

University of Alberta

**Structural and Biochemical Analysis of the Essential Spliceosomal Protein
Prp8**

by

Dustin Blaine Ritchie

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Biochemistry

©Dustin Blaine Ritchie
Spring 2010
Edmonton, Alberta

Permission is hereby granted to the University of Alberta Libraries to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only. Where the thesis is converted to, or otherwise made available in digital form, the University of Alberta will advise potential users of the thesis of these terms.

The author reserves all other publication and other rights in association with the copyright in the thesis and, except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatsoever without the author's prior written permission.

Examining Committee

Dr. Andrew MacMillan, Biochemistry

Dr. Mark Glover, Biochemistry

Dr. Richard Fahlman, Biochemistry

Dr. George Owtrim, Biological Sciences

Dr. Benoit Chabot, Microbiology & Infectious Diseases, University of Sherbrooke

Dedicated to my father Blaine Ritchie.

While here he instilled in me,

Passion is a pursuit of answers

Not of possessions.

Abstract:

More than 90% of human genes undergo a processing step called splicing, whereby non-coding introns are removed from initial transcripts and coding exons are ligated together to yield mature messenger RNA. Roughly 50% of human genetic diseases correspond to aberrant splicing. Splicing is catalyzed by an RNA/protein machine called the spliceosome. RNA components of the spliceosome are at least partly responsible for splicing catalysis. In addition, *in vitro* analyses implicate an essential and very highly conserved protein, Prp8, in orchestrating key steps in spliceosome assembly and possibly catalysis. Interestingly, mutant alleles of Prp8 are the cause of *retinitis pigmentosa*, an inherited form of retinal degeneration.

A key goal is elucidation of the precise role of Prp8 in the spliceosome by high resolution structural analysis. The large size of Prp8 and its insolubility hinder progress in this regard. Instead, structural understanding of Prp8 can be gained by investigating domains in isolation; however there is only limited information as to what domain boundaries are and few hints about the functional relevance of putative domains. Here we have further defined the previously proposed domain IV in Prp8, and identified the domain IV core. Structural determination of the domain IV core reveals an RNase H fold, which could not be predicted based on primary sequence alone. RNase H recognizes A-form nucleic acid duplexes, which strongly suggests the domain IV core interacts with double-stranded RNA in the context of the spliceosome. Characterizing the binding preferences of the domain IV revealed the highest affinity is for a 4-helix junction structure adopted by the very RNAs at the spliceosome active site. Our characterization of the protein/RNA binding interface by complementary footprinting techniques currently provides the best model of how RNA interacts with an essential protein component at the heart of the spliceosome.

Acknowledgements:

My life has changed in seven years of graduate school. I started graduate school with virtually no knowledge about what it means to conduct health research. Thanks to the welcoming environment that was in place in the MacMillan laboratory and in the Department of Biochemistry in general I was able to overcome the challenges associated with starting a new life. Particularly, Oliver Kent was my first mentor and showed me the proper way to think about a research project. He gave me advice that will stick with me forever. I've leaned heavily on his legacy in the preparation of this thesis.

My future wife Renee has been with me every step of the way since we first met. The pressures of competitive research have often forced me to expose my fears and insecurities to Renee like I could never do with anyone else. When I doubt myself, she makes me believe in myself because she believes in me. When it seems like the obstacles in place are too overwhelming to overcome, only she can remind me that I'm destined for even greater accomplishments than these. Of course it is her love and devotion that makes me twice as strong, such that now I see real possibilities that before I could only dream of. The successes I've enjoyed symbolize just as much sacrifice on her part as on mine, and I know that together we are going to achieve all of our goals. I am very excited about our future endeavours together.

Dr. Andrew MacMillan was a natural fit for my PhD supervisor. He lives and breathes Biochemistry, but instantly seemed to take interest in me as a person as well. Whether it was skiing, golf, or being a fan of Bob Dylan, Andrew obviously had a keen sense of the things that defined me as an individual. He has been nothing but supportive of me and my research efforts, such that over the course of several years I've learned to approach projects in a creative and independent way. Whether I've been working on a project that was flourishing or failing, he had the time to make sure I was going to get the most reward for my efforts. Today I am proud to call Andrew my friend, and I know he will be a valuable acquaintance for the rest of my career.

My mother Cathy and brother Josh are permanent fixtures in my life. When I first moved to Edmonton the support I received from my family was essential for transition into the life of a graduate student. It was hard to move still farther away from them, but our visits during holidays

kept us close, and gave them a chance to see that what I was doing was important for achieving my goals. Now my immediate family has grown to include Graham Clarke, and his son Aquilla and daughter Anastasia. I welcome their continued support with open arms. I have also found an invaluable support system in Renee's parents Deb and Jean-Guy Morissette. We've relied on their help and support several times already, and it is clear that they will always be there for us. As well, I've found a new golf partner for life in Jean-Guy, my future father-in-law. My grandparents Al and Beth Ritchie also continue to be instrumental in my life, as they always have been, especially during my undergraduate degree in Kamloops, BC.

I want to acknowledge the professors who are participating in my defence: Dr. Mark Glover, Dr. Richard Fahlman, and Dr. George Owtrim. Their advice at my candidacy exam has stayed with me to this day, and I know their input at my PhD defence will have even greater impact. I especially want to thank Dr. Benoit Chabot for agreeing to come from University of Sherbrooke to be my external examiner. I know he will provide a thorough review of my thesis, and hopefully will agree to serve as a powerful reference for the next stages of my career.

It is imperative that I thank both current and past members of the MacMillan laboratory. Matt Schellenberg has been a close collaborator for the majority of my graduate career, and in large part my success is due to his crystallography prowess. As well, Emily Gesner and Tao Wu are close friends and make day-to-day life in the laboratory more fun and exciting. Past members Dr. Steven Chaulk and Kaari Chilibeck were and continue to be close friends.

Finally, I've made a large number of friends in the Department of Biochemistry. Dave Arthur (my roommate of two years), Scott Williams, Ross Edwards, Dave Hau and Ryan Hoffman... All people who I relied on for support in my early Edmonton days. On a daily basis I depend on the following people for support outside of the lab and away from science in general: John Paul Glaves and Mel Shaw, Ian Robertson, Olivier Julien, Ross Edwards, and Susan Smith. When I'm old and grey I'm going to call up John Paul for a game of early morning golf... Today it still seems like we have an infinite number of chances to shoot that seemingly elusive round of under par golf.

Table of Contents:

Dedication	<i>i</i>
Abstract	<i>ii</i>
Acknowledgement	<i>iii</i>
Table of Contents	<i>v</i>
List of Figures	<i>viii</i>
List of Tables	<i>x</i>
List of Abbreviations	<i>xi</i>
Introduction. Pre-mRNA splicing and spliceosome assembly.	1
I-1. Introduction to pre-mRNA splicing.	2
I-1.1. Pre-mRNA splicing.	2
I-1.2. Assembly of the spliceosome.	3
I-2. A structural approach to studying the spliceosome.	5
I-2.1. Polypyrimidine tract recognition by U2AF in E complex.	7
I-2.2. Protein recognition mediated by U2AF homology motifs.	10
I-2.3. SF1 and branch point sequence recognition.	13
I-2.4. U1 snRNP: a high resolution view.	14
I-2.5. Structures related to the A complex.	18
I-2.6. Structure of RNA at the spliceosome active site.	20
I-3. Structural studies of the catalytic spliceosome.	26
I-4. References.	29
Chapter 1. The essential spliceosomal protein Prp8.	37
1-1. Introduction	38
1-1.1. Discovery of Prp8.	38
1-1.2. Domain structure of Prp8.	39
1-1.3. A conserved RRM.	40
1-1.4. Domain 3 (3'SS fidelity domain)	41
1-1.5. Domain IV (5'SS fidelity domain)	42
1-1.6. C-terminal MPN domain	44
1-2. Prp8 and its interactions.	46
1-2.1. Prp8 and kinases.	46
1-2.2. Prp8, transcription and chromatin remodelling factors.	48
1-2.3. Prp8 in a U5 snRNP assembly intermediate.	50
1-3. A protein factor in the spliceosome active site.	51
1-3.1. Cross-linking Prp8 to pre-mRNA.	52
1-3.2. Cross-linking Prp8 to snRNA.	54
1-3.3. Genetic interactions: spliceosomal RNA and <i>prp8</i> alleles.	55
1-4. References.	60
Chapter 2. Structural investigation of Prp8 domain IV.	67
2-1. Introduction.	68
2-1.1. Analysis of the C-terminal region of Prp8.	68
2-1.2. C-terminal region of Prp8 at the spliceosome active site.	69

2-2. Results and Discussion.	71
2-2.1. Structural overview of hPrp8 domain IV core.	71
2-2.2. RNase H homology of the domain IV core.	77
2-2.3. An RNase H domain in the spliceosome active site.	80
2-3. Methods.	82
2-3.1. Identification and expression of an hPrp8 domain IV core.	82
2-3.2. Crystallization.	83
2-3.3. Data collection and processing.	83
2-3.4. Model building and refinement.	84
2-4. References.	85
 Chapter 3. Protein/RNA interactions of Prp8 domain IV.	 89
3-1. Introduction.	90
3-1.1. Evidence for RNA binding by the domain IV core.	90
3-2. Results and Discussion.	91
3-2.1. Identifying RNA binding partners of hPrp8 domain IV.	91
3-2.2. Mapping Prp8/RNA interactions.	93
3-2.3. Implications of the core structure for interaction with the spliceosome active site.	99
3-3. Methods.	101
3-3.1. RNA binding by hPrp8.	101
3-3.2. Limited proteolysis.	102
3-3.3. RNase protection.	102
3-3.4. Mapping four-helix junction crosslink.	102
3-4. References.	104
 Appendix I. Analysis of a stable hPrp8 fragment.	 106
I-1. Dissecting the domain IV core: 1831-1990.	107
I-2. Methods.	108
 Appendix II. Characterization of a U2AF independent commitment complex (E') in the mammalian spliceosome assembly pathway.	 110
II-1. Introduction.	111
II-2. Results.	113
II-2.1. Identification of a novel prespliceosomal complex.	113
II-2.2. Relationship of E' to the prespliceosome.	117
II-2.3. Detection of BBP in the E' complex.	120
II-2.4. Directed hydroxyl radical probing of pre-mRNA structure in the E' complex.	121
II-3. Discussion.	125
II-4. Methods.	131
II-4.1. Synthesis of pre-mRNA substrates.	131
II-4.2. E and E' complex formation.	131
II-4.3. Gel filtration column purification of E' complex.	132

II-4.4. Splicing assays.	133
II-4.5. UV cross-linking and immunoprecipitation.	133
II-4.6. Hydroxyl radical probing of E and E' structure.	134
II-5. References.	136
Appendix III. Crystal structure of a core spliceosomal protein interface.	140
III-1. Introduction.	141
III-2. Results.	142
III-2.1. Structure of the p14-SF3b155 complex.	142
III-2.2. RNA protein interactions in a minimal complex.	145
III-2.3. Comparison with canonical and pseudoRRMs.	148
III-2.4. p14 within SF3b and the U11/U12 di-snRNP.	150
III-3. Discussion.	153
III-4. Methods.	155
III-4.1. Determination of SF3b155 p14 interaction domain.	155
III-4.2. Protein expression and purification.	156
III-4.3. Crystallization.	156
III-4.4. Data collection and processing.	157
III-4.5. Model building and refinement.	157
III-4.6. Crosslink mapping.	158
III-4.7. Comparison of X-ray and cryo-EM structures.	159
III-5. References.	161

List of Figures:

Introduction.	
I-1. Splicing of pre-mRNA by the spliceosome.	6
I-2. Early recognition of the polypyrimidine tract by U2AF.	12
I-3. Early branch sequence recognition by SF1.	14
I-4. Crystal structure of U1 snRNP.	16
I-5. Mechanism for spliceosome and group II splicing.	22
I-6. Proposed structures of U2/U6 snRNA in the spliceosome active site.	25
I-7. Comparison of group II intron and spliceosomal active site RNA structures.	25
I-8. Stereo views of the surface representation of the refined three-dimensional structure of C complex.	27
Chapter 1.	
1-1. Prp8 contains a conserved RRM.	41
1-2. Prp8 has been identified in several distinct complexes.	50
1-3. Genetic interaction between the <i>U4-cs1</i> mutation and conditional mutations in U6 snRNA involved in catalytic activation.	58
1-4. Scheme for progression of the pre-mRNA splicing progress, highlighting the Prp16-dependent transition between the two catalytic steps.	59
Chapter 2.	
2-1. U6/U2 base-paired complex that catalyzes a reaction similar to the first step of splicing.	70
2-2. Structure of hPrp8 domain IV.	73
2-3. Secondary-structure diagram and sequence conservation of Prp8 domain IV core.	75
2-4. Mapping of yeast mutant alleles and a 5'SS cross-link in the context of the Prp8 domain IV core.	76
2-5. Comparison of the Prp8 domain IV core with RNase H folds.	78
2-6. Analysis of the putative conserved binding site in Prp8 domain IV core.	80
Chapter 3.	
3-1. RNA binding by hPrp8 domain IV core.	93
3-2. Footprinting analysis of a Prp8/RNA complex.	94
3-3. RNase footprinting of a Prp8/RNA complex.	95
3-4. Mapping four-helix junction crosslink.	97
3-5. Representations of the proposed Prp8 domain IV core RNA binding surface.	98
Appendix I.	
I-1. Structure of hPrp8 domain IV 1831-1990.	109
Appendix II.	
II-1. Identification of a novel prespliceosome complex (E').	115

<i>II-2.</i> Formation of the E' complex requires SF1, and the complex contains U1 snRNP.	117
<i>II-3.</i> Commitment of pre-mRNA to splicing in the absence of U2AF.	119
<i>II-4.</i> E' complex contains SF1 at the branch point.	121
<i>II-5.</i> Directed hydroxyl radical cleavage of pre-mRNA within the ATP-independent prespliceosome complexes.	124
<i>II-6.</i> Assembly of the E complex proceeds through early recognition of the 5'SS/branch region followed by recruitment of U2AF to PPT.	130
<i>Appendix III.</i>	
<i>III-1.</i> Structure of p14·SF3b155 peptide complex.	144
<i>III-2.</i> Details of p14·SF3b155 interface.	146
<i>III-3.</i> Mapping of the p14·branch nucleotide interaction.	149
<i>III-4.</i> Comparison between p14·SF3b155 peptide surface and structures of U1A and U2AF65 RRM3.	151
<i>III-5.</i> p14·SF3b155 peptide structure within SF3b and U11/U12 snRNP.	153

List of Tables:

Chapter 1.

Table 1-1. Mutant yeast alleles mapping to Prp8 domain IV core and corresponding phenotypes.	44
---	----

Chapter 2.

Table 2-1. Mutant yeast alleles mapping to Prp8 domain IV core and corresponding phenotypes.	72
Table 2-2. Crystallographic data collection and refinement statistics.	74

List of Abbreviations:

Δ	delta, depleted
A	adenosine
a.a.	amino acid
AG	3' splice site dinucleotide
ATP	adenosine triphosphate
BBP	branch binding protein
BPS	branch point sequence
C	cytosine
<i>cat</i>	catalytic activation
CBP	cap-binding protein
CC	commitment complex
CP	creatine phosphate
cs	cold-sensitive
CTD	carboxy-terminal domain
CN-Br	cyanogen bromide
DTT	dithiolthreitol
E complex	early complex
EDTA	ethylene diamine tetraacetic acid
eIF3	eukaryotic initiation factor 3
Fe-BABE	bromoacetamidobenzyl-EDTA
G	guanidine
GST	glutathione S-transferase
hnRNP	heterogeneous nuclear ribonucleoprotein
ISL	internal stem loop
K_d	dissociation constant
KH	hnRNP K homology
MBP	maltose-binding protein
NE	nuclear extract
nt	nucleotide
NTC	nineteen complex
NCoA62	nuclear coactivator-62 kDa
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PDB	protein databank
PPT	poly-pyrimidine tract
PRP	pre-mRNA processing
PSF	PPT-binding protein-associated splicing factor
PTB	poly-pyrimidine tract binding protein
pre-mRNA	precursor messenger-RNA
r.m.s.d.s	root-mean-square deviations
RNAP II	RNA polymerase II
RP	<i>retinitis pigmentosa</i>
RRM	RNA recognition motif

RS	arginine-serine rich domain
RT	reverse transcriptase
SDS	sodium dodecyl sulfate
Se-Met	selenomethionine
SLII	stem loop II
snRNA	small nuclear RNA
snRNP	small nuclear ribonucleoprotein particle
SS	splice site
TEV	tobacco etch virus
TOPII α	topoisomerase II α
tRNA	transfer RNA
ts	temperature sensitive
U	uridine
U2AF	U2 auxiliary factor
UHM	U2AF homology motif
VDR	vitamin D receptor
WT	wild type

Introduction⁽¹⁾
Pre-mRNA splicing and spliceosome assembly

¹ Adapted from Ritchie et al., (2009). *Biochim. Biophys. Acta.* **1789**, 624-633.

Introduction: Pre-mRNA splicing and spliceosome assembly

I-1. Introduction

I-1.1. Pre-mRNA splicing

Over 90% of eukaryotic genes are initially expressed as precursor-messenger RNAs (pre-mRNAs) which contain coding exon sequences interrupted, or split, by non-coding intron sequences. The “split gene” structure was first discovered by the Sharp and Roberts laboratories during the mapping of adenoviral gene structure (Berget et al., 1977; Chow et al., 1977). Single-stranded viral DNA was digested with endonuclease and subsequently hybridized with viral mRNA. In some cases, the mRNA molecules contained 5' and 3' tails that were not hybridized to the DNA. Longer pieces of single-stranded DNA were tested, and regions of complementarity formed between viral DNA and the mRNA tails, while the forked structures of intervening DNA sequences were looped out as observed by electron microscopy. Sharp was able to reason that the looped-out regions of non-complementary DNA represent sequences that are removed during maturation of the pre-mRNA, a process termed splicing (Berget et al., 1977). The split gene structure was subsequently found to be common to most eukaryotic genes.

Intron removal and exon ligation occurs via two sequential phosphotransesterification reactions (Figure I-1). Introns are defined by conserved sequence elements at the sites of splicing chemistry. Important consensus sequences of introns include the 5' splice site (5'SS), 3'SS, branch site, and a poly-pyrimidine tract (PPT) just upstream of the 3'SS. The first

transesterification involves attack at the 5'SS by the 2' hydroxyl of a conserved adenosine at the intron branch site, resulting in a free 5' exon and a branched lariat intermediate containing a 2'-5' phosphodiester linkage. Subsequently, the free 5' exon attacks the 3'SS via a second transesterification reaction yielding ligated exons and a lariat-structure intron (Figure I-1A,B). The spliceosome, a 60S biochemical machine composed of upwards of three hundred protein factors and five small nuclear RNAs (snRNAs) (Jurica and Moore, 2003; Rappsilber et al., 2002; Zhou et al., 2002), catalyzes these two steps of splicing to produce a mature mRNA (Burge et al., 1999; Krämer, 1996; Staley and Guthrie, 1998). The spliceosome forms anew upon each pre-mRNA via stepwise assembly of the small nuclear ribonucleoprotein particles (snRNPs): U1, U2, U4, U5, and U6, named for the specific snRNA associated with each particle (Burge et al., 1999; Staley and Guthrie, 1998). Spliceosome assembly proceeds through the E, A, B, and C complexes and is guided by conserved sequences within the intron and sequences within the pre-mRNA by a mechanism which is highly conserved from humans to yeast (Figure I-1C).

I-1.2. Assembly of the spliceosome

The initial steps of spliceosome assembly involve the ATP independent formation of the Early (E' (see Appendix II) and E; mammalian) or Commitment Complex (CC; yeast) on the pre-mRNA substrate. In mammals, this complex includes U1 snRNP tightly associated with the 5' end of the intron mediated by a base-pairing interaction between U1 snRNA and the 5'SS (Jamison et al., 1992; Legrain et al.,

1988; Seraphin and Rosbash, 1989). The interaction between the 5'SS and U1 snRNA is aided by the cap-binding protein CBP80 as well as U1 snRNP associated proteins in yeast (Puig et al., 1999; Zhang & Rosbash, 1999). Additional non-snRNP protein factors such as U2AF and SF1 are also associated with the pre-mRNA in E complex and serve to define the 3'SS (Gaur et al., 1995; Krämer and Utans, 1991). U1 snRNP may be involved in recruiting U2AF to the PPT through indirect interactions with the SR protein SC35. Far Western analysis has shown that SC35 bridges the U1 snRNP specific protein U1-70K and U2AF35 (Wu and Maniatis, 1993) thus serving to bridge the 5' and 3' ends of the intron. Subsequently, the ATP dependent, stable association of U2 snRNP with the pre-mRNA results in the formation of A complex. Here, a duplex structure between U2 snRNA and the pre-mRNA branch sequence extrudes a bulged adenosine, which is believed to be the mechanism by which it is selected as the nucleophile for the first step of splicing (Query et al., 1994). Association of the U5·U4/U6 tri-snRNP with the pre-mRNA defines the B complex (reviewed in: Burge et al., 1999; Staley and Guthrie, 1998). Within the tri-snRNP and B complex the U4 and U6 snRNAs are extensively base-paired to each other in an arrangement that is mutually exclusive with the RNA structure in the activated spliceosome (Madhani and Guthrie, 1992).

The transition from the pre-spliceosomal B complex to the mature C complex is marked by extensive rearrangements of RNA/RNA and RNA/protein structures which drive spliceosome assembly and ultimately catalytic activation. Both U1 and U4 snRNP disassociate from the pre-mRNA during the transition

from B complex to C complex. After disassociation of U1 and U4, a protein complex including Prp19 (NTC standing for nineteen complex) has a role in stabilizing U5 and U6 binding to the spliceosome (Chan et al., 2003). Moreover, the U1/5'SS association is replaced by a base-pairing interaction between the U6 snRNA ACAGAGA box and the 5'SS, mediated by the helicase Prp28 (Staley and Guthrie, 1999). In addition, the extensive U4/U6 base-pairing in the U4/U6·U5 tri-snRNP is replaced with a U2/U6 base-pairing network which is believed to form the active site of the spliceosome (Staley and Guthrie, 1998). The idea that the U2/U6 structure is responsible for splicing catalysis is based partly on mutagenesis data that identified two evolutionarily conserved domains in U6 snRNA essential for catalysis: the ACAGAGA box (Lesser and Guthrie, 1993) that interacts with the 5'SS and the AGC triad (Hilliker and Staley, 2004). In addition, the internal stem loop of U6 (U6 ISL) coordinates a metal ion important for catalysis (Huppler et al., 2002; Yean et al., 2000) consistent with the possibility that the spliceosome is a ribozyme. As well, biochemical and genetic studies have implicated a role for U5 snRNA in tethering the 5' and 3' exons for the second transesterification reaction (Newman et al., 1995).

I-2. A structural approach to studying the spliceosome

The evidence for RNA in the spliceosome active site suggests that the snRNAs catalyze the splicing reaction, which would make the spliceosome a ribozyme; however, protein factors essential for splicing point to a cooperative role for the two different types of components. The spliceosome is reminiscent in terms of its

large size and RNA•protein composition with the ribosome (Burge et al., 1999; Krämer, 1996; Staley and Guthrie, 1998). Indeed, the advancement of our understanding of the ribosome by high resolution structure determination (Ban et al., 2000; Yusupov et al., 2001) suggests that a similar structural analysis of the spliceosome is a worthwhile goal. Such an achievement would greatly enhance our understanding of splicing on multiple levels including both the regulation and basic chemical mechanism of pre-mRNA splicing.

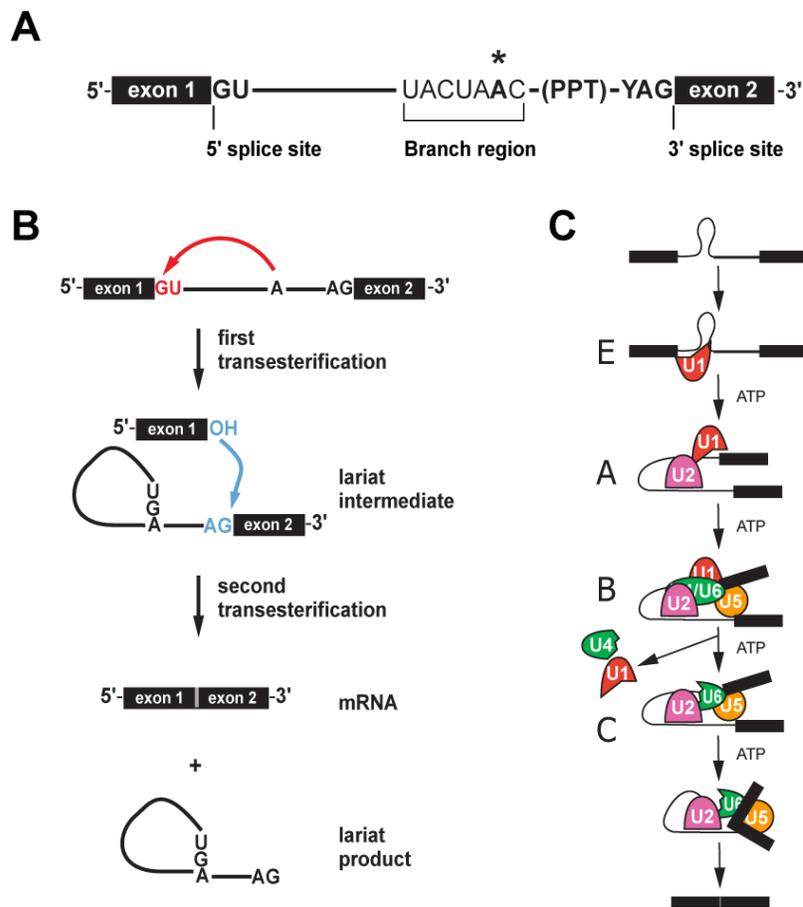


Figure I-1. Splicing of pre-mRNA by the spliceosome. A) Intron structure highlighting conserved sequences at the 5'SS and 3'SS, the optimal branch sequence, and metazoan polypyrimidine tract. The preferred branch adenosine is indicated (*). B) Sequential transesterification reactions catalyzed by the spliceosome. C) Stepwise assembly of the spliceosome on a pre-mRNA showing sequential association of U1, U2, U4/U6•U5 snRNPs with the intron. Shown are the formation of the E (commitment), A, and B complexes through to the mature spliceosome in C complex.

Complicating structural analysis of spliceosomal assemblies is the complexity of the splicing machinery not only in terms of sheer number of factors but also the nature of their interaction and dynamic association with each other and with the pre-mRNA substrate. Over and above non-trivial issues of abundance and purification, the dynamic nature of the spliceosome and the substrate dependence of its assembly have posed major difficulties in approaching structural studies. Current high resolution structural analysis of the spliceosome has relied primarily on a dissection of the complex into functionally important subunits amenable to analysis by X-ray or NMR. With one exception, structural descriptions of higher order complexes have been restricted to EM studies as outlined here. This chapter highlights advances in our understanding of spliceosomal structure with an emphasis on what has been learned with respect to splice site recognition during the dynamic process of spliceosome assembly as well as the catalytic RNA components at the heart of the splicing machinery.

I-2.1. Polypyrimidine tract recognition by U2AF in E complex

The U2 auxiliary factor (U2AF) is a heterodimer consisting of 55 and 35 kDa subunits. The large subunit (U2AF65) binds the polypyrimidine tract at the 3'SS, interacts with other E complex spliceosomal proteins (Selenko et al., 2003) and later the U2 snRNP component SF3b155 (Gozani et al., 1998). U2AF65 contains an N-terminal RS domain followed by three regions originally described as RRM (RNA recognition motifs). However, only the first two of these domains are

involved in RNA binding; the third represents a general protein interaction domain referred to as a UHM (U2AF Homology Motif; see below).

An RRM is a nucleic acid binding domain found in all organisms and typically binds single-stranded RNA, usually by recognition of a specific nucleotide sequence. RRM is a subclass of the ferredoxin fold (Carte et al., 2008) that contain a four-stranded β -sheet buttressed by two α -helices in a β - α - β - β - α - β arrangement; they are distinguished by the presence of two amino acid motifs, RNP1 and RNP2, featuring conserved aromatic residues. High resolution structures of RRM domains both alone and bound to RNA have been described (Allain et al., 2000; Deo et al., 1999; Handa et al., 1999; Oubridge et al., 1994; Wang and Hall, 2001). These reveal that single-stranded RNA typically binds on the face of the β -sheet, and that the two RNP motifs within the β -sheet are important for this interaction (Maris et al., 2005). In particular, a tyrosine or phenylalanine within the RNP motif makes a stacking interaction with bound nucleotidyl bases while the identity of the nucleotide is determined by hydrogen bonding interactions. Structural analyses of RNA binding by the poly-pyrimidine tract binding protein (PTB) and the alternative splicing factor Fox-1 reveal further complexity in RNA recognition by RRM domains (Auweter et al., 2006; Oberstrass et al., 2005). These include the substitution of RNP aromatic•RNA contacts with a separate set of hydrophobic interactions (Oberstrass et al., 2005) and distinct features with respect to the participation of RRM loops in RNA binding (Auweter et al., 2006).

Proper recognition of the polypyrimidine tract is essential for correct identification of the 3'SS; however, this sequence is of a variable length, and often interrupted by purine nucleotides. Analysis of the recent crystal structure of U2AF65 RRM1,2 bound to a polyuridine RNA as well as the high resolution structure of RRM1 alone suggests how U2AF65 is able to bind these disparate sequences (Sickmier et al., 2006; Thickman et al., 2007). Each RRM interacts with three or four uridine nucleotides; a combination of hydrogen bonding through waters and conformationally flexible amino acid residues would allow hydrogen-bonding interactions with the occasional non-uridine nucleotide in a target sequence. The crystal lattice of U2AF65 bound to RNA has an unexpected structure wherein the RNA is bound by both RRM1 and RRM2 but with each domain contributed from a different protein molecule. Based on this, Kielkopf and coworkers propose a model for the U2AF65 pyrimidine tract interaction in which the RRMs are in close proximity in a relatively condensed structure (Figure I-2A). Solution data of the protein alone obtained by SAXS (Jenkins et al., 2008) and NMR (Kent, Spyropoulos, and MacMillan unpublished) are more consistent with an extended structure for RRM1-RRM2. A model of RNA binding consistent with these observations may also be derived from the crystal structure (Figure I-2B). An intriguing possibility is that the two models represent distinct binding modes to accommodate pyrimidine tracts of varying length. This may be an important aspect of U2AF's function in bending the pre-mRNA to bring the branch point sequence and 3'SS in close proximity to each other as part

of an early organization of the pre-mRNA substrate during spliceosome assembly (Kent et al., 2003).

Early recognition of the 3'SS AG dinucleotide is mediated by U2AF35, the small subunit of the U2AF heterodimer (Merendino et al., 1999; Wu et al., 1999; Zorio and Blumenthal, 1999), but there is no high resolution structural data with respect to this interaction. This splice site recognition is coupled to pyrimidine tract binding by virtue of the U2AF35•U2AF65 pairing which is characterized by the tight interaction of a peptide from U2AF65 with the U2AF35 UHM.

I-2.2. Protein recognition mediated by U2AF homology motifs

In contrast to classical RRMs, U2AF homology motifs (UHM) have been shown to function as a protein-binding module (Kielkopf et al., 2004). The RNP1 and RNP2 sequence motifs are not conserved in these domains and the observation that they do not bind RNA is further explained by the occlusion of the canonical RNA binding surface by a C-terminal extension of the RRM positioned over the β -sheet as well as the overall negative charge of the resulting surface.

UHM•partner interaction occurs on the opposite side from the canonical RNA binding surface utilizing a conserved R-X-F motif, where X is any amino acid, found in the loop between α -helix two and β -strand four, and a conserved glutamate on helix 1 which forms a salt bridge with the arginine side chain. The UHM consensus ligand is a peptide containing the [RK]-X-[RK]-W sequence (Kielkopf et al., 2004) with the tryptophan aromatic buried in a binding pocket formed between the UHM phenylalanine and salt bridge. This interaction mode

was first observed in the X-ray structure of the RRM of U2AF35 bound to a short peptide from near the N-terminus of U2AF65 (Kielkopf et al., 2001). This UHM sandwiches W92 from the peptide ligand between F135 and the salt bridge between E88 and R133 (Figure I-2C). The U2AF35•U2AF65 interface further includes an interaction, not typical of other structures, between W134 of U2AF35 and a hydrophobic pocket formed by amino acids 95-104 of U2AF65 (Kielkopf et al., 2001).

Several other structures describing similar UHMs have been described which highlight the importance of this module in spliceosome assembly (Corsini et al., 2007; Corsini et al., 2009; Kielkopf et al., 2001). In the E complex, a UHM•ligand interaction between the N-terminus of SF1 and the third RRM of U2AF65 tethers the two proteins together. During subsequent steps of spliceosomal assembly the U2AF65-SF1 interaction is likely replaced by a U2AF65•SF3b155 interaction when SF1 is displaced from the branch region as SF3b155 contains a sequence similar to that of the UHM interacting SF1 peptide (Thickman et al., 2006). Furthermore, SF3b155 likely interacts with multiple UHMs since several UHM ligand motifs are found in the SF3b155 N-terminal region and these have been shown to be bound preferentially by different UHM proteins (Corsini et al., 2007; Thickman et al., 2006).

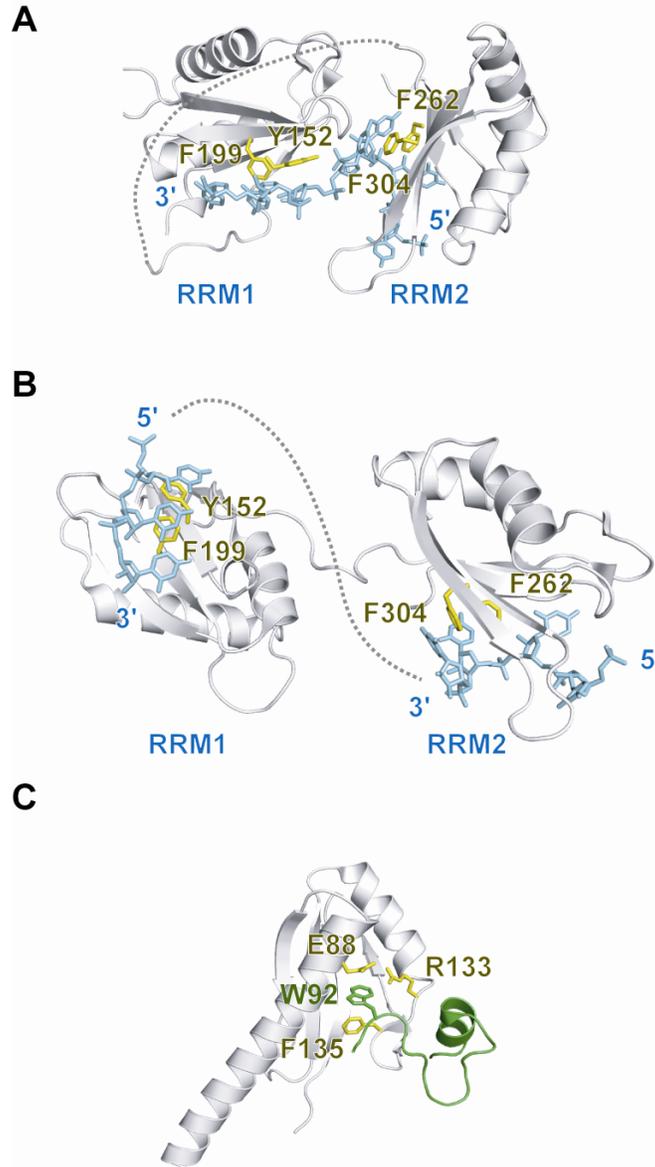


Figure I-2. Early recognition of the polypyrimidine tract by U2AF. Two possible modes of interaction based on the X-ray structure of U2AF65 RRM1 and RRM2 (grey) bound to a seven nucleotide polyuridine RNA (cyan) (Sickmier et al., 2006). A) Model based on interaction with a single RNA with RRM1 and RRM2 contributed from separate molecules as proposed in (Sickmier et al., 2006); dotted line indicates missing polypeptide linker between RRM1 and RRM2. B) Model based on a single protein interacting with two separate RNAs; dotted line indicates missing polynucleotide linker between bound uridines. Aromatic amino acids of the RNP1 and RNP2 motifs are indicated in yellow. C) UHM-ligand interaction. A tryptophan within the ligand peptide (green) is sandwiched between F377 and a salt bridge formed between E329 and R375 (yellow) of the SPF45 UHM domain.

I-2.3. SF1 and branch point sequence recognition

The branch region of the pre-mRNA substrate, like the splice sites, is recognized several times during the course of spliceosome assembly. In the commitment complex, the association of SF1 with the branch sequence is mediated by the protein's KH domain.

The hnRNP K homology (KH) domain is a ~70 amino acid single-stranded nucleic acid binding domain. Akin to the RRM, it is found in a diverse variety of organisms and is typically present in one or more copies within a protein. The KH domain family is further defined by the type I eukaryotic and type II prokaryotic variants (Grishin, 2001). The type I KH domain folds into a three-stranded anti-parallel β -sheet abutted by three α -helices. This structure represents a binding cleft which typically interacts with four nucleotides in a single-stranded extended conformation.

The NMR structure of SF1 bound to a ten nucleotide branch sequence RNA shows that it associates with the pre-mRNA via an extended type 1 KH domain in such a way that it identifies the branch sequence by hydrogen bonding with the nucleotides – especially with the invariant branch adenosine which is buried in a pocket within the protein (Liu et al., 2001) (Figure I-3A). The branch adenosine is specifically recognized on the Watson-Crick face by hydrogen bonding with the peptide backbone of I177 (Figure I-3B). The branch point sequence is subsequently recognized by U2 snRNA which forms an imperfect duplex from which the branch adenosine is extruded (Query et al., 1994); the

conformation in which SF1 holds the branch point sequence likely templates the formation of this duplex.

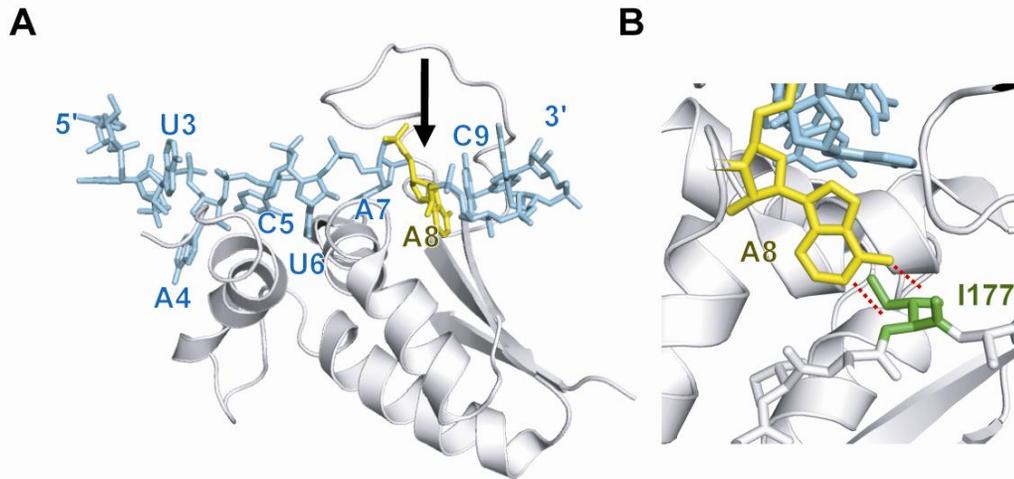


Figure I-3. Early branch sequence recognition by SF1. A) NMR structure of the SF1 KH domain (grey) (Liu et al., 2001) bound to the branch point sequence RNA (cyan) with branch adenosine (yellow) buried within a pocket (arrow). B) The branch adenosine A8 forms Watson-Crick type hydrogen bonds with the peptide backbone of I177 (green) of SF1.

