

University of Alberta

Recognition of Catalytic Structures by the Spliceosomal Proteins

SF3b14 and Prp8

by

Matthew Jeffrey Schellenberg

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

©Matthew Schellenberg

Spring 2011

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.



Library and Archives
Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

ISBN: 978-0-494-91439-7

Our file Notre référence

ISBN: 978-0-494-91439-7

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

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 thought to be responsible for splicing catalysis, however, the vast majority of the mass of the spliceosome is composed of protein. Two such proteins, SF3b14 and Prp8 have been shown to crosslink to the sites of splicing chemistry within an intron. This thesis describes the results of structural and biochemical characterization of these two proteins.

The protein SF3b14 forms a complex with SF3b155. This interface modifies the RRM domain of protein so that it is capable of recognising a specific RNA structure in the spliceosome – the BPS•U2 snRNA duplex. We describe a low resolution model of this protein•RNA complex using SAXS and disulfide tethering. The resulting model allows us to propose a function for the SF3b14 protein.

The protein Prp8 is a large (~220 kDa) protein which is highly conserved from yeast to human. Despite this, very little is known with respect to the structure of this protein. 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. We characterize metal ion binding by this domain, and investigate the effect of disruption of this metal site by mutation. Furthermore, we demonstrate that metal ion binding is regulated by a conformation change of this domain, which we are able to link to a conformational change of the spliceosome during catalysis.

Table of Contents:

Abstract	<i>i</i>
Table of Contents	<i>iii</i>
List of Figures	<i>vi</i>
List of Tables	<i>viii</i>
List of Abbreviations	<i>ix</i>
Chapter 1. Introduction to structural biology of pre-mRNA splicing.	1
1-1. Introduction.	2
1-1.1. Pre-mRNA splicing.	2
1-1.2. Assembly of the spliceosome.	3
1-2. A structural approach to studying the spliceosome.	7
1-2.1. The RRM domain.	8
1-2.2. Recognition of pre-mRNA by hnRNP A1.	10
1-2.3. pre-mRNA recognition by Fox-1.	12
1-2.4. Polypyrimidine tract recognition by U2AF in E complex.	13
1-2.5. Protein Recognition mediated by U2AF homology motifs.	15
1-2.6. SF1 and branch point sequence recognition.	18
1-2.7. U1 snRNP: a high resolution view	19
1-2.8. Structure of RNA at the spliceosome active site	23
1-3. Crosslinking to spliceosomal proteins.	28
1-4. References.	28
Chapter 2. Crystal structure of a core spliceosomal protein interface.	41
2-1. Introduction	42
2-2. Results	46
2-2.1. Structure of the p14•SF3b155 complex.	46
2-2.2. Comparison with canonical and pseudo-RRMs.	51
2-2.3. RNA protein interactions in a minimal complex.	53
2-2.4. Adenine recognition by the p14•SF3b155 complex.	57
2-2.5. p14 within SF3b and the U11/U12 di-snRNP.	59
2-3. Discussion.	62
2-4. Methods.	64
2-4.1. Identification and alignment of p14 ortholog sequences.	64
2-4.2. Determination of SF3b155 p14 interaction domain.	64
2-4.3. Protein expression and purification.	65
2-4.4. Crystallization.	66
2-4.5. Data collection and processing.	66
2-4.6. Model building and refinement.	67
2-4.7. Crosslink Mapping.	68
2-4.8. Comparison of X-ray and cryo-EM structures.	69
2-5. References.	70

