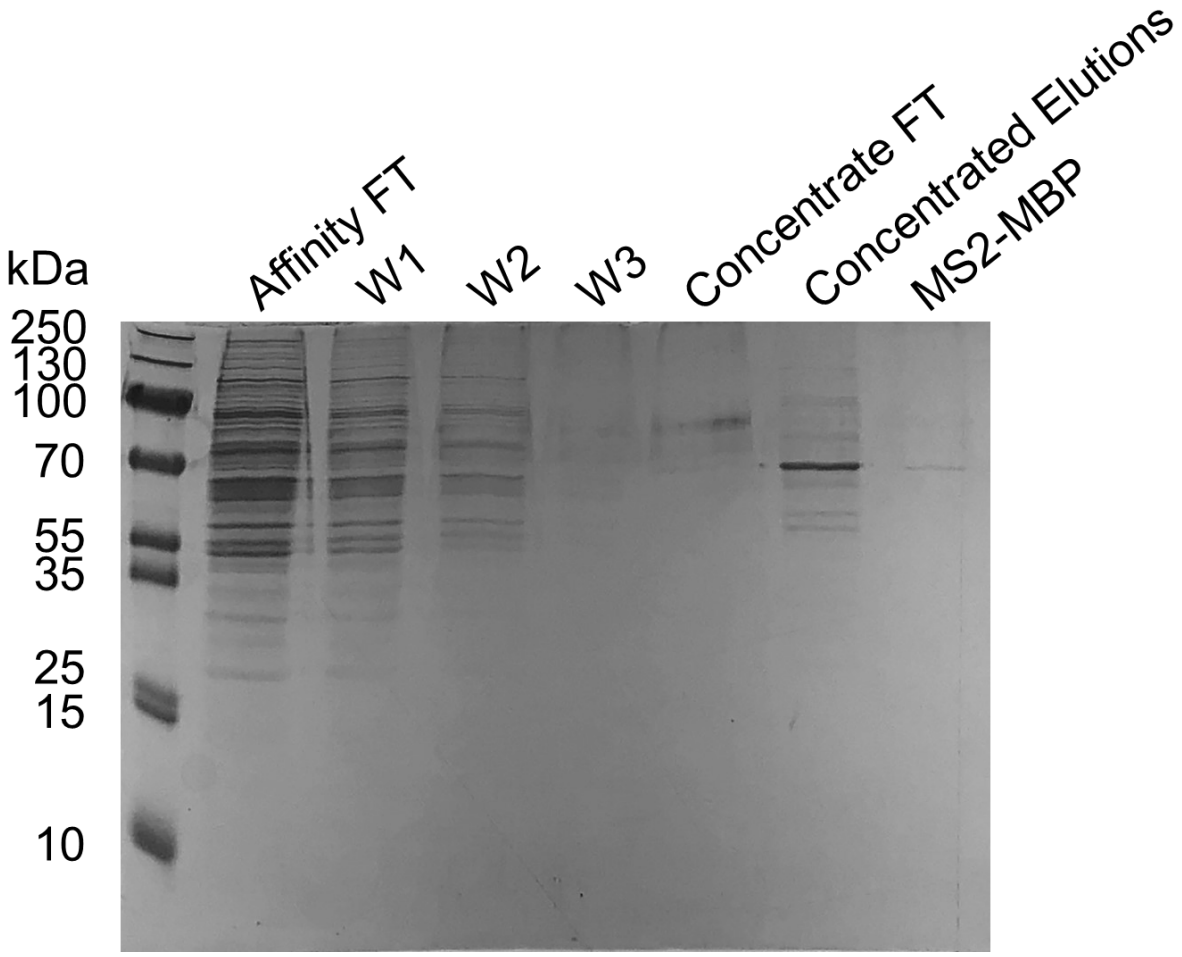

Figure 3-12. SDS-PAGE analysis of A3´ complex pull-down fractions eluted with 10 mM maltose. Representative SDS-PAGE of WT *S. pombe* A3´ complex pull-down samples. Complexes were eluted with 10 mM maltose. Proteins were visualized by colloidal Coomassie staining.