I-2.4. U1 snRNP: a high resolution view

Initial identification of the 5'SS sequence involves the binding of U1 snRNP mediated by base-pairing with the 5' end of U1 snRNA and interaction with the U1 snRNP protein component U1C. The structures of some of the individual components of U1 snRNP have been described, including a co-structure of the first RRM of U1A bound to the U1 snRNA stem loop II (SLII) (Oubridge et al., 1994), the second RRM of U1A (Lu and Hall, 1997), part of U1C (Muto et al., 2004), and the heterodimer of the D1 and D2 Sm core proteins (Kambach et al., 1999). A cryo-EM study of immuno-purified U1 snRNP with a resolution to ~10 Å established the overall relationship of the snRNP components (Stark et al.,

2001) as well as confirming the proposed donut arrangement of the Sm core proteins (Kastner et al., 1990). All of these studies have been crowned by the recent description of the nearly complete U1 snRNP crystal structure at a resolution of 5.5 Å (Pomeranz Krummel et al., 2009).

In the crystal structure of human U1 snRNP, the U1 snRNA is clearly defined with an overall structure similar to that proposed based on the cryo-EM study (Stark et al., 2001). A four-helix junction composed of helices I-III and H of U1 snRNA, the threading of the RNA through the Sm core, and helix IV are all apparent in the structure (Figure I-4A). Interestingly, within the crystal lattice, the 5' end of U1 snRNA, which base pairs with the 5'SS of the pre-mRNA during spliceosome assembly, interacts with the equivalent RNA of an adjacent monomer. This forms a duplex which is directly analogous to the U1 snRNA•5'SS pairing and is useful for interpretation of the function of the U1C protein.

The Sm core proteins form a heptameric ring structure and the seven nucleotides of the Sm binding site are threaded through this ring. The mounting of the U1 snRNA four-helix junction on one side of the Sm core is mediated by interactions between two of the helices from the junction with the N-terminal helical extensions of Sm D2 and Sm B (Figure I-4B). Thus, the structure suggests that the Sm core acts as a platform for complex RNA structures; this is likely to be a common feature of the other spliceosomal snRNPs as well because complex RNA structures are always found at the 5' side of the Sm core binding site (Pomeranz Krummel et al., 2009).

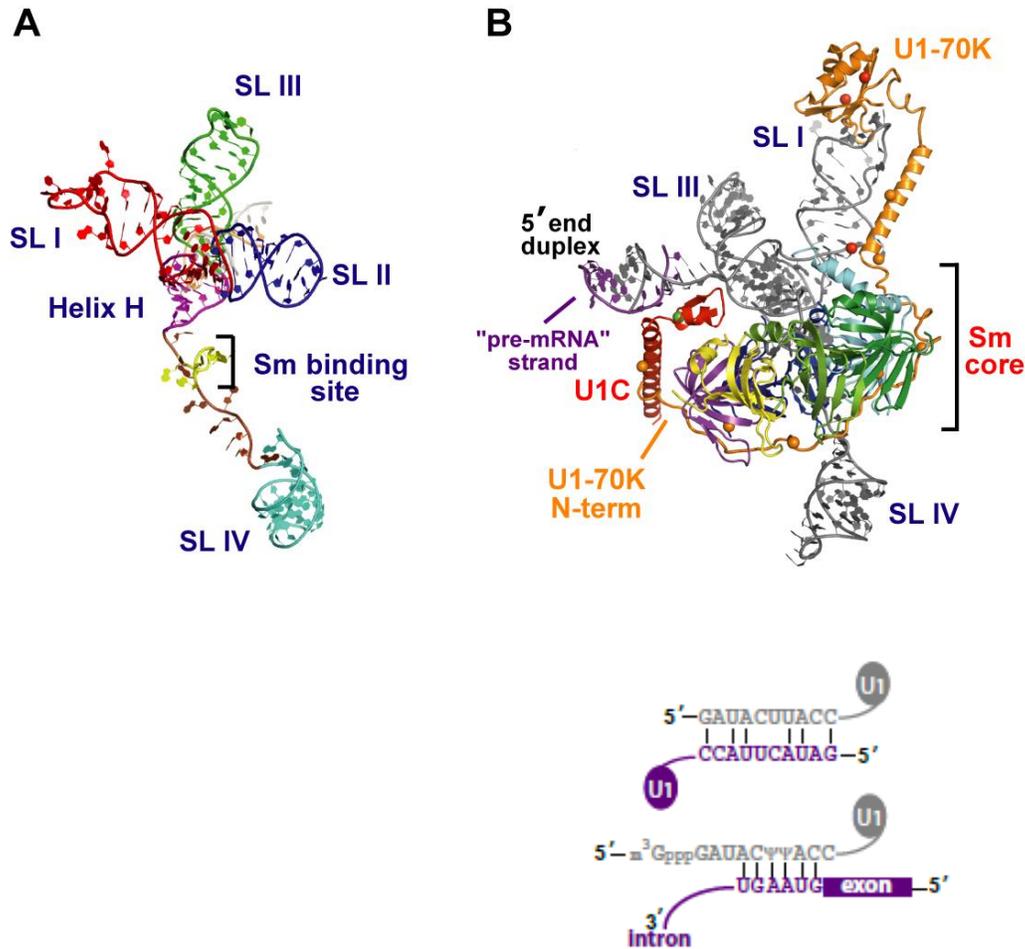


Figure I-4. Crystal structure of U1 snRNP. A) Conformation of U1 snRNA within the snRNP. The RNA forms a four-helix junction based on SLI-III and Helix H. The Sm binding site is 3' to this junction followed by Helix IV. B) (upper) The Sm protein core forms a platform upon which the four helix junction is mounted. U1C is tethered to the particle through interactions with both U1-70K and the Sm core. The U1-70K RRM binds SLI and its N-terminal 100 amino acids wrap around the Sm core to interact with U1C. (lower) Secondary structure diagram of the interaction between adjacent U1 snRNAs in the crystal (upper) shows the similarity to the spliceosomal U1 snRNA•5'SS pairing (lower). (Figures courtesy K. Nagai, adapted from Pomeranz Krummel et al., 2009).

The human U1C protein is known to contain a zinc finger structure (Muto et al., 2004), and yeast U1C has been proposed to directly interact with the 5'SS (Du and Rosbash, 2002). Integration of U1C into the U1 snRNP particle is known to be dependent on the N-terminal region of U1-70K and the Sm core

domain (Nelissen et al., 1994). The structure shows an extended segment, helix B, of U1C is responsible for interaction with these proteins (Figure I-4B). As described above, U1 snRNA•5'SS pairing is mimicked in the crystal lattice and the zinc finger of U1C can be seen interacting with this duplex. Due to the limited resolution of this structure, specific details of this interaction are not visible. Nonetheless, U1C is positioned along the minor groove of the RNA duplex, including the location corresponding to the base pairs with the invariant GU dinucleotide which defines the 5'SS. This suggests that the function of U1C may be to communicate to the rest of the snRNP that the correct 5'SS interaction has been formed.

The U1-70K protein contains a central RRM domain which binds the end of SLI in U1 snRNP and density corresponding to this interaction is seen in the crystal structure. The N-terminal 100 amino acids of this protein has no predicted domain structure and can be seen to extend along SLI, around the Sm core to where the N-terminus of U1-70K forms the binding site for U1C. This striking encirclement of the U1 snRNP likely functions to stabilize the snRNP structure like wrapping a string around a package (Figure I-4B).

A number of features left undefined in structural studies of individual components are resolved in the snRNP structure. For example, the N-terminal extension of the D2 Sm protein was not visible in the smaller D1/D2 heterodimer structure, the structure of U1C in isolation, although similar, is not identical and difficult to interpret in the absence of the U1 snRNA duplex observed in the U1 snRNP structure, and of course no meaningful structural information for the N-

terminus of U1-70K could be attained because its conformation is dependent on its context within the larger particle.

The crystal structure of the U1 snRNP also allows re-interpretation of older biochemical data such as hydroxyl radical footprinting data showing that the 5' end of U2 snRNA is in close proximity to U1 snRNP (Donmez et al., 2007). Mapping of strong cleavage sites from that study suggests that the 5' end of U2 snRNA is very near the center of the 4-helix junction of U1 snRNA which is considerably less splayed than proposed in the original model (Stark et al., 2001).

I-2.5. Structures related to the A complex

An EM structure of affinity purified spliceosomal A complex has been described (Behzadnia et al., 2007). Owing to the low resolution of the model — $\sim 40 \text{ \AA}$ — features corresponding to known proteins cannot be identified; the overall shape is slightly elongated, reminiscent of the cryo-EM structure of the U11/U12 di-snRNP (see appendix III) (Golas et al., 2005) (the U1 and U2 homologs from the minor spliceosome the cellular machinery responsible for splicing of the rare — less than 1% — class of U12 introns with atypical splice sites) (Patel and Steitz, 2003).

A salt-dissociable component of the 17S U2 snRNP, SF3b, can be purified from HeLa cell nuclear extracts and has been shown to be necessary for pre-mRNA splicing. SF3b is a 450 kDa multi-protein complex containing seven polypeptides ranging in molecular weight from 10 to 150 kDa; this includes the factors SF3b10, SF3b14, SF3b14b, SF3b49, SF3b130, SF3b145, and SF3b155.

The cryo-EM structure of SF3b has also been described (see appendix III) (Golas et al., 2003). The locations of SF3b49 and SF3b14 could be identified based on the fact that they contain RRM domains. The C-terminal region of SF3b155 contains α -helical HEAT repeats, visible in the EM structure, which form a ladder structure and may represent a scaffold for assembly of other proteins within the particle.

Progression of spliceosome assembly to the A complex involves recruitment of U2 snRNP to the branch point sequence where the U2 snRNA forms an imperfect duplex that bulges out the branch adenosine and selects it as the nucleophile for the first transesterification of splicing (Query et al., 1994). An analogous bulged duplex structure is formed in the self-splicing group II introns (Schlatterer et al., 2006; Zhang and Doudna, 2002) and is believed to be essential for catalysis. Several structures of a branch point sequence with a bulged adenosine have been described (Berglund et al., 2001; Lin and Kielkopf, 2008; Newby and Greenbaum, 2002). They differ slightly with respect to whether the preferred branch adenosine or the adjacent adenosine is bulged out as well as in the conformation of the bulged residue, which ranges from flipped out in an extended conformation to slightly extruded from the helix. These differing conformations imply a flexibility that is presumably fixed within the context of the spliceosome. However, it has also been shown that the conserved pseudouridine within U2 snRNA that lies opposite the branch nucleotide stabilizes the extruded conformation of the bulged adenosine, perhaps by extra

hydrogen bonding to the N5 position of the pseudouridine (Lin and Kielkopf, 2008; Newby and Greenbaum, 2002).

Formation of the bulged pre-mRNA•U2 snRNA duplex occurs within the context of an SF3b branch region interaction the function of which is poorly understood. A 14 kDa protein subunit of the SF3b particle, p14 (also known as SF3b14) directly contacts the branch adenosine in the A complex (MacMillan et al., 1994; Query et al., 1997) and this interaction persists into the mature spliceosome although recent work has shown that the intimate association of SF3b with the pre-mRNA is disrupted at or after the first step of splicing (Bessonov et al., 2008). A structural analysis of p14 in complex with another SF3b component, SF3b155, with investigations into the interaction between p14 and the branch adenosine is presented in Appendix III.

I-2.6. Structure of RNA at the spliceosome active site

Once assembled the spliceosome active site is believed to be composed of a complex network of snRNA interactions. Intriguingly, a protein-free preparation of U2/U6 snRNA based on the human complex is capable of stimulating a slow, inefficient reaction that mimics the first step of splicing (Valadkhan and Manley, 2001; Valadkhan et al., 2007). More recently, the protein-free U2/U6 system has been modified to perform a ligation reaction that is chemically identical to the second step of splicing by the spliceosome and group II introns (Valadkhan et al., 2009). The two-step reaction is performed on two short RNA substrates. Initial cleavage of the substrates is performed by hydrolysis, with subsequent ligation of

the two RNA substrates into a linear product that resembles fully spliced mRNA. The RNA-catalyzed ligation reaction is very sensitive to changes in the sequence of both the AGC triad and the ACAGAGA box of U6 snRNA, similar to requirements of the spliceosome. Catalysis of a splicing-like reaction in the absence of proteins argues strongly that the spliceosome is a ribozyme (Valadkhan et al., 2009). Speculation regarding the individual roles of RNA and protein in splicing catalysis has centered on the high evolutionary conservation between the group II self-splicing introns and the spliceosome. The chemistry of intron removal from group II RNAs is identical to that of pre-mRNA splicing indicating that the snRNAs of the spliceosome may be directly involved in catalysis. Indeed, similarities in the two systems extend to the reversibility of both transesterification reactions (Augustin et al., 1990; Morl and Schmelzer, 1990; Tseng and Cheng, 2008).

The catalytic strategy of group II introns most likely involves a two-metal ion mechanism as proposed for both the group II system as well as the spliceosomal reaction on the basis of the mechanism of DNA and RNA polymerases in phosphoryl-transfer reactions (Figure I-5) (Steitz and Steitz, 1993). Domain 5 of group II introns is a metal-binding platform that coordinates Mg^{2+} ions that are likely responsible for activating the first step nucleophile as well as for stabilizing the oxyanion leaving group during the first step of the splicing reaction (Toor et al., 2008). In the 3.1 Å crystal structure of a group II intron from *O. iheyensis*, tertiary RNA contacts stabilize the arrangement of the juxtaposed exon-binding sequences 1 and 3 which interact with the 5' and 3' exons

respectively, over the bulge of domain 5 which binds the essential Mg^{2+} ions (Toor et al., 2008). The two metal ions are 3.9 Å apart matching the ideal distance noted for the active sites of analogous protein enzymes (Steitz and Steitz, 1993).

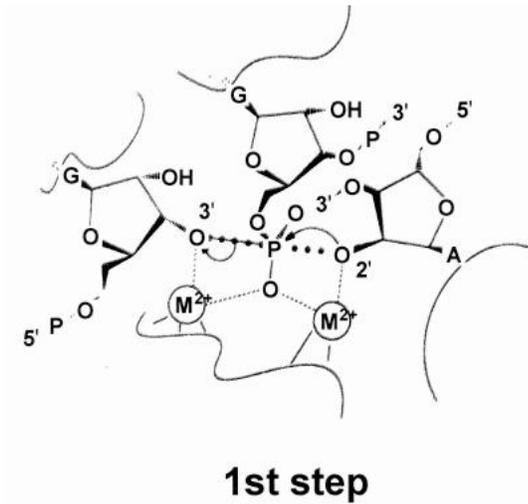


Figure I-5. Mechanism for spliceosome and group II splicing. Two Mg^{2+} ions responsible for activating the first step nucleophile as well as for stabilizing the oxyanion leaving group during the first step of splicing by group II introns and possibly by the spliceosome are positioned 3.9 Å apart. (Adapted from Steitz and Steitz, 1993).

The NMR data with respect to the yeast U2/U6 RNA duplex which mimics the snRNA arrangement at the spliceosome active site reveals a complex fold resembling the 4-helix junction adopted by other RNAs such as the Hairpin ribozyme (Figure I-6) (Rupert and Ferre-D'Amare, 2001; Sashital et al., 2004). The intermolecular base-pairing between U2 and U6 forms helices I-III; a key feature specific to the U2/U6 complex is the U6 ISL, which forms through intramolecular base-pairing when U6 snRNA is unwound from the U4/U6 duplex. The identification of the U2/U6 complex as a 4-helix junction has important implications for how several crucial RNA elements in the U2/U6 duplex are

arranged for catalysis. Models of 4-helix junction folds predict that the relative orientation of the U6 ISL with respect to the 5'SS in the U2/U6 structure is expected to juxtapose essential sequence elements within the ISL with the 5'SS and branch point sequence (Figure I-6) (Sashital et al., 2004) analogous to the situation observed in the crystal structure of the *Oceanobacillus iheyensis* group II intron (Toor et al., 2008). An unpaired residue, U80 in the yeast ISL, serves as a binding site for a catalytically essential metal (Yean et al., 2000) analogous to the metal-binding site reported for domain 5 of group II introns (Toor et al., 2008). Domain 5 and the U6 ISL are believed to be structurally and mechanistically analogous (Sigel et al., 2000).

The solution structure of a U2/U6 model duplex includes an extension of the U6 ISL involving the invariant AGC triad of U6 snRNA which was previously believed to make base-pairing interactions with U2 snRNA residues to form U2/U6 helix Ib. The U2 snRNA residues instead participate in intramolecular base-pairing interactions to form part of U2 helix stem I (Figure I-6). The extension of the U6 ISL to include the AGC triad increases its similarity with domain 5, as well as the similarity between the yeast U2/U6 structure and the structure previously predicted for the mammalian U2/U6 duplex (Sashital et al., 2004).

Recent work challenges the relevance of the 4-helix junction model of U2/U6 structure at the spliceosome active site (Mefford and Staley, 2009). As noted above, the 4-helix model predicts the formation of an extended U6 ISL as opposed to the mutually exclusive U2/U6 helix Ib. Studies aimed at elucidating a

role for helix Ib during splicing catalysis utilized mutant pre-mRNA substrates rate-limiting for either the first or second step of splicing. Mutations in U2/U6 that weaken either helix Ib or the 4-helix junction conformation exacerbated the effect of mutant substrates rate-limiting for either step. Intriguingly, only restoration of the helix Ib conformation but not the 4-helix junction eliminated this exacerbation of the splicing defect, providing compelling evidence that a U2/U6 structure that includes helix Ib is important at both steps of splicing catalysis (Figure I-6) (Mefford et al., 2009). This work does not rule out the possibility that a 4-helix junction conformation exists at other stages of spliceosome assembly, including between the two steps of splicing.

In the crystal structure of the group II intron, tertiary interactions cause a bend in the domain 5 helix, as well as an unusual kink in the bulge region responsible for coordination of the proposed catalytic magnesium ions. As well, a tertiary interaction between domain 5 and the linker between domains 2 and 3, J2/3, which is analogous to the highly conserved ACAGAGA sequence of U6 snRNA, brings together catalytically essential sequences (Figure I-7) (Toor et al., 2008). These features are not present in domain 5 studied in isolation (Sigel et al., 2004; Zhang and Doudna, 2002), so it seems likely that analogous structural features might be relevant to the U6 ISL in the context of the spliceosome perhaps mediated in this case by interactions with protein components. The leading candidate for a protein affecting splicing catalysis by modulating the RNA structures at the active site is Prp8 (the subject of subsequent chapters of this thesis) (Collins and Guthrie, 2000).

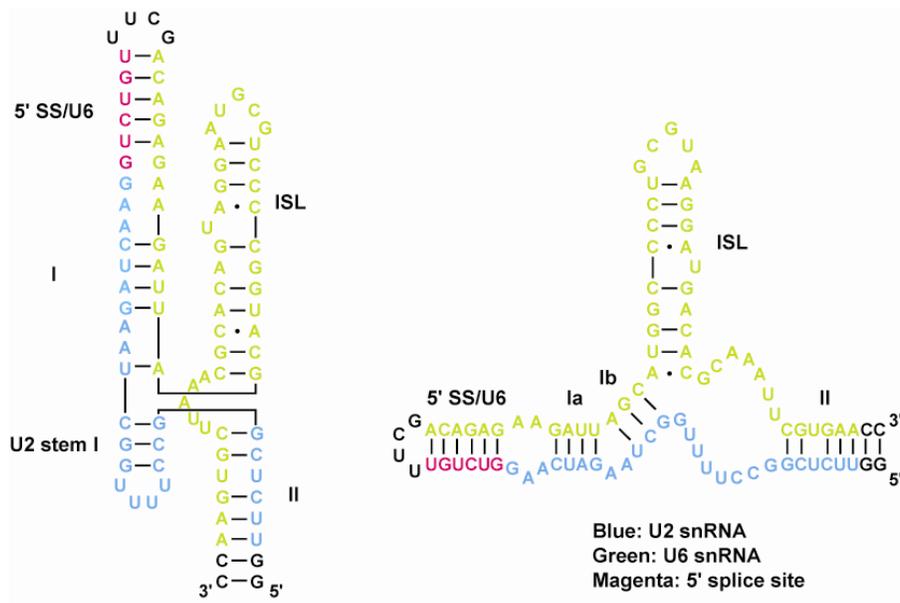


Figure I-6. Proposed structures of U2/U6 snRNA in the spliceosome active site. Two mutually exclusive structures have been proposed for the U2/U6 snRNA structure at the spliceosome active site. (Left) A 4-helix junction structure has been described which is expected to juxtapose the metal-binding site (bulged U) in the U6 ISL with the 5'SS helix (based on Sashital et al., 2004). (Right) Recent work suggests that a U2/U6 duplex that includes helix Ib is the relevant RNA structure at the spliceosome active site (Mefford and Staley, 2009). Sequences shown are for the mammalian system.

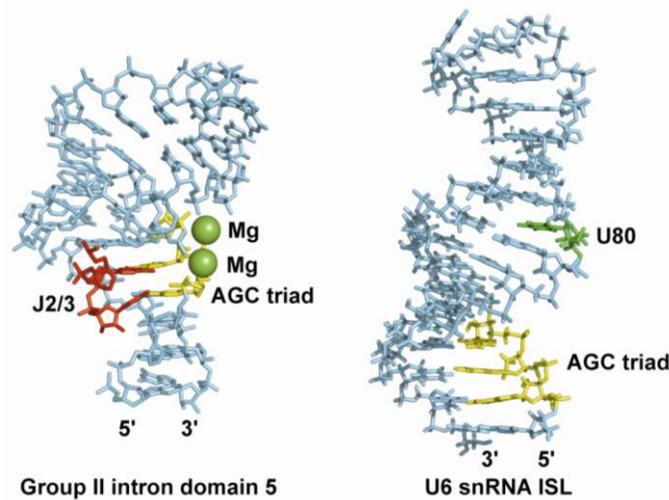


Figure I-7. Comparison of group II intron and spliceosomal active site RNA structures. The crystal structure of domain 5 from the *Oceanobacillus iheyensis* group II intron (Toor et al., 2008) (left) alongside the NMR structure of yeast U6 snRNA ISL (right) (Sashital et al., 2004). The location of two catalytic magnesium ions are shown as green spheres in the domain 5 structure and the U80 nucleotide implicated in magnesium binding by the U6 ISL is highlighted (green). Tertiary RNA contacts between domain 5 and the J2/3 region (red) important for catalysis are shown. The catalytic triad is shown for both domain 5 and U6 ISL (yellow).

I-3. Structural studies of the catalytic spliceosome

While detailed pictures of the components behind spliceosome activation and catalysis are provided by high resolution structural techniques, EM has provided useful information about the relative locations of sub-complexes in the larger catalytic C complex which contains the U2, U5 and U6 snRNPs and the Prp19 complex (Jurica et al., 2004). An EM structure of the spliceosome at ~30 Å resolution reveals a complex with dimensions of 270 X 240 Å arranged in three distinct domains (Figure I-8). The C complex, therefore, is roughly the same size as purified tri-snRNPs, a surprising observation considering how many proteins have been identified as associated with the catalytic spliceosome (Jurica et al., 2002).

In the C complex EM structure a roughly cylindrical domain is separated from two additional, distinct domains creating a large cavity in the centre of the structure. The open arrangement of the three domains has been interpreted to represent structural flexibility- a feature that might enable the many structural rearrangements that must occur during splicing. The assignment of sub-complexes to the density associated with the individual domains is not unambiguous although several models based on the masses and structures of individual components are possible (Jurica et al., 2004). These ambiguities highlight the challenge of obtaining additional high resolution structures of both lower and higher order complexes in order to interpret the existing EM snapshots of spliceosome assembly.

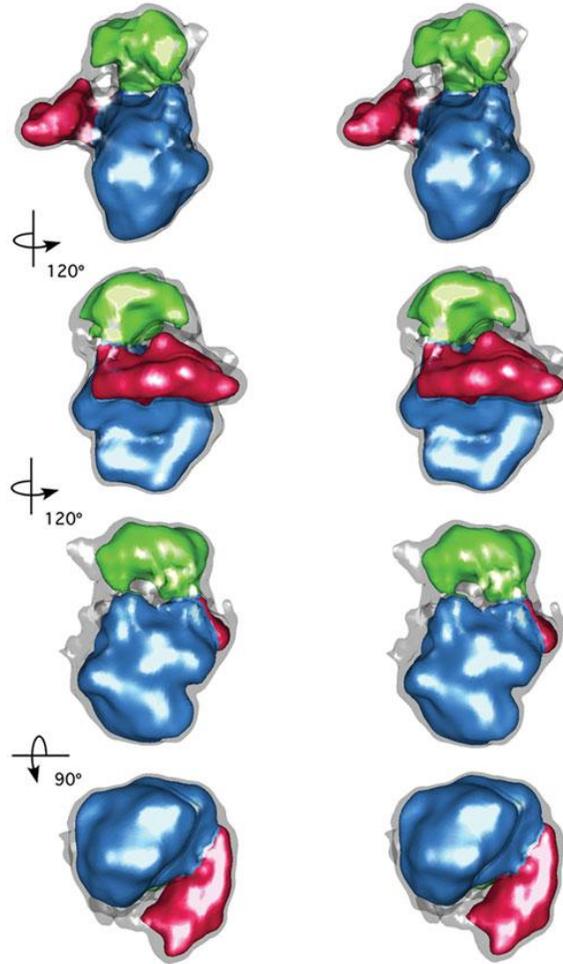


Figure I-8. Stereo views of the surface representation of the refined three-dimensional structure of C complex. Orientations correspond to rotations (120° or 90°) around the axes indicated. The gray contour represents a complex of ~ 2.6 MDa (summed molecular masses of U2, U5 and U6 snRNPs and the Prp19 complex). Three major domains of the structure are indicated by the colored regions (Adapted from Jurica et al., 2004).

While high resolution structures of spliceosomal complexes have been elusive, the landmark X-ray analysis of U1 snRNP represents the culmination of efforts to understand higher order spliceosomal structure in the context of an RNP. Furthermore, high resolution structural analyses of possible RNA and protein constituents of the spliceosome active site have recently yielded exciting results. But these structures are tantalizingly incomplete posing more questions than they answer. Is the spliceosome a ribozyme or an RNPzyme? What is the

molecular nature of spliceosomal rearrangements at the active site between the two steps of splicing? The solution to these puzzles could very well provide another link in the connection between the RNA and protein catalyzed worlds. This thesis highlights structural and biochemical studies of a potential active site protein essential for spliceosome function, Prp8. The results presented bring into focus the importance of potential interactions between RNA and protein at the heart of the spliceosome.

I-4. References

Allain, F.H.T., Bouvet, P., Dieckmann, T., and Feigon, J. (2000). Molecular basis of sequence-specific recognition of pre-ribosomal RNA by nucleolin. *EMBO J.* **19**, 6870-6871.

Augustin, S., Muller, M.W., and Schweyen, R.J. (1990). Reverse self-splicing of group II intron RNAs in vitro. *Nature* **343**, 383-386.

Auweter, S.D., Fasan, R., Reymond, L., Underwood, J.G., Black D.L., Pitsch S., and Allain F.H.T. (2006). Molecular basis of RNA recognition by the human alternative splicing factor Fox-1. *EMBO J.* **25**, 163-173.

Ban, N., Nissen, P., Hansen, J., Moore, P.B., and Steitz T.A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* **289**, 905-920.

Behzadnia, N., Golas, M.M., Hartmuth, K., Sander, B., Kastner, B., Deckert, J., Dube, P., Will, C.L., Urlaub, H., Stark, H., and Lührmann, R. (2007). Composition and three-dimensional EM structure of double affinity-purified, human prespliceosomal A complexes. *EMBO J.* **26**, 1737-1748.

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**, 3171-3175.

Berglund, J.A., Rosbash, M., and Schultz, S.C. (2001). Crystal structure of a model branchpoint-U2 snRNA duplex containing bulged adenosines. *RNA* **7**, 682-691.

Bessonov, S., Anokhina, M., Will, C.L., Urlaub, H., and Lührmann, R. (2008). Isolation of an active site step I spliceosome and composition of its RNP core. *Nature* **452**, 846-850.

Burge, C.B., Tuschl, T., and Sharp, P.A. (1999). *In The RNA World 2nd edn.* (eds. Gesteland, R.F., Cech, T.R., Atkins, J.F.) 525-560 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York).

Carte, J., Wang, R., Li, H., Terns, R.M., and Terns, M.P. (2008). Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev.* **22**, 3489-3496.

Chan, S.P., Kao, D.I., Tsai, W.Y., and Cheng, S.C. (2003). The Prp19p associated complex in spliceosome activation. *Science* **302**, 279-282.

- Chow, L.T., Gelinas, R.E., Broker, T.R., and Roberts, R.J. (1977). An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell* **12**, 1-8.
- Collins, C.A., and Guthrie, C. (2000). The question remains: is the spliceosome a ribozyme? *Nat. Struct. Biol.* **7**, 850–854.
- Corsini, L., Bonna, S., Basquin, J., Hothorn, M., Scheffzek, K., Valcárcel, J., and Sattler, M. (2007). U2AF-homology motif interactions are required for alternative splicing regulation by SPF45. *Nat. Struct. Mol. Biol.* **14**, 620-629.
- Corsini, L., Hothorn, M., Stier, G., Rybin, V., Scheffzek, K., Gibson, T.J., and Sattler, M. (2009). Dimerization and protein binding specificity of the U2AF homology motif of the splicing factor Puf60. *J. Biol. Chem.* **284**, 630-639.
- Deo, R.C., Bonanno, J.B., Sonenberg, N., and Burley, S.K. (1999). Recognition of polyadenylate RNA by the poly(A)-binding protein. *Cell* **98**, 835-845.
- Donmez, G., Hartmuth, K., Kastner, B., Will, C.L., and Luhrmann, R. (2007). The 5' end of U2 snRNA is in close proximity to U1 and functional sites of the pre-mRNA in early spliceosomal complexes. *Mol. Cell* **25**, 399-411.
- Du, H., and Rosbash, M. (2002). The U1 snRNP protein U1C recognizes the 5' splice site in the absence of base pairing. *Nature* **419**, 86-90.
- Gaur, R.K., Valcarcel, J., and Green, M.R. (1995). Sequential recognition of the pre-mRNA branch point by U2AF65 and a novel spliceosome-associated 28-kDa protein. *RNA* **1**, 407-417.
- Golas, M.M., Sander, B., Will, C.L., Luhrmann, R., and Stark, H. (2003). Molecular architecture of the multiprotein splicing factor SF3b. *Science* **300**, 980-984.
- Golas, M.M., Sander, B., Will, C.L., Luhrmann, R., and Stark, H. (2005). Major conformational change in the complex SF3b upon integration into the spliceosomal U11/U12 di-snRNP as revealed by electron cryomicroscopy. *Mol. Cell* **17**, 869-883.
- Gozani, O., Potashkin, J., and Reed, R. (1998). A potential role for U2AF-SAP 155 interactions in recruiting U2 snRNP to the branch site. *Mol. Cell. Biol.* **18**, 4752-4760.
- Grishin, N.V. (2001). KH domain: one motif, two folds. *Nucleic Acids Res.* **29**, 638-643.

- Handa, N., Nureki, O., Kurimoto, K., Kim, I., Sakamoto, H., Shimura, Y., Muto, Y., and Yokoyama, S. (1999). Structural basis for recognition of the tra mRNA precursor by the Sex-lethal protein. *Nature* **398**, 579-585.
- Hilliker, A. K., and Staley, J. P. (2004). Multiple functions for the invariant AGC triad of U6 snRNA. *RNA* **10**, 921-928.
- Huppler, A., Nikstad, L.J., Allmann, A.M., Brow, D.A., and Butcher, S.E. (2002). Metal binding and base ionization in the U6 RNA intramolecular stem-loop structure. *Nat. Struct. Biol.* **9**, 431-435.
- Jamison, S.F., Crow, A., and Garcia-Blanco, M.A. (1992). The spliceosome assembly pathway in mammalian extracts. *Mol. Cell. Biol.* **12**, 4279-4287.
- Jenkins, J.L., Shen, H., Green, M.R., and Kielkopf, C.L. (2008). Solution conformation and thermodynamic characteristics of RNA binding by the splicing factor U2AF65. *J. Biol. Chem.* **283**, 33641-33649.
- Jurica, M.S., Licklider, L.J., Gygi, S.P., Grigorieff, N., and Moore, M.J. (2002). Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis. *RNA* **8**, 426-439.
- Jurica, M.S., and Moore, M.J. (2003). Pre-mRNA splicing: awash in a sea of proteins. *Mol. Cell* **12**, 5-14.
- Jurica, M.S., Sousa, D., Moore, M.J., and Grigorieff, N. (2004). Three-dimensional structure of C complex spliceosomes by electron microscopy. *Nat. Struct. Mol. Biol.* **11**, 265-269.
- Kambach, C., Walke, S., Young, R., Avis, J.M., de la Fortelle, E., Raker, V.A., Lührmann, R., Li, J., and Nagai, K. (1999). Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. *Cell* **96**, 375-387.
- Kastner, B., Back, M., and Luhrmann, R. (1990). Electron microscopy of small nuclear ribonucleoprotein (snRNP) particles U2 and U5: evidence for a common structure-determining principle in the major U snRNP family. *Proc. Natl. Acad. Sci. U.S.A.* **87**, 1710-1714.
- Kielkopf, C.L., Lücke, S., and Green, M.R. (2004). U2AF homology motifs: protein recognition in the RRM world. *Genes Dev.* **18**, 1513-1526.
- 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**, 595-605.

- 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**, 50572-50577.
- Kramer, A. (1996). The structure and function of proteins involved in mammalian pre-mRNA splicing. *Ann. Rev. Biochem.* **65**, 367-409.
- Kramer, A., and Utans, U. (1991). Three protein factors (SF1, SF3, and U2AF) function in pre-splicing complex formation in addition to snRNPs. *EMBO J.* **10**, 1503-1509.
- Legrain, P., Seraphin, B., and Rosbash, M. (1988). Early commitment of yeast pre-mRNA to the spliceosome pathway. *Mol. Cell. Biol.* **8**, 3755-3760.
- Lesser, C. F., and Guthrie, C. (1993). Mutations in U6 snRNA that alter splice site specificity: implications for the active site. *Science* **262**, 1982–1988.
- Lin, Y., and Kielkopf, C.L. (2008). X-ray structures of U2 snRNA-branchpoint duplexes containing conserved pseudouridines. *Biochemistry* **47**, 5503-5514.
- Liu, Z., Luyten, I., Bottomley, M.J., Messias, A.C., Houngninou-Molango, S., Sprangers, R., Zanier, K., Kramer, A., and Sattler, M. (2001). Structural basis for recognition of the intron branch site RNA by splicing factor 1. *Science* **294**, 1098-1102.
- Lu, J., and Hall, K.B. (1997). Tertiary structure of RBD2 and backbone dynamics of RBD1 and RBD2 of the human U1A protein determined by NMR spectroscopy. *Biochemistry* **36**, 10393-10405.
- 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**, 3008-3020.
- Madhani, H.D., and Guthrie, C. (1992). A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. *Cell* **71**, 803–817.
- Maris, C., Dominguez, C., and Allain, F.H. (2005). The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J.* **272**, 2118-2131.
- Mefford, M.A., and Staley, J.P. (2009). Evidence that U2/U6 helix I promotes both catalytic steps of pre-mRNA splicing and rearranges in between these steps. *RNA* **15**, 1386-1397.

- Merendino, L., Guth, S., Bilbao, D., Martinez, C., and Valcarcel, J. (1999). Inhibition of msl-2 splicing by Sex-lethal reveals interaction between U2AF35 and the 3' splice site AG. *Nature* **402**, 838-841.
- Morl, M., and Schmelzer, C. (1990). Integration of group II intron b11 into a foreign RNA by reversal of the self-splicing reaction in vitro. *Cell* **60**, 629-636.
- Muto, Y., Pomeranz Krummel, D., Oubridge, C., Hernandez, H., Robinson, C.V., Neuhaus, D., and Nagai, K. (2004). The structure and biochemical properties of the human spliceosomal protein U1C. *J. Mol. Biol.* **341**, 185-198.
- Nelissen, R.L., Will, C.L., van Venrooij, W.J., and Lührmann, R. (1994). The association of the U1-specific 70K and C proteins with U1 snRNPs is mediated in part by common U snRNP proteins. *EMBO J.* **13**, 4113-4125.
- Newby, M.I., and Greenbaum, N.L. (2002). Sculpting of the spliceosomal branch site recognition motif by a conserved pseudouridine. *Nat. Struct. Biol.* **9**, 958-965.
- Newman, A.J., Teigelkamp, S., and Beggs, J.D. (1995). snRNA interactions at 5' and 3' splice sites monitored by photoactivated crosslinking in yeast spliceosomes. *RNA* **1**, 968-980.
- Oberstrass, F.C., Auweter, S.D., Erat, M., Hargous, Y., Henning, A., Wenter, P., Reymond, L., Amir-Ahmady, B., Pitsch, S., Black, D.L., and Allain, F.H.T. (2005). Structure of PTB bound to RNA: specific binding and implications for splicing regulation. *Science* **309**, 2054-2057.
- Oubridge, C., Ito, N., Evans, P.R., Teo, C.H., and Nagai, K. (1994). Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature* **372**, 432-438.
- Patel, A.A., and Steitz, J.A., (2003). Splicing double: insights from the second spliceosome. *Nat. Rev. Mol. Cell Bio.* **4**, 960-970.
- Pomeranz Krummel, D.A., Oubridge, C., Leung, A.K., Li, J., and Nagai, K. (2009). Crystal structure of human spliceosomal U1 snRNP at 5.5 Å resolution. *Nature* **458**, 475-480.
- Puig, O., Gottschalk, A., Fabrizio, P., and Séraphin, B. (1999). Interaction of the U1 snRNP with nonconserved intronic sequences affects 5' splice site selection. *Genes Dev.* **13**, 569-580.
- Query, C.C., McCaw, P.S., and Sharp, P.A. (1997). A minimal spliceosomal complex A recognizes the branch site and polypyrimidine tract. *Mol. Cell. Biol.* **17**, 2944-2953.