Chapter 3. RNA binding by the spliceosomal protein p14.	75
3-1. Introduction	76
3-2. Results and discussion.	79
3-2.1. Interrogating the p14•RNA complex using disulfide tethering strategy.	79
3-2.2. SAXS analysis of tethered p14•SF3b155 peptide•RNA complex.	86
3-2.3. Implications for the role of p14 in spliceosome assembly and activation.	91
3-2.4. <i>in vivo</i> depletion of p14.	94
3-3. Methods.	96
3-3.1. Protein expression and purification.	96
3-3.2. RNA synthesis.	96
3-3.3. Protein-RNA disulfide bond formation.	97
3-3.4. Protein-RNA disulfide reduction.	98
3-3.5. Small angle X-ray scattering.	98
3-3.6. <i>in vivo</i> depletion of p14 and SF3b155 using siRNAs.	100
3-3.7. Western blotting and RT-PCR.	101
3-4. References.	102
Chapter 4. The structure of Prp8 Domain IV.	110
4-1. Introduction	111
4-1.1. The highly conserved protein Prp8.	111
4-1.2. Prp8 crosslinks to important nucleotides in the spliceosome.	112
4-1.3. Prp8 protein domain structure.	114
4-1.4. A putative RRM in Prp8.	115
4-1.5. The MPN domain of Prp8.	116
4-1.6. Prp8 domain 3.	117
4-1.7. Prp8 Domain IV.	118
4-1.8. Domain IV as a target for structural studies.	118
4-2. Results and Discussion	120
4-2.1. Structural overview of hPrp8 Domain IV.	120
4-2.2. RNase H homology of the Domain IV core.	125
4-2.3. An RNase H domain in the spliceosome active site.	128
4-2.4. Identifying RNA binding partners of hPrp8 Domain IV.	130
4-2.5. Mapping Prp8/RNA interactions.	132
4-2.6. Implications of the core structure for the spliceosome active site.	133
4-3. Methods.	136
4-3.1. Identification and expression of an hPrp8 Domain IV core.	136
4-3.2. Crystallization.	136
4-3.3. Data collection and processing.	137
4-3.4. Model building and refinement.	137
4-3.5. RNA binding by hPrp8.	138
4-3.6. RNAs used in gel mobility shift with Prp8 Domain IV core.	138
4-3.7. Limited Proteolysis.	138

4-4. References.	139
Chapter 5. Metal-ion coordination mediated by PRP8 conformational change is coupled to pre-mRNA exon ligation.	145
5-1. Introduction	146
5-1.1. Prp8 alleles affecting the 1 st and 2 nd step of splicing.	146
5-1.2. C-terminal region of Prp8 at the spliceosome active site.	150
5-2. Results and Discussion	151
5-2.1. A conformational change unmasks a Mg ²⁺ binding site.	151
5-2.2. Disruption of Mg ²⁺ ion binding by Prp8 Domain IV impairs splicing.	154
5-2.3. PRP8 D1853C is not a first-step allele.	157
5-2.4. First- and second-step PRP8 alleles favour one of two conformations observed in the crystal structure of Prp8 RNase H domain.	159
5-2.5. The Prp8 Domain IV conformation change in the spliceosome.	165
5-3. Methods.	167
5-3.1. Protein expression, purification, and crystallization.	167
5-3.2. Data collection and processing .	167
5-3.3. Model building and refinement.	168
5-3.4. Creation of mutant <i>S. cerevisiae</i> strains splicing extracts, and splicing.	168
5-4. References.	169
Chapter 6. Summary and conclusions.	172
6-1. Summary and conclusions	173
6-1.1. The spliceosomal protein p14 plays an essential role in the spliceosome.	173
6-1.2. Domain IV of Prp8.	175
Appendix I. Context-dependent remodelling of structure in two large protein fragments	178
I-1. Introduction	179
I-2. Results.	182
I-2.1. Identification, purification, and characterization of Prp8 Δ D4 and D4.	182
I-2.2. Overview of Δ D4 and D4 structures.	183
I-2.3. Conservation of a helix-strand interaction between Δ D4 and D4.	186
I-2.4. Structural plasticity in the linking regions of Δ D4 and D4.	187
I-2.5. C-terminal conservation of tertiary structure in D4 and Δ D4.	189
I-2.6. Analysis of Δ D4/D4 structure and stability using CD spectroscopy.	189
I-3. Discussion	192
I-4. Methods	195
I-4.1. Identification and expression of a Prp8 domain IV core.	195
I-4.2. Crystallization.	196