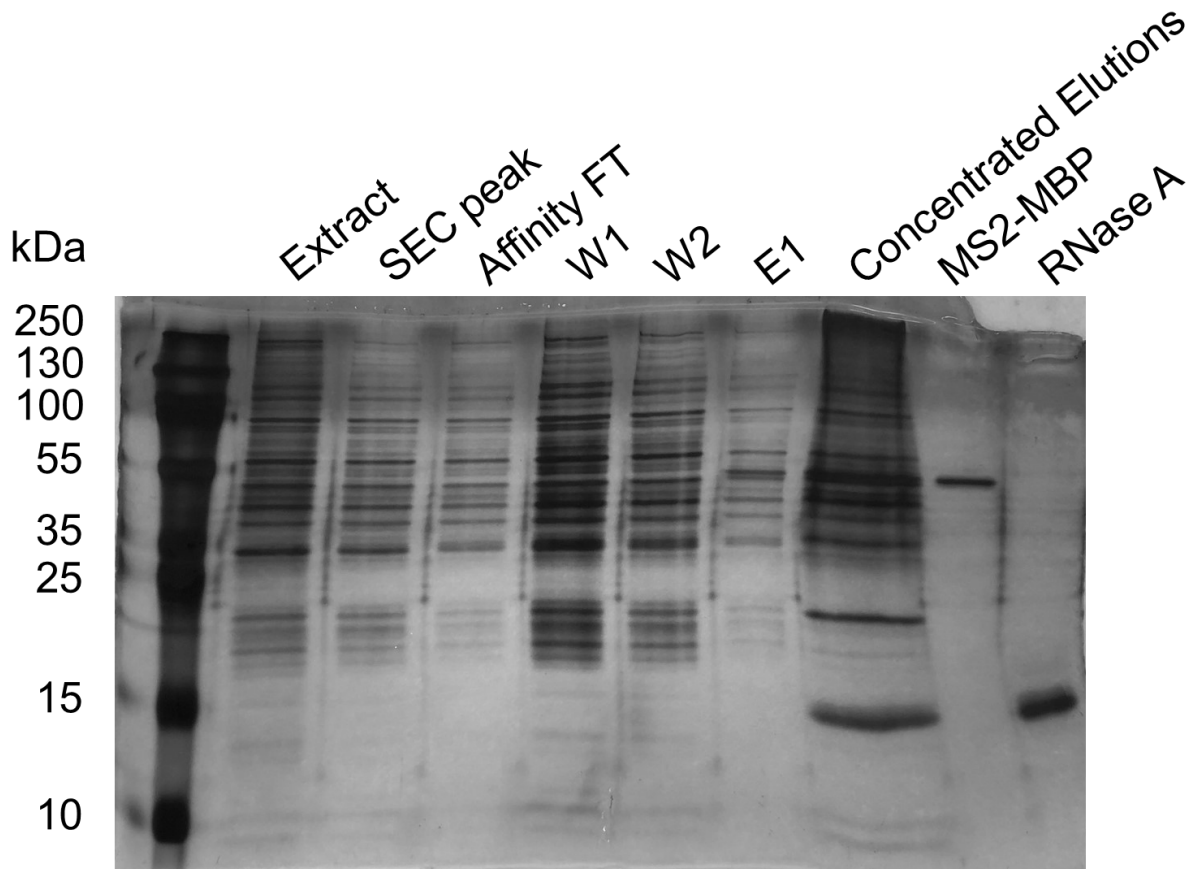

Figure 3-13. SDS-PAGE analysis of A3´ complex pull-down fractions eluted with RNase A. Representative SDS-PAGE of WT *S. pombe* A3´ complex pull-down samples. Complexes were eluted via RNase A mediated RNA degradation in this example. Similar protein enrichment was observed upon elution with RNase A relative to 10 mM maltose. SEC peak indicates a size exclusion chromatography complex elution peak sample. Proteins were visualized by silver stain.

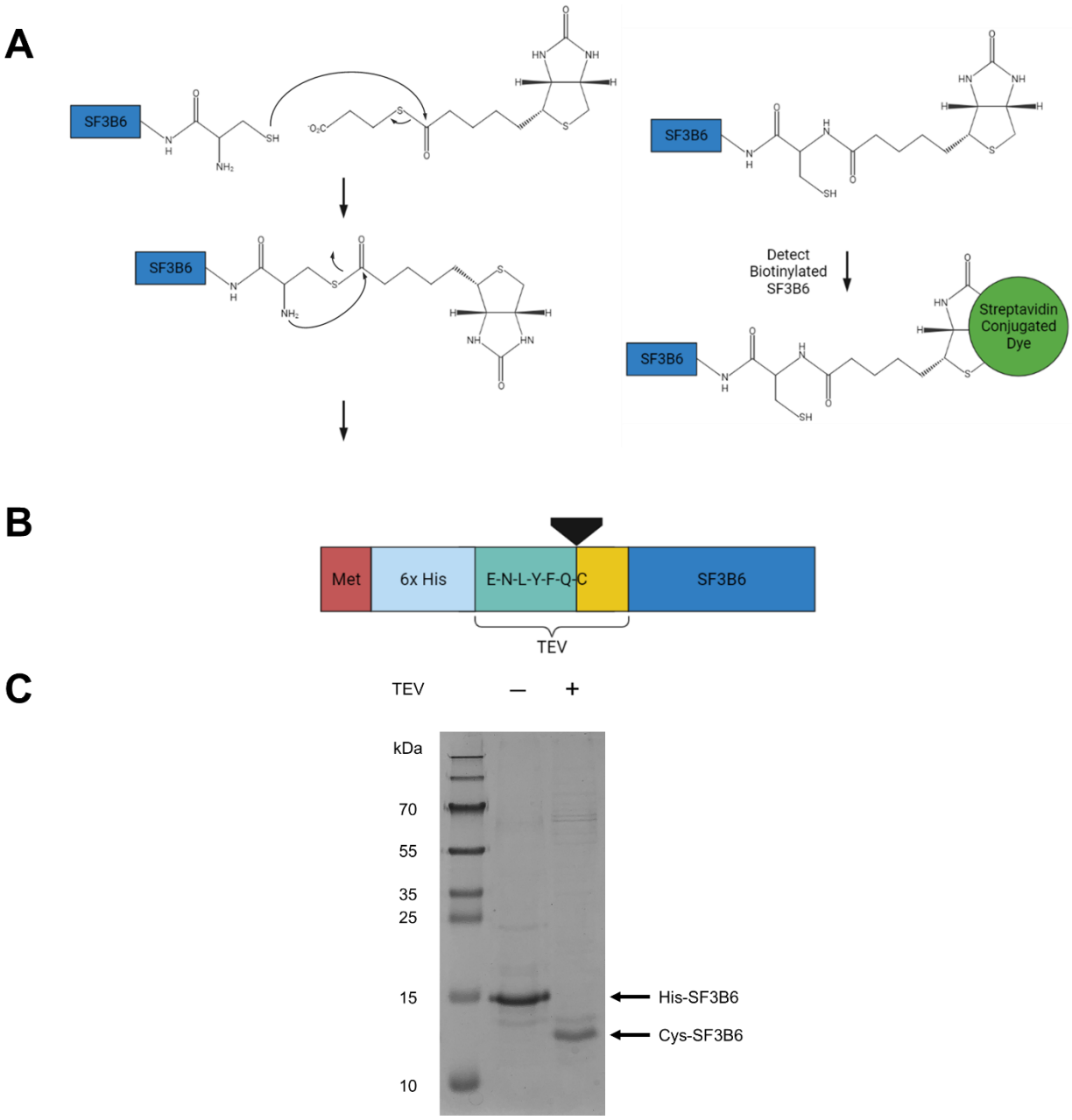

Figure 3-14: Generation of SF3B6 with an N-terminal Cys for site specific derivatization. (A) Example derivatization of Cys-SF3B6 with biotin-3-MPA. The rearrangement following the initial thiotransesterification reaction yields a stable amide linkage. Modification of internal Cys residues can be reversed in the presence of excess reducing agent. Biotinylated SF3B6 could be visualized with a streptavidin conjugated dye. (B) Schematic representation of precursor SF3B6 including an N-terminal His_6 tag followed by TEV recognition sequence. TEV cleavage leaves an N-terminal Cys residue. (C) SDS-PAGE of the SF3B6 construct before and after TEV cleavage.

Table 3-1. Long A3´-MS2 RNA detected throughout A3´ complex amylose affinity pull-down. Purification samples were quantified by Geiger counting.

Purification Sample	Activity (cpm)
Flow Through	180
Resin pre-elution	1800
Washes	60
Elution 1	1200
Elution 2	540
Elution 3	210
Resin post-elution	60

Materials and Methods

Cloning. DNA sequences of the short A3′ SF3B6 intron and the A3′ SF3B6 intron with BamHI recognition sequences on both the 5′ and 3′ ends were generated by PCR using overlapping oligos as template, then cloned into the BamHI sites of pMJ12 (gift from Melissa Jurica, UC Santa Cruz) and verified by Sanger sequencing. Also generated by PCR, the coding sequence of *S. pombe* SF3B6 engineered with an N-terminal TEV cleavage sequence of E-N-L-Y-F-Q↓C was cloned into the BamHI site of pET Duet-1 and verified by Sanger sequencing.