- Query C.C., Moore M.J., and Sharp, P.A. (1994). Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model. *Genes Dev.* **8**, 587–597.
- Rappsilber, J., Ryder, U., Lamond, A.I., and Mann, M., (2002). Large-scale proteomic analysis of the human spliceosome. *Genome Res.* **12**, 1231-1245.
- Rupert, P.B., and Ferre-D’Amare, A.R. (2001). Crystal structure of a hairpin ribozyme-inhibitor complex with implications for catalysis. *Nature* **410**, 780-786.
- Sashital, D.G., Cornilescu, G., and Butcher, S.E. (2004). U2-U6 RNA folding reveals a group II intron-like domain and a four-helix junction. *Nat. Struct. Mol. Biol.* **11**, 1237-1242.
- Schellenberg, M.J., Edwards, R.A., Ritchie, D.B., Kent, O.A., Golas, M.M., Stark, H., Luhrmann, R., Glover, J.N., and MacMillan, A.M., (2006). Crystal structure of a core spliceosomal protein interface. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 1266-1271.
- Schlatterer, J.C., Crayton, S.H., and Greenbaum, N.L. (2006). Conformation of the group II intron branch site in solution. *J. Am. Chem. Soc.* **128**, 3866-3867.
- 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**, 965-976.
- Seraphin, B., and Rosbash, M. (1989). Identification of functional U1 snRNA-pre-mRNA complexes committed to spliceosome assembly and splicing. *Cell* **59**, 349-358.
- Sickmier, E.A., Frato, K.E., Shen, H., Paranawithana, S.R., Green, M.R., and Kielkopf, C.L. (2006). Structural basis for polypyrimidine tract recognition by the essential pre-mRNA splicing factor U2AF65. *Mol. Cell* **23**, 49-59.
- Sigel, R.K., Sashital, D.G., Abramovitz, D.L., Palmer, A.G., Butcher, S.E., and Pyle, A.M. (2004). Solution structure of domain 5 of a group II intron ribozyme reveals a new RNA motif. *Nat. Struct. Mol. Biol.* **11**, 187-192.
- Sigel, R.K., Vaidya, A., and Pyle, A.M. (2000). Metal ion binding sites in a group II intron core. *Nat. Struct. Biol.* **7**, 1111-1116.
- Staley, J.P., and Guthrie, C. (1998). Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell* **92**, 315-326.

- Staley, J.P., and Guthrie, C. (1999). An RNA switch at the 5' splice site requires ATP and the DEAD box protein Prp28. *Mol. Cell* **3**, 55–64.
- Stark, H., Dube, P., Lührmann, R., and Kastner, B. (2001). Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle. *Nature* **409**, 539-542.
- Steitz, T.A., and Steitz, J.A. (1993). A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 6498-6502.
- Thickman K.R., Sickmier E.A., and Kielkopf C.L. (2007). Alternative conformations at the RNA-binding surface of the N-terminal U2AF(65) RNA recognition motif. *J. Mol. Biol.* **366**, 703-710.
- Thickman, K.R., Swenson, M.C., 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**, 664-683.
- Toor, N., Keating, K.S., Taylor, S.D., and Pyle, A.M. (2008). Crystal structure of a self-spliced group II intron. *Science* **320**, 77-82.
- Tseng, C.K., and Cheng, S.C. (2008). Both catalytic steps of nuclear pre-mRNA splicing are reversible. *Science* **320**, 1782-1784.
- Valadkhan, S., and Manley, J.L. (2001). Splicing-related catalysis by protein-free snRNAs. *Nature* **413**, 701-707.
- Valadkhan, S., Mohammadi, A., Wachtel, C., and Manley, J.L. (2007). Protein-free spliceosomal snRNAs catalyze a reaction that resembles the first step of splicing. *RNA* **13**, 2300-2311.
- Valadkhan, S., Mohammadi, A., Jaladat, Y., and Geisler, S. (2009). Protein-free small nuclear RNAs catalyze a two-step splicing reaction. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 11901-11906.
- Wang, X., and Hall, T.M.T. (2001). Structural basis for recognition of AU-rich element RNA by the HuD protein. *Nat. Struct. Biol.* **8**, 141-145.
- Wu, J.Y., and Maniatis, T. (1993). Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell* **75**, 1061-1070.
- 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**, 832-835.

Yean, S.L., Wuenschell, G., Termini, J., and Lin, R.J., (2000). Metal-ion coordination by U6 small nuclear RNA contributes to catalysis in the spliceosome. *Nature* **408**, 881-884.

Yusupov, M.M., Yusupova, G.Zh., Baucom, A., Lieberman, K., Earnest, T.N., Cate, J.H.D., and Noller, H.F. (2001). Crystal structure of the ribosome at 5.5 Å resolution. *Science* **292**, 883-896.

Zhang, L., and Doudna, J.A. (2002). Structural insights into group II intron catalysis and branch-site selection. *Science* **295**, 2084-2088.

Zhang, D., and Rosbash, M. (1999). Identification of eight proteins that crosslink to pre-mRNA in the yeast commitment complex. *Genes Dev.* **13**, 581-592.

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

Zorio, D.A., and Blumenthal, T. (1999). Both subunits of U2AF recognize the 3' splice site in *Caenorhabditis elegans*. *Nature* **402**, 835-838.

Chapter 1
The essential spliceosomal protein Prp8

Chapter 1: The essential spliceosomal protein Prp8

1-1. Introduction

1-1.1. Discovery of Prp8

Prp8 was first identified in a screen for temperature sensitive (ts) mutations in *S. cerevisiae*. Ten complementation groups, *rna2-rna11*, of mutations were identified that caused accumulation of RNA following a shift from the permissive (23°C) to the restrictive (36°C) temperature (Hartwell 1967; Hartwell et al., 1970). Many of these RNA genes were later renamed PRP genes, to indicate their involvement with pre-mRNA processing (Vijayraghavan et al., 1989). At the restrictive temperature the *rna8-1* (now *prp8-1*) cells displayed accumulation of RPL59 (ribosomal protein 59) pre-mRNA (Larkin and Woolford, 1983). As well, extract from *prp8-1* cells was defective for splicing actin pre-mRNA and was unable to support formation of active spliceosomes. Splicing was restored upon addition of extracts from other *rna* mutants (Lustig et al., 1986). These results identified the PRP8 gene product as an essential splicing factor involved early in the splicing process.

In mammalian nuclear extract, UV-crosslinking of protein to body-labelled adenovirus pre-mRNA identified a 220 kDa protein that binds pre-mRNAs under splicing conditions (Garcia-Blanco et al., 1990). The cross-link was specific as determined by competition assays with unlabelled RNA, and was sensitive to pre-mRNA mutations that impair splicing. Specifically, the cross-link was sensitive to both a point mutation at the 3' splice site (3'SS) and to multiple point mutants in the poly-pyrimidine tract (PPT). Finally, rabbit antibodies

against yeast Prp8 were able to immunoprecipitate the UV-crosslinked 220 kDa protein, suggesting that the protein is the functional human homologue of Prp8 in yeast. These experiments marked the discovery of Prp8 in mammalian extract (Garcia-Blanco et al., 1990).

For the remainder of this thesis, the human protein will be distinguished from the yeast (*S. cerevisiae*) protein by hPrp8 or yPrp8 where appropriate. Please note that hPrp8 is often referred to as p220 in the literature. If a statement is equally relevant to both homologues, Prp8 will be used. This chapter provides an overview of the extensive research efforts that have gone into the biochemical and structural characterization of the essential splicing protein, Prp8.

1-1.2. Domain structure of Prp8

With 61% sequence identity between yeast and humans, Prp8 is one of the most highly conserved nuclear proteins known. Prp8 is also among the largest spliceosomal factors with most orthologs being around 2400 amino acids in length. Prp8's high conservation in both sequence and size underscores its importance to spliceosome function. A key goal of researchers is elucidation of the precise role of Prp8 in the spliceosome by high resolution structural analysis. The large size of Prp8 and its insolubility hinder progress in this regard. Instead, structural understanding of Prp8 can be gained by investigating domains in isolation; however, with few structural domains predicted by primary sequence analysis there is only limited information as to what domain boundaries are and

few hints about the functional relevance of putative domains (Grainger and Beggs, 2005).

1-1.3. A conserved RRM

Grainger and Beggs (2005) offered the first report of an RRM conserved among all Prp8 orthologs (Figure 1-1). Genetic interactions between the RRM of yPrp8 and both U4 snRNA and Brr2, the RNA helicase implicated in U4/U6 unwinding, have been identified. Specifically, mutations that reside mostly in the β -sheet region of the RRM are able to suppress the *U4-cs1* (U4-cold sensitive 1) and/or *brr2-1* mutations, both of which affect U4/U6 unwinding (Kuhn et al. 1999, 2002; Kuhn and Brow 2000). The Prp8 RRM is unique among other RRMs in a stretch of 3 amino acids, KDM, which is the location of three mutations that suppress *U4-cs1*. Also of interest is the location of the KDM tripeptide in a loop between β -strands 2 and 3, a region important for determining specificity of RNA-binding. It is not clear whether the Prp8 RRM directly contacts RNA, or exerts its effect on U4/U6 unwinding through another mechanism. For example, it is possible that the observed genetic interactions reflect an indirect role for Prp8 whereby the activity of another protein involved in U4/U6 unwinding is modulated. Consistent with this, there are examples of RRMs that interact with other protein molecules instead of RNA (Kielkopf et al. 2004). Thus, the classification of Prp8's RRM as an RNA interaction domain, a protein interaction domain, or a combination of both, for example p14 (see appendix III) (Schellenberg et al., 2006), awaits further study.

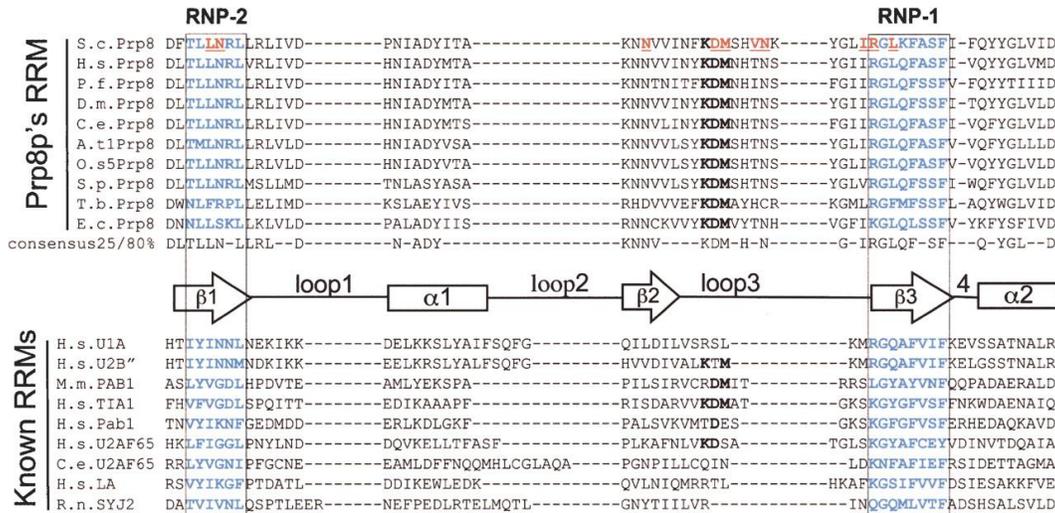


Figure 1-1. Prp8 contains a conserved RRM. The alignment between 10 different Prp8 sequences and nine known RRM domains is shown. The conserved RNP-1 and RNP-2 motifs are boxed and highlighted in bold blue type. A number of mutations are known in yPrp8 that suppress *brr2-1* and/or *U4-cs1* mutations and these are highlighted by bold red underlined text. A unique feature of Prp8's RRM is the conserved amino acid triplet KDM in the center of the domain. The RRM secondary structure based on the U1A crystal structure is shown between the primary sequences. (S.c.) *Saccharomyces cerevisiae*; (H.s.) *Homo sapiens*; (P.f.) *Plasmodium falciparum*; (D.m.) *Drosophila melanogaster*; (C.e.) *Caenorhabditis elegans*; (A.t1.) *Arabidopsis thaliana* chromosome 1; (O.s5) *Oryza sativa* chromosome 5; (S.p.) *Schizosaccharomyces pombe*; (T.b.) *Trypanosoma brucei*; (E.c.) *Encephalitozoon cuniculi*. (Adapted from Grainger and Beggs, 2005).

1-1.4. Domain 3 (3'SS fidelity domain)

Domain 3 (1372-1660 in yeast) was originally defined by a cluster of amino acids identified in a screen for suppressors of 3'SS mutations (Umen and Guthrie, 1996). Further analysis has revealed a number of *yprp8* alleles in domain 3 that suppress not only mutations at the 3'SS but also mutations at both the 5'SS and branch point (see below) (Collins and Guthrie 1999; Siatecka et al., 1999; Ben-Yehuda et al., 2000; Dagher and Fu 2001; Query and Konarska 2004). Despite being defined in terms that don't truly reflect the complexity of its function, domain 3 is a very interesting target of study. Grainger and Beggs have split domain 3 into two regions, the more C-terminal domain 3.2 (1547-1660 in yeast)

being most highly conserved (72% identity between yeast and human) and containing all but two of the pre-mRNA suppressor alleles. Despite high sequence conservation and clear functional importance, there are no clues in the primary sequence as to the 3-dimensional structure of domain 3.

Genetic interactions have also implicated domain 3.2 as being involved in an intramolecular interaction with a more N-terminal region of yPrp8 (Kuhn and Brow, 2000). Specifically, a slow growth phenotype caused by a H659P mutation is suppressed by the domain 3.2 mutation L1634F. Interestingly, the sequences around these mutations consist of a pattern of leucine and isoleucine residues suggestive of a leucine zipper motif (Landschulz et al., 1988). Secondary structure predictions suggest that the two regions of yPrp8 fold into α -helices with the hydrophobic residues clustered on one face in both cases. When the helices are modeled anti-parallel to each other to form a coiled-coil structure, H659 and L1634 are in close proximity consistent with the possibility that their genetic interaction reflects a physical one. The slow growth phenotype could be explained by H659P causing a local distortion of α -helical structure, while suppression caused by replacement of L1634 with Phe is a result of more hydrophobic contacts available to stabilize the coiled-coil structure (Kuhn and Brow, 2000).

1-1.5. Domain IV (5'SS fidelity domain)

Similar to domain 3, domain IV was originally defined by the observed clustering of yPrp8 alleles that suppress pre-mRNA mutations (Table 1-1) (Grainger and Beggs, 2005). In addition, Konarska and coworkers have proteolytically mapped

a cross-link formed between the 5'SS and hPrp8 during spliceosome assembly to a short stretch of amino acids in domain IV (discussed below), making it an interesting target to study. The stretch of five amino acids QACLK (h1894–1898) in hPrp8 that cross-link to the 5'SS are not conserved in yeast, the homologous sequence being SAAMS (y1966–1970) (Reyes et al., 1999). Unpublished work by Konarska and colleagues indicates that hPrp8 chimeras containing the SAAMS sequence can still cross-link to the 5'SS, indicating that the 5 amino acids identified are not essential for the cross-link, but are nearby. Thus, it is likely that other residues in domain IV are responsible for interacting with the 5'SS (Grainger and Beggs, 2005).

In addition to contacting the 5'SS, yeast alleles identified in screens for factors affecting poly-pyrimidine tract recognition (discussed below) are located in domain IV (Table 1-1) (Umen and Guthrie, 1995a). One of these same alleles, *prp8-101* (E1960K), also causes reduced cross-linking efficiency to the 3'SS (Umen and Guthrie, 1995b). The same mutation is synthetically lethal with mutations in the second step splicing factors Prp16, Prp17, Prp18, and Slu7 (Umen and Guthrie, 1995b; James et al. 2002). Clearly, an understanding of how mutations in domains IV and 3 can suppress RNA mutations that block both spliceosome assembly and catalysis is important for understanding Prp8 function. A powerful framework for understanding these affects could be provided by high resolution structural studies, but there are no clues in the primary sequence as to the three-dimensional structures of these domains. Our structural and biochemical studies of domain IV are the basis of chapters 2 and 3.

Table 1-1. Mutant yeast alleles mapping to Prp8 domain IV core and corresponding phenotypes⁽¹⁾.

Yeast Allele (Human)	Yeast Phenotype
V1860N (V1788)	suppresses U4cs ²
V1860D (V1788)	suppresses U4cs, synthetic lethal U6 U-A ³
T1861P (T1789)	suppresses U4cs, synthetic lethal U6 U-A, suppresses splice site mutations, synthetic lethal Prp28 Δ ⁴
V1862D (I1790)	suppresses U4cs, synthetic lethal U6 U-A
V1862A/Y (I1790)	suppresses U4cs
K1864E (K1792)	suppresses splice site mutations
N1869D (N1797)	suppresses splice site mutations
V1870N (L1798)	suppresses splice site mutations
I1875T (I1803)	suppresses U4cs
N1883D (N1811)	synthetic lethal Prp28 Δ
E1960K (E1888)	alternate 3' splice site choice
E1960G (E1888)	alternate 3' splice site choice
T1982A (T1910)	suppresses splice site mutations

[1] As reviewed in (Grainger and Beggs, 2005); [2] U4 cs (cold-sensitive) results in a hyperstabilized U4-U6; [3] U6 U-A corresponds to mutation of U6 at the base of the internal stem loop that stabilizes its catalytic conformation; [4] Prp28 is the U5 snRNP component required for resolving the 5'SS·U1 snRNA interaction.

1-1.6. C-terminal MPN domain

Prp8 contains an MPN domain (also known as Mov34, JAB, MPN+, PAD-1, or JAMM domain) at its C-terminus. Other proteins have been identified with MPN domains at their N-terminus, including proteasome regulatory subunits, eukaryotic initiation factor 3 (eIF3) subunits, the signalosome, and regulators of transcription factors. The typical model of an MPN domain is the proteasome component Rpn11. Crystal structures of MPN domains have been solved and reveal a zinc binding site consisting of 5 conserved coordinating ligands (Tran et al., 2003; Ambroggio et al., 2004). The Zn²⁺-coordinating residues include a conserved glutamate and a JAMM motif with the consensus H-x-H-x[7]-S-xx-D. The corresponding sequence in Prp8 is different, consisting of a glutamine instead of the conserved glutamate and a consensus H-x-Q-x[7]-S-xx-D. This indicates

that Prp8's MPN domain would likely not coordinate metal, which has been verified by high resolution structural analysis (Pena et al., 2007; Zhang et al., 2007). The first clues for function of Prp8's MPN domain came from work by Sontheimer and colleagues. Namely, an *in vitro* interaction between ubiquitin and a yPrp8 fragment containing the MPN domain with affinity similar to that seen for ubiquitin-binding domains in other proteins was observed (Bellare et al., 2006).

The physiological relevance of the interaction between the C-terminus of yPrp8 and ubiquitin has begun to be elucidated. Work by Sontheimer and colleagues has indicated a role for ubiquitin in the maintenance of U5·U4/U6 tri-snRNP levels. Extract prepared from yeast expressing triple-HA tagged Prp8 was incubated with either WT ubiquitin or the mutant I44A, predicted to disrupt ubiquitin-protein interactions. Reverse transcription reactions were performed on anti-HA immunoprecipitates to determine how the levels of snRNAs are affected by mutant ubiquitin. Only U5 snRNA was detected after incubation with mutant ubiquitin, indicative of a tri-snRNP assembly defect. The reduction in tri-snRNP levels is ATP-dependent suggesting that modulation of a phase of cycling between the tri-snRNP and dissociable components that requires energy is responsible for the effect of I44A ubiquitin. Further experiments by Sontheimer's group indicate that the reduction in tri-snRNP levels after incubation with I44A is caused by accelerated U4/U6 snRNA unwinding. Prp8 is the most likely target for ubiquitin-conjugation (Bellare et al., 2006), and it possible that an ubiquitin-conjugated form of yPrp8 is important for repression of U4/U6 unwinding. Consistent with a role for yPrp8 in modulating U4/U6 unwinding (Kuhn et al.,

1999; van Nues and Beggs, 2001) (see below) an effect on intermolecular interactions (perhaps with Brr2) involving ubiquitin-conjugated γ Prp8 might control the timing of U4/U6 unwinding (Bellare et al., 2008).

1-2. Prp8 and its interactions

Prp8 is a component of both the U5 snRNP (Stevens et al., 2001) and the U5·U4/U6 tri-snRNP (Gottschalk et al., 1999; Stevens and Abelson, 1999; Schneider et al., 2002). The interactions between Prp8 and other components of these snRNPs have been extensively characterized and are discussed in later chapters of this thesis. This section will broadly cover the observed interactions between Prp8 and other protein factors, the identities of which suggest a role for Prp8 in various stages of spliceosome assembly and activation. As well, Prp8 has been identified in association with non-slicing factors, pointing to possible functions outside of the spliceosome.

1-2.1. Prp8 and kinases

hPrp8 has been purified as a component of a complex containing the SR protein kinase, SRPK1 (Lee et al., 2004). This complex, which was termed the toposome, also contained a protein known to be involved in unlinking mitotic chromosomes, topoisomerase II α (TOP II α). The association of hPrp8 and TOP II α suggests both an alternative role for hPrp8 in mitosis (Lee et al., 2004) and for TOP II in splicing (Ajuh et al. 2000).

SRPK1 is the only SR protein kinase family member that is conserved between humans and *S. cerevisiae* (Sky1 in yeast). The RNA-binding protein with a role in mRNA export, Npl3, has been identified as a substrate for Sky1 in yeast (Siebel et al., 1999), and there is genetic data consistent with a role in Sky-regulated phosphorylation affecting yPrp8's role in 3'SS recognition. For example, deletion of *SKY1* is not lethal by itself, but synthetic lethality has been observed between *sky1Δ* and three *yprp8* alleles in domain 3. These *yprp8* alleles are able to suppress 3'SS mutations (Dagher and Fu 2001), and *sky1Δ* is able to suppress the same 3'SS mutations. *SKY1* deletion is also synthetic lethal with deletion of the second step splicing factor Prp17, as is the combination of *yprp17Δ* and one of the domain 3 mutations. These genetic interactions suggest that yPrp8 domain 3 and Sky1 function in a common pathway of 3'SS recognition during the second step of splicing. It is therefore tempting to speculate that Prp8's role in 3'SS recognition is regulated by phosphorylation of domain 3; however, the absence of a good consensus sequence for phosphorylation by the SR protein kinase family in domain 3 (Wang et al., 1998) suggests any effect is likely to be indirect.

There is also a reported link between *S. pombe* Prp8 and the kinase Prp4. Specifically, a mutation in the *S. pombe* Prp8 MPN domain suppresses a ts mutation of Prp4 kinase, which could indicate a role for Prp4 kinase in controlling the assembly of the catalytic spliceosome (Schmidt et al., 1999). The human ortholog of Prp4 kinase, PRP4K, is known to associate with U5 snRNP proteins as well as with chromatin remodelling factors (Dellaire et al., 2002) (see below).

Thus, there is evidence that Prp8 can be affected by the activity of kinases, but the functional significance of any interactions remains unknown.

1-2.2. Prp8, transcription and chromatin remodelling factors

hPrp8 has been co-purified with a component of the conserved SWI/SNF remodelling complex, BRG1, as has PRP4K. PRP4K has also been observed in direct association with the human U5 snRNP proteins Prp8, Brr2, and Prp6. The association of PRP4K with both BRG1 and splicing factors suggests a role for PRP4K in coupling transcriptional regulation to splicing via the modulation of chromatin structure (Dellaire et al., 2002).

Consistent with a more specific role for hPrp8 in transcriptional activation, it has been identified as a component in complexes isolated by purification of a single component carrying a recombinant tag such as glutathione S-transferase (GST). Specifically, SKIP (alternatively known as Nuclear coactivator-62 kDa (NCoA62)) (yeast Prp45) is a transcriptional co-activator that interacts with the vitamin D receptor (VDR) and acts as a remodelling protein upon recruitment to vitamin D-responsive elements in promoters of target genes (Baudino et al., 1998; Zhang et al., 2001, 2003; Barry et al., 2003). Adding GST-SKIP to HeLa extract allowed the purification of a complex containing hPrp8, PSF (poly-pyrimidine tract-binding protein-associated splicing factor), Prp28, Brr2, and Snu114 (Figure 1-2). Thus, there is evidence to suggest the association of hPrp8, most likely indirect, with nuclear receptor promoter regions of the vitamin D response element (Zhang et al., 2003). These observations were corroborated by the

copurification of SKIP with catalytic spliceosomes and with the U5/Prp19 complex (Makarov et al., 2002), making SKIP a leading candidate for a link between VDR-mediated transcription and pre-mRNA splicing. The other spliceosomal proteins implicated in interactions with the vitamin D response element were also present in the Makarov complex: hPrp8, PSF, Brr2, and Snu114 (Figure 1-2). In addition, the yeast homolog of SKIP, Prp45, has a known role in splicing (Albers et al. 2003).

There is growing evidence that pre-mRNA splicing is coupled to RNA polymerase II (RNAP II) transcription. In particular, the carboxy-terminal domain (CTD) of RNA polymerase II is located directly adjacent to the exit groove for the growing pre-mRNA (Cramer et al., 2001) making it an ideal spot for the accumulation of factors that affect processing of the nascent transcript. Interestingly, antibodies specific for the hyperphosphorylated CTD can immunoprecipitate both U5 snRNP and U5·U4/U6 tri-snRNP, indicating a stable interaction with RNA polymerase II that includes hPrp8 (Chabot et al., 1995; Vincent et al., 1996). Later analysis of purified transcription complexes revealed the presence of RNAPII, PSF, p54^{nrb}, hPrp8, and all five UsnRNAs, suggesting that the entire spliceosome can be associated with elongating polymerase (Figure 1-2) (Kameoka et al., 2004). PSF and p54^{nrb} are multifunctional proteins which, in addition to interacting with the CTD of RNAPII, interact with each other and with U5 snRNA stem 1b (Peng et al., 2002). Stem 1b of U5 snRNA has also been demonstrated to interact with yPrp8 (Dix et al., 1998).

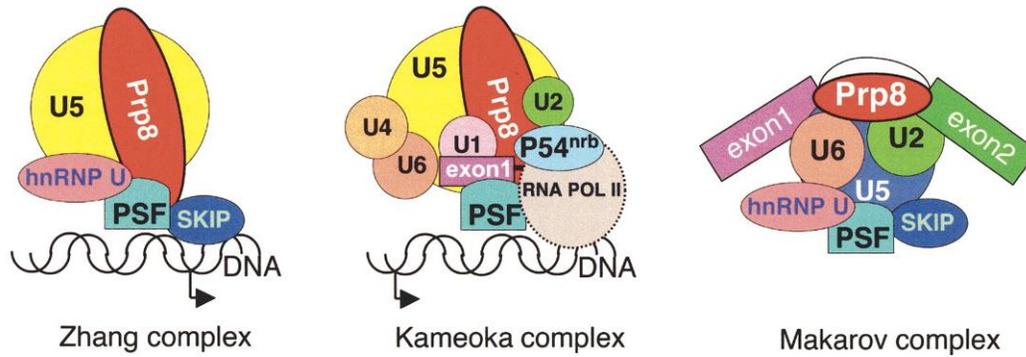


Figure 1-2. hPrp8 has been identified in several distinct complexes. Three purified hPrp8-containing protein complexes: the activation complex present at vitamin D-responsive elements (Zhang et al., 2003), the transcription elongation complex (Kameoka et al., 2004) and the spliceosome (Makarov et al., 2002) are shown. (Adapted from Grainger and Beggs, 2005).

1-2.3. *Prp8 in a U5 snRNP assembly intermediate*

Several lines of evidence indicate a physical interaction between the U1 and U5 snRNPs. In particular, a yeast U1/U5 snRNP complex has been identified which could indicate a collaboration between the two snRNPs early in spliceosome assembly (Neubauer et al., 1997; Ruby, 1997). Two interesting features of the U1/U5 snRNP complex are both the absence of Brr2 and the presence of a distinctive protein, Aar2. The Aar2 protein was originally characterized as having a role in splicing of *MATa1* pre-mRNA, but not in splicing of actin pre-mRNA (Nakazawa et al., 1991). Genetic depletion of Aar2 leads to defects in splicing of additional precursors, U3A and U3B (Gottschalk et al., 2001). Aar2 is not part of the U5·U4/U6 tri-snRNP or the spliceosome, consistent with a possible role in snRNP biogenesis. The different biochemical compositions between the U1/U5 snRNP and the U5 snRNP, in particular the absence of Brr2, suggest that the U1/U5 complex represents a step in the U5 snRNP assembly pathway. The Aar2-U5 snRNP contains yPrp8, Snu114, and the Sm proteins, indicating that these

factors may be sufficient to provide a scaffold for later association of additional snRNP factors (Gottschalk et al., 2001).

Consistent with an interaction between the U5 and U1 snRNPs, an interaction between the N terminus of yPrp8 (1-349) and the U1 snRNP protein Prp40 has been demonstrated (Abovich and Rosbash, 1997; Gottschalk et al., 1998). In addition, two other U1 snRNP proteins have been observed to interact with pieces of yPrp8 by yeast-two-hybrid assays. Specifically, Prp39 interacts with an N-terminal fragment of yPrp8 (1-263) (van Nues and Beggs, 2001) while Snp1, the yeast homolog of human U1-70K, interacts with a 28 amino acid region of yPrp8 (1166-1193). The latter interaction was confirmed by co-immunoprecipitation (Awasthi et al., 2001).

1-3. A protein factor in the spliceosome active site

The U5 snRNP protein Prp8 has long been known to be intimately associated with key components of the spliceosome including the pre-mRNA substrate and snRNAs present at the active site (Collins and Guthrie, 2000). A large number of mutant alleles of Prp8 have been characterized in yeast and the variety of associated phenotypes attest to the central role of this factor in spliceosome assembly and possibly catalysis (see table 1-1 for a summary of alleles located in domain IV) (Grainger and Beggs, 2005). These include suppression of the effects of hyper-stabilized U4/U6 as well as suppression of splicing defects due to mutations at both splice sites and the branch point sequence (Query and Konarska, 2004; Kuhn et al., 1999; Collins and Guthrie, 1999). A summary of both cross-

linking data and genetic interactions are offered below, and implications for Prp8's role in affecting the spliceosome active site are offered.

1-3.1. Cross-linking Prp8 to pre-mRNA

Cross-linking studies in yeast and HeLa nuclear extract revealed direct contact between Prp8 and the 5'SS, 3'SS, and branch point sequences (Maroney et al., 2000; Reyes et al., 1999; MacMillan et al., 1994; Teigelkamp et al., 1995). The Prp8/5'SS interaction occurs before the first step of splicing in B complex (Reyes et al., 1999; Maroney et al., 2000). In contrast, interaction with the 3'SS occurs concomitant with or subsequent to step 1 (Teigelkamp et al., 1995). As Prp8 is the only spliceosomal protein that directly contacts all reactive sites of the intron, it is the leading candidate for a protein cofactor directly contributing to splicing catalysis (Collins and Guthrie, 2000).

The region of hPrp8 involved in cross-linking the 5'SS has been mapped by proteolytic methods (Reyes et al., 1999; Turner et al., 2006). A short oligo containing a consensus 5'SS sequence A₅G/GUAAGUAdTdC₃ (/ denotes 5'SS) that supports trans-splicing in HeLa extract was used in cross-linking experiments (Reyes et al., 1999). The cross-link to hPrp8 is very specific as substitutions of the highly conserved U at position + 2 in the intron with either T or 5-IodoU abolishes the cross-link, which is correlated with a decrease in splicing. Moreover, mutagenic experiments on the 5'SS sequence confirm that the cross-link only occurs in the context of a 5'SS sequence that is capable of making crucial interactions with other spliceosomal components, in particular U6 snRNA.

The hPrp8/5'SS cross-link was mapped using proteolytic techniques to amino acids 1894-1898 (Reyes et al., 1999).

Interestingly, more recent cross-linking experiments carried out in yeast followed by proteolytic mapping suggest different regions of the protein are responsible for interaction with the 5'SS than previously identified (Turner et al., 2006). By randomly inserting the tobacco etch virus (TEV) protease site into the yPRP8 gene it is possible to proteolytically map cross-links between yPrp8 and spliceosomal RNA with TEV protease and thereby identify discrete regions of yPrp8 responsible for interacting with specific spliceosomal RNAs. By this method the 5'SS-interacting regions of yPrp8 were identified as three non-contiguous regions comprising amino acids 871-970, 1281-1413, and 1503-1673, which are all quite N-terminal to the five amino acids previously implicated in 5'SS cross-linking in the human system (Turner et al., 2006). The difference may be explained by cross-linking being carried out in spliceosomal complexes at different stages of assembly. In fact, Konarska's cross-link likely reflects an interaction present in B complex (Reyes et al., 1999) while Newman's cross-link is more indicative of interactions occurring in the catalytic C complex (Turner et al., 2006). Interestingly, results of cross-link mapping directed from the BS overlapped exactly with those from the 5'SS (Turner et al., 2006), consistent with cross-linking being carried out in a complex where the 5'SS and BS have already been juxtaposed.

1-3.2. Cross-linking Prp8 to snRNA

In addition to directly contacting all three sites involved in splicing chemistry, cross-linking experiments have identified direct interactions between Prp8 and both U5 and U6 snRNAs. Early cross-linking studies identified the U5 and U6 snRNA components at the spliceosome active site, implicating them as directly affecting catalysis (Sontheimer and Steitz, 1993). More specifically, an invariant U-rich loop of U5 snRNA (loop I) interacts with the 5' exon, while the invariant ACAGA sequence of U6 snRNA interacts with the 5'SS region. As the U5 and U6 snRNAs are strongly implicated as directors of splicing catalysis there is significant interest in identifying factors that interact with U5 and U6.

Cross-linkers have also been site-specifically introduced into U5 and U6 snRNAs in both yeast and HeLa cells (Dix et al., 1998; Urlaub et al., 2000; Vidal et al., 1999). Interactions with Prp8 were observed when cross-linkers were introduced into the invariant loop I of U5 snRNA in the context of yeast U5 snRNP or human tri-snRNP (Dix et al., 1998; Urlaub et al., 2000). The invariant loop of U5 snRNA is believed to interact with the 5' exon during catalysis (Sontheimer and Steitz, 1993). The similarities in cross-linking between U5 snRNA and Prp8 for yeast and human systems suggest that not only is Prp8 highly conserved, but also its interactions with U5 snRNA (Urlaub et al., 2000). Similar experiments have been done identifying interactions between yPrp8 and U6 snRNA (Vidal et al., 1999). Introducing a cross-linker at U6-U54 revealed an interaction with yPrp8 in the context of purified U5·U4/U6 tri snRNP.

Similar to proteolytic mapping of the 5'SS cross-link using randomly inserted TEV protease recognition sequences in the yPRP8 open reading frame (ORF), a cross-link to both the U6 snRNA and U5 snRNA have been mapped in the context of purified U5·U4/U6 tri-snRNPs (Turner et al., 2006). Cross-linking was directed from the invariant U-rich loop of U5 snRNA that interacts with the 5' exon. Cross-links to U5 snRNA mapped to yPrp8 amino acids 770-871 as well as 1281-1413. The cross-link to U6-U54 was similarly mapped, implicating amino acids 1503-1673 as directing the yPrp8 interaction with U6 snRNA. In addition to providing a basis for dissecting yPrp8 in terms of RNA-interacting regions, the identities of putative functional domains can be inferred by large stretches of amino acids that don't tolerate insertion of TEV protease recognition sequences.

1-3.3. Genetic interactions: spliceosomal RNA and prp8 alleles

In addition to physically contacting all 3 sites of splicing chemistry, genetic studies in yeast have identified *prp8* alleles that suppress splicing defects caused by mutations at the 5'SS, 3'SS, or BS (Collins and Guthrie, 1999; Siatecka et al., 1999; Query and Konarska, 2004). It is interesting that alleles identified as suppressors of BS mutations are also able to suppress mutations at either the 5'SS or 3'SS (Query and Konarska, 2004). Similarly, alleles previously identified as suppressors of splice site mutations (Collins and Guthrie, 1999; Siatecka et al., 1999) are able to suppress branch site mutations, arguing that the observed phenotype reflects disruption of an interaction with indirect effects on splicing

catalysis rather than disruption of direct interactions with the pre-mRNA (Query and Konarska, 2004).

Consistent with a role for Prp8 in interacting with pre-mRNA during splicing catalysis, an allele of yPrp8, *prp8-101*, was identified as causing abnormal recognition of the 3'SS. The pre-mRNA construct used performs *cis* competition of 3'SS sequences, a U-rich 3'SS used upon proper PPT recognition, and an A-rich 3'SS. Utilization of the incorrect A-rich 3'SS results in an in-frame CUP1 reporter gene leading to a growth phenotype that denotes increased copper-resistance (Umen and Guthrie, 1995a). Interestingly, primer extension of RNA from yeast strains carrying *prp8-101* as their sole copy of Prp8 indicates an accumulation of lariat intermediate for pre-mRNA substrates with BS mutations, indicating improvement of the first step of splicing (Liu et al., 2007). The *prp8-101* allele is distinct from splice and branch site suppressors as they act by improving the second step of splicing (Query and Konarska, 2004).

In addition to the physical interaction with U6U54 observed by cross-linking, genetic interactions have been observed between yPrp8 and U4/U6 snRNAs (Kuhn et al., 1999) which are base-paired in the U4/U6·U5 tri-snRNP prior to spliceosome activation. Genetic studies have utilized a *cs* mutant of U4 snRNA, *U4-cs1*, blocking catalytic activation but not assembly of the spliceosome at low temperature. This particular mutation is a triple nucleotide substitution extending the base-pairing interaction of U4/U6 and masking the invariant ACAGAGA sequence of U6 making it unavailable for interaction with the 5'SS in the activated spliceosome. It is suggested U4/U6 unwinding is inhibited when the

U6/5'SS interaction is not formed. Furthermore, observed suppression of *U4-cs1* by a *yprp8* allele led to the proposal Prp8 proofreads recognition of the 5'SS by the U6 ACAGAGA-box (Kuhn et al., 1999). Consistent with a role for Prp8 in formation of the activated RNA structures, genetic observations have been observed between yPrp8 and the U1/5'SS unwindase Prp28. Specifically, the *prp-28* mutation that displays a *cs* phenotype, as well as enhances the *U4-cs1* phenotype, is suppressed by a *yprp8* allele near the N-terminus, L280P, that also suppresses *U4-cs1* (Kuhn et al., 2002).

A search for suppressors of *U4-cs1* identified over 40 mutations in yPrp8, *prp8-cat* mutations (*catalytic activation*), containing single amino acid substitutions mapping to 5 discreet regions a-e (Kuhn and Brow, 2000). The *U4-cs1* suppression phenotype implicates these residues as positive effectors of catalytic activation. A mutation of U6 snRNA, *U6-UA*, causes a slow growth phenotype itself and rescues *U4-cs1* lethality at 20°C, likely by stabilizing the free form of U6 snRNA (Figure 1-3). Interestingly, the *U6-UA* mutation is synthetic lethal with three adjacent residues at the C-terminus of region e in yPrp8, but does not interact genetically with other mutations in regions a-e (Kuhn et al., 2002). The observed genetic interaction between three amino acids 1860-1862 and the *U6-UA* mutation that hyperstabilizes the U6 ISL formed upon U4/U6 unwinding suggests a specific interaction, likely important for activation of the first step catalytic centre.