I-4.3. Data collection and processing.	196
I-4.4. Model building and refinement.	197
I-4.5. CD spectroscopy.	197
I-5. References	198

List of Figures:

Chapter 1.	
1-1. Splicing of pre-mRNA by the spliceosome.	4
1-2. Structure of hnRNP A1 bound to DNA.	10
1-3. Structure of Fox-1 bound to RNA.	13
1-4. Early recognition of the polypyrimidine tract by U2AF.	16
1-5. Early branch sequence recognition by SF1.	19
1-6. Crystal structure of U1 snRNP.	21
1-7. Mechanism for spliceosome and group II intron splicing.	24
1-8. Comparisons of group II intron and spliceosomal active site RNA structures.	26
Chapter 2.	
2-1. Crosslinking of proteins to the branch nucleotide.	43
2-2. Sequence alignment of p14 orthologs.	45
2-3. Structure of p14•SF3b155 peptide complex.	48
2-4. Details of the p14•SF3b155 peptide interface.	50
2-5. Comparison between p14•SF3b155 peptide surface and structures of U1a and U2AF65 RRM 3.	52
2-6. Mapping of the p14•branch nucleotide interaction.	55
2-7. Recognition of adenosine by p14.	58
2-8. p14•SF3b155 peptide structure within SF3b and U11/U12 snRNP.	60
Chapter 3.	
3-1. Protein-protein and protein-RNA interactions of p14 in the human spliceosome.	78
3-2. Disulfide crosslinking strategy for interrogating the p14•SF3b155 peptide•RNA complex.	80
3-3. p14•SF3b155 peptide•RNA disulfide formation and reduction.	82
3-4. Small angle X-ray scattering analysis of p14•SF3b155 peptide and tethered p14•SF3b155 peptide•RNA complexes	88
3-5. <i>in vivo</i> knockdown of p14 and SF3b155 proteins using RNAi at 66 or 96 hours.	95
Chapter 4.	
4-1. Structure of hPrp8 Domain IV.	121
4-2. Secondary-structure diagram and sequence conservation of Prp8 Domain IV.	123
4-3. Mapping of yeast alleles and a 5'SS cross-link in the context of the Prp8 Domain IV core.	124
4-4. Comparison of the Prp8 Domain IV core with RNase H folds.	126

4-5. Analysis of the putative conserved binding site in Prp8 Domain IV core.	128
4-6. RNA binding by the hPrp8 Domain IV core.	131
4-7. Footprinting analysis of a Prp8/RNA complex.	133
4-8. Representations of the proposed Prp8 domain IV core RNA binding surface.	134
Chapter 5.	
5-1. First- and second-step alleles of PRP8.	149
5-2. Conformational plasticity in the PRP8 RNase H domain.	155
5-3. The PRP8 D1853C mutation impairs the second step of splicing.	158
5-4. Structure and phenotype of V1860D/V1788D Prp8.	160
5-5. Structure of the T1789P Prp8 Domain IV.	162
5-6. Structure of the L1798N and N1797D Prp8 Domain IV mutants.	164
Chapter 6.	
6-1. Model of p14•RNA complex.	174
6-2. The two-step model of the spliceosome active site.	155
Appendix I.	
I-1. Structural analysis of D4 and Δ D4.	184
I-2. Conservation of a helix-strand tertiary interaction in D4 and Δ D4	188
I-3. Remodeling of the helical bundle protein interface.	190
I-4. Solution structure and stability of Δ D4 and D4.	191

List of Tables:

Chapter 2.

Table 2-1. Crystallographic data collection and refinement statistics 47

Chapter 3.

Table 3-1. Parameters and quality indicators derived from scattering data. 87

Table 2-2. Crystallographic data collection and refinement statistics. 74

Chapter 4.