Pre-mRNA ^{32}P Body-Labelling. Radioactive transcriptions were performed similarly to cold transcriptions as described previously, but with 0.5 mM of each unlabelled NTP, 60 µM of cold NTP of that which was used to label, and 2 µCi/µL of α-[^{32}P]-ATP or UTP. RNAs were excised from gels and extracted as above. However, radioactive bands were visualized by autoradiography rather than by UV shadowing.

S. pombe **Extract Preparation.** Start-up cultures were inoculated with single colony isolates of previously streaked WT or ΔSF3B6 JK484 *S. pombe* cells and grown in YES media at 30°C overnight. YES cultures (2L) were inoculated with these start-ups and grown to an OD_{600} of 2 to 2.5, then harvested by centrifugation for five minutes at 1600 x g and 4°C. Cell pellets were washed with 50 mL of cold ddH_2O, resuspended with 50 mL of cold AGK buffer (10 mM HEPES pH 7.9, 1.5 mM $MgCl_2$, 200 mM KCl, 0.5 mM DTT, 1 mM PMSF, 10% glycerol), then pelleted by centrifugation. Cell pellets were resuspended with 3.75 mL of AGK buffer per 1 L of culture, then transferred dropwise through an 18- or 22-gauge needle into liquid nitrogen. The frozen cell droplets were ground in liquid N_2 with a mortar and pestle for approximately 40 minutes to a fine powder. Cell powders were thawed on ice with stirring, then centrifuged for 30 minutes at 18000 x g and 4°C. Avoiding the pellet and film at the top of the sample, the supernatant was extracted and centrifuged for 1 hour at 100 000 x g at 4°C. The resulting supernatant was dialyzed twice against 2 L of chilled buffer D (20 mM HEPES pH 7.9, 0.2 mM EDTA, 50 mM KCl, 0.5 mM DTT, 1 mM PMSF, 20% glycerol) for 1.5 hours each at 4°C using 6-8 kDa molecular weight cut off dialysis membrane (Fischer). Total protein concentration was

determined via Bradford Assay. Extract aliquots were snap frozen in liquid nitrogen and stored at -80°C.

MS2-MBP Expression and Purification. An overnight culture of BL21 Gold cells transformed with the expression vector were used to inoculate LB cultures (2L) containing the appropriate antibiotic and grown at 37°C until reaching an OD_{600} of 0.5. Cultures were then cooled at 4°C for 1 hour without shaking. Subsequently, protein expression was induced by the addition of IPTG to a final concentration of 0.5 mM, followed by incubation at 15°C overnight. Cells were harvested by centrifugation for 30 minutes at 3000 x g and 4°C. Cell pellets were resuspended in 30 mL of MS2-MBP lysis buffer (20 mM HEPES pH 7.9, 400 mM NaCl, 1 mM EDTA, 1 mM PMSF, 1 mg/mL lysozyme, 5 mM BME) per liter of cells and stirred for 1 hour at 4°C. Subsequently, the lysate was sonicated on ice for 6 cycles of 30 seconds with 0.5 second pulses at 80% amplitude, followed by 1 minute of rest. The lysate was cleared by centrifugation at 14 500 x g and 4°C, then the supernatant was incubated with 2 mL of amylose resin previously equilibrated with MS2-MBP wash buffer (20 mM HEPES pH 7.9, 50 mM NaCl, 1 mM EDTA, 5 mM BME) for 1 hour at 4°C with gentle mixing. After allowing binding to occur, the resin-lysate mixture was transferred to a gravity flow column and a flowthrough fraction was collected. The resin was washed with 3 x 10 mL of MS2-MBP lysis buffer, then with 2 x 10 mL of MS2-MBP wash buffer, and finally eluted with 5 x 10 mL of MS2-MBP elution buffer (20 mM HEPES pH 7.9, 50 mM NaCl, 1 mM EDTA, 10 mM maltose, 5 mM BME). Elutions were concentrated over a 30 kDa molecular weight cut off ultrafiltration unit (Amicon), then purified by gel filtration on a Superdex-75 26/60 column (GE Healthcare) equilibrated with MS2-MBP wash buffer lacking imidazole. Peak fractions were analyzed by SDS-PAGE and concentrated. Glycerol was added to a final concentration of 10%, then the concentrate was quantified by absorbance at 280 nm. Nucleic acid contamination was assessed by a ratio of absorbance at 280 and 260 nm. The A_{280} / A_{260} was found to be ~1.7, indicative of protein purified from nonspecific nucleic acid binding partners.

A3′ Complex Formation and EMSAs. Complex assembly reactions were prepared to contain pre-mRNAs (50×10^3 cpm per gel lane), 20 mM KOAc, 2 mM $MgOAc_2$, 0.05 mg/mL yeast

tRNA, 0.1 mM DTT, 0.5 U/µL RNaseOut (Thermo), 1 mM ATP, 5 mM creatine phosphate, and 40% WT or ΔSF3B6 *S. pombe* extract based on a previous characterization of *S. pombe* A complex (Huang *et al.* 2002). Reactions were carried out at 30°C for 0 to 60 minutes, stopped on ice with the addition of native loading dye and immediately loaded onto a 1.2 % agarose, 1X TBE gel and subjected to electrophoresis at 70 V for 4 hours at 4°C, or adjusted to 0.5 mg/mL heparin and loaded onto a 4% (80:1 acrylamide:bisacrylamide) 50 mM Tris-glycine gel and subjected to electrophoresis at 400 V for three hours at 4°C. A phosphor screen was exposed to gels dried under vacuum at 80°C and scanned using a Typhoon 9400 Variable Mode Imager (GE Healthcare).

RNase H and Antisense Oligonucleotide snRNA Depletions. For RNase H digestions, RNase H (NEB) was added to a concentration of 5 U/µL with 0.25 µg of antisense oligonucleotide per microliter of extract and incubated at 30°C for 30 minutes prior to complex assembly. Nucleotides 1-14 of the U1 snRNA were targeted by U1 RNase H antisense oligo (5′-CATGCCAGGTAAGT-3′), nucleotides 28-42 of the U2 snRNA were targeted by U2 RNase H antisense oligo (5′-AGATACTACACTTGA-3′), nucleotides 106-130 of the U3 snRNA were targeted by U3 RNase H antisense oligo (5′- CGTCAGAAAACACCAGCTGCCCTAC -3′), and nucleotides 21-69 of the U6 snRNA were targeted by U6 RNase H antisense oligo (5′-TGCAGGGGCCATGCTAATCTTCTCTGTATCGTTTCAATTTGACCAAAGT-3′). Experimental conditions and oligonucleotide conditions were based on a similarly conducted analysis of WT *S. pombe* A complex (Huang *et al.* 2002).