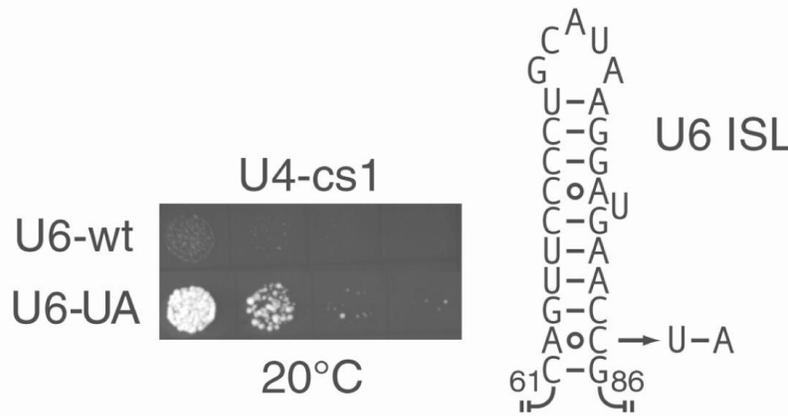


Figure 1-3. Genetic interaction between the *U4-cs1* mutation and conditional mutations in U6 snRNA involved in catalytic activation. The *U6-UA* mutation suppresses the cold-sensitive growth defect of *U4-cs1* at 20°C. (Right) A schematic of the ISL of U6 RNA. The *U6-UA* mutation is indicated with an arrow. (Adapted from Kuhn et al., 2002).

The yeast allele *prp8-21* (amino acid not defined) was identified in a screen for factors that are synthetic lethal with an 11 nt substitution in U2 snRNA stem I that is predicted to affect the formation of U2/U6 structures present in the activated spliceosome (Xu et al., 1998). In a WT background, *prp8-21* confers a ts phenotype. The same yPrp8 allele is also synthetic lethal with mutations in the conserved loop I of U5 snRNA, and causes a severe growth defect in the presence of *prp16-1* (Frank et al., 1992; Xu et al., 1998). Both U5 loop I and Prp16, an RNA-dependent ATPase, function to promote the second step of splicing by tethering the exons and by hydrolyzing ATP to remodel the spliceosome active site after the first step, respectively.

Finally, interactions between *prp8*, *prp16*, and U6 snRNA alleles point to a cooperative role between these three factors in facilitating the rearrangement that takes place between the two steps of splicing. Specifically, the combination of mutant pre-mRNAs inefficient for both steps of splicing with *prp8* alleles which improve the second step of splicing and *prp16* or U6 snRNA alleles that

improve the first step of splicing results in suppression of both steps (Query and Konarska, 2004). The model proposed by Query and Konarska suggests that the first and second steps of the spliceosome require two different active site conformations, which are in kinetic competition (Figure 1-4). It is not clear exactly how Prp8 would modulate the transition between the two steps, but the genetic data are consistent with Prp8 affecting multiple steps of spliceosome assembly through physical interactions with either snRNAs or multiple ATPases. Thus, there are a number of complementary observations from both cross-linking and genetic experiments that are consistent with multiple roles for Prp8 in structuring of the spliceosome active site (Grainger and Beggs, 2005).

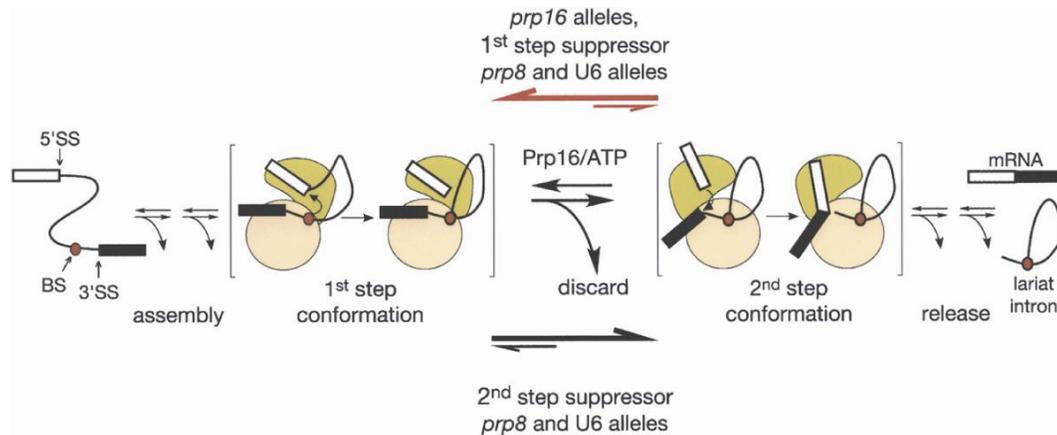


Figure 1-4. Scheme for progression of the pre-mRNA splicing process, highlighting the Prp16-dependent transition between the two catalytic steps. The first and second catalytic steps require different conformations of the spliceosome, the equilibrium between these conformations being modulated by interactions of Prp16, yPrp8, and U6 snRNA. As indicated by the black and red arrows, certain alleles of these factors improve the first step and inhibit the second, and other alleles act oppositely. The Prp16 ATPase facilitates the transition between the first and second steps and, as a result, provides an opportunity for rejection of substrates that do not efficiently proceed to the second step (i.e., discard). (Adapted from Konarska and Query, 2005).

1-4. References.

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

Ambroggio, X.I., Rees, D.C., and Deshaies, R.J. (2004). JAMM: A metalloprotease-like zinc site in the proteasome and signalosome. *PLoS Biol.* **2**, 0113–0119.

Ajuh, P., Kuster, B., Panov, K., Zomerdijk, J.C., Mann, M., and Lamond, A.I. (2000). Functional analysis of the human CDC5L complex and identification of its components by mass spectrometry. *EMBO J.* **19**, 6569–6581.

Albers, M., Diment, A., Muraru, M., Russell, C.S., and Beggs, J.D. (2003). Identification and characterization of Prp45p and Prp46p, essential pre-mRNA splicing factors. *RNA* **9**, 138–150.

Awasthi, S., Palmer, R., Castro, M., Mobarak, C.D., and Ruby, S.W. (2001). New roles for the Snp1 and Exo84 proteins in yeast premRNA splicing. *J. Biol. Chem.* **276**, 31004–31015.

Barry, J.B., Leong, G.M., Church, W.B., Issa, L.L., Eisman, J.A., and Gardiner, E.M. (2003). Interactions of SKIP/NCoA-62, TFIIB, and retinoid X receptor with vitamin D receptor helix H10 residues. *J. Biol. Chem.* **278**, 8224–8228.

Baudino, T.A., Kraichely, D.M., Jefcoat Jr., S.C., Winchester, S.K., Partridge, N.C., and MacDonald, P.N. (1998). Isolation and characterization of a novel coactivator protein, NCoA-62, involved in vitamin D-mediated transcription. *J. Biol. Chem.* **273**, 16434–16441.

Bellare, P., Kutach, A.K., Rines, A.K., Guthrie, C., and Sontheimer, E.J. (2006). Ubiquitin binding by a variant Jab1/MPN domain in the essential pre-mRNA splicing factor Prp8p. *RNA* **12**, 292–302.

Bellare, P., Small, E.C., Huang, X., Wohlschlegel, J.A., Staley, J.P., and Sontheimer, E.J. (2008). A role for ubiquitin in the spliceosome assembly pathway. *Nat. Struct. Mol. Biol.* **15**, 444–451.

Ben-Yehuda, S., Russell, C.S., Dix, I., Beggs, J.D., and Kupiec, M. (2000). Extensive genetic interactions between PRP8 and PRP17/CDC40, two yeast genes involved in pre-mRNA splicing and cell cycle progression. *Genetics* **154**, 61–71.

Chabot, B., Bisotto, S., and Vincent, M. (1995). The nuclear matrix phosphoprotein p255 associates with splicing complexes as part of the [U4/U6.U5] tri-snRNP particle. *Nucleic Acids Res.* **23**, 3206–3213.

Collins, C.A., and Guthrie, C. (1999). Allele-specific genetic interactions between Prp8 and RNA active site residues suggest a function for Prp8 at the catalytic core of the spliceosome. *Genes Dev.* **13**, 1970–1982.

Collins, C.A., and Guthrie, C. (2000). The question remains: Is the spliceosome a ribozyme? *Nat. Struct. Biol.* **7**, 850–854.

Cramer, P., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 Å resolution. *Science* **292**, 1863–1876.

Dagher, S.F., and Fu, X.D. (2001). Evidence for a role of Sky1p-mediated phosphorylation in 3' splice site recognition involving both Prp8 and Prp17/Slu4. *RNA* **7**, 1284–1297.

Dellaire, G., Makarov, E.M., Cowger, J.J., Longman, D., Sutherland, H.G., Luhrmann, R., Torchia, J., and Bickmore, W.A. (2002). Mammalian PRP4 kinase copurifies and interacts with components of both the U5 snRNP and the N-CoR deacetylase complexes. *Mol. Cell. Biol.* **22**, 5141–5156.

Dix, I., Russell, C.S., O'Keefe, R.T., Newman, A.J., and Beggs, J.D. (1998). Protein–RNA interactions in the U5 snRNP of *Saccharomyces cerevisiae*. *RNA* **4**, 1675–1686.

Frank, D., Patterson, B., and Guthrie, C. (1992). Synthetic lethal mutations suggest interactions between U5 small nuclear RNA and four proteins required for the second step of splicing. *Mol. Cell. Biol.* **12**, 5197–5205.

Garcia-Blanco, M.A., Anderson, G.J., Beggs, J., and Sharp, P.A. (1990). A mammalian protein of 220 kDa binds pre-mRNAs in the spliceosome: a potential homologue of the yeast PRP8 protein. *Proc. Natl. Acad. Sci. U.S.A.* **87**, 3082–3086.

Gottschalk, A., Kastner, B., Luhrmann, R., and Fabrizio, P. (2001). The yeast U5 snRNP coisolated with the U1 snRNP has an unexpected protein composition and includes the splicing factor Aar2p. *RNA* **7**, 1554–1565.

Gottschalk, A., Neubauer, G., Banroques, J., Mann, M., Luhrmann, R., and Fabrizio, P. (1999). Identification by mass spectrometry and functional analysis of novel proteins of the yeast [U4/U6.U5] trisnRNP. *EMBO J.* **18**, 4535–4548.

Gottschalk, A., Tang, J., Puig, O., Salgado, J., Neubauer, G., Colot, H.V., Mann, M., Séraphin, B., Rosbash, M., Luhrmann, R., and Fabrizio, P. (1998). A comprehensive biochemical and genetic analysis of the yeast U1 snRNP reveals five novel proteins. *RNA* **4**, 374–393.

- Grainger, R.J., and Beggs, J.D. (2005). Prp8 protein: at the heart of the spliceosome. *RNA* **11**, 533-557.
- Hartwell, L.H. (1967). Macromolecule synthesis in temperature-sensitive mutants of yeast. *J. Bacteriol.* **93**, 1662–1670.
- Hartwell, L.H., McLaughlin, C.S., and Warner, J.R. (1970). Identification of ten genes that control ribosome formation in yeast. *Mol. Gen. Genet.* **109**, 42–56.
- Hinz, M., Moore, M.J., and Bindereif, A. (1996). Domain analysis of human U5 RNA. Cap trimethylation, protein binding, and spliceosome assembly. *J. Biol. Chem.* **271**, 19001–19007.
- James, S.A., Turner, W., and Schwer, B. (2002). How Slu7 and Prp18 cooperate in the second step of yeast pre-mRNA splicing. *RNA* **8**, 1068–1077.
- Kameoka, S., Duque, P., and Konarska, M.M. (2004). p54(nrb) associates with the 5' splice site within large transcription/splicing complexes. *EMBO J.* **23**, 1782–1791.
- Kielkopf, C.L., Lucke, S., and Green, M.R. (2004). U2AF homology motifs: Protein recognition in the RRM world. *Genes Dev.* **18**, 1513–1526.
- Konarska, M.M., and Query, C.C. (2005). Insights into the mechanisms of splicing: more lessons from the ribosome. *Genes Dev.* **19**, 2255-60.
- Kuhn, A.N., and Brow, D.A. (2000). Suppressors of a cold-sensitive mutation in yeast U4 RNA define five domains in the splicing factor Prp8 that influence spliceosome activation. *Genetics* **155**, 1667–1682.
- Kuhn, A.N., Li, Z., and Brow, D.A. (1999). Splicing factor Prp8 governs U4/U6 RNA unwinding during activation of the spliceosome. *Mol. Cell* **3**, 65–75.
- Kuhn, A.N., Reichl, E.M., and Brow, D.A. (2002). Distinct domains of splicing factor Prp8 mediate different aspects of spliceosome activation. *Proc. Natl. Acad. Sci.* **99**, 9145–9149.
- Larkin, J.C. and Woolford Jr., J.L. (1983). Molecular cloning and analysis of the CRY1 gene: A yeast ribosomal protein gene. *Nucleic Acids Res.* **11**, 403–420.
- Landschulz, W.H., Johnson, P.F., and McKnight, S.L. (1988). The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science* **240**, 1759–1764.

- Lee, C.G., Hague, L.K., Li, H., and Donnelly, R. (2004). Identification of toposome, a novel multisubunit complex containing topoisomerase II. *Cell Cycle* **3**, 638–647.
- Liu, L., Query, C.C., and Konarska, M.M. (2007). Opposing classes of prp8 alleles modulate the transition between the catalytic steps of pre-mRNA splicing. *Nat. Struct. Mol. Biol.* **14**, 519-526.
- Lustig, A.J., Lin, R.J., and Abelson, J. (1986). The yeast RNA gene products are essential for mRNA splicing in vitro. *Cell* **47**, 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**, 3008–3020.
- Makarov, E.M., Makarova, O.V., Urlaub, H., Gentzel, M., Will, C.L., Wilm, M., and Luhrmann, R. (2002). Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. *Science* **298**, 2205–2208.
- Maroney, P.A., Romfo, C.M., and Nilsen, T.W. (2000). Functional recognition of 5' splice site by U4/U6.U5 tri-snRNP defines a novel ATP-dependent step in early spliceosome assembly. *Mol. Cell* **6**, 317–328.
- Nakazawa, N, Harashima, S, and Oshima, Y. (1991). AAR2, a gene for splicing pre-mRNA of the MATa1 cistron in cell type control of *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **11**, 5693–5700.
- Neubauer, G., Gottschalk, A., Fabrizio, P., Séraphin, B., Lührmann, R., and Mann, M. (1997). Identification of the proteins of the yeast U1 small nuclear ribonucleoprotein complex by mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 385–390.
- Pena, V., Liu, S., Bujnicki, J.M., Luhrmann, R., and Wahl, M.C. (2007). Structure of a multipartite protein-protein interaction domain in splicing factor Prp8 and its link to Retinitis Pigmentosa. *Mol. Cell* **25**, 615-624.
- Peng, R., Dye, B.T., Perez, I., Barnard, D.C., Thompson, A.B., and Patton, J.G. (2002). PSF and p54nrb bind a conserved stem in U5 snRNA. *RNA* **8**, 1334–1347.
- Query, C.C., and Konarska, M.M. (2004). Suppression of multiple substrate mutations by spliceosomal prp8 alleles suggests functional correlations with ribosomal ambiguity mutants. *Mol. Cell* **14**, 343–354.

Reyes, J.L., Gustafson, E.H., Luo, H.R., Moore, M.J., and Konarska, M.M. (1999). The C-terminal region of hPrp8 interacts with the conserved GU dinucleotide at the 5' splice site. *RNA* **5**, 167–179.

Ruby, S.W. (1997). Dynamics of the U1 small nuclear ribonucleoprotein during yeast spliceosome assembly. *J. Biol. Chem.* **272**, 17333–17341.

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

Schmidt, H., Richert, K., Drakas, R.A., and Kaufer, N.F. (1999). spp42, identified as a classical suppressor of prp4–73, which encodes a kinase involved in pre-mRNA splicing in fission yeast, is a homologue of the splicing factor Prp8p. *Genetics* **153**, 1183–1191.

Schneider, C., Will, C.L., Makarova, O.V., Makarov, E.M., and Luhrmann, R. (2002). Human U4/U6.U5 and U4atac/U6atac.U5 trisnRNPs exhibit similar protein compositions. *Mol. Cell. Biol.* **22**, 3219–3229.

Siatecka, M., Reyes, J.L., and Konarska, M.M. (1999). Functional interactions of Prp8 with both splice sites at the spliceosomal catalytic center. *Genes Dev.* **13**, 1983–1993.

Siebel, C.W., Feng, L., Guthrie, C., and Fu, X.D. (1999). Conservation of a yeast protein kinase specific for the SR family of splicing factors. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 5440–5445.

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

Stevens, S.W., and Abelson, J. (1999). Purification of the yeast U4/U6.U5 small nuclear ribonucleoprotein particle and identification of its proteins. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 7226–7231.

Stevens, S.W., Barta, I., Ge, H.Y., Moore, R.E., Young, M.K., Lee, T.D., and Abelson, J. (2001). Biochemical and genetic analyses of the U5, U6, and U4/U6 x U5 small nuclear ribonucleoproteins from *Saccharomyces cerevisiae*. *RNA* **7**, 1543–1553.

Teigelkamp, S., Newman, A.J., and Beggs, J.D. (1995). Extensive interactions of PRP8 protein with the 5' and 3' splice sites during splicing suggest a role in stabilization of exon alignment by U5 snRNA. *EMBO J.* **14**, 2602–2612.

- Tran, H.J., Allen, M.D., Lowe, J., and Bycroft, M. (2003). Structure of the Jab1/MPN domain and its implications for proteasome function. *Biochemistry* **42**, 11460–11465.
- Turner, I.A., Norman, C.M., Churcher, M.J., and Newman, A.J. (2006). Dissection of Prp8 protein defines multiple interactions with crucial RNA sequences in the catalytic core of the spliceosome. *RNA* **12**, 375–386.
- Umen, J.G., and Guthrie, C. (1995)a. A novel role for a U5 snRNP protein in 3' splice site selection. *Genes Dev.* **9**, 855–868.
- Umen, J.G., and Guthrie, C. (1995)b. Prp16p, Slu7p, and Prp8p interact with the 3' splice site in two distinct stages during the second catalytic step of pre-mRNA splicing. *RNA* **1**, 584–597.
- Umen, J.G., and Guthrie, C. (1996). Mutagenesis of the yeast gene PRP8 reveals domains governing the specificity and fidelity of 3' splice site selection. *Genetics* **143**, 723–739.
- Urlaub, H., Hartmuth, K., Kostka, S., Grelle, G., and Luhrmann, R. (2000). A general approach for identification of RNA–protein crosslinking sites within native human spliceosomal small nuclear ribonucleoproteins (snRNPs). Analysis of RNA–protein contacts in native U1 and U4/U6.U5 snRNPs. *J. Biol. Chem.* **275**, 41458–41468.
- van Nues, R.W., and Beggs, J.D. (2001). Functional contacts with a range of splicing proteins suggest a central role for Brr2p in the dynamic control of the order of events in spliceosomes of *Saccharomyces cerevisiae*. *Genetics* **157**, 1451–1467.
- Vidal, V.P., Verdone, L., Mayes, A.E., and Beggs, J.D. (1999). Characterization of U6 snRNA–protein interactions. *RNA* **5**, 1470–1481.
- Vincent, M., Lauriault, P., Dubois, M.F., Lavoie, S., Bensaude, O., and Chabot, B. (1996). The nuclear matrix protein p255 is a highly phosphorylated form of RNA polymerase II largest subunit which associates with spliceosomes. *Nucleic Acids Res.* **24**, 4649–4652.
- Vijayraghavan, U., Company, M., and Abelson, J. (1989). Isolation and characterization of pre-mRNA splicing mutants of *Saccharomyces cerevisiae*. *Genes Dev.* **3**, 1206–1216.
- Wang, H.Y., Lin, W., Dyck, J.A., Yeakley, J.M., Songyang, Z., Cantley, L.C., and Fu, X.D. (1998). SRPK2: A differentially expressed SR protein-specific kinase

involved in mediating the interaction and localization of pre-mRNA splicing factors in mammalian cells. *J. Cell. Biol.* **140**, 737–750.

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

Zhang, C., Baudino, T.A., Dowd, D.R., Tokumaru, H., Wang, W., and MacDonald, P.N. (2001). Ternary complexes and cooperative interplay between NCoA-62/Ski-interacting protein and steroid receptor coactivators in vitamin D receptor-mediated transcription. *J. Biol. Chem.* **276**, 40614–40620.

Zhang, C., Dowd, D.R., Staal, A., Gu, C., Lian, J.B., Van Wijnen, A.J., Stein, G.S., and MacDonald, P.N. (2003). Nuclear coactivator-62 kDa/Ski-interacting protein is a nuclear matrix-associated coactivator that may couple vitamin D receptor-mediated transcription and RNA splicing. *J. Biol. Chem.* **37**, 35325–35336.

Zhang, L., Shen, J., Guarnieri, M.T., Heroux, A., Yang, K., and Zhao, R. (2007). Crystal structure of the C-terminal domain of splicing factor Prp8 carrying retinitis pigmentosa mutants. *Protein Sci.* **16**, 1024–1031.

Chapter 2⁽¹⁾
Structural investigation of Prp8 domain IV

¹ Adapted from Ritchie et al., (2008). *Nat. Struct. Mol. Biol.* **15**, 1199-1205.

The work presented in this chapter represents a collaboration between the authors on the paper. D.R. performed all of the experiments and analysis except solving the crystal structure which was performed by M.S. E.G., S.R., and D.S. assisted in experiments not described in the chapter and provided helpful discussions.

Chapter 2: Structural investigation of Prp8 domain IV

2-1. Introduction

2-1.1. Analysis of the C-terminal region of Prp8

The essential spliceosomal protein Prp8 has been a target of extensive biochemical investigations because of its importance to splicing as well as its very high sequence conservation across diverse species. Of considerable interest are the details of protein/protein interactions between Prp8 and other U5 snRNP proteins. Using the chaotropic salt sodium thiocyanate to preserve only the strongest protein-protein interactions in human U5 snRNP, hPrp8 is observed to interact very strongly with the U5 snRNP component Snu114 (U5-116K in humans) (Achsel et al., 1998). Snu114 is a GTPase with homology to the elongation factor EF-2, a GTPase that functions to permit the translocation step during protein synthesis by the ribosome. When the amount of chaotropic salt is reduced, hPrp8 can also be purified in association with Brr2 (h200K) and a human 40-kDa protein (Cwf17/Spf38 in *S. pombe*).

Yeast-two-hybrid analysis suggests that both Brr2 and Snu114 interact with both the N- and C-termini of yPrp8 (Boon et al., 2006; Pena et al., 2007; Zhang et al., 2007). The very C-terminus of Prp8 contains an MPN domain which normally correlates with ubiquitin hydrolase activity; however, the crystal structure of the Prp8 MPN domain suggests that the active site is not conserved (Pena et al., 2007; Zhang et al., 2007). The 35 C-terminal amino acids of Prp8 extend from the MPN domain as an unstructured tail, and are the site of 16 mutations that cause *retinitis pigmentosa* (RP) in humans, a retinal degeneration disease that causes blindness (Mordes et al., 2006; Boon et al., 2007). Several of the RP mutations have the effect of weakening the interaction between Brr2 and yPrp8, indicating a causative factor in retinitis pigmentosa may be defective U5

snRNP maturation (Boon et al., 2007). The functions of Brr2 and Snu114 along with their observed interactions with yPrp8 implicate all 3 factors as modulators of spliceosome activation: Brr2 is the ATP-dependent RNA helicase required for unwinding the U4/U6 duplex (Laggerbauer et al., 1998; Raghunathan & Guthrie, 1998), while Snu114 modulates Brr2 activity (Small et al., 2006).

Work by Guthrie and coworkers demonstrates that a yPrp8 construct consisting of the ~600 C-terminal amino acids interacts with Brr2 and stimulates its ATP-dependent U4/U6 unwinding activity. In contrast, Brr2's ATPase activity is inhibited by the yPrp8 C-terminal fragment, suggesting that yPrp8 may have a role in regulating the timing of U4/U6 unwinding by Brr2. RP mutations were introduced into the system, and some were observed to reduce yPrp8's effects on U4/U6 unwinding while not weakening the interaction between yPrp8 and Brr2 (Maeder et al., 2009). While physical and biochemical interactions have been observed which are indicative of a crucial role for the C-terminal region of Prp8 in spliceosome assembly, structural understanding of this region has been provided by high resolution structures of both the MPN domain and of domain IV (see below) in isolation.

2-1.2. C-terminal region of Prp8 at the spliceosome active site

Two sequential transesterification reactions, catalyzed by the spliceosome, are required in the splicing of pre-mRNAs. The assembled spliceosome contains a U2/U6 snRNA structure which has been proposed to form the active site for catalysis of the transesterifications (Figure 2-1) (Brow, 2002; Valadkhan and Manley, 2001; Valadkhan et al., 2007, 2009). Although pre-mRNA splicing is believed to be intrinsically RNA catalyzed, there is evidence to suggest an intimate interaction between spliceosomal proteins and the active site of the

spliceosome (Collins and Guthrie, 2000). Most notably in an RNA•protein crosslinking study, hPrp8 has been shown to directly contact the 5'SS, while the yeast homolog has been cross-linked to U6 snRNA in the context of the U5•U4/U6 tri-snRNP (Reyes et al., 1999; Vidal et al., 1999). As well, mutant Prp8 alleles in yeast strongly suggest interactions of this factor with both substrate and snRNA catalytic structures in the spliceosome (Grainger and Beggs, 2005). Thus, the catalytic heart of the spliceosome includes protein as well as RNA components.

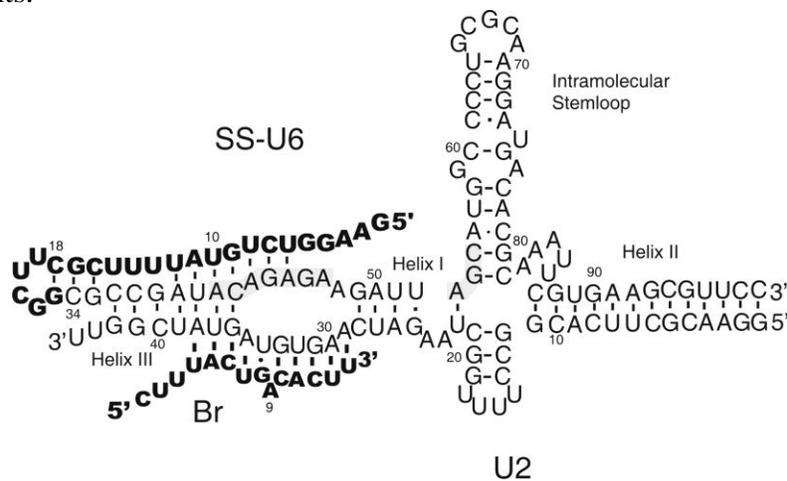


Figure 2-1. U6/U2 base-paired complex that catalyzes a reaction similar to the first step of splicing. The substrates are shown in bold letters. Base-pairing interactions between Br and U2 and the bulged A are shown. SS is covalently attached to the 5' end of U6 via a linker sequence and a hairpin (nt 11–25 of the bold region of SS-U6). The positions of the U6 intramolecular stem-loop and U6/U2 helices I, II, and III are shown. The highlighted areas mark the invariant domains of U6. The numbering of the U6 part of SS-U6 and U2 refers to human numbering. (Adapted from Valadkhan et al., 2007).

Crosslinking studies in HeLa and yeast nuclear extracts have revealed direct contacts between Prp8 and the 5'SS and 3'SS as well as the branch region (Teigelkamp et al., 1995; MacMillan et al., 1994). The hPrp8•5'SS interaction was detected before the first step of splicing in B complex (Reyes et al., 1999). In contrast, interaction with the 3'SS occurs concomitant with or subsequent to the first step of splicing. As Prp8 is the only spliceosomal protein that directly

contacts all reactive sites of the intron, in addition to forming interactions with putative snRNA catalytic structures, it is the leading candidate for a protein cofactor directly contributing to splicing catalysis (Collins and Guthrie, 2000).

A large number of mutant alleles of Prp8 have been characterized in yeast and the variety of associated phenotypes attest to the central role of this factor in spliceosome assembly and possibly catalysis (Grainger and Beggs, 2005). Alleles corresponding to suppression of a cs phenotype related to the transition between U4/U6 and U2/U6 pairing, as well as exhibiting synthetic lethality with a single nucleotide mutation affecting the U6 ISL within the catalytically active form of U6, are associated with domain IV (Table 2-1) (Kuhn et al., 2002). In addition, the region of hPrp8 involved in cross-linking to the 5'SS has been mapped to a short peptide sequence within domain IV (Figure 2-2A) by proteolytic methods (Reyes et al., 1999).

The preponderance of evidence from genetic studies in yeast and crosslinking studies in the mammalian system argues for the importance of domain IV with respect to the function of Prp8 in spliceosome assembly and/or catalysis. This chapter describes the high resolution X-ray structure of a stable core of domain IV. Analysis of this structure, combined with extant genetic and biochemical data, as well as the characterization of novel mutant alleles in yeast suggests an intimate association with the heart of the spliceosome.

2-2. Results and Discussion

2-2.1. Structural overview of hPrp8 domain IV core

Initially we solved the crystal structure of a stable fragment of hPrp8, amino acids 1831-1990, identified by limited proteolysis (Figure 2-2B; see appendix 1). The structure was not considered relevant to spliceosomal function because it was a

dimer, and it is known that Prp8 exists in only one copy per U5 snRNP; however, analysis of the structure was the basis for designing later clones containing N-terminal extensions. Ultimately, we were able to crystallize a selenomethionine (Se-Met) substituted domain corresponding to amino acids 1769-1990 of hPrp8 and determined its structure to 1.85 Å resolution using MAD methods (PDB ID code 3ENB) (Figure 2-2C; Table 2-2).

Table 2-1. Mutant yeast alleles mapping to Prp8 domain IV core and corresponding phenotypes⁽¹⁾.

Yeast Allele (Human)	Yeast Phenotype
V1860N (V1788)	suppresses U4cs ²
V1860D (V1788)	suppresses U4cs, synthetic lethal U6 U-A ³
T1861P (T1789)	suppresses U4cs, synthetic lethal U6 U-A, suppresses splice site mutations, synthetic lethal Prp28Δ ⁴
V1862D (I1790)	suppresses U4cs, synthetic lethal U6 U-A
V1862A/Y (I1790)	suppresses U4cs
K1864E (K1792)	suppresses splice site mutations
N1869D (N1797)	suppresses splice site mutations
V1870N (L1798)	suppresses splice site mutations
I1875T (I1803)	suppresses U4cs
N1883D (N1811)	synthetic lethal Prp28Δ
E1960K (E1888)	alternate 3'SS choice
E1960G (E1888)	alternate 3'SS choice
T1982A (T1910)	suppresses splice site mutations

[1] As reviewed in (Grainger and Beggs, 2005); [2] U4 cs (cold-sensitive) results in a hyperstabilized U4-U6; [3] U6 U-A corresponds to mutation of U6 at the base of the internal stem loop that stabilizes its catalytic conformation; [4] Prp28 is the U5 snRNP component required for resolving the 5'SS·U1 snRNA interaction.

The structure of the domain IV core is bipartite consisting of an N-terminal sub-domain (amino acids 1769-1887) with an RNase H fold and a tightly packed C-terminal cluster of five helices (amino acids 1900-1990). The RNase H fold exhibits a characteristic five-stranded parallel/anti-parallel β-sheet, buttressed by two α-helices. This structural homology was not predicted by analysis of primary sequence because of a seventeen amino acid insertion (amino acids 1787-

1803) that in itself is a striking feature of the core domain. In one of the two monomers found in the asymmetric unit, amino acids 1787-1803 are well structured forming a two-stranded anti-parallel β -sheet finger while in the second this region was disordered. The juxtaposition of the N-terminal RNase H homology domain with the C-terminal helices establishes a shallow channel across the core approximately 20 Å in width and 30 Å in length. The floor of this channel is in part formed by an α -helix (α 5: amino acids 1893-1898); this along with a directly N-terminal eight amino acid loop and C-terminal β -strand form a linker between the two sub-domains of the core.

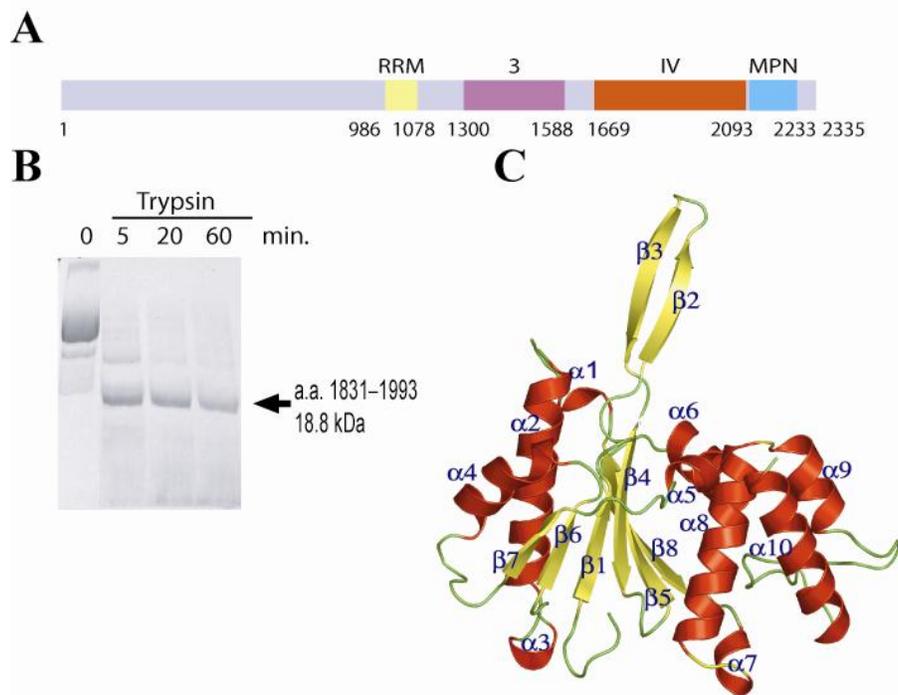


Figure 2-2. Structure of hPrp8 domain IV. A) Primary structure of hPrp8. Predicted domains include those inferred from primary-sequence homologies (a central RRM and C-terminal Jab1/MPN domain) and from the identification of mutant alleles in yeast, coupled with tolerance of amino acid insertions and sequence conservation across species. B) Partial tryptic proteolysis of hPrp8 1689–2001 expressed and purified from *E. coli*. C) Ribbon diagram of the hPrp8 domain IV core, amino acids 1769–1990. α -helices and β -strands are colored red and yellow, respectively.

Table 2-2. Crystallographic data collection and refinement statistics.

Data collection	Native		Se-Met	
space group	P2 ₁ 2 ₁ 2 ₁		P2 ₁ 2 ₁ 2 ₁	
cell dimensions				
<i>a</i> , <i>b</i> , <i>c</i> (Å)	75.397, 78.063, 93.575		76.047, 78.003, 93.008	
α , β , γ (Å)	90, 90, 90		90, 90, 90	
	<i>Native</i>	<i>peak</i>	<i>inflection</i>	<i>remote</i>
Wavelength	1.115872	0.979625	0.979741	1.019951
Resolution (Å)	1.85	2.3	2.3	2.3
R _{sym} or R _{merge}	4.4 (40.8)	5.1 (26.4)	5.1 (26.5)	4.5 (22.2)
I/ σ I	27.9 (3.4)	18.8 (3.9)	19.0 (4.0)	20.8 (5.0)
Completeness (%)	99.7 (99.6)	99.1 (98.0)	99.2 (98.0)	99.2 (98.3)
Redundancy	3.9 (3.5)	2.9 (2.8)	2.9 (2.8)	2.9 (2.8)
Mosaicity	0.27	0.5	0.5	0.5
Refinement				
Resolution (Å)	60-1.85			
No. reflections	47833			
R _{work} /R _{free}	0.195/0.232			
No. atoms	3890			
Protein	3422			
Ligand				
Waters	468			
<i>B</i> -factors				
Protein	24.2			
Ligand				
Waters	40.2			
R.m.s. deviations				
Bond lengths (Å)	0.01			
Bond angles (°)	1.18			

Peak, inflection, and remote wavelength data were collected on a single Se-met substituted crystal. Refinement was performed against the native dataset from an un-substituted crystal. Numbers in parenthesis refer to the highest resolution shell (1.92-1.85 Å for native dataset, and 2.38-2.30 Å for Se-met datasets).

The very high sequence similarity of Prp8 domain IV core between humans and yeast (69% identical; the identities between the human versus *C. elegans* or *D. melanogaster* core sequence are 90% and 97% respectively) allows interpretation of the hPrp8 structure in light of known yeast alleles (Figure 2-3; Table 2-1). All but one of the Prp8 domain IV core alleles characterized

correspond to amino acids on one face of the core structure (Figure 2-4). An allele synthetically lethal with the failure to resolve the 5'SS•U1 snRNA interaction (due to deletion of Prp28) is found on the surface of the RNase-H homology β -sheet (Grainger and Beggs, 2005). An allele corresponding to a spatially proximate residue is associated in yeast with aberrant 3'SS choice (Umen and Guthrie, 1995). Intriguingly, a cluster of characterized yeast alleles map to the two-strand β -finger. These include suppressors of mutations at the splice sites/branch region as well as suppressors of hyperstabilization of U4/U6 that are also synthetically lethal with a mutation that affects the catalytically active form of U6 snRNA. One of the yeast alleles, corresponding to a mutation of T1789 (T1861P in yeast), is expected to disrupt the β -sheet hinting at conformational flexibility or change within the context of the spliceosome.



Figure 2-3. Secondary-structure diagram and sequence conservation of Prp8 domain IV core. The secondary structure diagram depicts α -helices (red) and β -strands (yellow) of the Prp8 domain IV core. Also shown is an alignment with core domain sequences from *S. cerevisiae*, *C. elegans* and *D. melanogaster* orthologs indicating sequence identity (*), conservation (:), and partial conservation (.). Positions of mutant yeast alleles identified within the core domain along with the corresponding residues in the human sequence are highlighted (red). The sequence identified as forming a crosslink with the 5'SS in spliceosomal B complex is shown in the human sequence (purple). Amino acids spatially conserved with respect to the RNase H metal binding site are indicated (blue) for both the human and yeast sequences.

Konarska and colleagues showed the direct interaction between nucleotides at the 5'SS and hPrp8 in B complex in 254 nm UV crosslinking experiments; using a combination of proteolysis and immunoprecipitation they were able to map this crosslink to within a sequence corresponding to amino acids 1894 to 1898 of domain IV (Reyes et al., 1999). The structure of the core domain shows that these amino acids correspond to the helix (α 5) that forms the base of the channel separating the RNase H homology domain and the C-terminal helical cluster (Figures 2-2C, 2-4).