Table 4-1. Mutant Prp8 alleles mapping to domain IV and corresponding phenotypes. 120

Table 4-2. Crystallographic data collection and refinement statistics. 122

Chapter 5.

Table 5-1. Mutant PRP8 alleles mapping to domain IV core and corresponding phenotypes. 148

Table 5-2. Crystallographic data collection and refinement statistics. 153

Appendix I.

Table I-1. Data collection, phasing and refinement statistics for $\Delta D4$. 185

List of Abbreviations:

Δ	delta, depleted
A	adenosine
aa	amino acid
ATP	adenosine triphosphate
BPS	branch point sequence
BS-C	ACT-1 pre-mRNA branch site A to C mutation
C	cytosine
CBP	cap-binding protein
CC	commitment complex
cs	cold-sensitive
CTD	carboxy-terminal domain
CNBr	cyanogen bromide
DTNB	5,5'-Dithio-Bis(2-Nitrobenzoic Acid)
DTT	dithiothreitol
E complex	early complex
EDTA	ethylene diamine tetraacetic acid
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
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PDB	protein databank
PPT	poly-pyrimidine tract
PRP	pre-mRNA processing
pre-mRNA	precursor messenger-RNA
r.m.s.d.	root-mean-square deviations
RP	<i>retinitis pigmentosa</i>
RRM	RNA recognition motif
RS	arginine-serine rich domain
RT	reverse transcriptase
SDS	sodium dodecyl sulfate
Se-Met	selenomethionine
SL	stem loop
snRNA	small nuclear RNA
snRNP	small nuclear ribonucleoprotein particle
SS	splice site

TEV	tobacco etch virus
<i>ts</i>	temperature-sensitive
U	uridine
U2AF	U2 auxiliary factor
UHM	U2AF homology motif
WT	wild type

Chapter 1⁽¹⁾
Introduction to structural biology of pre-mRNA splicing

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

1-1. Introduction

1-1.1. 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). The split gene structure was subsequently found to be common to most eukaryotic genes, and transcription of genes replicates this structure in pre-mRNAs.

An accurate and efficient mechanism exists in eukaryotic cells, whereby the introns are excised and the flanking exons are ligated together to produce a mRNA which encodes a functional protein. This process is called splicing, and is catalyzed by a large and complex macromolecular machine called the spliceosome (Kramer, 1996; Burge et al., 1999), which recognizes and splices introns defined by conserved sequence elements. The 5' end of the intron is defined by the 5' splice site (5'SS) sequence, which contains a conserved GU dinucleotide. The 3' splice site (3'SS) sequence is more complex; it is defined by the branch point sequence (BPS) which is followed by the poly-pyrimidine tract (PPT) and a conserved YAG trinucleotide at the 3'SS (Figure 1-1A). In many eukaryotes (especially complex multi-cellular ones) a combination of splicing enhancers or repressors are present in the vicinity of splice sites (within both introns and exons), which are recognized by proteins that aid the spliceosome in splice site selection (Luco et al., 2010). Once the spliceosome has identified an intron in a

pre-mRNA, it catalyses two phospho-transesterification reactions to “splice out” the intron sequence. 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 1-1B).

The spliceosome is 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), and 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 discrete intermediate complexes, and is guided by conserved sequences within the intron recognized by both proteins and RNAs in a mechanism which is highly conserved from human to yeast (Figure 1-1C).