U1 and U2 snRNA RT-PCRs. WT and ΔSF3B6 A3′ complex forming reactions with A3′-MS2 RNA were prepared and resolved by 1.2% agarose gel electrophoresis as previously described, except with low melting point agarose. H and A3′ complexes were visualized and excised. Complex containing gel pieces were melted at 65°C for 10 minutes, then diluted with 3 mL of 10 mM Tris pH 8.0 and 1 mM EDTA. RNAs were purified by phenol/chloroform extraction, followed by ethanol precipitation and subsequent resuspension in 15 µL DEPC treated water. The SuperScript III First-Strand Synthesis System for RT-PCR (Thermo) was used to generate cDNA for either the U1 or U2 snRNAs using 2 uL of each resuspended RNA and 2.5 µM of

either U1 RT antisense oligonucleotide (5′- TTGCCCCAAATGAGGGACGAAC -3′) or U2 RT antisense oligonucleotide (5′- TTCGGCGTCGCTTGCCAG -3′). cDNAs were PCR amplified with U1 sense oligonucleotide (5′- ACTTACCTGGCATGAGTTTCTGCAGC -3′) or U2 sense oligonucleotide (5′- TATTCTCTCTTTGCCTTTTGGCTTAGATCAAGTGT -3′) with the corresponding RT antisense oligonucleotides used previously. PCR products were resolved by 1% agarose gel electrophoresis at 200 mA for 30 min, then visualized by ethidium bromide staining.

A3′ Complex Pull-downs. 10-100 nM cold A3′-MS2 or long A3′-MS2 RNA doped with $200 - 500 \times 10^3$ cpm of radiolabelled RNA was incubated with 1-10 μM MS2-MBP on ice for 30 minutes to allow binding, then used in a 0.2-20 mL A3′ complex forming reaction prepared as previously described. Complex forming reactions were incubated at 30℃ for 30 minutes, then immediately loaded onto a 5 mL S-200 or S-500 column equilibrated with pull-down (PD) wash buffer (20 mM HEPES pH 7.9, 250 mM KCl, 5 mM EDTA, 1 mM DTT). Size exclusion fractions (0.5 mL) were collected and approximately quantified using a Geiger counter. Complex containing peak fractions were pooled and incubated with 1-3 mL of amylose resin previously equilibrated with PD wash buffer for 1 hour at 4℃ with gentle mixing. After allowing binding to occur, the resin was washed with 3 x 3 mL PD wash buffer, followed by elution with 3 x 900 μL PD elution buffer (20 mM HEPES pH 7.9, 250 mM KCl, 5 mM EDTA, 1 mM DTT, 10 mM maltose) via batch affinity chromatography. Elutions were concentrated over a 100 kDa molecular weight cut off ultrafiltration unit (GE Healthcare) and analyzed by 1.2% native agarose gel electrophoresis as previously described, or by 16% SDS-PAGE. Pull-down SDS-PAGE gels were visualized either using the Pierce™ Silver Stain for Mass Spectrometry kit (Thermo), or by colloidal Coomassie staining.

Cys-SF3B6 Expression and Purification. Cells overexpressing precursor SF3B6 (Figure 3-14 B) were harvested by centrifugation for 30 minutes at 3000 x g and 4℃. Cell pellets were resuspended in 30 mL of SF3B6 lysis buffer (20 mM Tris pH 8.0, 100 mM NaCl, 20 mM imidazole, 1 mM PMSF, 1 mg/mL lysozyme, 5 mM BME) per liter of cells and stirred for 1 hour at 4℃. Subsequently, the lysate was sonicated on ice for 6 cycles of 30 seconds with 0.5 second

pulses at 80% amplitude, with 1 minute of rest between each cycle. The lysate was cleared by centrifugation at 14 500 x g at 4℃, then the supernatant was incubated with 2 mL of Ni-NTA resin previously equilibrated with SF3B6 wash buffer (20 mM Tris pH 8.0, 100 mM NaCl, 20 mM imidazole, 5 mM BME) for 1 hour at 4℃ with gentle mixing. After allowing binding to occur, the resin-lysate mixture was transferred to a gravity flow column and a flowthrough fraction was collected. The resin was washed with 5 x 10 mL of SF3B6 wash buffer. For the purification of precursor SF3B6, the protein was eluted with 3 x 10 mL of SF3B6 elution buffer (20 mM Tris pH 8.0, 100 mM NaCl, 250 mM imidazole, 5 mM BME). Elutions were pooled and dialyzed twice against 2 L of chilled dialysis buffer (20 mM Tris pH 8.0, 100 mM NaCl, 5 mM BME, 10% glycerol) for 1.5 hours each at 4℃ using 6-8 kDa molecular weight cut off dialysis membrane (Fischer). For the purification of Cys-SF3B6, precursor SF3B6 was cleaved from the affinity column via overnight incubation with TEV protease at 4℃ with gentle shaking. The cleaved product was washed from the resin with 4 x 5 mL SF3B6 wash buffer, then dialyzed twice against 2 L of chilled dialysis buffer as before. Proteins were quantified by absorbance at 280 nm, then snap frozen as aliquots in liquid nitrogen and stored at -80℃.

Chapter 4

Conclusions and Future Directions

Chapter 4. Conclusions and Future Directions.

Towards the Reconstitution of an SF3B6-SF3B1-U2AF Complex.

Although each of the generated SF3B6-SF3B1 peptide and U2AF59-U2AF23 constructs were predicted to be capable of interaction based on sequence and interaction motifs, a stable interaction was not clearly observed upon *in vitro* reconstitution and gel filtration analysis. Such reconstitutions depend on the SF3B1-U2AF59 ULM-UHM interaction, for which a K_d of ~2.8 μM has been determined in the human system (Thickman *et al.* 2006). If these human and *S. pombe* SF3B1-U2AF59 interactions are of similar affinity, then it is possible that binding affinity has limited the successful reconstitution of an SF3B6-SF3B1-U2AF complex purified by gel filtration. For instance, the seemingly low proportion of complexed SF3B6-SF3B1-U2AF observed in Figure 2-3 and the split elution of U2AF59 observed in Figure 2-8 may be explained by protein concentrations that are insufficient for stable binding. Furthermore, the mobility shift analysis of SF3B6-SF3B1 175-291 and U2AF59 93-161-U2AF23 titrated against the A3′ intron RNA suggests the possible formation of a super-shifted ternary complex at protein concentrations ≥50 μM. The identity of the proteins in these gel shifts including the potential super-shift could be determined with a UV crosslinking experiment. Complexes of these proteins crosslinked to radiolabelled pre-mRNAs by UV light irradiation could be resolved by native-PAGE for extraction for subsequent SDS-PAGE analysis. Extracted complexes corresponding to each mobility shift could be resolved into bands corresponding to the size of either the SF3B1 peptide, SF3B6, U2AF59 or U2AF23. If necessary, pre-mRNAs could be RNase digested prior to SDS-PAGE such that the apparent size of each protein is not meaningfully perturbed.