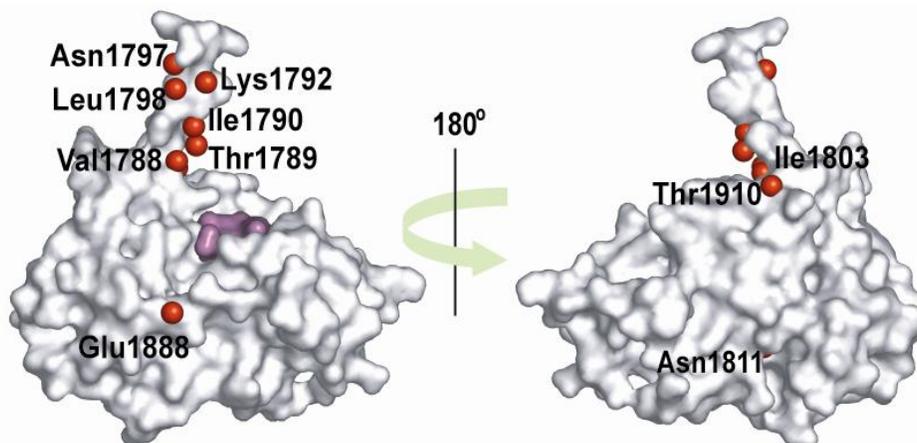


Figure 2-4. Mapping of yeast mutant alleles and a 5'SS cross-link in the context of the Prp8 domain IV core. Thirteen mutant alleles, corresponding to mutation at ten positions in yeast Prp8, map to the surface of the Prp8 domain IV core. All of the wild-type residues are identical or conservative substitutions in humans, and all except one (Prp8 Thr1910) correspond to amino acid side chains positioned on one face of the core domain structure. Shown are the positions of amino acids corresponding to mutant alleles (red; phenotypes summarized in Table 2-1) and the region cross-linked to the 5'SS in the B complex (purple).

After we solved the crystal structure of hPrp8 domain IV core, another group published the crystal structures of both a human construct comprised of amino acids 1755-2016 (PDB ID code 3E9L) and two yeast constructs consisting of 1827-2092 (PDB ID code 3E9P) and 1836-2092 (PDB ID code 3E9O) (Pena et

al., 2008). The two yeast structures superimposed with pairwise root-mean-square deviations (r.m.s.d.s) of 0.8-1.0 Å, and they align to the hPrp8 1755-2016 structure with r.m.s.d.s of 1.2-1.7 Å over approximately 230 Ca atoms, indicating that the structures are very similar. The only differences between the yeast and human structures are the relative orientations of the β-finger to the RNase H β-sheet and the C-terminal helical bundle to the β-sheet, indicating flexibility of these elements (Pena et al., 2008).

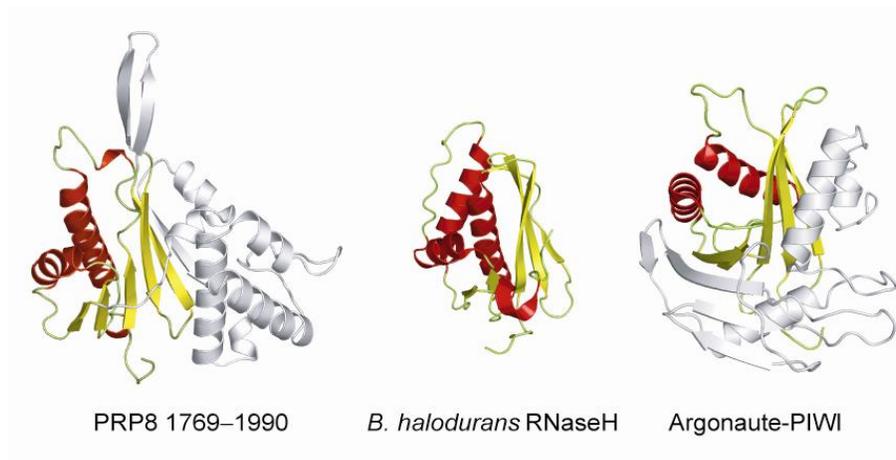
2-2.2. RNase H homology of the domain IV core

Analysis of the structure of the core of hPrp8 domain IV revealed two subdomains: a C-terminal helical assembly and an N-terminal portion with structural homology to RNase H (Figure 2-5A). Although RNase H binds RNA•DNA hybrids, catalyzing hydrolysis of the RNA strand, the general specificity of binding is for an A-form duplex; for example, the PIWI domain of Argonaute proteins also contains an RNase H fold and is predicted to bind a siRNA/mRNA duplex and cleave the mRNA strand (Figure 2-5A) (Song et al., 2004).

The RNase H homology within the domain IV core structure includes conserved secondary and tertiary structure but little relationship at a primary sequence level. Catalysis of RNA cleavage by RNase H like enzymes is proposed to involve a two metal mechanism in which divalent magnesium ions, bound at adjacent sites separated by ~4 Å, promote hydrolysis by a combination of activation of a water nucleophile and transition state stabilization (Nowotny et al.,

2005). Inspection of the domain IV core structure showed that only one of these canonical metal-binding sites is present with coordinating side chains – two aspartates and a threonine – spatially conserved with respect to Mg^{2+} coordinating residues within the RNase H fold (Figure 2-5B).

A



B

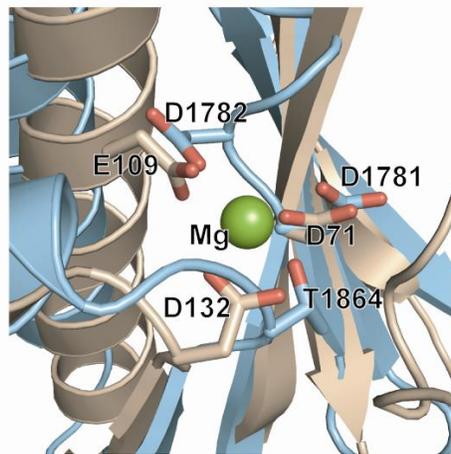


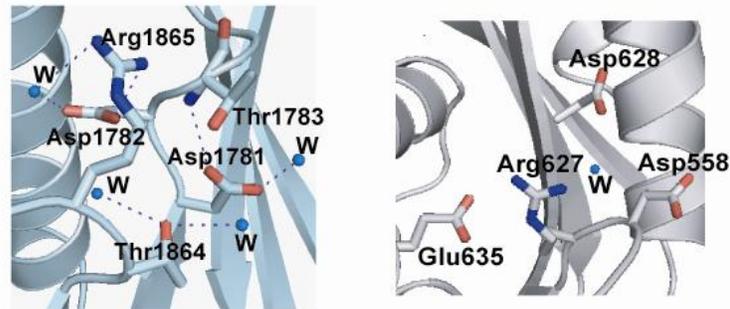
Figure 2-5. Comparison of the Prp8 domain IV core with RNase H folds. A) Structure of the hPrp8 domain IV core alongside *Bacillus halodurans* RNase H (Nowotny et al., 2005) and the PIWI domain of *P. furiosus* Argonaute (Song et al., 2004), with the RNase H fold indicated in red and yellow. B) Alignment of hPrp8 domain IV core (cyan) with RNase H (gold) reveals the conservation of a putative metal-binding site in hPrp8 that corresponds to a functionally critical site in RNase H. The RNase H DDE triad consisting of residues D71, D132, and E109 binds Mg^{2+} (green) and corresponds with a DDT triad in hPrp8 composed of residues D1781, D1782, and T1864 although no metal is observed in the hPrp8 structure.

We were unable to observe an Mg^{2+} ion in the core domain site despite the fact that crystals were grown in 100-200 mM $MgCl_2$, and that Mg^{2+} is observed bound at another site bridging a symmetry related glutamate between neighbouring molecules in the crystal lattice. Inspection of the site shows that in place of bound metal, the spatially conserved aspartates are stabilized by a network of water molecules (Figure 2-6A). As well, the side chain of Arg1865 - part of the loop between β_6 and α_4 - is hydrogen-bonded to Asp1782 effectively blocking the potential metal binding site. This is directly analogous to the RNase H fold of the *P. furiosus* Argonaute PIWI domain in which an arginine residue is positioned in a similar fashion (Figure 2-6A) (Song et al., 2004); another example of this structural feature is found in the Tn5 transposase (Steiniger-White et al., 2004).

The importance of this putative metal-binding site for hPrp8 function has been investigated by mutating the corresponding residues in yeast and testing their effects on viability (Figure 2-6B) (Pena et al., 2008). All mutants were associated with dramatic growth defects. D1853/1781 (yeast/human) was lethal on conversion to alanine and produced both cs and ts phenotypes on mutation to asparagine. R1937/1865 was similarly lethal on mutation to alanine and conditionally lethal at low and high temperatures on change to a lysine. D1854/1782 was cs and ts on mutation to asparagine or alanine. Similarly, T1855/1783A and T1936/1864A were both cs and ts. The association of severe growth defects with mutations of amino acids at the putative metal-binding site underscores the importance of these residues to Prp8 function (Pena et al., 2008),

although it remains an open question whether these residues are coordinating an essential metal ion.

A



B

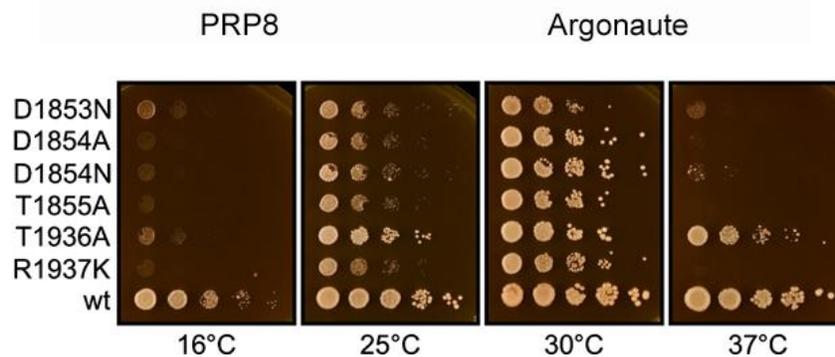


Figure 2-6. Analysis of the putative conserved binding site in Prp8 domain IV core. A) The potential metal binding site in the core domain (left) includes four hydrogen-bonding water molecules and is blocked by the side chain of Arg1865 hydrogen bonded to Asp1782 in a similar fashion to the interaction observed in the analogous *P. furiosus* PIWI domain site (right). B) Cell viability assay monitoring the effects of exchanging invariant Prp8 residues (yeast residue numbers) D1853, D1854, T1855, T1936 and R1937 as indicated. After selection of clones, the culture and serial dilutions were spotted and grown at the indicated temperatures for 2 days. (B is adapted from Pena et al., 2008, and thus highlights an experiment conducted by a different group).

2-2.3. An RNase H domain in the spliceosome active site

The identification of an RNase H domain in Prp8 provides a structural basis for the genetic and cross-linking data that implicate Prp8 in spliceosome activation and possibly catalysis. The genetic interactions between Prp8 and U6 ISL are suggestive of a direct interaction in the spliceosome active site. There are

numerous examples of *cs* suppression implicating Prp8 in U4/U6 unwinding, but synthetic lethality with the mutually exclusive U6 ISL has rarely been observed for *prp8* alleles (Kuhn et al., 2002). The location of three adjacent alleles (human 1788-1790) that are unique because they not only suppress *U4-cs1* but are also synthetic lethal with *U6-UA*, the mutation that hyper-stabilizes the U6 ISL, now have a structural basis to help understand their function. In particular, the residues lie at the base of the flexible β -finger, close to helix $\alpha 5$ that cross-links to the 5'SS. The relative location of these structural features is consistent with the proposed role for domain IV in coupling U4/U6 unwinding with the rearrangements at the 5'SS (Kuhn et al., 1999). Perhaps Prp8 is in position to recognize formation of a proper ACAGAGA-box/5'SS duplex, then trigger complete unwinding of the U4/U6 duplex.

The U6 ISL is an interesting RNA candidate for direct interaction with Prp8 because it makes contact with a metal ion essential for splicing catalysis (Yean et al., 2000). The analogous structural feature in the Group II intron, domain V, clearly coordinates two metal ions in perfect position to catalyze splicing (Toor et al., 2008). An open question is whether Prp8 can coordinate a metal residue, perhaps in combination with the U6 ISL. In the context of an active spliceosome, U6 snRNA not only forms an intramolecular ISL but also interacts with the 5'SS and extensively with U2 snRNA. The network of RNA interactions provides the basis for splicing chemistry (Brow, 2002; Valadkhan et al., 2007). If Prp8 indeed physically interacts with the U6 ISL in an activated

spliceosome, it must be in proximity to RNA structures at the active site, and therefore in a position to modulate catalysis.

The exact role Prp8 would be playing in splicing catalysis is a mystery. Perhaps Prp8 is stabilizing a RNA structure at the active site, analogous to RNA tertiary interactions stabilizing the catalytic conformation of Group II introns (Toor et al., 2008). Indeed, an RNase H domain provides an excellent platform for interactions with multiple strands of RNA, consistent with Prp8's multiple roles in mediating rearrangements in RNA structure on the way to spliceosome activation. Investigating the interactions between Prp8 domain IV and RNA constructs modeled after the spliceosome active site are the basis of the next chapter.

2-3. Methods

2-3.1. Identification and expression of an hPrp8 domain IV core

A cDNA representing an N-terminal extension of hPrp8 1831-1990 (see Appendix 1) encoding hPrp8 amino acids 1769-1990, core domain IV, was cloned into the *EcoRI* and *HindIII* sites of pMALc2x (NEB) using PCR primers to insert a TEV protease cleavage site between maltose-binding protein (MBP) and the core domain. Mutagenesis of the core domain was carried out by PCR and confirmed by sequencing. The resulting MBP•core domain fusion proteins were expressed in *E. coli* and purified by sequential amylose resin, anion exchange, and size exclusion chromatography.

Forward PCR primer sequence (a.a. 1769 START): 5' to 3' direction:

GCGCGCGAATTCGAAAATTTGTATTTTCAAGGTGAGCTCTTCTCC

Reverse PCR primer sequence (a.a. 1990 END):

5'- GCGCGC AAG CTT TCA TCA GTC AGC CAA GAT CAG -3'

2-3.2. Crystallization

Crystals of hPrp8¹⁷⁶⁹⁻¹⁹⁹⁰ were grown at 23°C using the hanging drop vapor diffusion technique. Crystals of native protein and protein containing Se-Met substitutions were grown by mixing one µl of 10 mg/ml protein solution (10 mM Tris, pH 8.0, 0.1 mM EDTA, 5 mM DTT, 0.02% NaN₃) with one µl of precipitant (2.5 M NaCl, 100 mM Tris, pH 7, 100 mM MgCl₂). Crystals were transferred to precipitant containing 20% glycerol and frozen in liquid nitrogen for data collection.

2-3.3. Data collection and processing

Data were collected at beam line 8.3.1 of the Advanced Light Source, Lawrence Berkeley National Laboratory. Anomalous data was collected from a single Se-methionine derivatized crystal; a three-wavelength MAD experiment was performed collecting data in an inverse-beam mode at the experimentally-determined Se/K edge, the inflection point, and a low energy remote wavelength. Data were processed and scaled using the HKL package (Table 2-2) (Otwinowski and Minor, 1997).

2-3.4. Model building and refinement

The program SOLVE (Terwilliger and Berendzen, 1999) was used to determine the positions of the four expected Se atoms. Initial phases to 2.3 Å were improved with density modification in RESOLVE (Terwilliger, 2000). An initial model automatically built by ARPWARP (Terwilliger, 2002) served as the basis for subsequent manual model building and refinement. Iterative cycles of refinement in REFMAC (Murshudov et al., 1997) against 1.85 Å data collected on a native crystal and manual model building using XFIT (McRee, 1999) was used to complete and refine the model. Refinement statistics for both structures are summarized in Table 2-2.

2-4. References

- Achsel, T., Ahrens, K., Brahms, H., Teigelkamp, S., and Luhrmann, R. (1998). The human U5–220kD protein (hPrp8) forms a stable RNA-free complex with several U5-specific proteins, including an RNA unwindase, a homolog of ribosomal elongation factor EF-2, and a novel WD-40 protein. *Mol. Cell. Biol.* **18**, 6756–6766.
- Boon, K.L., Norman, C.M., Grainger, R.J., Newman, A.J., and Beggs, J.D. (2006). Prp8p dissection reveals domain structure and protein interaction sites. *RNA* **12**, 198–205.
- Boon K.L., Grainger R.J., Ehsani P., Barrass J.D., Auchynnikava T., Inglehearn C.F., and Beggs J.D. (2007). prp8 mutations that cause human retinitis pigmentosa lead to a U5 snRNP maturation defect in yeast. *Nat. Struct. Mol. Biol.* **14**, 1077–1083.
- Brow, D.A. (2002). Allosteric cascade of spliceosome activation. *Annu. Rev. Genet.* **36**, 333–360.
- Collins, C.A., and Guthrie, C. (1999). Allele-specific genetic interactions between Prp8 and RNA active site residues suggest a function for Prp8 at the catalytic core of the spliceosome. *Genes & Dev.* **13**, 1970–1982.
- Collins, C.A., and Guthrie, C. (2000). The question remains: Is the spliceosome a ribozyme? *Nat. Struct. Biol.* **7**, 850–854.
- Grainger, R.J., and Beggs, J.D. (2005). Prp8 protein: at the heart of the spliceosome. *RNA* **11**, 533–557.
- Krämer, A. (1996). The structure and function of proteins involved in mammalian pre-mRNA splicing. *Annu. Rev. Biochem.* **65**, 367–409.
- Kuhn, A.N., Li, Z., and Brow, D.A. (1999). Splicing factor Prp8 governs U4/U6 RNA unwinding during activation of the spliceosome. *Mol. Cell* **3**, 65–75.
- Kuhn, A.N., Reichl, E.M., and Brow, D.A. (2002). Distinct domains of splicing factor Prp8 mediate different aspects of spliceosome activation. *Proc. Natl. Acad. Sci.* **99**, 9145–9149.
- Laggerbauer, B., Achsel, T., and Luhrmann, R. (1998). The human U5–200kD DEXH-box protein unwinds U4/U6 RNA duplexes in vitro. *Proc. Natl. Acad. Sci.* **95**, 4188–4192.

- 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**, 3008–3020.
- Maeder, C., Kutach, A.K., and Guthrie, C. (2009). ATP-dependent unwinding of U4/U6 snRNAs by the Brr2 helicase requires the C terminus of Prp8. *Nat. Struct. Mol. Biol.* **16**, 42–48.
- McRee, D.E. (1999). XtalView/Xfit: a versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **125**, 156–165.
- Mordes, D., Luo, X., Kar, A., Kuo, D., Xu, L., Fushimi, K., Yu, G., Sternberg, P., and Wu, J.Y. (2006). Pre-mRNA splicing and retinitis pigmentosa. *Mol. Vis.* **12**, 1259–1271.
- Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255.
- Nowotny, M., Gaidamakov, S.A., Crouch, R.J., and Yang, W. (2005). Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis. *Cell* **121**, 1005–1016.
- Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.
- Pena, V., Liu, S., Bujnicki, J.M., Luhrmann, R., and Wahl, M.C. (2007). Structure of a multipartite protein-protein interaction domain in splicing factor Prp8 and its link to Retinitis Pigmentosa. *Mol. Cell* **25**, 615–624.
- Pena, V., Rozov, A., Fabrizio, P., Luhrmann, R., and Wahl, M.C. (2008). Structure and function of an RNase H domain at the heart of the spliceosome. *EMBO J.* **27**, 2929–2940.
- Query, C.C., and Konarska, M.M. (2004). Suppression of multiple substrate mutations by spliceosomal prp8 alleles suggests functional correlations with ribosomal ambiguity mutants. *Mol. Cell* **14**, 343–354.
- Raghuathan, P.L., and Guthrie, C. (1998). RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2. *Curr. Biol.* **8**, 847–855.
- Reyes, J.L., Gustafson, E.H., Luo, H.R., Moore, M.J., and Konarska, M.M. (1999). The C-terminal region of hPrp8 interacts with the conserved GU dinucleotide at the 5' splice site. *RNA* **5**, 167–179.

- Small, E.C., Leggett, S.R., Winans, A.A., and Staley, J.P. (2006). The EF-G-like GTPase Snu114p regulates spliceosome dynamics mediated by Brr2p, a DExD/H box ATPase. *Mol. Cell* **23**, 389–399.
- Song, J.J., Smith, S.K., Hannon, G.J., and Joshua-Tor, L. (2004). Crystal structure of Argonaute and its implications for RISC slicer activity. *Science* **305**, 1434–1437.
- Steiniger-White, M., Rayment, I., and Reznikoff, W.S. (2004). Structure/function insights into Tn5 transposition. *Curr. Opin. Struct. Biol.* **14**, 50–57.
- Teigelkamp, S., Newman, A.J., and Beggs, J.D. (1995). Extensive interactions of PRP8 protein with the 5' and 3' splice sites during splicing suggest a role in stabilization of exon alignment by U5 snRNA. *EMBO J.* **14**, 2602–2612.
- Terwilliger, T.C. (2000). Maximum-likelihood density modification. *Acta Crystallogr. D Biol. Crystallogr.* **56**, 965–972.
- Terwilliger, T.C. (2002). Automated main-chain model-building by template-matching and iterative fragment extension. *Acta Crystallogr. D Biol. Crystallogr.* **59**, 38–44.
- Terwilliger, T.C., and Berendzen, J. (1999). Automated MAD and MIR structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **55**, 849–861.
- Toor, N., Keating, K.S., Taylor, S.D., and Pyle, A.M. (2008). Crystal structure of a self-spliced group II intron. *Science* **320**, 77–82.
- Turner, I.A., Norman, C.M., Churcher, M.J., and Newman, A.J. (2006). Dissection of Prp8 protein defines multiple interactions with crucial RNA sequences in the catalytic core of the spliceosome. *RNA* **12**, 375–386.
- Umen, J.G., and Guthrie, C. (1995). A novel role for a U5 snRNP protein in 3' splice site selection. *Genes & Dev.* **9**, 855–868.
- Valadkhan, S., and Manley, J.L. (2001). Splicing-related catalysis by protein-free snRNAs. *Nature* **413**, 701–707.
- Valadkhan, S., Mohammadi, A., Wachtel, C., and Manley, J.L. (2007). Protein-free spliceosomal snRNAs catalyze a reaction that resembles the first step of splicing. *RNA* **13**, 2300–2311.
- Valadkhan, S., Mohammadi, A., Jaladat, Y., and Geisler, S. (2009). Protein-free small nuclear RNAs catalyze a two-step splicing reaction. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 11901–11906.

Vidal, V.P., Verdone, L., Mayes, A.E., and Beggs, J.D. (1999). Characterization of U6 snRNA–protein interactions. *RNA* **5**, 1470–1481.

Yean, S.L., Wuenschell, G., Termini, J., and Lin, R.J. (2000). Metal-ion coordination by U6 small nuclear RNA contributes to catalysis in the spliceosome. *Nature* **408**, 881-884.

Zhang, L., Shen, J., Guarnieri, M.T., Heroux, A., Yang, K., and Zhao, R. (2007). Crystal structure of the C-terminal domain of splicing factor Prp8 carrying retinitis pigmentosa mutants. *Protein Sci.* **16**, 1024-1031.

Chapter 3⁽¹⁾
Protein/RNA interactions of Prp8 domain IV

¹ Adapted from Ritchie et al., (2008). *Nat. Struct. Mol. Biol.* **15**, 1199-1205.

The work presented in this chapter represents collaboration between the authors on the paper. D.R. performed all of the experiments and analysis except footprinting by partial proteolysis which was performed by M.S. E.G., S.R., and D.S. assisted in experiments not described in the chapter and provided helpful discussions.

Chapter 3: Interactions between RNA and Prp8 domain IV

3-1. Introduction

3-1.1. Evidence for RNA binding by the domain IV core

There are a number of lines of evidence indicating that a role of Prp8 is interaction with spliceosomal RNA. In particular, domains 3 and IV both contain numerous amino acids that have been identified in screens for alleles that suppress mutations in the pre-mRNA substrate. The suppressor alleles do not discriminate between mutations at different sites in the pre-mRNA. Instead, a cross-suppression phenotype is observed whereby mutations in domain 3 or IV suppress mutations at all of the 5'SS, 3'SS, and BS. The cross-suppression phenomenon argues against a direct interaction between Prp8 and the pre-mRNA being affected, but rather a loosening of interactions around the active site (Query and Konarska, 2004). Nonetheless, cross-linking data do demonstrate that Prp8 is in intimate contact with the sites of splicing chemistry both before and during catalysis (Maroney et al., 2000; Reyes et al., 1999; MacMillan et al., 1994; Teigelkamp et al., 1995). The genetic and cross-linking data support a model whereby Prp8 is involved in multiple interactions that are important for sculpting the spliceosome active site.

In addition to genetic interactions with the pre-mRNA, domain IV is the location of alleles that suppress the cold-sensitive mutation *U4-cs1*. The extended network of U4/U6 base-pairing in U5·U4/U6 tri-snRNP caused by *U4-cs1* sequesters the ACAGAGA box of U6 snRNA, making it unavailable to interact with the 5'SS. In effect, *U4-cs1* stalls spliceosome assembly at lower

temperatures by inhibiting unwinding of U4/U6 snRNA. The identification of mutations in Prp8 that suppress this defect led to a model of spliceosome activation whereby U4/U6 unwinding is triggered upon proper recognition of the U6/5'SS structure by Prp8 (Kuhn et al., 1999). Thus, characterization of mutant yeast alleles of Prp8 suggests an interaction of the domain IV core with snRNA and/or conserved regions of the pre-mRNA substrate during spliceosome assembly and through the transesterification steps of splicing. The work of Konarska and colleagues demonstrated a direct interaction between this domain and the 5'SS in B complex (Reyes et al., 1999) but there is no additional direct evidence of RNA binding by this domain of hPrp8. This chapter highlights what is currently known about the RNA-binding properties of hPrp8 domain IV.

3-2. Results and Discussion

3-2.1. Identifying RNA binding partners of hPrp8 domain IV

In order to test the hypothesis that the core domain interacts with either spliceosomal or substrate RNA, we carried out gel mobility shift assays of the core domain, as well as several mutants, with a variety of RNAs. Because of the lack of data with respect to core domain•RNA interactions, we carried out studies to examine the interaction of this domain with RNAs differing in both sequence and proposed structure (Figure 3-1). Electrophoretic mobility shift assays with a panel of RNAs assayed against wild-type hPrp8 domain IV core and several mutants revealed dissociation constants in the range of 20 to >300 nM. We examined binding of a number of other core domain mutants corresponding to

known yeast alleles (hPrp8: V1788D, T1789P, V1862D); all of these either abrogated or severely diminished binding to the RNAs tested and implicate the b-finger in RNA binding.

We tested the binding of the core domain with a variety of RNAs including single-stranded oligonucleotides representing the 5'SS or 3'SS, short duplex RNAs, a bulged RNA mimicking the pre-mRNA•U2 snRNA duplex, and several more complex RNAs based on proposed structures in the mature spliceosome (Figure 3-1 and data not shown). We were unable to observe any sequence-specificity in these binding studies. However, the core domain showed greater affinity for duplex RNAs over single-stranded molecules (dissociation constants of ~100 mM vs >300 mM respectively; Figure 3-1).

Extensive studies, especially in yeast, have suggested a model for the interaction between the U2 and U6 snRNAs and between the 5'SS and U6 snRNA at the catalytic site of the spliceosome. Butcher and coworkers have elaborated this model to suggest that the associated RNAs fold into a four-helix junction that can be roughly mapped onto the hairpin ribozyme structure (Rupert and Ferre-D'Amare, 2001; Sashital et al., 2004). Based on this model, we examined the affinity of the core domain for an RNA representing the four-helix assembly in the mammalian spliceosome (U2/U6/5'SS); the affinity of the core domain for this RNA (~20 mM) was 5-10 fold greater than observed for simple single- or double-stranded RNAs. We also synthesized a variant of the hairpin ribozyme and observed that this RNA was bound by the core domain with essentially the same affinity as the U2/U6/5'SS model (Figure 3-1).

hPrp8 1769-1990

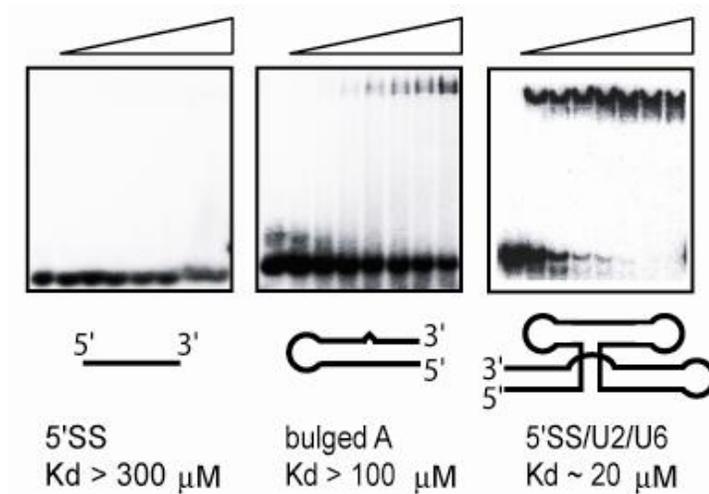


Figure 3-1. RNA binding by hPrp8 domain IV core. Gel mobility-shift assays for RNA binding to the core domain showing the affinity of three RNAs incubated with increasing concentrations (2–100 μ M) of protein. Left to right: an 11-nucleotide RNA representing a pre-mRNA 5'SS, a 14-base-pair, 32-nucleotide bulged duplex representing interaction between the branch region and U2 snRNA, and a 92-nucleotide four-helix junction RNA modeling interaction between U2 snRNA, U6 snRNA and the 5'SS (Sashital et al., 2004).

3-2.2. Mapping Prp8/RNA interactions

The binding affinity of the four-helix junction for the domain IV core was tighter than that observed for any other construct tested: we therefore investigated the complex formed between this RNA and the core using a combination of partial proteolysis and mild RNase treatment (Figures 3-2, -3).

Incubation of domain IV core with either a small amount of trypsin or endoprotease Arg-C resulted in the formation of a ~20 kDa polypeptide corresponding to the C-terminal product of cleavage at Arg1832 within the loop connecting α 1 and α 2 (Figure 3-2). Similarly, treatment of the protein with chymotrypsin yielded a slightly larger ~21 kDa fragment representing the C-

terminal product of cleavage at Trp1827 just C-terminal to a1 (Figure 3-2). When these proteolytic treatments were repeated in the presence of saturating four-helix-junction RNA, cleavage at Arg1832 was reduced three-fold while that at Trp1827 was not affected (Figure 3-2). Despite their proximity, the two residues in question are oriented towards opposite surfaces of the core domain, with Arg1832 disposed towards the canonical RNA-binding face of the RNase H fold. Together, these results thus argue for an intimate association of the bound RNA across the RNase-H fold (consistent with protection from cleavage of Arg1832 but not Trp1827).

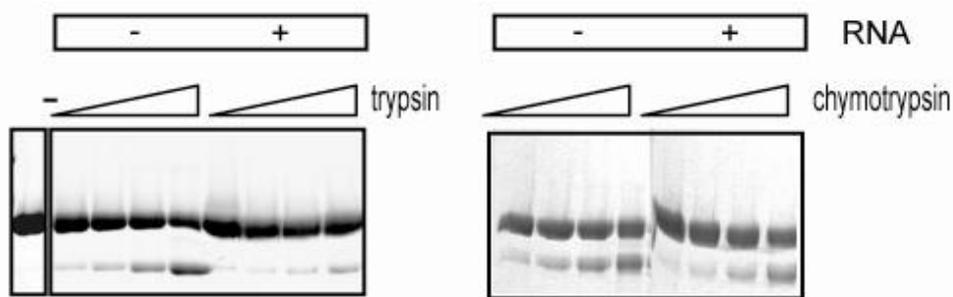


Figure 3-2. Footprinting analysis of a Prp8/RNA complex. Partial proteolysis of the core domain alone and in the presence of saturating four-helix-junction RNA showing protection against cleavage by trypsin at Arg1832 upon binding but not against cleavage by chymotrypsin at Trp1827.

We investigated interactions between the four-helix junction RNA and core domain using mild RNase treatment as a probe. Treatment of the free RNA with RNase T1 (Figure 3-3) yielded strong cleavages at G28 and G29, as well as at G33, lesser cleavages at sites 3' to this on the opposite strand (G42, G44, G47), and a major cleavage within the loop containing G65; cleavages at other positions were significantly weaker consistent with the proposed secondary structure of this RNA (Figure 3-3) (Sashital et al., 2004). In the presence of saturating protein,

most of the positions described above were strongly protected from cleavage, consistent with either occlusion by interaction with the core domain or stabilization of secondary structure (G33, G42, G47); an exception was noted at the G65 position where more moderate protection was seen. We repeated this analysis using RNase A and observed specific protection from cleavage at a subset of U or C positions upon binding of the four-helix junction RNA to the core domain (Figure 3-3).

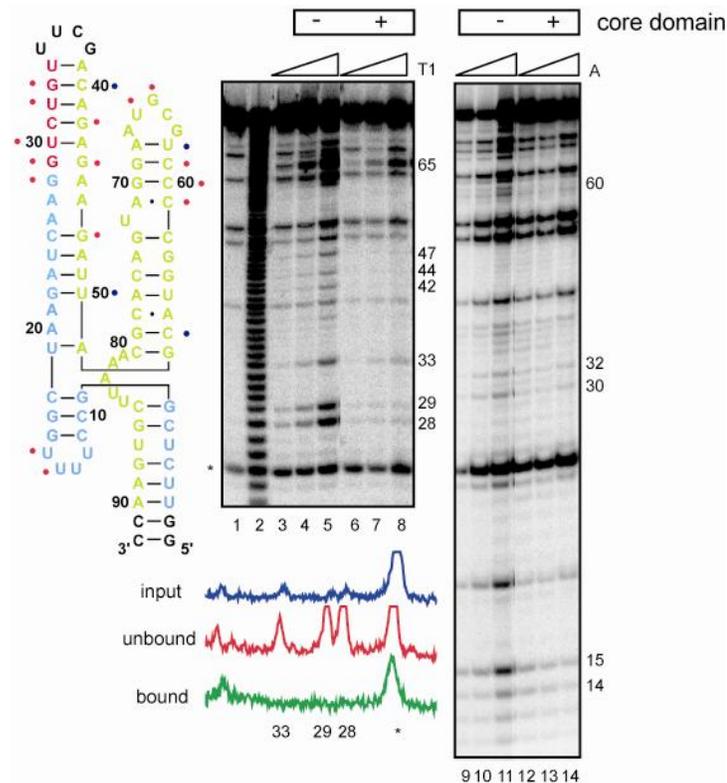


Figure 3-3. RNase footprinting of a Prp8/RNA complex. Partial digestion of ^{32}P end-labeled four-helix-junction RNA using RNases T1 and A alone and in the presence of saturating core domain. Lane 1, input; lane 2, base hydrolysis ladder; lanes 3–5, incubation of unbound RNA in the presence of increasing RNase T1; lanes 6–8, incubation of RNA in the presence of core domain and increasing RNase T1; lanes 9–11, incubation of unbound RNA in the presence of increasing RNase A; lanes 12–14, incubation of RNA in the presence of core domain and increasing RNase A. A densitometric trace highlights a subset of RNase T1 protections. Also shown is the proposed secondary structure of the four-helix-junction RNA (blue nucleotides correspond to U2 snRNA, green to U6 snRNA and magenta to 5'SS nucleotides). RNase cleavages abrogated by binding to the core domain are indicated (red circles) as are those unaffected by binding (blue circles).

We investigated the interaction of the four-helix junction RNA with the core domain by mapping a crosslink formed between this RNA and the protein (Figure 3-4). A complex formed between the core domain and RNA containing a single ^{32}P label at G28 (corresponding to the 5'SS) was irradiated at 254 nm. The resulting crosslink (Figure 3-4A) was digested with RNase A and treated with cyanogen bromide. Cyanogen bromide (CN-Br) cleavage of the PRP8 domain IV core can yield three products corresponding to complete digestion (46, 76, and 100 aa) as well as two corresponding to partial digestion ($46+76 = 122$ aa; $46 + 100 = 146$ aa). For digestion of the crosslink, two bands are observed (Figure 3-4C, lane 2 blue and red arrows). The lower (blue arrow) migrates at ~ 3.9 kDa consistent with the fragment corresponding to 46 aa (+ negative charges from the appended five nucleotides). This is also consistent with the spacing from this to the upper band (red arrow) in comparison to the protein-alone profile (Figure 3-4B). Less complete digestion with CN-Br produces a profile where the upper band is much more prominent (data not shown) showing that this is a precursor to the lower band and consistent with the identification of the upper as 122 aa and the lower as 46 aa. Inspection of the core domain structure shows that this fragment maps to the structure formed from a4 through a6 on the surface corresponding to the RNA binding face of RNase H and including the a5 helix to which Konarska and coworkers mapped the 5'SS crosslink (Figure 3-4D) (Reyes et al., 1999).

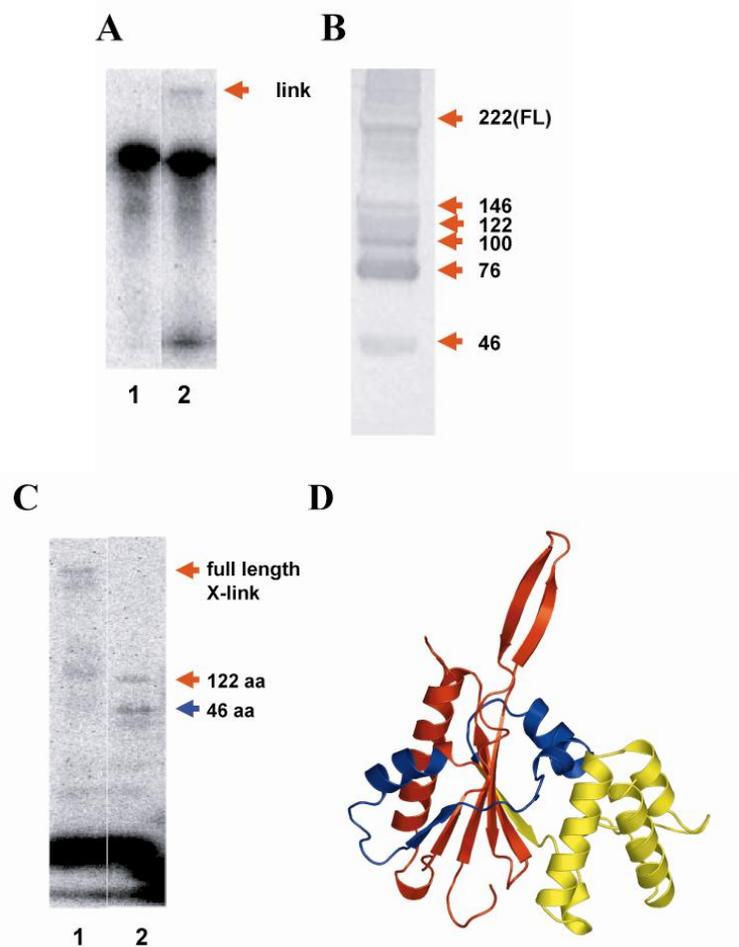
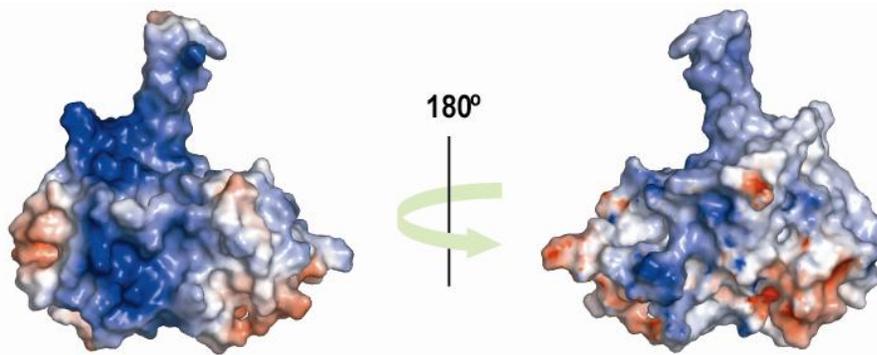


Figure 3-4. Mapping four-helix junction crosslink. A) Crosslinking of U2/U6/5'SS four-helix junction to core domain. Lane 1: RNA site-specifically labeled at G28, lane 2: labeled RNA incubated with core domain and irradiated at 254 nm. B) Silver-stained SDS-PAGE analysis of CN-Br cleavage of core domain showing fragments corresponding to partial (146 and 122 aa) and complete (100, 76, and 46 aa) cleavage. C) Mapping of four-helix junction RNA crosslink. Lane 1: crosslinking reaction digested with RNase A, lane 2: crosslinking reaction digested with RNase A followed by treatment with CN-Br. Red arrow indicates partial cleavage to 122 aa fragment, blue arrow indicates complete cleavage to 46 aa fragment. D) Ribbon diagram of core domain showing CN-Br fragments (red: 100 aa, blue: 46 aa, green, 76 aa).