1-1.2. Assembly of the spliceosome

The initial steps of spliceosome assembly involve the ATP independent formation of the early (E; mammalian) or commitment complex (CC; yeast) on the pre-mRNA substrate. In mammals, this complex includes U1 snRNP tightly

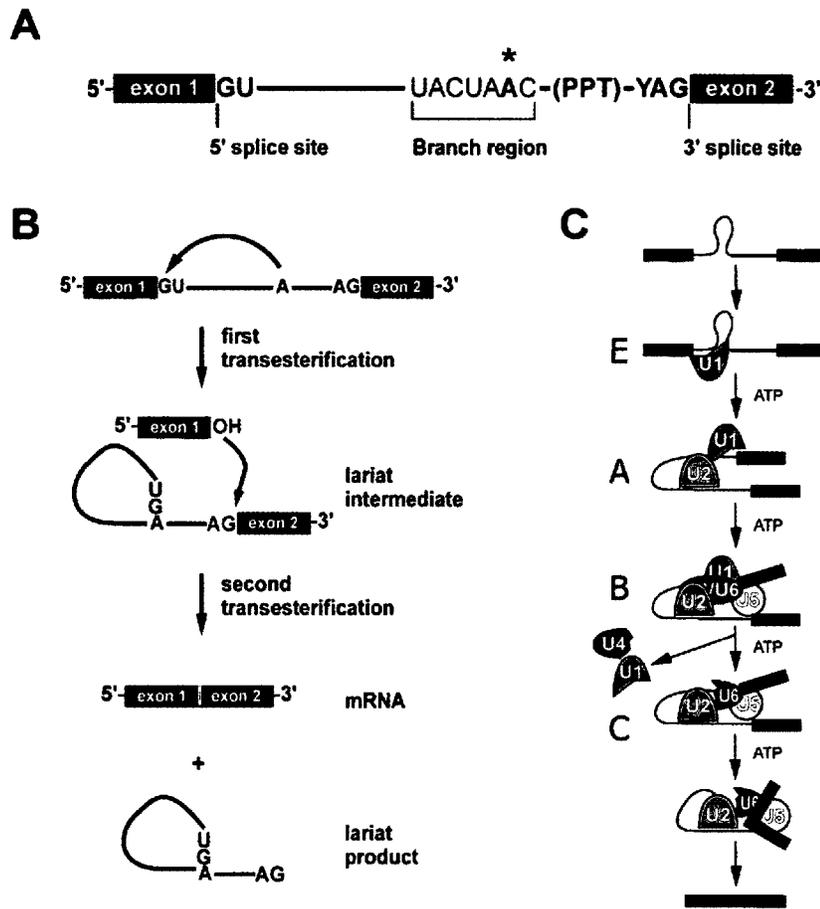


Figure 1-1. Splicing of pre-mRNA by the spliceosome. A) Intron structure highlighting conserved sequences at the 5'SS and 3'SS, the optimal branch sequence, and metazoan polypyrimidine tract. The preferred branch adenosine is indicated (*). B) Sequential transesterification reactions catalyzed by the 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.

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 cap-binding protein CBP80 as well as U1 snRNP associated proteins aid the interaction between the 5'SS and U1 snRNA (Puig et al., 1999; Zhang & Rosbash, 1999). The 3'SS is identified by the U2

PRP8 Metal Binding Coupled to Exon Ligation

promote hydrolysis by activation of a water nucleophile combined with transition state stabilization.²² Inspection of the PRP8 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. Despite the fact that our original crystals were grown in 100-200 mM $MgCl_2$, we did not detect a metal ion in the PRP8 RNase H domain site. Instead, the spatially conserved aspartates are stabilized by a network of water molecules and the side chain of Arg1865 is hydrogen-bonded to Asp1782 possibly blocking the metal binding site (Fig. 2b). This has previously been observed in the RNase H fold of the Argonaute PIWI domain and also in the Tn5 transposase.²³