An SF3B6-SF3B1-U2AF complex may be sufficiently stabilized for size exclusion chromatography with the introduction of a common pre-mRNA binding partner, such as the U9, U12 or A3′ RNAs. Accordingly, a possible solution to the purification of a reconstituted complex may be by the equimolar addition of a pre-mRNA to a set of SF3B6-SF3B1 peptide and U2AF59-U2AF35 complexes. A single monodisperse size exclusion peak of a pre-mRNA stabilized SF3B6-SF3B1-U2AF complex would have potential for structural characterization at atomic resolution by X-ray crystallography or in solution by small angle X-ray scattering (SAXS). A structural analysis of such a complex may provide insight into the molecular mechanism of 3′ SS recognition. Even if SF3B6-SF3B1-U2AF complexed on a pre-mRNA is

still not found to elute as a single monodisperse size exclusion peak but the previously described crosslinking experiment demonstrates the formation of a ternary complex, it still may be possible to generate crystals at sufficiently high protein concentrations.

Biochemical Characterization of the A3′ Complex in the Presence and Absence of SF3B6. Gel shift analysis of *S. pombe* splicing extracts demonstrated the formation of an A3′ complex in the absence of the key spliceosomal component SF3B6 (Figure 3-2). The U2 snRNA dependent formation of this ΔSF3B6 A3′ complex (Figure 3-5) was consistent with that of WT A3′ complexes (Huang *et al.* 2002), suggesting U2 snRNP recruitment to pre-mRNA substrates in the absence of SF3B6. While the antisense oligonucleotide and RNase H targeted degradation of the BPS binding region of the U2 snRNA appeared to eliminate the ΔSF3B6 A3′ complex (Figure 3-5), this experiment could be reinforced with a complementary Northern blot analysis. Aliquots of the extracts subjected to targeted snRNA degradation could be taken for phenol/chloroform extraction of RNA, then resolved by PAGE and analyzed by Northern blot with specific antisense oligonucleotides to assess the extent of degradation. This experiment would control for the completeness of each snRNA degradation. Furthermore, the ΔSF3B6 A3′ complex was found to be stable upon titration of excess competitor A3′ RNA to an extent comparable to that of the WT A3′ complex (Figures 3-3 and 3-4). Interestingly, the ΔSF3B6 complex was often found to appear as a tighter, more distinct gel shift that persisted through heparin equilibration unlike the WT A3′ complex (Figure 3-2).

An A3′ complex was not successfully purified using the MS2 aptamer-tagged pre-mRNA and MBP-affinity chromatography strategy previously described (Figure 3-7). Components of the A complex, such as the SF3B subunits, were not identified upon mass spectrometry analysis of any of the previously described purification attempts. Native-PAGE analysis suggested the peaks S-200 fractions of the crude size exclusion purification step to contain non-specific proteins bound to the A3′-MS2 RNA (Figure 3-10). Protein-RNA complexes collected at this stage were not found to correspond to control A3′ complex gel shifts, but rather appeared as a non-specific smear similar to that of the H complex. SDS-PAGE analysis of the MBP-affinity step indicated the enrichment of numerous proteins in the elution (Figures 3-12 and 3-13). However, as components of the spliceosome could not be consistently identified by mass spectrometry analysis of these elutions, many of these enriched proteins may correspond to non-specific

binding partners. Importantly, the MS2-MBP fusion protein was observed to elute upon the addition of maltose (Figure 3-12) and was identified in subsequent mass spectrometry analyses. Furthermore, the MBP-affinity step was RNA dependent, as elution could be achieved by RNase A degradation (Figure 3-13). Radiolabelled long A3′ RNA was also found to remain bound to the amylose resin until elution with maltose (Table 1). Altogether, these results suggested that the MS2-MBP and MS2 aptamer-tagged pre-mRNAs were functioning as intended in the affinity purification step. Therefore, the limitation of the approach seemed to be the capacity of the A3′ complex to withstand the purification procedure.

An A3′ complex purification using the previously described strategy (Figure 3-7) may be achievable upon further optimization of purification conditions. However, a potentially simpler approach to confirming the association of SF3B with the ΔSF3B6 A3′ complex may be to utilize chemically-derivatized protein constructs. An SF3B6 construct was successfully purified and cleaved with TEV protease to yield the full-length protein with an introduced N-terminal cysteine (Figure 3-14). The N-terminal cysteine residue of this SF3B6 construct can be site-specifically derivatized using established chemistry (Erlanson *et al.* 1996; Kent *et al.* 2003) with several thioester reagents. As previously discussed, SF3B6 may be derivatized with biotin (Figure 3-14 A) for the detection of SF3B and A3′ complex formation in the absence of SF3B6. Additionally, SF3B6 may be derivatized with the powerful photo-crosslinker benzophenone. Benzophenone derivatized SF3B6 could be added back into ΔSF3B6 extract for complex formation and subjected to UV irradiation to generate crosslinks to nearby factors. Subsequently, the A3′ complex could be resolved by native-PAGE and extracted. Crosslinks in the extracted A3′ complex could then be identified by SDS-PAGE gel shift or by mass spectrometry (O' Reilly and Rappsilber, 2018) for the detection and identification of nearby A3′ complex components. Finally, SF3B6 could also be derivatized with EDTA-3-MPA, which could be similarly added back to ΔSF3B6 complex forming reactions and used to map pre-mRNA contacts by hydroxy radical footprinting (Grewal *et al.* 2017). The derivatization of SF3B6 with reagents such as these would enable a more thorough biochemical characterization of protein and RNA contacts made by SF3B6 within the *S. pombe* A3′ complex.

Chapter 5

Supplementary Information

Chapter 5. Supplementary Information.