The overall electrostatics of the core domain surface matching the RNA binding face of RNase H are consistent with RNA-binding (Figure 3-5A). The observed protections from proteolysis combined with extant crosslinking data (Reyes et al., 1999), as well as the crosslink mapping reported here, also strongly implicate a surface comprising the RNase H fold as well as the narrow channel

across the core domain as being involved in RNA binding (Figure 3-5B). This conclusion is supported by the fact that the majority of the core domain yeast alleles related to interaction with RNA map to this face of the domain (Figure 2-4). Finally, the RNase protection results show that specific regions of the four-helix junction RNA, including a model of the 5'SS, are protected upon interaction with the core domain (Figure 3-3).

A



B

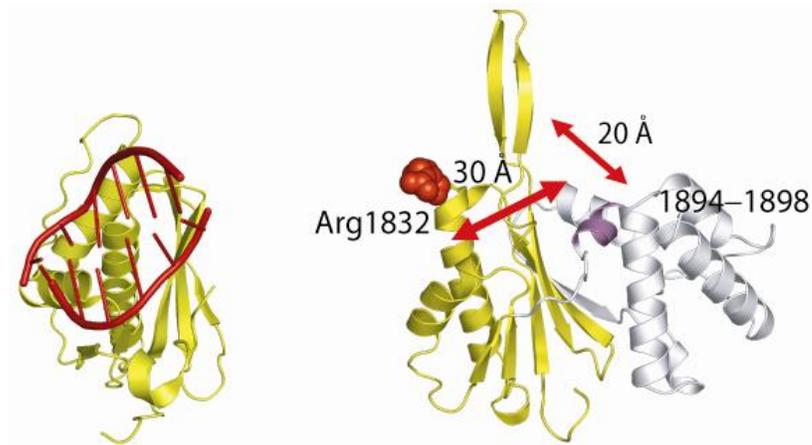


Figure 3-5. Representations of the proposed Prp8 domain IV core RNA binding surface. A) Surface electrostatics for two surfaces of the Prp8 domain IV core. Blue, basic; red, acidic; white, neutral. B) Comparison of RNA binding by RNase H and proposed core binding surface. Left, structure of an A-form duplex bound to *B. halodurans* RNase H (Nowotny et al., 2005) compared to the core domain fold. Right, proposed RNA binding surface of the core domain showing the location of the 5'SS crosslink (purple), dimensions of the associated channel and the position of Arg1832 (red), which is protected upon RNA binding. The RNase H fold is shown in yellow.

3-2.3. Implications of the core structure for interaction with the spliceosome active site

The structure of the hPrp8 domain IV core, containing an RNase H fold, suggests direct interaction with RNA components of the splicing machinery. This inference is consistent with crosslinking of this domain to the 5'SS (Reyes et al., 1999), previously characterized yeast alleles that demonstrate genetic interactions with snRNA/pre-mRNA sequences (Query and Konarska, 2004; Kuhn et al., 1999; Collins and Guthrie, 1999), and the RNA-binding studies reported here. The mapped location of the crosslink in a channel across the core structure, in close spatial proximity to the residues corresponding to the majority of yeast mutants suggests that RNA is bound across the corresponding surface. The dimensions of this channel are appropriate for binding an extended RNA duplex structure and the protection from proteolysis observed in this study suggests that this binding mode extends across the RNase H fold.

A large body of evidence convincingly argues that the catalytic core of the spliceosome is formed by U2/U6 snRNA structures and that the spliceosome is therefore, like the ribosome, a ribozyme (Nissen et al., 2000). In marked contrast to the ribosome however, there is also considerable evidence that protein components of the spliceosome are directly associated with key catalytic snRNA components as well as substrate pre-mRNA sequences throughout both spliceosome assembly and the transesterification steps of splicing.

Ribozymes characterized to date have been shown to employ a variety of mechanisms to promote catalysis including the use of divalent metal ions as

cofactors or to stabilize RNA structure (Doherty and Doudna, 2000). The observed RNase H fold of the core domain is an RNA binding surface. The absence of a bound metal in this structure and the fact that the putative site is blocked by a basic side chain in the core domain structure suggests that this domain has evolved away from such a metal-binding function. Alternatively, association of snRNA/pre-mRNA with the core domain, involving a displacement of this residue could position a bound metal ion at this site contributing, for example, to stabilization of RNA structure in the assembled spliceosome.

The exact identity of the RNA bound by the core domain remains unresolved. With respect to this question, the multiple phenotypes of mutant alleles characterized in yeast suggest the possibility of distinct RNA-binding functions at various stages of spliceosome assembly through catalysis. Indeed, an interaction between a C-terminal (RNase H + MPN domains) Prp8 construct and the U4/U6 snRNA structure has been reported, with an affinity similar to what we have observed for the 5'SS/U2/U6 4-helix junction (Maeder et al., 2009). The fact that distinct alleles have been characterized mapping to Prp8 domain 3, but corresponding to phenotypes associated with domain IV mutants, suggests that a true understanding of Prp8•RNA interaction will also involve high-resolution structural analysis of domain 3. Nevertheless, the gel mobility shift assays reported here are strongly supportive of RNA binding by this domain and a modest though significant hierarchy of affinities was observed. The specificity for a four-helix junction RNA including protection from RNase treatment argues

that the core domain recognizes a complex RNA structure such as this at the heart of the spliceosome.

3-3. Methods

3-3.1. RNA binding by hPrp8

Binding reactions (10 mM Tris, pH 8.0, 50 mM KCl, 5 mM MgCl₂, 10% glycerol (v/v)) containing 5' ³²P-labelled RNAs (see Supplementary Material) and protein concentrations ranging from 2-100 μM were incubated at room temperature for 30 min and then resolved by native PAGE (6%, 89:1 acrylamide/bisacrylamide; 50 mM Tris-glycine) at 110 V for 3 h. tRNA (2 μg) was added as a non-specific competitive inhibitor of binding.

RNAs used in gel mobility shift with PRP8 domain IV core

The following RNAs were synthesized by T7 transcription, end-labeled using γ-³²P-ATP, and used in binding assays with PRP8 domain IV core.

1) 5'SS pre-mRNA mimic

5'-AAGGUAAGUAdT-3' (Reyes et al., 1999)

2) Branch Point Sequence-U2 snRNA mimic (based on Schellenberg et al., 2006)

5'-GGGCGGUGGUGCCCUGGUGGGUGCUGACCGCCC-3'

3) 5'SS/U2/U6 4-helix junction model (based on Sashital et al. 2004)

5'-GGU UCU CGG CCU UUU GGC UAA GAU CAA GGU CUG UUU CGA
CAG AGA AGA UUA GCA UGG CCC CUG CGU AAG GAU GAC ACG CAA
AUU CGU GAA CC-3'

3-3.2. Limited proteolysis

hPrp8 domain IV core (100 pmol) was incubated in 4 μ l of buffer (10 mM Tris, pH 8.0, 50 mM KCl, 5 mM MgCl₂, 5 mM β -ME, 10% glycerol) with or without the four-helix junction RNA (125 pmol) for 20 min at 22°C, followed by addition of 1, 3, 10, or 30 ng of trypsin, or 10, 30, 100, 300 μ g of chymotrypsin. Samples were incubated at 22°C for 30 min, quenched by addition of 6 μ L denaturing SDS-PAGE loading dye (50 mM Tris pH 6.8, 16% glycerol, 3.2% SDS, 8% β -ME, 7M urea), and analysed by 16% SDS-PAGE.

3-3.3. RNase protection

RNase cleavage assays were performed on 5' end-labeled RNA either in the presence or absence of a saturating amount of hPrp8 domain IV core. RNA was digested with RNase T1 (2 U) for 2, 5, or 20 min or with RNase A (10⁻⁵ U) for 1, 3, or 6 min. A base hydrolysis ladder was produced by incubating 5' end-labeled RNA in 50 mM sodium carbonate, pH 9.2, 1 mM EDTA with 2 μ g tRNA at 55°C for 15 min. Reactions were extracted with phenol/chloroform, ethanol precipitated and separated on a 15% (19:1, polyacrylamide:bisacrylamide) denaturing sequencing gel run for 2 hrs at 70W.

3-3.4. Mapping four-helix junction crosslink

Protein RNA crosslinking

U2/U6/5'SS four-helix junction RNA (100,000 cpm), site-specifically labeled with ³²P 5' to G28, was incubated with PRP8 domain IV core under binding

conditions at room temperature for 20 min followed by irradiation at 254 nm with a 5 W UV lamp on ice for 15 min. Crosslinking to RNA was analyzed by 16% (29:1) SDS-PAGE (Figure 3-4A).

Mapping of crosslink

The CN-Br cleavage profile for hPrp8 domain IV core was generated by treatment of 0.5 μ l of 10 mg/mL protein with 2 μ l of 70 mg ml⁻¹ CN-Br in 80% formic acid for 2 hr. followed by analysis on a 14% (15:1) Tris-tricine gel (Figure 3-4B). For crosslink mapping, an aliquot of the crosslinking reaction (10, 000 cpm) was digested with RNase A (1 U) for 20 min at room temperature followed by treatment with CN-Br as described above. Following cleavage, the reactions were dried, washed with acetone to remove residual formic acid, then acetone precipitated. The resulting samples were separated on 14% (15:1) Tris-tricine gels run at 20W for 5 hours.

3-4. References.

Collins, C.A., and Guthrie, C. (1999). Allele-specific genetic interactions between Prp8 and RNA active site residues suggest a function for Prp8 at the catalytic core of the spliceosome. *Genes Dev.* **13**, 1970–1982.

Doherty, E.A., and Doudna, J.A. (2000). Ribozyme structures and mechanisms. *Annu. Rev. Biochem.* **69**, 597–615.

Kuhn, A.N., Li, Z., and Brow, D.A. (1999). Splicing factor Prp8 governs U4/U6 RNA unwinding during activation of the spliceosome. *Mol. Cell* **3**, 65–75.

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**, 3008–3020.

Maeder, C., Kutach, A.K., and Guthrie, C. (2009). ATP-dependent unwinding of U4/U6 snRNAs by the Brr2 helicase requires the C terminus of Prp8. *Nat. Struct. Mol. Biol.* **16**, 42–48.

Maroney, P.A., Romfo, C.M., and Nilsen, T.W. (2000). Functional recognition of 5' splice site by U4/U6.U5 tri-snRNP defines a novel ATP-dependent step in early spliceosome assembly. *Mol. Cell* **6**, 317–328.

Nissen, P., Hansen, J., Ban, N., Moore, P.B., and Steitz, T.A. (2000). The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**, 920–930.

Query, C.C., and Konarska, M.M. (2004). Suppression of multiple substrate mutations by spliceosomal prp8 alleles suggests functional correlations with ribosomal ambiguity mutants. *Mol. Cell* **14**, 343–354.

Reyes, J.L., Gustafson, E.H., Luo, H.R., Moore, M.J., and Konarska, M.M. (1999). The C-terminal region of hPrp8 interacts with the conserved GU dinucleotide at the 5' splice site. *RNA* **5**, 167–179.

Rupert, P.B., and Ferre-D'Amare, A.R. (2001). Crystal structure of a hairpin ribozyme-inhibitor complex with implications for catalysis. *Nature* **410**, 780–786.

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

Schellenberg, M.J., Edwards, R.A., Ritchie, D.B., Kent, O.A., Golas, M.M., Stark, H., Luhrmann, R., Glover, J.N., MacMillan, A.M. (2006). Crystal structure of a

core spliceosomal protein interface. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 1266-1271.

Teigelkamp, S., Newman, A.J., and Beggs, J.D. (1995). Extensive interactions of PRP8 protein with the 5' and 3' splice sites during splicing suggest a role in stabilization of exon alignment by U5 snRNA. *EMBO J.* **14**, 2602–2612.

Appendix I
Analysis of a stable hPrp8 fragment

I-1. Dissecting the domain IV core: 1831-1990

A stable fragment of hPrp8 comprising amino acids 1831-1990 was originally identified as a stably folded fragment by limited proteolysis of a larger domain IV construct 1689-1996 that was deaggregated using the detergent dodecyl maltopyranoside. The fragment 1831-1990 did not require detergent for deaggregation, but eluted from a S200 size exclusion chromatography column as an extended dimer in buffer containing 20 mM Tris, pH 8.0 and 100 mM KCl. The protein fragment had the unusual property of immediately precipitating when it was warmed above 4°C, and instantly going back into solution when it was re-cooled. We were able to crystallize an Se-Met substituted 1831-1990 fragment of hPrp8 at 4°C and determined its structure to 1.85 Å resolution using MAD methods. The structure was immediately disregarded as irrelevant to any function related to splicing because it was a dimer, and it is known that Prp8 exists in only one copy per U5 snRNP. Nonetheless, analysis of the structure, especially the dimerization interface composed largely of intertwined β-strands, proved to be an important starting point for the design of future constructs in our attempts to define the hPrp8 domain IV core. Specifically, the interacting β-strands near the N-terminus of each monomer point to missing β-strand structure at the N-terminus of the construct that could serve to stabilize folding of domain IV if included.

The structure of hPrp8 1831-1990 is comprised of two interacting molecules each consisting of two domains. The C-terminal domain (1921-1990) of each monomer contains 4 α-helices (α3-α6 in one molecule, α2-α5 in the other molecule). The N-terminal domain (1831-1920) is extended including two β-

strands (β 2- β 3) (1873-1885) and another β -strand (β 5) (1914-1918) that approach the helical bundle. Two additional β -strands (β 1 and β 4) (1861-1865 and 1897-1901) and two α -helices (α 1 and α 2) (1842-1851 and 1854-1859) are more distal relative to the helical bundle. The N-terminal domain of one monomer includes only helix 1, while the sequence corresponding to helix 2 is a loop. Helix α 4 of one molecule and the corresponding helix from the other molecule (α 3) are packed very close together in space, with the closest distance between helices from different monomers being around 7 Å. In addition, β -strand 5 from each molecule interacts with the β -strands 2 and 3 from the other molecule which results in formation of two β -sheets that stack one on the other, each with three antiparallel β -strands. The rest of the dimerization interface represents an unusual β -strand interchange, where β -strands 1 and 4 from each monomer are interdigitated (Figure I-1).

I-2. Methods

A GST•hPrp8 fusion protein encompassing domain IV residues 1689-1996 was expressed in *E. coli*, purified, and subjected to limited proteolysis with trypsin resulting in the identification of a stable ~18.8 kDa (hPrp8 1831-1993) fragment by MALDI mass spectrometry. We trimmed off the three C-terminal most amino acids of this construct which included a tyrosine residue. A cDNA representing hPrp8 1831-1990 was cloned into the *EcoRI* and *HindIII* sites of pMALc2x (NEB) using PCR primers to insert a TEV protease cleavage site between MBP and hPrp8 fragment 1831-1990.

Forward PCR primer sequence (a.a. 1831 START): 5' to 3' direction:

GCGCGCGAATTCGAAAATTTGTATTTTCAAGGTAAGCGTTTGGGGCAG

Reverse PCR primer sequence (a.a. 1990 END):

5'- GCGCGC AAG CTT TCA TCA GTC AGC CAA GAT CAG -3'

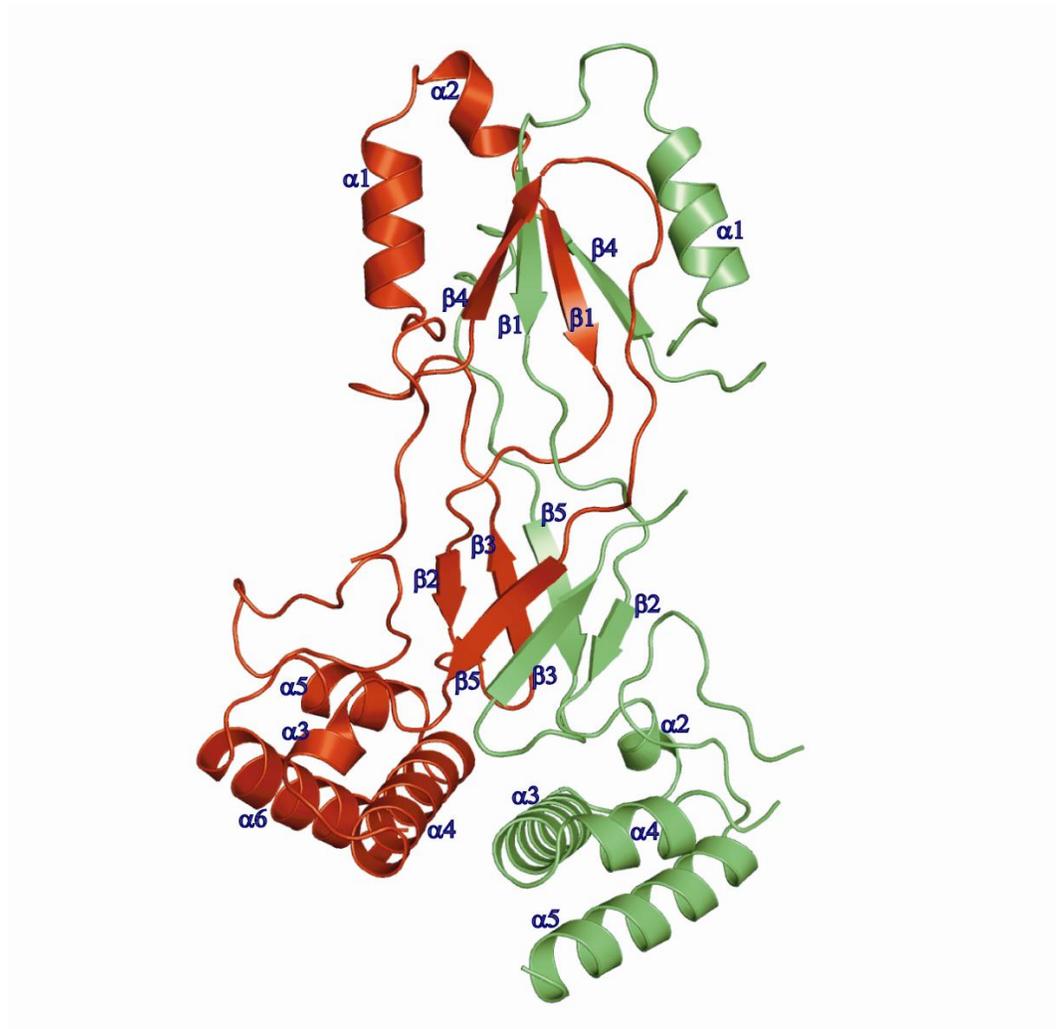


Figure I-1. Structure of hPrp8 domain IV 1831-1990. Ribbon diagram of the hPrp8 domain IV fragment amino acids 1831–1990. Different monomers are colored red and green to highlight the unusual dimerization interface.

Appendix II⁽¹⁾

**Characterization of a U2AF independent commitment complex (E')
in the mammalian spliceosome assembly pathway**

¹ Adapted from Kent et al., (2005). *Mol. Cell. Biol.* **25**, 233-240.

The work presented in this appendix represents a collaboration between the authors on the paper. O.K. performed all of the experiments except purification of the E' complex by size exclusion chromatography, which was designed and performed by D.R.

II-1. Introduction

Early recognition of pre-mRNA substrates by the splicing machinery proceeds through the formation of the ATP-independent CC in yeast or E complex in mammals (Jamison et al., 1992; Legrain et al., 1998; Seraphin and Rosbash, 1989). In both yeast and mammals, the 5'SS is initially recognized by U1 snRNP; the branch region and 3'SS are defined by the association of the branch binding protein (BBP) and Mud2 (yeast) or SF1 and the heterodimer U2AF (mammals). In the presence of ATP, U2 snRNP becomes stably associated with the pre-mRNA through a base pairing interaction between U2 snRNA and the branch region; subsequent recruitment of the U4/U6/U5 tri-snRNP and several snRNA rearrangements result in the formation of the mature spliceosome (Staley and Guthrie, 1998).

The mammalian E complex can be visualized either by gel filtration or by native gel electrophoresis (Das and Reed, 1999; Michaud and Reed, 1993). This complex contains U1 snRNP tightly associated with the 5'SS and SF1 and U2AF bound to the branch region, PPT, and 3'SS. As well, the E complex contains members of the SR protein family of splicing factors (Eperon et al., 1993; Fu, 1993), including SC35, which has been shown to bridge U1 snRNP and U2AF35, the small subunit of U2AF (Wu and Maniatis, 1993). In studies of mammalian prespliceosome assembly, it has been shown that SR proteins facilitate the earliest recognition of the 5'SS and 3'SS and promote formation of ATP-independent prespliceosomal complexes (Fu, 1993; Staknis and Reed, 1994).

In yeast, two CCs, CC1 and CC2, have been identified in U2 snRNP-depleted extracts by gel electrophoresis (Legrain et al., 1998; Seraphin and Rosbash, 1989; Zhang and Rosbash, 1999). Formation of the faster-mobility complex CC1 is dependent on the presence of a 5'SS and U1 snRNP. The lower-mobility complex CC2 is dependent on the presence of a 5'SS as well as a functional branch point sequence (BPS). Several components of CC2 have been identified. These include the splicing factor Mud2p, which has been shown to interact with the highly conserved yeast BPS and U1 snRNP (Abovich and Rosbash, 1997; Zhang and Rosbash, 1999). Mud2p has been identified as the yeast homolog of U2AF65, the large subunit of U2AF that interacts with the PPT 3' to the highly degenerate mammalian BPS (Zamore et al., 1992). The yeast BBP has been found to interact directly with Mud2p, which parallels the association of the mammalian splicing factor SF1 with U2AF65 (Berglund et al., 1998; Rutz and Seraphin, 1999). Taken together, these results suggest similar structural organizations of the yeast and mammalian commitment complexes.

It has been demonstrated that the initial recognition of the pre-mRNA and formation of the CC in both the yeast and mammalian systems are dependent on U1 snRNP. Depletion experiments have demonstrated that the stable association of U2 snRNP with the pre-mRNA requires the presence of U1 snRNP (Barabino et al., 1990). There is good evidence that U1 snRNP is involved in recruiting U2AF to the PPT, which in turn promotes the association of U2 snRNP with the branch region (Cote et al., 1995; Li and Blencowe, 1999). The stable association of U1 snRNP with the 5'SS involves base pairing interactions between the pre-

mRNA and the 5' end of U1 snRNA. Recently, it has been suggested that in yeast this RNA-RNA interaction is preceded by the sequence-specific recognition of the 5'SS by the U1 snRNP protein U1-C (Du and Rosbash, 2002; Will et al., 1996).

Here, we report the identification of a novel prespliceosomal complex designated E', which forms in U2AF-depleted HeLa nuclear extracts. The E' complex formation is dependent on SF1 and on U1 snRNA-5'SS base pairing and commits the pre-mRNA to the splicing pathway. RNA-protein cross-linking combined with immunoprecipitation demonstrates that the branch binding protein SF1 is associated with the pre-mRNA within the E' complex. The structure of complexes formed in wild-type and U2AF-depleted extracts has been probed with RNAs site-specifically derivatized with the hydroxyl radical source (*S*)-1-[*p*-(bromoacetamido)benzyl]-EDTA-Fe (Fe-BABE) (Kent and MacMillan, 2002). These experiments indicate that the proximity of the 5'SS and branch region that is observed in the E complex is also found in the E' complex; this proximity is dependent upon the presence of SF1 but not a functional branch point sequence and supports a model whereby U1 snRNP alone or in conjunction with SF1 directly recruits U2AF to the pre-mRNA substrate.

II-2. Results

II-2.1. Identification of a novel prespliceosomal complex

Work by Reed and coworkers has demonstrated that a complex formed on pre-mRNAs corresponding to the previously characterized E complex can be resolved from the nonspecific H complex by native agarose gel electrophoresis (Das and

Reed, 1999; Michaud and Reed, 1993). As part of a series of investigations of the earliest steps of spliceosome assembly, we specifically depleted HeLa nuclear extracts of U2AF (MacMillan et al., 1997; Zamore et al., 1992) and analyzed complex formation on the PIP85.B pre-mRNA by native agarose gel. Following incubation of pre-mRNA in depleted extract in the absence of ATP for 60 min at 30°C, we were able to observe the formation of a complex of slower mobility than H, distinct from E, which we designated E' (Figure II-1A; compare lanes 2 and 6). Formation of both the E and the E' complexes involves U1 snRNA-5'SS base pairing, since preincubation of extract with an antisense 2'-O-methyl oligonucleotide complementary to the 5' end of U1 snRNA abolished complex formation (Figure II-1A, lanes 3 and 7). Similarly, pretreatment of extract with RNase H and a DNA oligonucleotide directed against the 5' end of U1 snRNA abolished the formation of both E and E' complexes (data not shown). Blocking the interaction of U2 snRNP with the pre-mRNA using an antisense oligonucleotide had no effect on either E or E' formation (Figure II-1A, lanes 4 and 8). To obtain further evidence of E' complex identity and stability, complexes formed on pre-mRNA substrates incubated in U2AF-depleted extract were isolated by gel filtration (Reed and Chiara, 1999), and fractions that were obtained were analyzed by native agarose gel (Figure II-1B). We observed two distinct peaks that represent the isolation of the E' complex from the previously identified H complex.

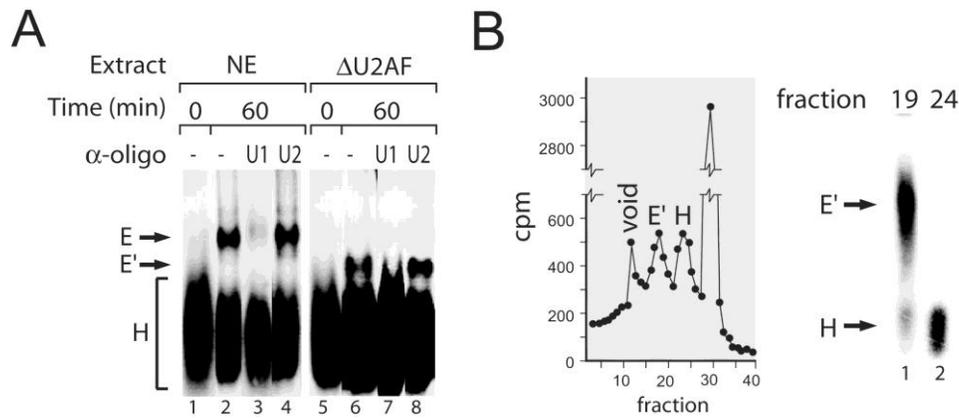


Figure II-1. Identification of a novel prespliceosome complex (E'). A) Native agarose gel shift analysis of complexes formed in nuclear extract (NE) and U2AF-depleted nuclear extract (Δ U2AF). Complex formation was analyzed in the presence and absence of anti-U1 and anti-U2 oligonucleotides. B) Gel filtration column purification of the E' complex from U2AF-depleted extract. (Left panel) Pre-mRNA elution profile of column-purified prespliceosome reaction. The void volume and E' and H complexes are indicated. (Right panel) Native agarose gel analysis of an aliquot of fraction 19 (E', lane 1) and fraction 24 (H, lane 2).

In order to examine the effect of the pre-mRNA branch region on E' complex formation, mutant RNAs were prepared in which the BPS was scrambled. E' complex was detected on a pre-mRNA substrate that contained a mutant BPS (Figure II-2A). This result is consistent with the observation that the formation of the E complex is not dependent on a functional BPS (Figure II-2A) (Champion-Arnaud et al., 1995). We next tested whether SF1 was required for E' complex formation by analyzing complex formation in SF1-depleted extracts. Depletion of SF1 from wild-type or U2AF-depleted nuclear extracts prevented formation of both the E and E' complexes, as assayed by native gel (Figure II-2B, lanes 2 and 4). We next asked whether recombinant SF1 could reconstitute E'

formation in U2AF- and SF1-depleted extract (Figure II-2C). Pre-mRNAs incubated under E' complex-forming conditions in U2AF- and SF1-depleted nuclear extract were unable to form E' complex (Figure II-2C, lane 2), but recombinant SF1-C4 (a functional SF1 mutant) (Rain et al., 1998) was able to restore E' formation (Figure II-2C, lane 3). Addition of recombinant U2AF65 had no effect on the doubly depleted extract (Figure II-2C, lane 4).

In order to examine the effect of the pre-mRNA sequence elements on E' complex formation, mutant RNAs were prepared in which one or more sequence elements were eliminated. We challenged E' complex formation on wild-type body-labeled pre-mRNA in U2AF-depleted nuclear extract that was preincubated with a 500-fold excess of cold mutant RNAs (Figure II-2D). Competitor RNAs containing a 5'SS inhibited E' formation, but no inhibition was observed with a substrate that lacked this sequence element (Δ 5'SS), indicating that a 5'SS is both necessary and sufficient for formation of the E' complex.

Since the E' complex requires a 5'SS to form, we wanted to confirm that the E' complex contains U1 snRNP. Therefore, both E and E' complexes were purified from agarose gels, and the isolated RNA was subjected to reverse transcription with the appropriate primers. U1 snRNA was found to be present in both the E and E' complexes (Figure II-2E), indicating that the E' complex contains U1 snRNP; we were unable to detect U2 snRNA in either complex (Figure II-2E).

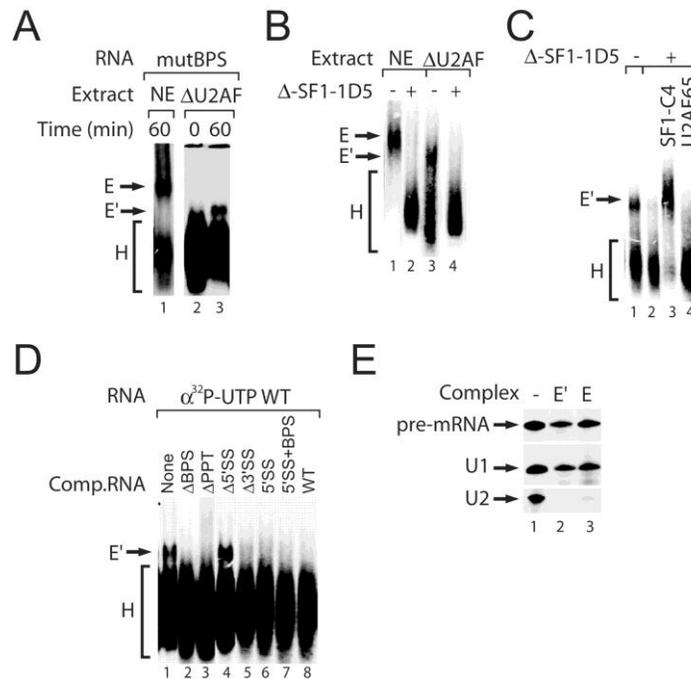


Figure II-2. Formation of the E' complex requires SF1, and the complex contains U1 snRNP. A) E' complex formation in the absence of a functional BPS within the pre-mRNA. Complexes were formed on pre-mRNAs containing a scrambled BPS (mutBPS) in wild-type nuclear extract (NE) or U2AF-depleted nuclear extract (Δ U2AF). B) E' complex formation requires the branch binding protein SF1. Native agarose gel shift of complexes formed in wild-type (NE) or U2AF-depleted (Δ U2AF) nuclear extract which had been SF1 immunodepleted (+). C) SF1 reconstitutes formation of E' in depleted extracts. Native agarose gel shift of complexes formed in U2AF-depleted nuclear extract which had been SF1 immunodepleted (+) and reconstituted with recombinant SF1-C4 mutant (lane 3) or U2AF65 (lane 4). D) E' complex formation is dependent on the presence of a 5'SS within the pre-mRNA. Complexes were formed on a body-labeled ($[\alpha\text{-}^{32}\text{P}]\text{UTP}$) wild-type (WT) pre-mRNA in extract saturated with unlabeled mutant competitor RNA. E) E' complex contains U1 snRNP. Shown are results from RT primer extension of RNAs purified from isolated complexes (lanes 2 and 3) and RNA isolated from an aliquot of precipitated nuclear extract (lane 1). The pre-mRNA, U1 snRNA, and U2 snRNA are indicated.

II-2.2. Relationship of E' to the prespliceosome

The formation of the E complex commits the bound pre-mRNA to the splicing pathway. Thus, prior formation of E complex at 30°C on a ^{32}P -labeled substrate results in efficient splicing of the labeled RNA even after the addition of a 500-fold excess of competitor RNA (Legrain et al., 1988). We tested the commitment ability of both wild-type and U2AF-depleted extracts by preincubation of pre-

mRNA in extract followed by a chase with extract containing an excess of competitor RNA. Direct incubation of a radiolabeled pre-mRNA substrate in extract in the presence of increasing concentrations of unlabeled competitor RNA resulted in the decrease of spliced products (Figure II-3A, lanes 1 to 4, and B). When E complex was formed by preincubation of substrate RNA in wild-type extract, followed by addition of an extract mix containing competitor, no inhibition of splicing was observed (Figure II-3A, lanes 5 to 8). Commitment correlated with E complex formation, since preincubation of a pre-mRNA in wild-type or U2AF-depleted extract on ice followed by a chase with wild-type extract mix containing competitor RNA resulted in significantly decreased splicing (Figure II-3A, lanes 9 to 12 and 13 to 16). We tested whether the formation of E' commits pre-mRNA to the splicing pathway by preincubation of labeled substrate RNA in U2AF-depleted extract under E' complex-forming conditions followed by a chase with wild-type extract containing competitor (Figure II-3A, lanes 17 to 20). Under these conditions, normal splicing was observed, indicating that the E' complex is a functional commitment complex (Figure II-3B).

We investigated the relationship between the E' and E complexes by performing E' on a substrate pre-mRNA, chasing with either wild-type extract or recombinant U2AF, and then analyzing the complexes by native agarose gel electrophoresis (Figure II-3C). Following the formation of the E' complex, addition of wild-type nuclear extract efficiently chased E' into the E complex (Figure II-3C, lane 2); however, addition of an aliquot of U2AF-depleted extract failed to shift E' (data not shown). Furthermore, addition of U2AF65 alone or

recombinant U2AF efficiently chased E' into the E complex (Figure II-3C, lanes 4 and 6), suggesting that E' is a functional precursor to E in the commitment of this pre-mRNA to the splicing pathway.

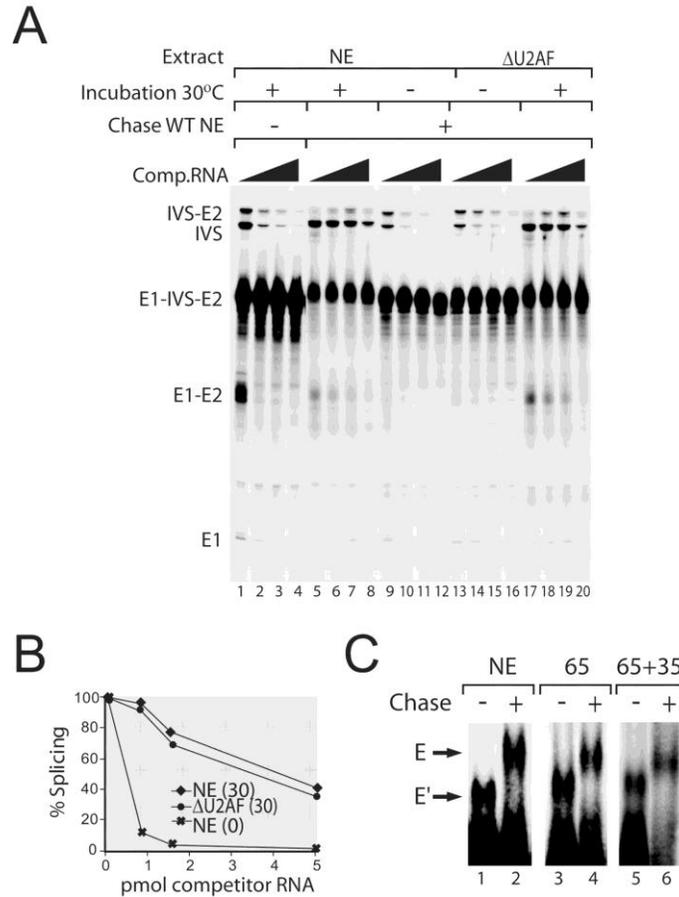


Figure II-3. Commitment of pre-mRNA to splicing in the absence of U2AF. A) The E' complex contains commitment ability. RNAs were preincubated in nuclear extract (NE) or U2AF-depleted nuclear extract (Δ U2AF) and chased with wild-type nuclear extract saturated with increasing competitor RNA (0.8, 1.4, or 5 pmol). Splicing substrates, products, and intermediates are indicated. B) Quantification of splicing products (from panel A) generated after the formation of commitment complexes at 30°C under E complex-forming conditions in wild-type extract [NE(30)] or E' complex-forming conditions in U2AF-depleted extract [Δ U2AF(30)] alongside a control reaction in wild-type extract incubated on ice [NE(0)]. C) The E' complex is a precursor to the E complex. Preformed E' complex chased with nuclear extract (NE; lanes 1 and 2), U2AF65 (lanes 3 and 4) or U2AF65/35 heterodimer (lanes 5 and 6) shifts E' complex to E complex.