If R1865 was blocking a canonical RNase H-like metal-binding site of PRP8, this site might be unmasked by displacement of this side chain, for example by RNA binding during spliceosome assembly/activation. In order to model this, we expressed, crystallized, and solved the 1.15 Å X-ray structure of the R1865A PRP8 mutant which mimics a displacement of R1865. This structure is remarkably similar to that of the wild-type protein including two monomers in the asymmetric unit but with the disrupted β -hairpin of monomer b significantly more ordered. One consequence of the structural rearrangement involved in the a to b transition is that D1782 moves closer to D1781 and T1783 is displaced ~ 4 Å upwards (Fig. 1b). Significantly, within monomer b of this structure, a single Mg^{2+} ion bound at high occupancy is observed in the canonical RNase H site. Coordination of this ion includes inner sphere contact with the side chain of D1781 and outer sphere coordination via five ordered waters to the carboxylate of D1782, carbonyls of D1782 and L1891, the amide carbonyl of Q1894, and the hydroxyl of T1864. We confirmed this assignment by soaking the R1865A crystals in $CoCl_2$ and observing strong anomalous diffraction from Co^{2+} bound in the RNase H site (Fig. 1). Interestingly, under

PRP8 Metal Binding Coupled to Exon Ligation

new crystallization conditions, we were ultimately able to observe partial occupancy of this site by Mg^{2+} , Mn^{2+} , or Co^{2+} in monomer b of wild-type protein; the weaker association may be explained by the proximity of the R1865 positive charge. It thus appears that it is the larger conformational change observed between a and b that is required for binding of divalent metal to the PRP8 RNase H site although rearrangement of R1865 might contribute to enhanced metal-binding within the context of the spliceosome.

It has been shown previously that mutation of the residues shown here to be metal coordinating is deleterious in yeast.^{20,21} For example, mutation of the outer sphere ligand, D1854A (human D1782A), results in severe growth defects at both 16° and 37° while the mutation of the inner sphere D1853 to alanine is lethal. In order to further investigate the role of metal-binding in PRP8 function, we mutated the inner sphere ligand D1853 to cysteine in yeast PRP8 based on the expectation that this mutation should reduce Mg^{2+} coordination at this site because the softer ligand strongly disfavours the hard metal. The D1853C allele confers a slow growth phenotype at 30°C and manifests a severe growth defect at 37°C. We examined *in vitro* splicing of an actin pre-mRNA comparing extracts prepared from wild-type and D1853C PRP8 yeast strains. The D1853C mutant extract forms spliceosomal complexes and is competent for the first step of splicing but exhibits a strong block to the second step. Attempted rescue of the second step in the presence of the thiophilic divalent manganese ion was unsuccessful in contrast to Mn^{2+} rescue of splicing in phosphorothioate-modified U6 snRNA where the substitution disrupts the critical U80 Mg^{2+} binding site. The impairment in D1853C function is not due to any gross structural change. We solved the structure of the corresponding mutation in human PRP8 (D1871C) and the architecture of the metal binding site is preserved. As expected,

PRP8 Metal Binding Coupled to Exon Ligation

Mg²⁺ does not bind to this site; however in contrast to the wild-type protein, we were unable to observe bound Mn²⁺ following crystal soaks.

Query and Konarska have suggested an elegant model whereby PRP8 regulates the equilibrium between two distinct spliceosomal conformations associated with the first and second transesterification steps respectively.¹⁸ Two sets of PRP8 alleles, designated as first- or second-step, are proposed to act by shifting this equilibrium to favour one step over the other. Because we only observe Mg²⁺ ion binding in monomer b in our crystal lattice, and abrogation of Mg²⁺ binding by the D1853C mutation in PRP8 causes a strong block to the second step of splicing, we determined the structures of some PRP8 point mutants that corresponded to first- and second-step alleles.

Because the largest group of suppressor alleles in the RNase H domain, including those designated as first or second step, map to the β -hairpin, we endeavoured to determine whether these are specifically associated with either of the two conformations observed crystallographically. We investigated mutants of the human protein corresponding to both first- and second-step suppressor alleles that crystallized as described above with one monomer of the asymmetric unit containing the β -finger and the second divalent metal-binding monomer featuring the displaced loop structure.