In addition to the human SF3B6-SF3B1 peptide structure (Schellenberg *et al.* 2006), structures of analogous *S. pombe* and *Candida albicans* (*C. albicans*) SF3B6-SF3B1 peptide complexes have also been solved by X-ray crystallography in our lab (Figure 5-1; Grewal *et al.* unpublished). By sequence alignment, *C. albicans* lacks a conserved aromatic residue in the pocket of RNP2 that has been implicated in branch point adenosine binding (Schellenberg *et al.* 2006; Schellenberg *et al.* 2011). More specifically, Y22 of human SF3B6 aligns with L10 of *C. albicans* SF3B6. Interestingly, a complex of *C. albicans* SF3B6 with residues 144-183 of SF3B1 was observed to form crystals with two distinct morphologies under differing conditions, one of which was optimized to yield a crystal structure (Grewal *et al.* unpublished). The RRM of *C. albicans* SF3B6 was found to superimpose well with that of other SF3B6 orthologues (Figure 5-1). The second condition (0.1M citric acid pH 5.0, 30% PEG 6000) yielded rod-shaped crystals (Figure 5-2). Efforts were made to optimize growth of these rod-shaped crystals to investigate if this alternative morphology might correspond to a distinct but functionally relevant conformation. Initial investigations have found these crystals to diffract X-rays. However, the crystals examined thus far have exhibited a high degree of mosaicity and the best of these diffractions yielded a structure at a resolution of ~4 Å. Consequently, the generation of these crystals must be further optimized to yield higher quality diffraction data for comparison to previously characterized structures. A structure of *C. albicans* SF3B6-SF3B1 peptide derived from these rod-shaped crystals solved to atomic resolution would provide the information necessary towards evaluating the structural significance of this non-conserved L10 residue.

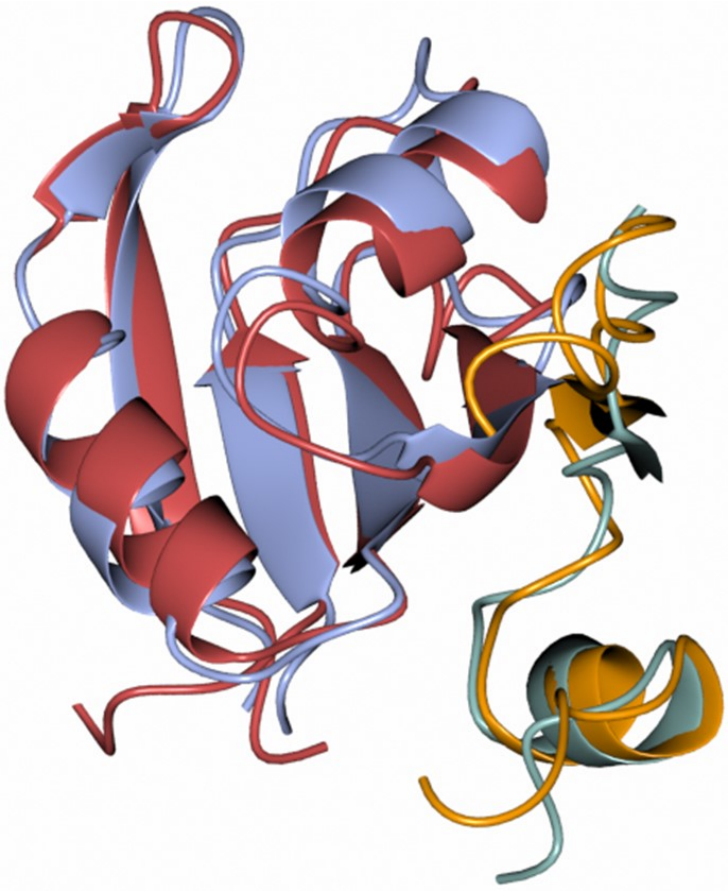

Figure 5-1. Superposition of *S. pombe* and *C. albicans* SF3B6-SF3B1 peptide complexes. *S. pombe* and *C. albicans* SF3B6-SF3B1 peptide crystal structures (Grewal *et al*., unpublished) superimposed. *S. pombe* SF3B6 is colored red, *S. pombe* SF3B1 peptide is colored orange, *C. albicans* SF3B6 is colored blue, and *C. albicans* SF3B1 peptide is colored teal.

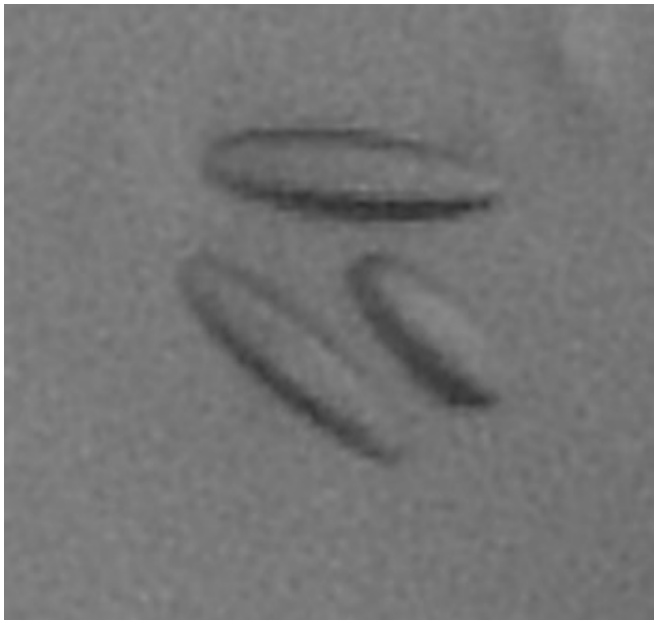

Figure 5-2. *C. albicans* SF3B6-SF3B1 peptide crystal morphology. Example photograph of the *C. albicans* SF3B6-SF3B1 144-183 complex crystal morphology when crystallized in 0.1M citric acid pH 5.0 and 30% PEG 6000.

Materials and Methods.

***C. albicans* SF3B6-SF3B1 Expression and Purification.** Overexpressed cells were harvested by centrifugation for 30 minutes at 3000 x g and 4°C. Cell pellets were resuspended in 30 mL of *C. albicans* lysis buffer (50 mM Tris pH 8.0, 1 M NaCl, 2M urea, 15% glycerol, 20 mM imidazole, 1 mM PMSF, 1 mg/mL lysozyme, 5 mM BME) per liter of cells and stirred for 1 hour at 4°C. Subsequently, the lysate was sonicated on ice for 6 cycles of 30 seconds with 0.5 second pulses at 80% amplitude, with 1 minute of rest between each cycle. The lysate was cleared by centrifugation at 14 500 x g at 4°C, then the supernatant was incubated with 2 mL of Ni-NTA resin previously equilibrated with *C. albicans* wash buffer (50 mM Tris pH 8.0, 1 M NaCl, 2M urea, 15% glycerol, 20 mM imidazole, 5 mM BME) for 1 hour at 4°C with gentle mixing. After allowing binding to occur, the resin-lysate mixture was transferred to a gravity flow column and a flowthrough fraction was collected. The resin was washed with 5 x 10 mL of wash buffer, then eluted with 3 x 10 mL of *C. albicans* elution buffer (50 mM Tris pH 8.0, 1 M NaCl, 2M urea, 15% glycerol, 250 mM imidazole, 5 mM BME). Elutions were concentrated over a 3 kDa molecular weight cut off ultrafiltration device (Amicon). Concentrates were purified by gel filtration on a Superdex-75 26/60 column (GE Healthcare) equilibrated with *C. albicans* wash buffer lacking imidazole. Peak fractions were analyzed by SDS-PAGE, then concentrated to 25 mg/mL as quantified via absorbance at 280 nm. Protein aliquots were snap frozen in liquid nitrogen and stored at -80°C.