II-2.3. Detection of BBP in the E' complex

Since pre-mRNAs containing a mutant BPS were incorporated into both the E and E' complexes (Figure *II-2A*), we wanted to determine whether SF1 associated with the pre-mRNA branch region in the absence of U2AF. We prepared a pre-mRNA modified at the BPS with the photo-cross-linker benzophenone (MacMillan et al., 1994) and containing a unique ^{32}P label to facilitate the analysis of cross-links (Figure *II-4A*). E' and E complexes were formed on derivatized pre-mRNAs with the same efficiency as that observed with unmodified pre-mRNAs (data not shown), and the resulting complexes were photolyzed to cross-link proteins bound proximal to the branch region (Figure *II-4B*). Immunoprecipitation with an antibody specific to SF1 enriched an 80-kDa cross-link, confirming the association of the BBP with the pre-mRNA in U2AF-depleted as well as wild-type HeLa extract (Figure *II-4B*, lanes 2 and 4). We determined the cross-links specific to E and E' by performing SDS-PAGE with cross-linked proteins eluted from purified complexes (Figure *II-4C*). In the E complex, three proteins with molecular weights of 80, 65, and 35 kDa were efficiently cross-linked to the branch region: these are SF1 and most likely the two subunits of U2AF, U2AF65 and U2AF35 (Figure *II-4C*, lane 2). The cross-link to the 3'SS binding protein U2AF35 is consistent with a U2AF-mediated bending of the pre-mRNA (Kent et al., 2003). In the E' complex, only the cross-link to the BBP SF1 was detected (Figure *II-4C*, lane 1). This result suggests that SF1 can associate with the pre-mRNA in the absence of U2AF.

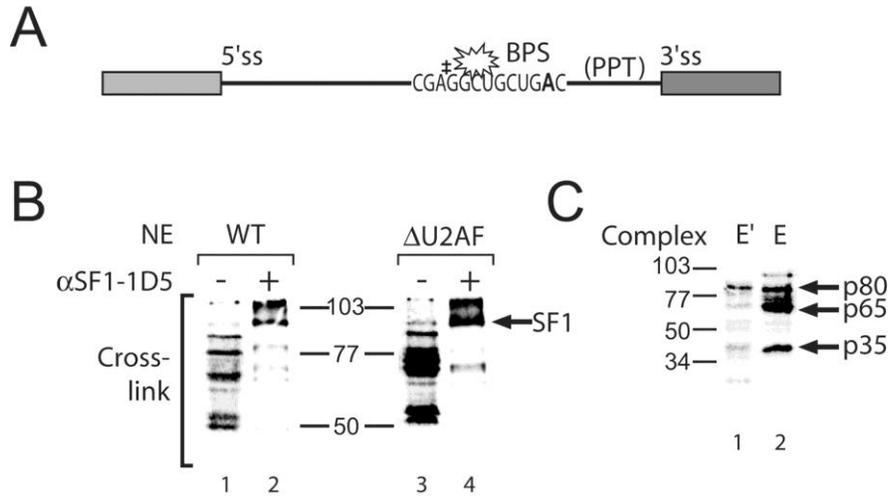


Figure II-4. E' complex contains SF1 at the branch point. A) Pre-mRNA containing a unique ^{32}P label (\ddagger) was site-specifically modified with benzophenone (star) at the BPS. B) Cross-linking SF1 to the branch region under E' complex conditions. Modified pre-mRNAs were incubated in wild-type (WT) or U2AF-depleted (ΔU2AF) nuclear extracts under E (lanes 1 and 2) and E' (lanes 3 and 4) complex-forming conditions, irradiated with UV light, treated with RNase A, and analyzed by SDS-PAGE before (lanes 1 and 3) and after (lanes 2 and 4) immunoprecipitation with anti-SF1 antibody (anti-SF1-1D5). The position of SF1 is indicated. C) E' complex contains SF1. Proteins cross-linked to the BPS were purified from agarose gel-isolated E' complex (lane 1) and E complex (lane 2). Indicated are molecular weight markers and cross-links to p80, p65, and p35 proteins.

II-2.4. Directed hydroxyl radical probing of pre-mRNA structure in the E' complex

We probed the structure of RNA bound in the E' complex using pre-mRNA that was site-specifically derivatized with a directed hydroxyl radical probe, Fe-BABE, tethered to the pre-mRNA (Kent and MacMillan, 2002). Diffusible radicals produced from a tethered iron-EDTA moiety are excellent probes of local structure, since they are capable of cleaving only the phosphodiester backbone within 10 to 20 Å from their site of generation. Using a wild-type RNA construct that had been derivatized with Fe-BABE at either the branch region (WT148) or the 3'SS (WT179), we probed complexes formed in wild-type and U2AF-depleted

extracts (Figure II-5A and B). Reactions were incubated under E or E' complex formation conditions (lacking ATP and MgCl₂), followed by Fe-BABE-mediated cleavage and analysis of cleavage products by RT primer extension. As we have previously reported, under E complex-forming conditions, probes tethered to either the branch region or 3'SS generated strong cleavages at the 5'SS (Figure II-5B, lanes 5 and 8), indicating the proximity of these sequence elements in the E complex. No cleavages at the 5'SS were observed from either derivatized pre-mRNA when incubated in nuclear extract on ice (Figure II-5B, lanes 4 and 7). Interestingly, strong 5'SS cleavages were observed under E' complex-forming conditions in U2AF-depleted extract using pre-mRNAs derivatized at the branch region (Figure II-5B, lane 6). In contrast, no 5'SS cleavage was observed under the same conditions using the probe derivatized at the 3'SS (Figure II-5B, lane 9). The same panel of cleavages was also generated independently of a functional branch region (Figure II-5C). Pre-mRNA derivatized analogously to WT148 but with a scrambled branch point sequence was incubated under E or E' complex-forming conditions and probed by Fe-BABE-mediated cleavage. As reported previously, 5'SS cleavages are still observed in substrates lacking a functional branch sequence (Figure II-5C, lane 3) (Kent and MacMillan, 2002). Similarly, the same pattern of cleavages was observed from the mutant branch region pre-mRNA that was incubated under E' complex-forming conditions (Figure II-5C, lane 4), suggesting that a functional branch region is not required for the proximity of the 5'SS-branch region and consistent with the weak dependence of E complex formation on the branch sequence (Champion-Arnaud et al., 1995).

We wanted to determine whether the observed cleavages at the 5'SS were dependent on the presence of the BBP SF1. Previously, it was found that the depletion of U1 snRNP from nuclear extract abolished the cleavages at the 5'SS that were observed with pre-mRNAs derivatized at either the branch region or the 3'SS (Kent and MacMillan, 2002). This result correlates well with the observation that U1 snRNP is required for E complex formation. Consistent with this result, experiments with branch region-derivatized RNAs in U1 snRNP- and U2AF-depleted extracts did not show cleavages at the 5'SS, and assembly of the E' complex was also abolished in this extract (data not shown). We then specifically depleted SF1 from wild-type and U2AF-depleted nuclear extracts and probed Fe-BABE-modified pre-mRNAs for complex formation and 5'SS cleavage. Although the E complex could not be detected on native gels in SF1-depleted extract, strong cleavages were still observed at the 5'SS (Figure II-5D, lane 2). However, depletion of both SF1 and U2AF abolished E' complex formation and severely diminished the observed cleavages at the 5'SS (Figure II-5D, lane 4). This result may be interpreted as a redundant requirement for SF1/U2AF in tethering the branch region and 5'SS; either of these factors bound to the branch region can interact with U1 snRNP at the 5'SS, resulting in the proximity of these two pre-mRNA sequence elements.

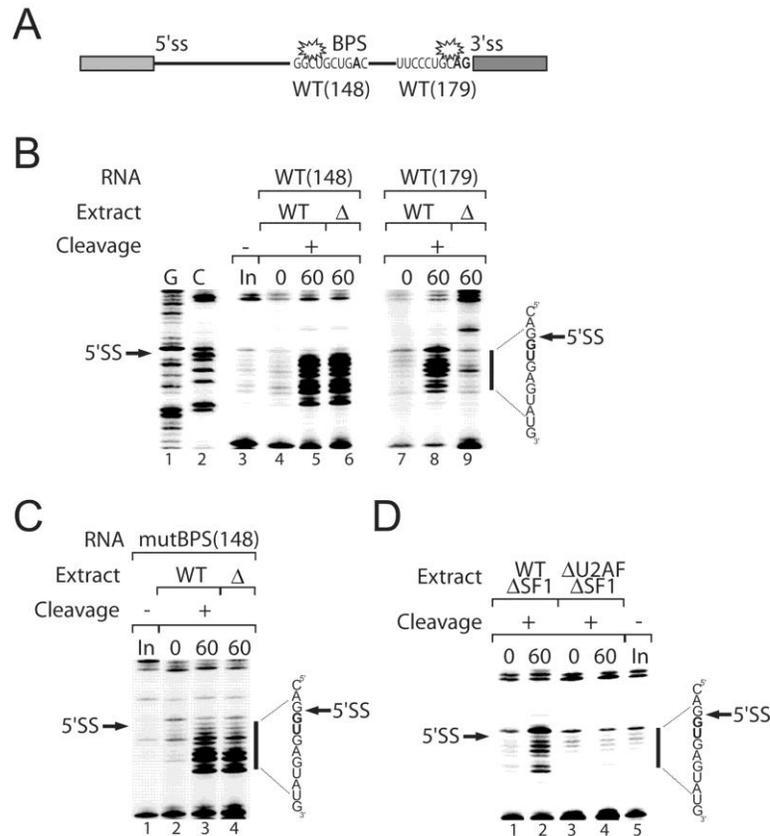


Figure II-5. Directed hydroxyl radical cleavage of pre-mRNA within the ATP-independent prespliceosome complexes. A) Pre-mRNA substrate derivatized with Fe-BABE (star) at the branch region [WT(148)] or 3'SS [WT(179)]. B) Analysis of pre-mRNA cleavage at the 5'SS in E and E' complexes. Reactions were performed in nuclear extract (WT) and U2AF-depleted (Δ) nuclear extract probed with WT148 (lanes 4 to 6) or WT179 (lanes 7 to 9) pre-mRNAs. Fe-BABE cleavage reactions were initiated after incubation at 30°C for 0 or 60 min. Reverse transcription of cleavage reactions were compared to input RNA (In; lane 3) for location of reverse transcription stops. G and C sequencing is shown (lanes 1 and 2). C) Analysis of pre-mRNA cleavage at the 5'SS with mutant BPS pre-mRNAs. Pre-mRNAs derivatized at position 148 containing a scrambled branch region were probed in wild-type (WT) and U2AF-depleted (Δ) nuclear extracts. Fe-BABE cleavage reactions were initiated after incubation at 30°C for 0 or 60 min. Cleavage reactions (lanes 2 to 4) were compared to input RNA (In; lane 1) for location of reverse transcription stops. D) Analysis of pre-mRNA cleavage at the 5'SS in SF1-depleted extracts. Reactions were performed in SF1-depleted nuclear extract (WT Δ SF1) or Δ U2AF SF1-depleted extract (Δ U2AF Δ SF1) and probed with WT148 pre-mRNA. Fe-BABE cleavage reactions were initiated after incubation at 30°C for 0 or 60 min. Reverse transcription lanes of cleavage reactions were compared to input RNA (In; lane 5) for location of reverse transcription stops. The location of and sequence around the 5'SS are indicated. Regions of significant cleavage are represented by the vertical bar.

II-3. Discussion

We have detected the formation of a novel CC, designated E', on pre-mRNA substrates that were incubated in U2AF-depleted extracts. The E' complex forms in the absence of ATP and contains U1 snRNP and SF1 (Figure II-6). E' complex formation requires a 5'SS and functional U1 snRNP as well as SF1. Our results suggest that E' is analogous to the yeast commitment complex CC1 and may be similar to the yeast CC1 in several ways: formation of both the E' complex and CC1 requires a functional 5'SS and the presence of U1 snRNP. The formation of the E' complex does not require a functional branch sequence mirroring the requirement for formation of CC1. Furthermore, since the overall pathways of spliceosome assembly in yeast and mammals are highly conserved (Abovich and Rosbash, 1997), it is likely that formation of the E complex and the yeast CC follow similar assembly pathways. The observation of E' in HeLa extracts suggests a further parallel with the yeast splicing system: the assembly of the mammalian spliceosome may proceed through the formation of two ATP-independent CCs.

Experiments with the yeast system have demonstrated that the commitment of pre-mRNAs to splicing is dependent on U1 snRNP and that the complexes formed in U2 snRNP-depleted extracts are intermediates in spliceosome assembly (Legrain et al., 1988, Seraphin and Rosbash, 1989). Similarly, our results suggest that commitment of mammalian pre-mRNAs to splicing is dependent upon the interaction of U1 snRNP with the 5'SS and the

interaction of SF1 at the branch region followed by the recruitment of U2AF to the PPT and 3'SS to give the E complex (Figure II-6).

Although no sequences 3' to the BPS in yeast have been shown to be essential for CC formation, the formation of the mammalian E complex is dependent on the presence of a functional PPT. However, the roles of SF1/U2AF and yeast BBP/Mud2p in CC formation may be more similar in the mammalian and yeast systems than previously reported. It has been reported that yeast commitment complex formation (CC2) can be detected in Mud2p extracts but not in extracts lacking BBP (Abovich and Rosbash, 1997; Abovich et al., 1994). This result and other data (Abovich and Rosbash, 1997) suggest that BBP is more important to commitment complex formation than Mud2p. Our results demonstrate the formation of the E' complex in U2AF-depleted extracts but not in SF1-depleted extracts. These observations of HeLa nuclear extract parallel the previously reported observations of yeast extract and suggest that not only are bridging interactions conserved between the yeast and mammalian systems (Abovich and Rosbash, 1997), but the temporal recognition of the pre-mRNA and the formation of the CCs may also be conserved. These observations are somewhat surprising, given that yeast splicing is dependent on a functional branch sequence and mammalian splicing is dependent on a functional PPT.

The precise determinants of commitment in the prespliceosome are not fully understood. The studies outlined here indicate that commitment occurs in the absence of U2AF but confirm that the commitment ability of E' requires U1 snRNP association with the 5'SS. Although the U1 snRNP-pre-mRNA

association in E' requires base pairing between the U1 snRNA and the 5'SS, actual commitment probably precedes this association. For example, a pre-mRNA that was incubated in extract where the 5' end of U1 snRNA has been removed by RNase H digestion does not splice when added to a reaction containing wild-type U1 snRNP, suggesting "commitment" in the digested extract (data not shown).

U2 snRNP is recruited to the CC through its association with U2AF independent of U1 snRNP at the 5'SS. U2 snRNP can be recruited to a minimal RNA substrate that contains a branch region and PPT forming a complex in nuclear extract, Amin, that resembles the A complex but lacks U1 snRNP (Query et al., 1997). However, this Amin complex rapidly dissociates in the presence of ATP, and therefore stable U2 snRNP recruitment to the assembling spliceosome must involve other factors that are not present in the Amin complex. A recent report has demonstrated that although commitment to splicing involves ATP-independent CC formation, commitment to splice site choice takes place in the prespliceosome A complex (Lim and Hertel, 2004). Therefore, commitment to spliceosome formation is a separate event from commitment to splice site choice, which may take place concomitant with U2 snRNP stabilization at the branch region.

The role of U1 snRNP in CC formation remains unclear. In the canonical pathway, prespliceosome formation is dependent on U1 snRNP to form the E complex (and the E' complex described here). However, a class of pre-mRNAs containing cis-acting elements that promote spliceosome formation in the absence of U1 snRNP has been identified (Crispino et al., 1996). In these pre-mRNAs, the

recruitment of U2 snRNP to the branch region involves additional factors: specifically, members of the SR protein family which can reconstitute splicing in U1 snRNP-depleted extracts (Crispino et al., 1994).

The SR proteins are essential splicing factors with characteristic RNA binding domains and a serine/arginine-rich motif which associates with the pre-mRNA at early stages and throughout spliceosome assembly (Fu, 1993; Roscigno and Garcia-Blanco, 1995; Staknis and Reed, 1994). The SR proteins have been shown to influence splice site selection (Eperon et al., 1993; Wu and Maniatis, 1993) as well as commit pre-mRNAs to the splicing pathway (Fu, 1993). Indeed, it has been shown that different SR proteins are capable of committing pre-mRNAs to splicing with pronounced substrate specificity (Fu, 1993). Furthermore, the role of SR proteins may be to provide the initial recognition of the 5'SS and 3'SS (Staknis and Reed, 1994). For example, SF2/ASF has been shown to increase U1 snRNP binding to the 5'SS (Eperon et al., 1993, Kohtz et al., 1994) and SC35, which bridges U1 snRNP and U2AF35 cross-links to the 3'SS in an E3' complex (Staknis and Reed, 1994; Wu and Maniatis, 1993). As well, SC35 has also been shown to activate 5'SS usage independent of functional U1 snRNP (Tarn and Steitz, 1994). SR proteins may contribute to the formation and commitment ability of the E' complex, such as initial U1 snRNP recruitment and later U2AF recruitment to the PPT to form the E complex.

The branch binding protein SF1 interacts specifically with the pre-mRNA branch region; however, its role in mammalian prespliceosome assembly has been difficult to define. Work from several groups suggests that SF1 contributes to

branch point definition and acts cooperatively with U2AF to define not only the branch region but also the 3'SS (Berglund et al., 1998; Rutz and Seraphin, 1999). Our results indicate that SF1 is involved in the formation of the E' complex and may help promote the stable association and proximity of the 5'SS and branch region.

A proximity of both the branch region and 3'SS with the 5'SS in the E complex has been previously observed by use of pre-mRNA probes that were modified with Fe-BABE (Kent and MacMillan, 2002). Similar experiments, under E' complex-forming conditions in the absence of U2AF, demonstrated that U2AF is required for 5'SS and 3'SS association in the E complex. However, our data indicate that the 5'SS and branch region are close to one another in E', and this proximity probably reflects bridging interactions between U1 snRNP and the BBP SF1 (Figure II-6) (Reed, 2000). Interactions between SF1, the mammalian Prp40-like protein FBP11, and U1 snRNP have been proposed to bridge the 5'SS and the branch region in the E complex (Abovich and Rosbash, 1997; Reed, 2000); a similar structural organization is likely present in the E' complex.

The fact that no 5'SS cleavages are observed in E' with a pre-mRNA probe that was derivatized at the 3'SS is consistent with the role of U2AF as a 3'SS recognition factor which brings that portion of the RNA into direct proximity with the branch region and hence the 5'SS (Kent et al., 2003). It has been demonstrated previously that U2AF35 interacts with the U1 snRNP specific protein U1-70K via an interaction with the SR protein SC35 that serves to bridge the two ends of the intron (Wu and Maniatis, 1993). It is possible that this

interaction serves to recruit U2AF to the PPT or that the proximity of the 5'SS and 3'SS is a consequence of the 5'SS-branch interaction that was first established in the E' complex.

U1 snRNP association with the pre-mRNA defines the 5'SS and the interaction of U2AF65 with the PPT, and specifically, the direct recognition of the AG dinucleotide by U2AF35 defines the 3'SS (Merendino et al., 1999; Wu et al., 1999; Zhang and Rosbash, 1999; Zorio and Blumenthal, 1999). The proximity of the 5'SS and branch region in the E' complex are consistent with a U1 snRNP- and SF1-facilitated recruitment of U2AF to the pre-mRNA (Cote et al., 1995; Li and Blencowe, 1999) and argue for a very early definition of the intron that precedes this association.

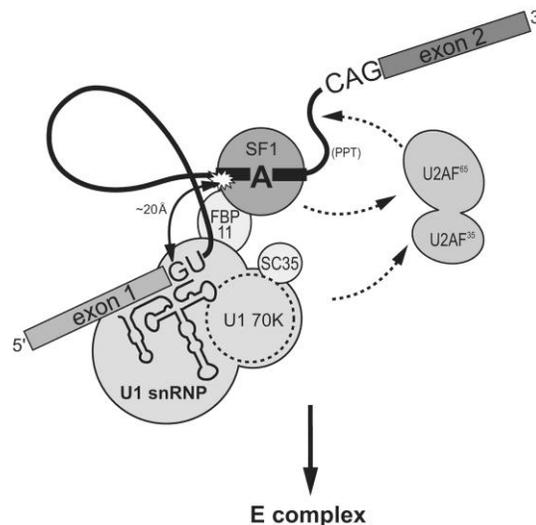


Figure II-6. Assembly of the E complex proceeds through early recognition of the 5'SS-branch region followed by recruitment of U2AF to the PPT. Pre-mRNAs are committed to prespliceosome assembly by the association of U1 snRNP with the 5'SS. SF1 interacts with the branch point sequence and may contact FBP11, providing a bridging interaction between SF1 and U1 snRNP. This interaction results in the proximity of the BPS and the 5'SS (curved arrow). U1 snRNP and/or SF1 recruit U2AF to the PPT (dotted arrows) resulting in structural reorganization of the 3'SS and E complex formation.

II-4. Methods

II-4.1. Synthesis of pre-mRNA substrates

Wild-type and mutant pre-mRNAs were synthesized by in vitro T7 transcription using double-stranded PCR products derived from the PIP85.B plasmid (Query et al., 1994) as templates. Transcriptions were carried out under standard conditions in the presence of 100 μ Ci of [α^{32} P]UTP; products were purified directly by 8% polyacrylamide gel electrophoresis (PAGE) (19:1).

In order to prepare derivatized pre-mRNAs, the 10-nucleotide oligomers 5'-GG(C*)UGCUGAC-3' and 5'-UUCCCUG(C*)AG-3', corresponding to the pre-mRNA branch site and the 3'SS, respectively, containing a site-specific N4-functionalized cytosine (C*), were synthesized, ligated into full-length pre-mRNAs, and derivatized with benzophenone or Fe-BABE as described previously (Kent and MacMillan, 2002; MacMillan et al., 1994). Branch sequence mutants were created in an analogous fashion with a 10-nucleotide oligomer, 5'-GGGUCGUCAG-3', corresponding to a scrambled branch region.

II-4.2. E and E' complex formation

Pre-mRNAs (50 x 10³ to 100 x 10³ cpm) were incubated in 10- μ l reactions containing 25% HeLa nuclear extract or U2AF-depleted HeLa nuclear extract (MacMillan et al., 1997; Zamore et al., 1992), 60 mM KCl, and 4 U of RNase inhibitor. Following incubation at 30°C for 25 to 60 min, reactions were immediately loaded onto a 1.2% agarose-0.5x Tris-borate-EDTA gel. Dried gels were exposed to a Molecular Dynamics phosphor screen and scanned with a

Molecular Dynamics Storm 840 Phosphorimager. Chase experiments were performed by performing E' complex in U2AF-depleted extract as described above followed by addition of an aliquot of the reaction to wild-type nuclear extract and further incubation at 30°C for various times. Competition experiments were performed as described above, except that an aliquot of the preformed complex was added to a reaction containing wild-type nuclear extract and 0.8, 1.4, or 5 pmol of competitor RNA, and the whole was further incubated at 30°C. Reactions blocking U1 or U2 snRNA were performed in extracts preincubated in the presence of 5 to 20 μ M 2'-O-methyl antisense oligonucleotide (Dharmacon) as described previously (Blencowe and Lamond, 1999). Add-back experiments were performed in U2AF-depleted extract with 60 μ M recombinant U2AF65 or U2AF65/35 dimer. SF1 add-back experiments contained 60 μ M recombinant SF1-C4 mutant (Rain et al., 1998). His6-tagged U2AF65, U2AF35, and SF1-C4 were overexpressed in Escherichia coli cells and purified by Ni-nitrilotriacetic acid chromatography as described elsewhere (Kent et al., 2003).

II-4.3. Gel filtration column purification of E' complex

Pre-mRNAs (50 x 10³ to 100 x 10³ cpm) were incubated under the E' complex-forming conditions described above in 100- μ l reactions. Reactions were loaded directly onto a Sephacryl S500 gel filtration column (40-ml column volume; Amersham Pharmacia Biotech) equilibrated with 20 mM Tris, pH 7.8, 0.1% Triton X-100, 60 mM KCl, and 2.5 mM EDTA (as described in reference Reed and Chiara, 1999). Columns were run at 50 μ l min⁻¹, and 40 to 50 1-ml fractions

were collected. An aliquot of each fraction was used for scintillation counting to generate an elution profile. Fractions were then concentrated to 20 μ l by using Ultrafree-MC centrifugal filters (molecular weight cutoff, 10,000; Millipore) and subjected to native agarose gels as described above.

II-4.4. Splicing assays

Pre-mRNAs (50 x 10³ to 100 x 10³ cpm) were incubated in 20- μ l reactions containing 40% HeLa nuclear extract, 2 mM MgCl₂, 60 mM KCl, 1 mM ATP, 5 mM creatine phosphate (CP), and 4 U of RNase inhibitor. Competition assays were performed in the presence of competitor RNA as described above. Commitment assays were performed by incubating pre-mRNA (50 x 10³ to 100 x 10³ cpm) under E or E' complex-forming conditions as described above. Aliquots were then added to a splicing mix containing ATP/CP and MgCl₂ with or without competitor RNA. Following incubation at 30°C, splicing reactions were digested with 40 μ g of proteinase K in the presence of 20 μ g of tRNA at 55°C for 20 min. Reactions were extracted with phenol-chloroform-isoamyl alcohol, ethanol precipitated, and then subjected to denaturing PAGE (15% acrylamide; 19:1). Dried gels were exposed to a Molecular Dynamics phosphor screen and scanned with a Molecular Dynamics Storm 840 Phosphorimager.

II-4.5. UV cross-linking and immunoprecipitation

Cross-linking was performed in 20- μ l reactions containing 25% HeLa nuclear extract or U2AF-depleted nuclear extract, which were incubated under E or E'

complex-forming conditions and then irradiated at 302 nm on ice for 20 min. RNase A (2.5 U) was added, and reactions were digested at 37°C for 30 min. For immunoprecipitations, photolyzed samples were then added to 10 µl of protein G-Sepharose beads bound to anti-SF1-1D5 antibodies in 400 µl of buffer containing 10 mM Tris-HCl, pH 8, 150 mM KCl, 20 mM MgCl₂, and 0.5 mM dithiothreitol. Samples were rotated at 4°C for 2 h, washed with buffer three times, and subjected to sodium dodecyl sulfate (SDS)-12% PAGE. For analysis of cross-links in E and E', complexes were purified by native agarose gel electrophoresis as described above. Cross-linked proteins were isolated from the gel by electroelution (200 V for 3 h), RNA was digested with RNase A, and the samples were subjected to SDS-12% PAGE. Dried gels were exposed to a Molecular Dynamics phosphor screen and scanned with a Molecular Dynamics Storm 840 Phosphorimager.

II-4.6. Hydroxyl radical probing of E and E' structure

Hydroxyl radical cleavage experiments were performed in 20-µl reactions containing 25% HeLa nuclear extract or U2AF-depleted extract that had been dialyzed into glycerol-free buffer D (20 mM HEPES, pH 7.9, 0.1 M KCl, 0.5 mM dithiothreitol). Wild-type and U2AF-depleted extracts were cleared of SF1 by passing the extract over protein G-Sepharose bound to anti-SF1-1D5 antibodies in buffer D. Clearing of the extract was performed twice with fresh protein G-Sepharose. Following SF1 depletion, extracts were dialyzed into glycerol-free buffer D; depletion was confirmed by Western analysis. Reactions probing E' and

E complexes (60 min at 30°C) contained 60 mM KCl and 4 U of RNase inhibitor but lacked ATP/CP and MgCl₂. Cleavage reactions were initiated by the addition of H₂O₂ and ascorbic acid as described elsewhere (Wilson and Noller, 1998). Reactions were quenched with thiourea-glycerol (final concentrations, 30 mM and 1%, respectively), treated with proteinase K, extracted with phenol-chloroform-isoamyl alcohol, and ethanol precipitated. Primer extension reactions were performed as described elsewhere (Stern et al., 1988) by using SuperScript reverse transcriptase (RT) (Gibco BRL). For analysis of the 5'SS, the primer (5'-TCGAGACGAGCTGACATC-3') that was complementary to the region immediately 5' of the branch site was used. Primer extension reactions contained approximately 1 to 5 pmol of RNA substrate and 1 μM DNA primer. Following extension, reactions were extracted with phenol-chloroform-isoamyl alcohol, ethanol precipitated, and then subjected to denaturing PAGE (8% acrylamide; 19:1). Dried gels were exposed to a Molecular Dynamics phosphor screen and scanned with a Molecular Dynamics Storm 840 Phosphorimager.

II-5. References

- Abovich, N., and M. Rosbash. 1997. Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. *Cell* **89**:403-412.
- Abovich, N., X. C. Liao, and M. Rosbash. 1994. The yeast MUD2 protein: an interaction with PRP11 defines a bridge between commitment complexes and U2 snRNP addition. *Genes Dev.* **8**:843-854.
- Barabino, S. M. L., B. J. Blencowe, U. Ryder, B. S. Sproat, and A. I. Lamond. 1990. Targeted snRNP depletion reveals an additional role for mammalian U1 snRNP in spliceosome assembly. *Cell* **63**:293-302.
- Berglund, J. A., N. Abovich, and M. Rosbash. 1998. A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. *Genes Dev.* **12**:858-867.
- Blencowe, B. J., and A. I. Lamond. 1999. Purification and depletion of RNP particles by antisense affinity chromatography. *Methods Mol. Biol.* **118**:275-287.
- Champion-Arnaud, P., O. Gozani, L. Palandjian, and R. Reed. 1995. Accumulation of a novel spliceosomal complex on pre-mRNAs containing branch site mutations. *Mol. Cell. Biol.* **10**:5750-5756.
- Cote, J., R. T. Beaudoin, and B. Chabot. 1995. The U1 small nuclear ribonucleoprotein/5' splice site interaction affects U2AF65 binding to the downstream 3' splice site. *J. Biol. Chem.* **270**:4031-4036.
- Crispino, J. D., B. J. Blencowe, and P. A. Sharp. 1994. Complementation by SR proteins of pre-mRNA splicing reactions depleted of U1 snRNP. *Science* **265**:1866-1869.
- Crispino, J. D., J. E. Mermoud, A. I. Lamond, and P. A. Sharp. 1996. Cis-acting elements distinct from the 5' splice site promote U1-independent pre-mRNA splicing. *RNA* **2**:664-673.
- Das, R., and R. Reed. 1999. Resolution of the mammalian E complex and the ATP-dependent spliceosomal complexes on native agarose mini-gels. *RNA* **5**:1504-1508.
- Du, H., and M. Rosbash. 2002. The U1 snRNP protein U1C recognizes the 5' splice site in the absence of base pairing. *Nature* **419**:86-90.

- Eperon, I. C., D. C. Ireland, R. A. Smith, A. Mayeda, and A. R. Krainer. 1993. Pathways for selection of 5' splice sites by U1 snRNPs and SF2/ASF. *EMBO J.* **12**:3607-3617.
- Fu, X. D. 1993. Specific commitment of different pre-mRNAs to splicing by single SR proteins. *Nature* **365**:82-85.
- Jamison, S. F., A. Crow, and M. A. Garcia-Blanco. 1992. The spliceosome assembly pathway in mammalian extracts. *Mol. Cell. Biol.* **12**:4279-4287.
- Kent, O. A., and A. M. MacMillan. 2002. Early organization of pre-mRNA during spliceosome assembly. *Nat. Struct. Biol.* **9**:576-581.
- Kent, O. A., A. Reayi, L. Foong, K. A. Chilibeck, and A. M. MacMillan. 2003. Structuring of the 3' splice site by U2AF65. *J. Biol. Chem.* **278**:50572-50577.
- Kohtz, J. D., S. F. Jamison, C. L. Will, P. Zuo, R. Luhrmann, M. A. Garcia-Blanco, and J. L. Manley. 1994. Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors. *Nature* **368**:119-124.
- Legrain, P., B. Seraphin, and M. Rosbash. 1988. Early commitment of yeast pre-mRNA to the spliceosome pathway. *Mol. Cell. Biol.* **8**:3755-3760.
- Li, Y., and B. J. Blencowe. 1999. Distinct factor requirements for exonic splicing enhancer function and binding of U2AF to the polypyrimidine tract. *J. Biol. Chem.* **274**:35074-35079.
- Lim, S. R., and K. J. Hertel. 2004. Commitment to splice site pairing coincides with A complex formation. *Mol. Cell* **15**:477-483.
- MacMillan, A. M., P. S. McCaw, J. D. Crispino, and P. A. Sharp. 1997. SC35-mediated reconstitution of splicing in U2AF-depleted nuclear extract. *Proc. Natl. Acad. Sci. USA* **94**:133-136.
- MacMillan, A. M., C. C. Query, C. R. Allerson, S. Chen, G. L. Verdine, and P. A. Sharp. 1994. Dynamic association of proteins with the pre-mRNA branch region. *Genes Dev.* **8**:3008-3020.
- Merendino, L., S. Guth, D. Bilbao, C. Martinez, and J. Valcarcel. 1999. Inhibition of msl-2 splicing by Sex-lethal reveals interaction between U2AF35 and the 3' splice site AG. *Nature* **402**:838-841.
- Michaud, S., and R. Reed. 1993. A functional association between the 5' and 3' splice sites is established in the earliest prespliceosome complex (E) in mammals. *Genes Dev.* **7**:1008-1020.

Query, C. C., P. S. McCaw, and P. A. Sharp. 1997. A minimal spliceosomal complex A recognizes the branch site and polypyrimidine tract. *Mol. Cell. Biol.* **17**:2944-2953.

Query, C. C., M. J. Moore, and P. A. Sharp. 1994. Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model. *Genes Dev.* **8**:587-597.

Rain, J. C., Z. Rafi, Z. Rhani, P. Legrain, and A. Kramer. 1998. Conservation of functional domains involved in RNA binding and protein-protein interactions in human and *Saccharomyces cerevisiae* pre-mRNA splicing factor SF1. *RNA* **4**:551-565.

Reed, R. 2000. Mechanisms of fidelity in pre-mRNA splicing. *Curr. Opin. Cell Biol.* **12**:340-345.

Reed, R., and M. D. Chiara. 1999. Identification of RNA-protein contacts within functional ribonucleoprotein complexes by RNA site-specific labeling and UV crosslinking. *Methods* **18**:3-12.

Roscigno, R. F., and M. A. Garcia-Blanco. 1995. SR proteins escort the U4/U6.U5 tri-snRNP to the spliceosome. *RNA* **1**:692-706.

Rutz, B., and B. Seraphin. 1999. Transient interaction of BBP/ScSF1 and Mud2 with the splicing machinery affects the kinetics of spliceosome assembly. *RNA* **5**:819-831.

Seraphin, B., and M. Rosbash. 1989. Identification of functional U1 snRNA-pre-mRNA complexes committed to spliceosome assembly and splicing. *Cell* **59**:349-358.

Staknis, D., and R. Reed. 1994. SR proteins promote the first specific recognition of pre-mRNA and are present together with the U1 small nuclear ribonucleoprotein particle in a general splicing enhancer complex. *Mol. Cell. Biol.* **14**:7670-7682.

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

Stern, S., D. Moazed, and H. Noller. 1988. Structural analysis of RNA using chemical and enzymatic probing monitored by primer extension. *Methods Enzymol.* **164**:481-489.

Tarn, W. Y., and J. A. Steitz. 1994. SR proteins can compensate for the loss of U1 snRNP functions in vitro. *Genes Dev.* **8**:2704-2717.

Will, C. L., S. Rumpler, J. Klein Gunnewiek, W. J. van Venrooij, and R. Luhrmann. 1996. In vitro reconstitution of mammalian U1 snRNPs active in splicing: the U1-C protein enhances the formation of early (E) spliceosomal complexes. *Nucleic Acids Res.* **24**:4614-4623.

Wilson, K. S., and H. F. Noller. 1998. Mapping the position of translational elongation factor EF-G in the ribosome by directed hydroxyl radical probing. *Cell* **92**:131-139.

Wu, J. Y., and T. Maniatis. 1993. Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell* **75**:1061-1070.

Wu, S., C. M. Romfo, T. W. Nilsen, and M. R. Green. 1999. Functional recognition of the 3' splice site AG by the splicing factor U2AF35. *Nature* **402**:832-835.

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

Zhang, D., and M. Rosbash. 1999. Identification of eight proteins that cross-link to pre-mRNA in the yeast commitment complex. *Genes Dev.* **13**:581-592.

Zorio, D. A., and T. Blumenthal. 1999. Both subunits of U2AF recognize the 3' splice site in *Caenorhabditis elegans*. *Nature* **402**:835-838.

Appendix III⁽¹⁾
Crystal structure of a core spliceosomal protein interface

¹ Adapted from Schellenberg et al., (2006). *Proc. Natl. Acad. Sci. U.S.A.* **103**, 1266–1271.

The work presented in this appendix represents a collaboration of the authors on the paper. The experiments presented were primarily conducted by M.S. D.R. originally cloned the p14 gene and performed initial expression tests.

III-1. Introduction

The dynamic nature of spliceosome assembly has been established by a wealth of studies that have described a complex series of RNA·RNA and RNA·protein associations during the course of assembly (Kramer, 1996; Staley and Guthrie, 1998; Burge et al., 1999). Site-specific modification of the pre-mRNA branch adenosine with benzophenone followed by photocrosslinking revealed the specific, sequential association of several factors with the branch adenosine from very early in spliceosome assembly (MacMillan et al., 1994). Of particular interest was a strong crosslink to p14 that appeared in A complex and persisted within the fully assembled spliceosome. Subsequent experiments showed that this protein crosslinked directly to the branch nucleotide in A through C complexes indicating an intimate association between protein and RNA at the heart of the mammalian spliceosome (Query et al., 1996). This finding contrasts with the ribosome in which high-resolution structural analysis has shown that the active site is exclusively composed of RNA with the closest ribosomal protein found ≈ 18 Å distant from the active site (Nissen et al., 2000).

p14 was initially isolated from purified mammalian spliceosomes and subsequently identified as a constituent of U2 snRNP and U12 snRNP, a component of the minor spliceosome responsible for splicing a relatively rare subset of introns (Will et al., 2001). p14 is an evolutionarily highly conserved protein with orthologs across diverse species; the human protein is a 125-aa polypeptide containing a central region with strong homology to the well characterized RRM domain (Kenan et al., 1991). In particular, p14 contains

consensus RNP1 and RNP2 motifs, which mediate RRM-RNA interactions, as evidenced by high-resolution structural analyses (Oubridge et al., 1994; Deo et al., 1999; Handa et al., 1999; Allain et al., 2000; Wang and Hall, 2001). Despite the presence of an RRM, the association of p14 with U2 snRNP is mediated, at least in part, through protein-protein interactions. Specifically, p14 is part of the heteromeric protein complex SF3b, a salt-dissociable component of U2 snRNP, and a strong interaction between p14 and the SF3b protein SF3b155 has been demonstrated (Will et al., 2001). The intimate association of p14 with the branch nucleotide in the fully assembled spliceosome suggests that this protein may contribute to the architecture of the active site.

To further define the role of p14 in the spliceosome, we have solved the high-resolution x-ray structure of human p14 in a complex with a portion of SF3b155 (PDB ID code 2F9D). Analysis of this structure combined with previous crosslinking data and analyses of extant RRM structures and cryo-EM data allow us to propose a model for branch region recognition in the fully assembled spliceosome.

III-2. Results

III-2.1. Structure of the p14-SF3b155 complex

Initial investigations of the properties of p14 suggested that the protein in isolation was partially unfolded as evidenced by poor solubility, aggregation, and ¹H NMR (M.J.S., L. Spyrocopolous, and A.M.M., unpublished data). We therefore focused further studies on the complex formed between p14 and

SF3b155 (Will et al., 2001). By using a pull-down assay to examine a series of deletion constructs, we were able to establish the existence of a strong interaction between p14 and a peptide representing amino acids 373–415 of SF3b155 (Figure III-1A). The complex formed between p14 and this peptide is very soluble and elutes by gel filtration at the volume expected for a 1:1 complex. We crystallized a selenomethionine-substituted complex and determined its structure to 2.5 Å resolution by using multiwavelength anomalous diffraction methods (Figure III-1B).

As predicted, p14 contains a central RRM domain spanning residues 20–91 (Figure III-1C). In addition, the C terminus of p14 contains two additional α -helices: a short 5-aa helix (aa 94–98) and a second 15-residue helix (aa 103–117). The bound SF3b155 peptide consists of a long N-terminal α -helix (aa 380–396) and a second shorter helix (aa 401–407). Interestingly, the C terminus of the SF3b155 fragment contains a short β -strand that interacts with β -3 of the p14 RRM β -sheet, which is connected to the shorter α -helix by a loop that makes extensive contacts with both the shorter C-terminal helix and RRM of p14. Amino Acids 1–11 of p14 and 373–376 of SF3b155 are disordered in the structure. Because the crosslinks formed between spliceosomal versus recombinant p14 and RNA differ in size by \approx 1 kDa, we speculate that spliceosomal p14 lacks \approx 10 aa of the predicted full-length protein; therefore, the former residues are probably not relevant to the structure of p14 within the spliceosome (M.J.S. and A.M.M., unpublished data).

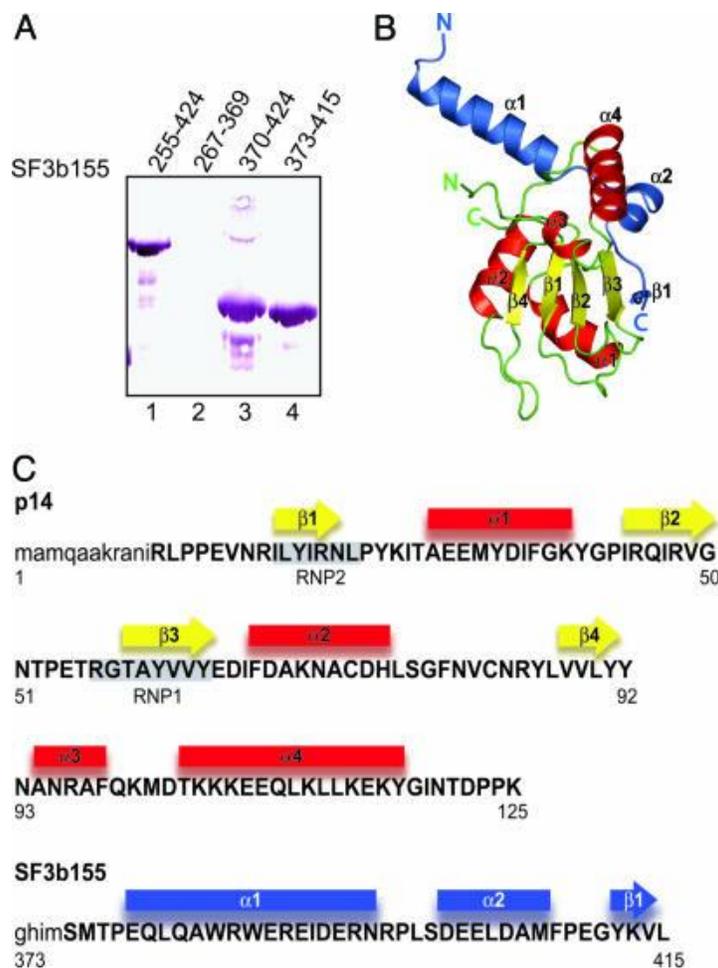


Figure III-1. Structure of p14-SF3b155 peptide complex. A) Determination of a minimal interaction surface between p14 and SF3b155. Pulldowns using amylose resin of the indicated GST-deletion constructs of SF3b155 after incubation with MBP-p14. Reactions were analyzed by 16% 29:1 SDS/PAGE using anti-GST horseradish peroxidase conjugate to detect SF3b155. B) Ribbon diagram of p14-SF3b155 peptide complex. α -Helices and β -strands of p14 are colored red and yellow, respectively; SF3b155 peptide is colored blue. C) Secondary structure diagram depicting α -helices (red) and β -strands (yellow) of p14. RNP motifs are highlighted with gray boxes. Secondary structural elements of SF3b155 peptide are colored blue.

The most striking feature of the p14-peptide complex is that the p14 β -sheet is occluded by one of the C-terminal helices of p14, the central helix of the SF3b155 peptide, and the loop connecting SF3b155 to the C-terminal β -strand, resulting in a buried surface area of $\approx 1,300 \text{ \AA}^2$. The interface between the two proteins is extensive and includes a hydrophobic core (Figure III-2A) surrounded

by a set of hydrogen bonds and salt bridges (Figure III-2B). The blocking of one face of p14 is significant because the four-stranded β -sheet of the canonical RRM represents the RNA-binding surface of the domain including the highly conserved RNP1 and RNP2 motifs (Kenan et al., 1991; Oubridge et al., 1994; Deo et al., 1999; Handa et al., 1999; Allain et al., 2000; Wang and Hall, 2001). In the p14-peptide complex, residues of RNP1 and RNP2 are largely buried. Intriguingly, a portion of RNP2 is exposed within a pocket on the otherwise occluded surface: a highly conserved aromatic residue within RNP2, Y22, forms the base of this pocket on the p14-peptide surface (Figure III-2C). The surface surrounding this pocket includes four basic residues: R24, R57, R96, and K100 (Figure III-2C). The side chains of two of these residues (R24 and R57) project from the surface of the p14 β -sheet, the third (R96) is found at the end of α_3 , and the fourth (K100) in the loop between α_3 and α_4 . The identities of R24 and R57 are highly conserved among p14 orthologs but not between p14 and other RRMs.

III-2.2. RNA protein interactions in a minimal complex

The direct interaction of p14 with the branch nucleotide bulged from a base-paired duplex is intriguing because in all previously characterized RNA-RRM complexes, single-stranded RNA binds across the β -sheet of the RRM (Kenan et al., 1991; Oubridge et al., 1994; Deo et al., 1999; Handa et al., 1999; Allain et al., 2000; Wang and Hall, 2001). Mobility shift assays using either p14 or the p14-peptide complex show only weak RNA-protein association ($>50 \mu\text{M}$ dissociation constants) and no preference for single-stranded, duplex, or bulged-

duplex RNA (M.J.S., A.M.M., unpublished data). These results probably reflect the cooperative nature of RNA·protein interaction within the spliceosome, including interactions of the pre-mRNA with SF3b155 (Gozani et al., 1998); whereas p14 is the only protein that directly interacts with the branch adenosine in the fully assembled spliceosome, as evidenced by photocrosslinking studies (Query et al., 1996), SF3b155 has been shown to directly interact with nucleotides at the -6 position, just 5' to the branch region·U2 duplex and at the +5 position, 3' to the branch (Gozani et al., 1998).

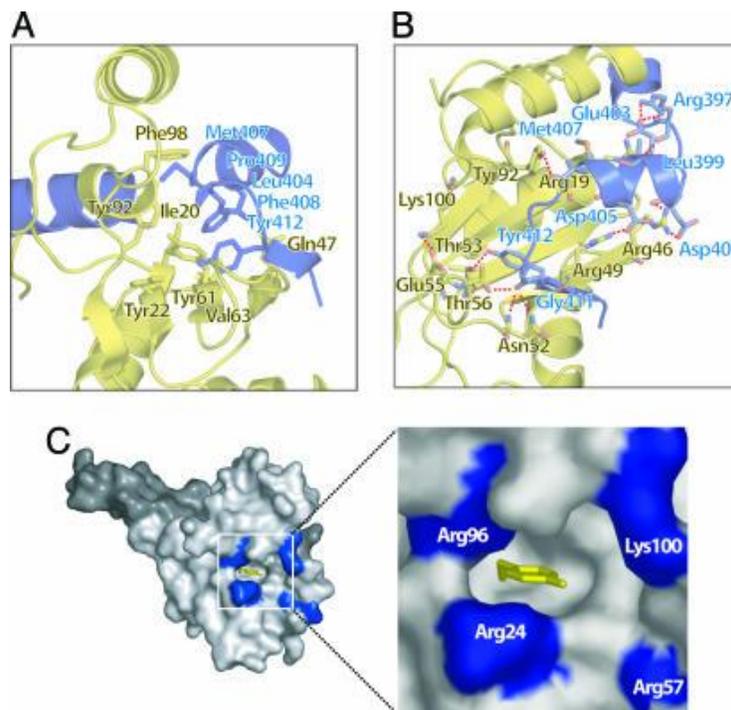


Figure III-2. Details of p14-SF3b155 interface. A) Hydrophobic core of the p14-SF3b155 interface. p14 is colored yellow, and SF3b155 peptide is colored blue. B) Hydrogen-bonding network and salt bridges surrounding the hydrophobic core. p14 is colored yellow, and SF3b155 peptide is colored blue. Hydrogen bonds involved in secondary structural elements are omitted. C) Surface representation of the p14-SF3b155 complex showing Y22 of RNP2 exposed within a surface pocket surrounded by conserved basic residues. p14 is shaded light gray, SF3b is shaded dark gray, Y22 is colored yellow, and R24, R57, R96, and K100 are colored blue.

To determine the part of p14 that interacts with the branch adenosine, we mapped the location of the crosslink formed between the branch nucleotide and p14. We first synthesized an RNA containing a hairpin representing the pre-mRNA branch region duplex; this RNA contained a single adenosine at the branch position and could thus be uniquely labeled at this position by carrying out transcriptions in the presence of [α - 32 P]ATP. We then performed crosslinking experiments in a minimal system containing this RNA and the p14-peptide complex. After crosslinking by irradiation at 254 nm, the reactions were digested with Nuclease P1 and analyzed by SDS/PAGE, which showed that the branch adenosine crosslinks to p14 but not the SF3b155 peptide (Figure *III-3A* and data not shown). By using cyanogen bromide and endoproteases Glu-C and Lys-C, we were able to establish that the branch nucleotide crosslinks to p14 between amino acids 16 and 29 (Figure *III-3A*); this region includes β -1 of the RRM and the RNP2 consensus region. Thus, the pre-mRNA branch interacts with a portion of the RRM of p14.

We next examined the interaction with RNA of a mutant p14, Y22M, in which the conserved aromatic of RNP2 has been changed to methionine. This mutation has little effect on the structure of the p14-peptide complex; the 3.0-Å X-ray structure of the Y22M p14-SF3b peptide complex (PDB ID code 2F9J) is essentially superimposable on the native complex, with the side chain of M22 forming the base of the pocket in the same fashion as Y22 in the wild-type. We performed crosslinking experiments in reactions containing the Y22M p14-peptide complex and the bulged duplex RNA and observed the formation of a

crosslink between the branch nucleotide and Y22M p14 (Figure III-3B). When we attempted to perform a crosslink mapping experiment, we observed that the crosslink was sensitive to cyanogen bromide, as evidenced by the disappearance of labeled protein upon treatment with the reagent (Figure III-3B). This result is consistent with the crosslink occurring between the branch nucleotide and the terminal methyl group of the side chain of M22 (Clement et al., 2005), which suggests that the conserved aromatic of RNP2 interacts directly with the bulged nucleotide (Figure III-3C). Given the fact that single-stranded RNA containing a single adenosine crosslinks to p14 in the same fashion, it is also possible that the p14-peptide complex can recognize a structure other than the bulged duplex; nevertheless, the crosslinking experiments strongly support specific interaction of this residue of RNP2 with an unpaired nucleotide at the branch position.

III-2.3. Comparison with canonical and pseudoRRMs

The structure of the p14-peptide complex is reminiscent of a number of other unusual RRM and RRM-like structures (Avis et al., 1996; Rupert et al., 2003; Kielkopf et al., 2001; Selenko et al., 2003). An α -helix C-terminal to the RRM of the spliceosomal U1A protein has been shown to adopt two different conformations with respect to the RNA-binding face of the RRM (Figure III-4A). In one orientation, the β -sheet of the RRM is masked (Avis et al., 1996); an $\approx 135^\circ$ rotation of this helix reveals the surface of the RRM (Rupert et al., 2003) in a structure similar to that observed in the U1A protein-RNA complex (Oubridge et al., 1994). In contrast to U1A, the C-terminal helices of p14 are held rigidly in

position through an extensive network of hydrophobic and hydrophilic interactions, which are both intramolecular and between p14 and SF3b (Figures III-2A and B and III-4A). It is unlikely that a rearrangement of structure similar to that proposed for U1A would expose the p14 RNP motifs for branch duplex recognition.

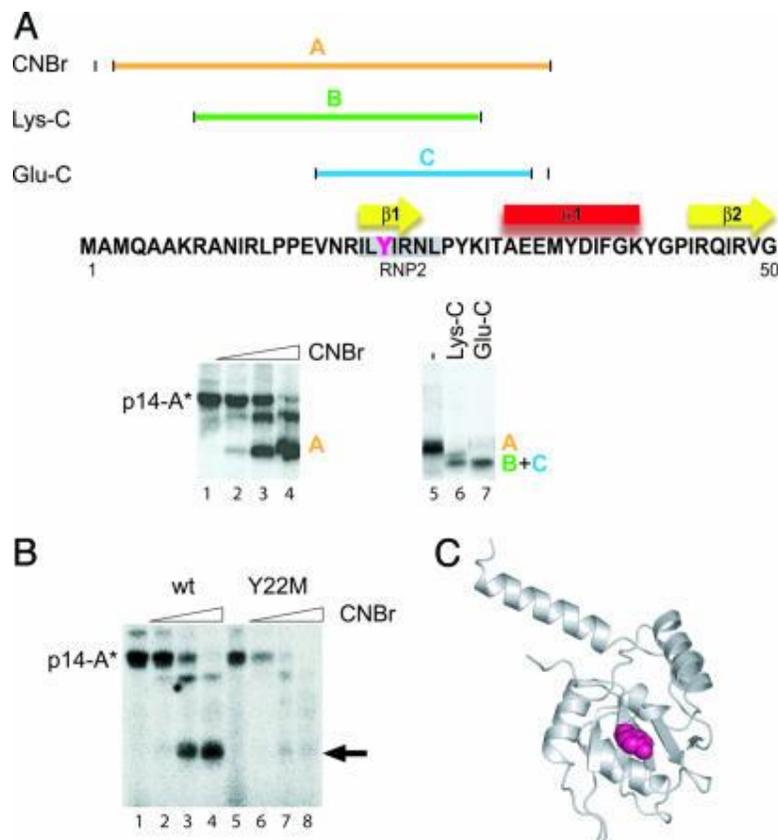


Figure III-3. Mapping of the p14-branch nucleotide interaction. A) Treatment of p14-SF3b155 peptide complex-RNA crosslink with nuclease P1 yields a single radiolabeled band on SDS/PAGE corresponding to p14 (lane 1). Cleavage with CNBr produces fragment A, which appears immediately (lane 2) and persists as a final digestion product (lanes 3 and 4). Further digestion of purified CNBr fragment A (lane 5) with endoproteinase Lys-C produces fragment B (lane 6). Alternatively, cleavage of fragment A with endoproteinase Glu-C produces fragment C. Cleavage with endoproteinase Glu-C of a $\Delta 1-11$ deletion mutant of p14 produces an identically migrating fragment, indicating that fragment C is not the N-terminal Glu-C fragment of p14 (data not shown). B) CNBr treatment of p14-RNA (lanes 1-4) and Y22M p14-RNA (lanes 5-8) crosslinks. C) Ribbon diagram of p14-SF3b155 peptide complex with Y22 depicted in a space-filling representation (purple).

A number of proteins contain atypical RRM motifs that differ from canonical RRM motifs in conserved sequence elements, such as the RNP motifs, in the length of conserved secondary structural elements, and in the predicted charge of the surface of the RRM β -sheet. Recent high-resolution structural studies of atypical RRM motifs from the splicing factors U2AF65 and U2AF35 provide a model of the interaction between U2AF65 and U2AF35 as well as that between U2AF65 and the splicing factor SF1; in both cases, the RRM serves as a scaffold for interactions with a partner protein mediated by contacts to the rear (α -helical) surface of the RRM (Kielkopf et al., 2001; Selenko et al., 2003). These pseudoRRMs have been labeled UHMs and are part of a larger family of putative protein-protein interaction domains (Kielkopf et al., 2004); thus, RRM or RRM-related motifs can serve as scaffolds for noncanonical ligand interactions. A comparison of both the p14 RNP sequence and the structure of the p14 RRM with standard RRM motifs and UHMs suggests that p14 is best classified as containing a canonical RRM. In addition, the overall positive charge of the p14 RRM and the p14-SF3b155 peptide complex is consistent with either the RRM surface or the complex representing an RNA interaction surface, in contrast, for example, to the UHM of U2AF65 (Figure III-4B).

III-2.4. p14 within SF3b and the U11/U12 di-snRNP

Recent cryo-EM studies have provided a 10-Å resolution model of SF3b alone and within the context of the U11/U12 di-snRNP, a component of the minor

spliceosome; p14 was modeled on the basis of its predicted RRM and assigned density in both structures (Golas et al., 2003; Golas et al., 2005).

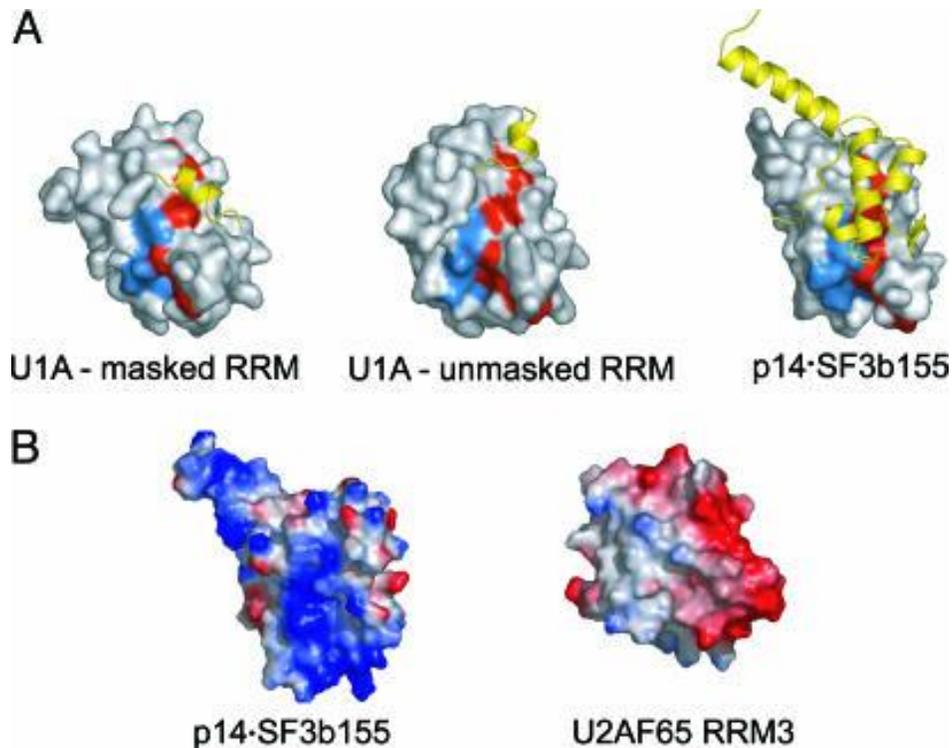


Figure III-4. Comparison between p14-SF3b155 peptide surface and structures of U1A and U2AF65 RRM3. A) (Left) NMR structure of free U1A with C-terminal helix (in yellow) positioned across RNA-binding surface (Avis et al., 1996). (Middle) X-ray structure of free U1A showing C-terminal helix rotated to unmask the RNA binding surface (Rupert et al., 2003). (Right) Structure of p14-SF3b155 complex showing C-terminal helices of p14 and Sf3b155 peptide (in yellow). RNP1 and RNP2 are colored red and blue, respectively. B) Representation of surface charges of p14-SF3b155 and U2AF homology motif (RRM3) of U2AF65. The color scheme is as follows: blue, basic; red, acidic; white, neutral.

Fitting of the x-ray structure of the p14-SF3b155 peptide into the EM density of isolated SF3b resulted in an excellent agreement with respect to the overall shape (Figure III-5A). The cocrystal structure and isolated SF3b show a globular domain corresponding to the RRM domain of p14 and two connecting

bridges corresponding to the N-terminal α -helix of the SF3b155 peptide and the C-terminal α -helix of the SF3b155 peptide plus the C-terminal α -helix of p14. Similarly, the cocrystal structure of the p14·SF3b155 peptide fits well into the outer globular domain of the U11/U12 di-snRNP previously suggested (Golas et al., 2005) to represent p14. In this fit, the β -sheet of the p14 RRM is oriented toward the outer surface of the U11/U12 di-snRNP, whereas the two α -helices of the p14 RRM are located toward the interior of the complex (Figure *III-5B*). The C-terminal α -helix of p14 is located on the outside, consistent with immunoprecipitation data in which the U11/U12 di-snRNP was precipitated by an anti-p14 antibody directed against the C terminus (Will et al., 2001). In contrast to isolated SF3b (Golas et al., 2003), where density representing p14 was found to be caged in a central cavity, within the U11/U12 structure a rotation of one SF3b shell half by 90°, combined with some smaller movements, opens up SF3b and reveals the surface of p14 (Golas et al, 2005). Overall, there are no indications in the EM data of isolated SF3b or the U11/U12 di-snRNP that the conformation of the p14·SF3b155 peptide interface is changed.

The positioning of p14 within the U11/U12 structure suggests that the pre-mRNA enters into a cleft on the surface of the U11/U12 di-snRNP to interact with p14 (Figure *III-5B*). This cleft is enough to accommodate even double-stranded RNA because the space between the outer wall of the U11/U12 di-snRNP and the p14 RRM is >2 nm wide. It is possible, however, that pre-mRNA association involves conformational changes involving the p14 region, particularly given the requirement for duplex formation between U12 (U2) snRNA and the branch

region. Because RNA density has not been assigned in the U11/U12 di-snRNP, it is not possible at this point to model the spatial relationship between p14 and the U12 snRNA.

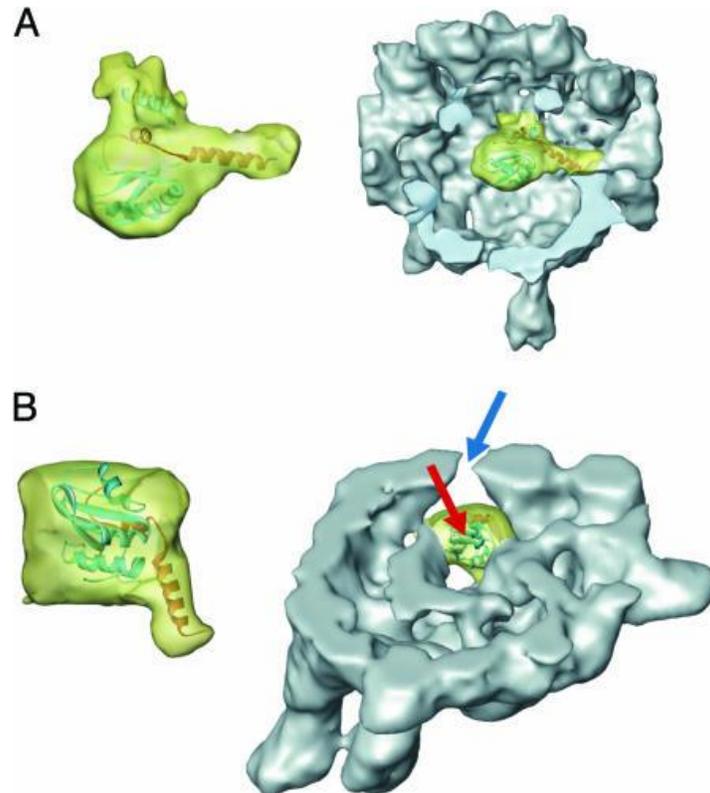


Figure III-5. p14-SF3b155 peptide structure within SF3b and U11/U12 snRNP. A) Fitting of p14-SF3b155 peptide density into the SF3b cryo-EM structure (Golas et al., 2003). (Left) cryo-EM density corresponding to p14-SF3b155 peptide (yellow), with the p14-SF3b155 x-ray structure represented as a ribbon diagram. (Right) cryo-EM density corresponding to p14-SF3b155 peptide within the context of a cutaway view of the overall SF3b cryo-EM structure. B) Fitting of p14-SF3b155 peptide density into the U11/U12 cryo-EM structure (Golas et al., 2005). (Left) cryo-EM density corresponding to p14-SF3b155 peptide (yellow), with the p14-SF3b155 x-ray structure represented as a ribbon diagram. (Right) cryo-EM density corresponding to p14-SF3b155 peptide within the context of a cut-away view of the overall U11/U12 cryo-EM structure. The red arrow indicates the proposed RNA-binding surface of p14-SF3b155; the blue arrow indicates a possible path of pre-mRNA to groove on U11/U12 surface.

III-3. Discussion

The results of the crosslinking experiments reported here suggest that the pre-mRNA branch adenosine interacts directly with the conserved aromatic of RNP2

in the RRM of the p14·SF3b155 peptide complex. High-resolution structures of RNA·RRM complexes show a conserved mode of RNA·protein interaction in which RNA is bound across the RRM β -sheet, making interactions with a variety of protein side chains, including conserved residues of RNP1 and RNP2. Because a significant portion of the p14 β -sheet is occluded by a C-terminal α -helix and portions of the SF3b peptide, either a rearrangement must occur upon RNA binding or the branch duplex interacts with p14·SF3b155 in a noncanonical fashion. Modeling of the p14-peptide structure into the cryo-EM structures suggests that no rearrangement occurs upon incorporation of SF3b into U12 (or U2) snRNP, although this could also occur later in spliceosome assembly.

It is important to note that p14 is unique among RRM-containing proteins characterized to date in that it interacts with a bulged duplex as opposed to single-stranded RNA. The canonical interaction between an RRM and RNA includes critical stacking interactions between nucleobases and the conserved aromatic residues of RNP1 and RNP2. Given the fact that the bulged branch duplex contains only one unpaired residue, it seems reasonable that only one of the two RNP aromatics would interact with bound RNA.

A bulged duplex model of the pre-mRNA·U2 structure that features an extruded adenosine at the branch position (Berglund et al., 2001) can be docked to the p14·SF3b155 interface by positioning the adenine base of the bulged nucleotide within the surface pocket to allow a stacking interaction with Y22; a small movement of the side chain of one of the solvent-exposed basic residues surrounding the pocket, R96, would allow this interaction. By anchoring the

branch nucleotide at this position, two orientations of the duplex rotated 180° with respect to one another are possible. Interestingly, either model positions the phosphate backbone of the duplex proximal to the conserved, positively charged residues of p14 that surround the pocket (Figure *III-2B*). Thus, specificity for the branch duplex could be determined by an adenosine-binding pocket and appropriately positioned basic residues. We propose that these features of the p14·SF3b155 interface present a surface for RNA recognition at the heart of the spliceosome.

III-4. Methods

III-4.1. Determination of SF3b155 p14 interaction domain

MBP-p14 fusion protein (300 pmol) was incubated with 300 pmol of GST-SF3b155 constructs and 10 µl of amylose beads (GE Healthcare) in buffer (10 mM Hepes, pH 7.9/60 mM KCl/2 mM MgCl₂/0.1 mM EDTA/0.5 mM DTT) for 30 min at 23°C. Beads were pelleted by centrifugation at 1,000 × g and washed twice with buffer after removal of supernatant. Protein was eluted with buffer containing 100 mM maltose, run on a 16% SDS/PAGE gel, and transferred to nitrocellulose (Millipore) for 3 h at 200 mA. Membranes were blocked in 5% milk powder and probed with a 1:5,000 dilution of anti-GST-horseradish peroxidase conjugate. After being washed, the membrane was developed by addition of 1 mg/ml 4-chloro-1-naphthol and 0.1% H₂O₂ in Tris-buffered saline with 20% methanol.

III-4.2. Protein expression and purification

DNA encoding full-length human p14 was cloned into the EcoRI and PstI sites of pMALc2x (NEB, Beverly, MA) by using PCR primers to insert a TEV protease cleavage site between MBP and p14. Mutagenesis of p14 was carried out by PCR and confirmed by sequencing. The resulting MBP·p14 fusion proteins were expressed in *Escherichia coli* and purified by sequential amylose resin and anion exchange chromatography. DNA encoding amino acids 373–415 of SF3b155 was cloned into the EcoRI and BamHI sites of pGEX6p1 (GE Healthcare) by using PCR primers to insert a TEV protease cleavage site between GST and SF3b 373–415. Fusion protein was expressed in *E. coli* and purified by glutathione Sepharose chromatography (GE Healthcare). After cleavage of the GST tag using TEV protease, the SF3b155 peptide was purified on a Superdex-75 26/60 column (GE Healthcare).

To prepare the complex, SF3b155 peptide was incubated with MBP·p14 followed by cleavage of the fusion protein with TEV protease. The p14·SF3b155 peptide complex was then purified by cation exchange on a Source15S HR10/10 column (GE Healthcare) followed by gel filtration with a Superdex-75 26/60 column (GE Healthcare).

III-4.3. Crystallization

Crystals of p14·SF3b155 peptide were grown at 23°C by using the hanging drop vapor diffusion technique. Crystals of native complex and complex containing selenomethionine-substituted SF3b155 peptide were grown by mixing 2 µl of 10

mg/ml protein solution (10 mM Tris, pH 8.0/60 mM KCl/1 mM EDTA/5 mM 2-mercaptoethanol/0.02% NaN₃) with 2 µl of precipitant (14–18% polyethylene glycol 3350/100 mM MOPS, pH 6.0/200 mM NaHCO₂). Y22M-containing crystals were grown by mixing 2 µl of protein solution with 2 µl of precipitant (1.8 M 60%/40% NaH₂PO₄/K₂HPO₄ and 500 mM β-alanine). Crystals were transferred to precipitant containing 20% glycerol and frozen in liquid nitrogen for data collection.

III-4.4. Data collection and processing

For the wild-type and mutant complexes, data were collected at beamline 8.3.1 of the Advanced Light Source at Lawrence Berkeley National Laboratory. Data for the wild-type complex were collected from a single selenomethionine derivatized crystal; a two-wavelength multi-wavelength anomalous diffraction experiment was performed, collecting data in an inverse-beam mode at the midpoint energy between the experimentally determined Se/K edge and the inflection point. The crystal was subsequently translated, and further data were collected at a lower energy wavelength, remote from the Se absorption edge. Data were processed and scaled with the HKL package (Otwinowski and Minor, 1997).

III-4.5. Model building and refinement

For the native complex, the program solve (Terwilliger and Berendzen, 1999) was used to determine the positions of four of the expected six Se atoms (the remaining two Se atoms were later found to be in a disordered region of the

structure). Initial phases to 3.1 Å were extended to 2.5 Å with maximum likelihood density modification in resolve (Terwilliger, 2000). An initial model automatically built by resolve (Terwilliger, 2002) served as the basis for model building and refinement. Iterative cycles of refinement in REF-MAC (Murshudov et al., 1997) against the low-energy remote data, manual model building using xfit (McRee, 1999), and automatic model building using resolve were used to complete and refine the model. Two-fold noncrystallographic symmetry restraints were maintained throughout refinement but were relaxed for regions of the structure showing deviation due to different packing environments. The structure of the Y22M mutant complex was solved by molecular replacement.

III-4.6. Crosslink mapping

RNA (5'-GGGCGGUGGUGCCCUGGUGGGUGCUGACCGCCC-3') was prepared and labeled at the branch nucleotide by T7 transcription [using 3,000 Ci/mmol (1 Ci = 37 GBq) [α -³²P]ATP] from a synthetic DNA, followed by purification on a 15% denaturing PAGE gel. Recombinant p14·SF3b155 373–415 (20 μM) was incubated with 1 pmol of RNA in buffer (10 mM Hepes, pH 7.9/60 mM KCl/2 mM MgCl₂/0.1 mM EDTA/0.5 mM DTT) in the presence of 1 μg of tRNA (Roche) for 40 min at 23°C. Reactions were irradiated on ice with a 5-W, 254-nm UV lamp (Ultraviolet Products, San Gabriel, CA) at a distance of 8 mm for 30 min. Nuclease P1 was added to a concentration of 50 ng/μl, and RNA was digested at 55°C for 2 h. For CNBr cleavage, reactions were brought to 70% formic acid and 50 mg/ml CNBr and incubated overnight at 23°C in the dark.

Formic acid was removed *in vacuo*, and protein was precipitated with acetone to remove traces of acid. For Glu-C digestions, the purified CNBr fragment was resuspended in 100 mM NH_4HCO_3 , pH 7.8, with 10% acetonitrile. Glu-C protease (500 ng) was added, and cleavage was allowed to proceed overnight at 30°C. For Lys-C digestions, the purified CNBr fragment was resuspended in 25 mM Tris, pH 8.5/1 mM EDTA/10% acetonitrile. Lys-C (100 ng) was added, and cleavage was allowed to proceed overnight at 37°C. Crosslinking and cleavage reactions were separated by 16% 15:1 SDS/PAGE and visualized using a Molecular Dynamics PhosphorImager.

III-4.7. Comparison of X-ray and cryo-EM structures

For fitting, the EM densities of isolated SF3b (Golas et al., 2003) and of the U11/U12 di-snRNP (Golas et al., 2005) were low-pass-filtered to their respective resolution. Thresholds of the EM densities were chosen based on the theoretical molecular mass of the particle. Fitting was performed manually by using the 3D alignment tool of the software AMIRADEV 2.3 (TGS Europe, Merignac, France). The cage-like structure of SF3b contains a limited number of larger-density elements that are potential locations for an RRM. Three of nine elements revealed good fits with an RRM; p14 was distinguished from the SF3b component SF3b49 based on the presence of two adjacent RRMs in SF3b49 and only one in p14. Subsequent to manual fitting of the x-ray structure of the p14/SF3b155 peptide, an exhaustive real-space refinement in a small translational and rotational

range was performed to optimize the fit. The quality of the fits was judged visually and by normalized cross-correlation coefficients.

III-5. References

- Allain, F.H.T., Bouvet, P., Dieckmann, T., and Feigon, J. (2000). Molecular basis of sequence-specific recognition of pre-ribosomal RNA by nucleolin. *EMBO J.* **19**, 6870–6871.
- Avis, J.M., Allain, F.H., Howe, P.W., Varani, G., Nagai, K., and Neuhaus, D. (1996). Solution structure of the N-terminal RNP domain of U1A protein: the role of C-terminal residues in structure stability and RNA binding. *J. Mol. Biol.* **257**, 398–411.
- Berglund, J.A., Rosbash, M., and Schultz, S.C. (2001). Crystal structure of a model branchpoint-U2 snRNA duplex containing bulged adenosines. *RNA* **7**, 682–691.
- Burge, C.B., Tuschl, T., and Sharp, P.A. (1999). in *The RNA World*, eds. Gesteland, R. F., Cech, T. R. & Atkins, J. F. (Cold Spring Harbor Lab. Press, Woodbury, NY), 2nd Ed., pp. 525–560.
- Clément, M., Martin, S.S., Beaulieu, M.E., Chamberland, C., Lavigne, P., Leduc, R., Guillemette, G., and Escher, E. (2005). Determining the environment of the ligand binding pocket of the human angiotensin II type I (hAT1) receptor using the methionine proximity assay. *J. Biol. Chem.* **280**, 27121–27129.
- Deo, R.C., Bonanno, J.B., Sonenberg, N., and Burley, S.K. (1999). Recognition of polyadenylate RNA by the poly(A)-binding protein. *Cell* **98**, 835–845.
- Golas, M.M., Sander, B., Will, C.L., Lührmann, R., and Stark, H. (2003). Molecular architecture of the multiprotein splicing factor SF3b. *Science* **300**, 980–984.
- Golas, M.M., Sander, B., Will, C.L., Lührmann, R., and Stark, H. (2005). Major conformational change in the complex SF3b upon integration into the spliceosomal U11/U12 di-snRNP as revealed by electron cryomicroscopy. *Mol. Cell* **17**, 869–883.
- Gozani, O., Potashkin, J., and Reed, R. (1998). A potential role for U2AF-SAP 155 interactions in recruiting U2 snRNP to the branch site. *Mol. Cell. Biol.* **18**, 4752–4760.
- Handa, N., Nureki, O., Kurimoto, K., Kim, I., Sakamoto, H., Shimura, Y., Muto, Y., and Yokoyama, S. (1999). Structural basis for recognition of the tra mRNA precursor by the Sex-lethal protein. *Nature* **398**, 579–585.
- Kenan, D.J., Query, C.C., and Keene, J.D. (1991). RNA recognition: towards identifying determinants of specificity. *Trends Biochem. Sci.* **16**, 214–220.

- Kielkopf, C.L., Lucke, S., and Green, M.R. (2004). U2AF homology motifs: protein recognition in the RRM world. *Genes Dev.* **18**, 1513–1526.
- 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**, 595–605.
- Krämer, A. (1996). The structure and function of proteins involved in mammalian pre-mRNA splicing. *Annu. Rev. Biochem.* **65**, 367–409.
- MacMillan, A.M., Query, C.C., Allerson, C.A., Chen, S., Verdine, G.L., and Sharp, P.A. (1994). Dynamic association of proteins with the pre-mRNA branch region. *Genes Dev.* **8**, 3008–3020.
- McRee, D. E. (1999). XtalView/Xfit--A versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **125**, 156–165.
- Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255.
- Nissen, P., Hansen, J., Ban, N., Moore, P.B., and Steitz, T.A. (2000). The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**, 920–930.
- Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.
- Oubridge, C., Ito, N., Evans, P.R., Teo, C.H., and Nagai, K. (1994). Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature* **372**, 432–438.
- Query, C.C., Strobel, S.A., and Sharp, P.A. (1996). Three recognition events at the branch-site adenine. *EMBO J.* **15**, 1392–13402.
- Rupert, P.B., Xiao, H., and Ferre-D'Amare, A.R. (2003). U1A RNA-binding domain at 1.8 Å resolution. *Acta Crystallogr. D* **57**, 1521–1524.
- 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**, 965–976.
- Staley, J.P., and Guthrie, C. (1998). Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell* **92**, 315–326.

Terwilliger, T.C. (2000). Maximum-likelihood density modification. *Acta Crystallogr. D* **56**, 965–972.

Terwilliger, T.C. (2002). Automated main-chain model-building by template-matching and iterative fragment extension. *Acta Crystallogr. D* **59**, 34–44.

Terwilliger, T.C., and Berendzen, J. (1999). Automated MAD and MIR structure solution. *Acta Crystallogr. D* **55**, 849–861.

Wang, X., and Hall, T.M.T. (2001). Structural basis for recognition of AU-rich element RNA by the HuD protein. *Nat. Struct. Biol.* **8**, 141–145.

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**, 4536–4546.