***C. albicans* SF3B6-SF3B1 Crystallization.** C. *albicans* SF3B6 concentrated to 25 mg/mL was crystallized via hanging drop vapour diffusion by mixing 1 µL protein with 0.75 µL reservoir solution (0.1M citric acid pH 5.0, 30% PEG 6000) at room temperature. Crystals were snap frozen in reservoir solution supplemented with 20% glycerol as a cryoprotectant.

References

References

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

Barbosa-Morais, N., L., Irimia, M., Pan, Q., Xiong, H., Y., Gueroussov, S., Lee., L., J., Slobodeniuc, V., Kutter, C., Watt, S., Colak, R., Kim, T., Misquitta-Ali, C., M., Wilson, M., D., Kim, P., M., Odom, D., T., Frey, B., J., and Blencowe, B., J. (2012) The evolutionary landscape of alternative splicing in vertebrate species. *Science*. **338(6114)**: 1587-1593.

Berget, S., M., Moore, C., and Sharp, P., A. (1977) Spliced segments at the 5′ terminus of adenovirus 2 late mRNA. *Proc. Natl. Acad. Sci. U. S. A.* **74(8)**: 3171-3175.

Burge, C., B., Tuschl, T., and Sharp, P., A. (1999) Splicing of precursors to mRNAs by the spliceosomes. *In The RNA World* **(Gesteland, R. et al., eds)**: 525–560. Cold Springs Harbour Laboratory Press.

Chen, M., and Manley, J., L. (2009) Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat. Rev. Mol. Cell Biol.* **10(11)**: 741-754.

Chen, S., Anderson, K., and Moore, M., J. (2000) Evidence for a linear search in bimolecular 3′ splice site AG selection. *Proc. Natl. Acad. Sci. U. S. A.* **97(2)**: 593-598.

Chow, L., T., Gelinas, R., E., Broker, T., R., and Roberts, R., J. (1977) *Cell.* **12(1)**: 1-8.

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

Erlanson, D., A., Chytil, M., and Verdine, G., L. (1996) The leucine zipper domain controls the orientation of AP-1 in the NFAT.AP-1.DNA complex. *Chem. Biol.* **3(12)**: 981-991.

Fair, B., J., and Pleiss, J., A. (2017) The power of fission: yeast as a tool for understanding complex splicing. *Curr. Genet.* **63(3)**: 375-380.

Fleckner, J., Zhang, M., Valcárcel, J., and Green, M., R. (1997) U2AF65 recruits a novel human DEAD box protein required for the U2 snRNP-branchpoint interaction. *Genes Dev.* **11(14)**: 1864-1872.

Gao, K., Masuda, A., Matsuura, T., and Ohno, K. (2008) Human branch point consensus sequence is yUnAy. *Nucleic Acids Res.* **36(7)**: 2257-2267.

Gozani, O., Potashkin, J., and Reed, R. (1998) A potential role of U2AF-SAP 155 interactions in recruiting U2 snRNP to the branch site. *Mol. Cell Biol.* **18(8)**: 4752-4760.

Grewal, C., S., Kent, O., A., and MacMillan, A., M. (2017) Radical probing of spliceosome assembly. *Methods.* **125**: 16-24.

Hammond, S., M., and Wood, M., J. (2011) Genetic therapies for RNA mis-splicing diseases. *Trends Genet.* **27(5)**: 196-205.

House, A., E., and Lynch, K., W. (2008) Regulation of alternative splicing: more than just the ABCs. *J. Biol. Chem.* **283(3)**: 1217-1221.

Huang, T., Vilardell, J., and Query, C., C. (2002) Pre-spliceosome formation in *S. pombe* requires a stable complex of SF1-U2AF(59)-U2AF(23). *EMBO.* **21(20)**: 5516-5526.

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

Kapust, R., B., Tözsér, J., Copeland, T., D., and Waugh, D., S. (2002) The P1′ specificity of tobacco etch virus protease. *Biochem. Biophys. Res. Commun.* **294(5)**: 949-955.

Kenan, D., J., Query, C., C., and Keene, J., D. (1991) RNA recognition: towards identifying determinants of specificity. *Trends Biochem Sci.* **16(6)**: 214-220.

Kent, O., A., Reayi, A., Foong, L., Chilibeck, K., A., and MacMillan, A., M. (2003) Structuring of the 3′ splice site by U2AF65. *J. Biol. Chem.* **278(50)**: 50572-50577.

Kielkopf, C., L., Rodionova, N., A., Green, M., R., and Burley, S., K. (2001) A novel peptide recognition mode revealed by the X-ray structure of a core U2AF35/U2AF65 heterodimer. *Cell.* **106(5)**: 595-605.

Kielkopf, C., L., Lücke, S., and Green, M., R. (2004) U2AF homology motifs: protein recognition in the RRM world. *Genes Dev.* **18(13)**: 1513-1526.

Konarska, M., M., Vilardell, J., and Query, C., C. (2006) Repositioning of the reaction intermediate within the catalytic center of the spliceosome. *Mol. Cell.* **21(4)**: 543-553.

Lamond, A., I., Konarska, M., M., and Sharp, P., A. (1987) A mutational analysis of spliceosome assembly: evidence for splice site collaboration during spliceosome formation. *Genes Dev.* **1(6)**: 532-543.

Loerch, S., and Kielkopf, C., L. (2016) Unmasking the U2AF homology motif family: a bona fide protein-protein interaction motif in disguise. *RNA.* **22(12)**: 1795-1807.

Lustig, A., J., Lin, R., J., and Abelson, J. (1986) The yeast RNA gene products are essential for mRNA splicing *in vitro*. *Cell*. **47(6)**: 953-963.

MacMillan, A., M., Query, C., C., Allerson, C., R., Chen, S., Verdine, G., L., and Sharp, P., A. (1994) Dynamic association of proteins with the pre-mRNA branch region. *Genes Dev.* **8(24)**: 3008-3020.

Madi, S. (2020) Characterizing the role of p14 in pre-mRNA splicing. *ERA*. DOI: https://doi.org/10.7939/r3-qfgj-tz78.

Matlin, A., J., Clark, F., and Smith, C., W., J. (2005) Understanding alternative splicing: towards a cellular code. *Nat. Rev. Mol. Cell Biol.* **6(5)**: 386-398.

Modrek, B., and Lee, C. (2002) A genomic view of alternative splicing. *Nat. Genet.* **30(1)**: 13-19.

O'Reilly, F., J., and Rappsilber, J. (2018) Cross-linking mass spectrometry: methods and applications in structural, molecular and systems biology. *Nat. Struct. Mol. Biol.* **25(11)**: 1000-1008.

Plaschka, C., Lin, P., Charenton, C., and Nagai, K. (2018) Prespliceosome structure provides insights into spliceosome assembly and regulation. *Nature.* **559(7714)**: 419-422.

Query, C., C., Strobel, S., A., and Sharp, P., A. (1996) Three recognition events at the branch-site adenine. *EMBO J.* **15(6)**: 1392-1402.

Ritchie, D., B., Schellenberg, M., J., and MacMillan, A., M. (2009) Spliceosome structure: piece by piece. *Biochim. Biophys. Acta.* **1789(9-10)**: 624-633.

Romfo, C., M., and Wise, J., A. (1997) Both the polypyrimidine tract and the 3′ splice site function prior to the first step of splicing in fission yeast. *Nucleic Acids Res.* **25(22)**: 4658-4665.

Rudner, D., Z., Kanaar, R., Breger, K., S., and Rio, D., C. (1998) Interaction between subunits of heterodimeric splicing factor U2AF is essential in vivo. *Mol. Cell Biol.* **18(4)**: 1765-1773.

Ruskin, B., Zamore, P., D., and Green, M., R. (1988) A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly. *Cell.* **52(2)**: 207-219.

Rutz, B., and Séraphin, B. (1999) The transient interaction of BBP/ScSF1 and Mud2 with the splicing machinery affects the kinetics of spliceosome assembly. *RNA.* **5(6)**: 819-831.

Schellenberg, M., J., Dul, E., L., and MacMillan, A., M. (2011) Structural model of the p14/SF3b155·branch duplex complex. *RNA.* **17(1)**: 155-165.

Schellenberg, M., J., Edwards, R., A., Ritchie, D., B., Kent, O., A., Golas, M., M., Stark, H., Lührmann, R., Glover, J., N., M., and MacMillan, A., M. (2006) Crystal structure of a core spliceosomal protein interface. *Proc. Natl. Acad. Sci. U. S. A.* **103(5)**: 1266-1271.

Selenko, P., Gregorovic, G., Sprangers, R., Stier, G., Rhani, Z., Krämer, A., and Sattler, M. (2003) Structural basis for the molecular recognition between human splicing factors U2AF65 and SF1/mBBP. *Mol. Cell.* **11(4)**: 965-976.

Sun, C. (2020) The SF3b complex: splicing and beyond. *Cell Mol. Life Sci.* **77(18)**: 3583-3595.

Thickman, K., R., Swenson, M., Kabogo, J., M., Gryczynski, Z., and Kielkopf, C., L. (2006) Multiple U2AF65 binding sites within SF3b155: thermodynamic and spectroscopic characterization of protein-protein interactions among pre-mRNA splicing factors. *J. Mol. Biol.* **356(3)**: 664-683.

Valcarcel, J., Gaur, R., K., Singh, R., and Green, M., R. (1996) Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA. *Science.* **273**: 1706-1709.

Vijayraghavan, U., Company, M., and Abelson, J. (1989) Isolation and characterization of pre-mRNA splicing mutants of *Saccharomyces cerevisiae*. *Genes Dev.* **3(8)**: 1206-1216.

Wahl, M., C., Will, C., L., and Lührmann, R. (2009) The spliceosome: design principles of a dynamic RNP machine. *Cell.* **136(4)**: 701-718.

Wentz-Hunter, K., and Potashkin, J. (1996) The small subunit of the splicing factor U2AF is conserved in fission yeast. *Nucleic Acids Res.* **24(10)**: 1849-1854.

Will, C., L., Schneider, C., MacMillan, A., M., Katopodis, N., F., Neubauer, G., Wilm, M., Lührmann, R., and Query, C., C. (2001) A novel U2 and U11/U12 snRNP protein that associates with the pre-mRNA branch site. *EMBO J.* **20(16)**: 4536-4546.

Will, C., L., and Lührmann, R. (2011) Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.* **3(7)**: Article a003707.

Wu, S., Romfo, C., M., Nilsen, T., W., and Green, M., R. (1999) Functional recognition of the 3′ splice site AG by the splicing factor U2AF35. *Nature.* **402(6763)**: 832-835.

Yoshida, H., Park, S., Oda, T., Akiyoshi, T., Sato, M., Shirouzo, M., Tsuda, K., Kuwasako, K., Unzai, S., Muto., Y., Urano, T., and Obayashi, E. (2015) A novel 3′ splice site recognition by the two zinc fingers in the U2AF small subunit. *Genes Dev.* **29(15)**: 1649-1660.

Zamore, P., D., and Green, M., R. (1989) Identificatin, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. *Proc. Natl. Acad. Sci. U. S. A.* **86(23)**: 9243-9247.

Zamore, P., D., Patton, J., G., and Green M., R. (1992) Cloning and domain structure of the mammalian splicing factor U2AF. *Nature.* **355(6361)**: 609-614.

Zhang, Z., Will., C., L., Bertram, K., Dybkov, O., Hartmuth, K., Agafonov, D., E., Hofele, R., Urlaub, H., Kastner, B., Lührmann, R., and Stark, H. (2020) Molecular architecture of the human 17S U2 snRNP. *Nature.* **583(7815)**: 310-313.

Zhang, X., Yan, C., Zhan, X., Li, L., Lei, J., and Shi, Y. (2018) Structure of the human activated spliceosome in three conformational states. *Cell Res.* **28(3)**: 307-322.

Zhou, Z., Licklider, L., J., Gygi, S., P., and Reed, R. (2002) Comprehensive proteomic analysis of the human spliceosome. *Nature.* **419(6903)**: 182-185.

Zorio, D., A., and Blumenthal, T. (1999) U2AF35 is encoded by an essential gene clustered in an operon with RRM/cyclophilin in Caenorhabditis elegans. *RNA.* **5(4)**: 487-494.