The yeast PRP8 allele V1860D has been proposed to be a first step allele, which we confirmed by observing an improvement of the first step of splicing of a BS-C pre-mRNA. We were able to solve a 1.95 Å crystal structure of the human V1788D mutant corresponding to this yeast allele. In this case, compared to the wild-type, monomer a contains an additional

PRP8 Metal Binding Coupled to Exon Ligation

hydrogen-bonding interaction between the side chains of D1788 and Y1786 which would argue that the yeast allele favours the β -finger/non-metal binding conformation.

In order to examine a second-step suppressor, we crystallized and solved the 2 Å structure of the human T1789P mutant corresponding to the yeast T1861P allele. As expected, the β -hairpin is disrupted in monomer a but the proline substitution is accommodated in metal-binding monomer b.

In the studies described here, we have observed crystallographically two distinct conformations of the PRP8 RNase H domain, one of which permits coordination of Mg^{2+} at the conserved RNase H metal-binding site. The structures of proteins corresponding to known suppressor alleles argue that the non-metal binding conformation corresponds to a first-step conformation and link divalent metal coordination by PRP8 to the second step of splicing. The latter hypothesis is supported by the results of mutagenesis of the inner-sphere D1853 in yeast.

The transition between first and second step states in the spliceosome is more complex than that described here as evidenced by alleles associated with domain 3 of PRP8. Nevertheless, the conformational change described here is most likely a representation of a switch between the two steps of splicing. The fact that it is coupled to binding of a required metal ion implicates it as a key regulatory mechanism in catalysis of the second step of splicing.

It has long been believed that the spliceosome is in essence a ribozyme based, in part, on the mechanistic similarity between self-splicing group II introns and the processing of nuclear pre-mRNAs.²⁴ There is also considerable evidence for the role of the spliceosomal U2/U6 snRNA structure and particularly U6 snRNA in catalysis of splicing.^{25,26} High resolution structural analysis of the group I and group II introns has supported a general two-metal ion

PRP8 Metal Binding Coupled to Exon Ligation

mechanism proposed for catalytic RNA for, including the spliceosome, based on analysis of phosphoryl transfer by protein enzymes.²⁷⁻²⁹ There is nevertheless scope for the involvement of more or fewer metal ions in similar enzymatic transformations. Berger and coworkers have recently proposed a mechanism for type II and IA topoisomerases where the transition state is stabilized by a single bound metal and conserved arginine.³⁰

Methods Summary

Acknowledgements

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported the NIH (GM033944 and GM053960 for N.O.; T32CA09592 for J.E.D.; GM08295 for B.H.S. and CA077373 for J.M.B.).

Author Contributions E. purified the complex, grew the crystals, and solved the structures. J.M.B. assisted with refinement and inspection of the molecular models. A.B.B. synthesized the phosphorothiolate reagent. J.E.D. assisted the design of the DNA substrate and optimizing cleavage conditions. N.O. and J.M.B. designed the experiments. All authors contributed to the manuscript.

Author Information Coordinates for the apo and Zn-bound complexes have been deposited in the RSCB PDB under the accession numbers 3L4J and 3L4K. Reprints and permissions information is available at www.nature.com/reprints. The authors

1-2.2. Recognition of pre-mRNA by hnRNP A1

hnRNP A1 is an abundant and well studied nuclear protein. Its function has been implicated in pre-mRNA splice site selection, transcription, mRNA export, as well as in the maintenance of telomeres (Ding et al., 1999 and references therein). The ability of hnRNP A1 to influence splice site choice by the spliceosome depends on its two RRM domains, which bind single-stranded nucleic acid (Mayeda et al., 1998). Because hnRNP A1 binding sites can reside in exons, the protein must be able to recognize a degenerate RNA target sequence so that a binding site within a protein coding exon does not place excessive restriction on the encoded amino acid sequence. A crystal structure of the two RRM domains of hnRNP A1 bound to single stranded DNA shows how hnRNP A1 can bind nucleic acid (Figure 1-2; Ding et al., 1999). The substrate contains

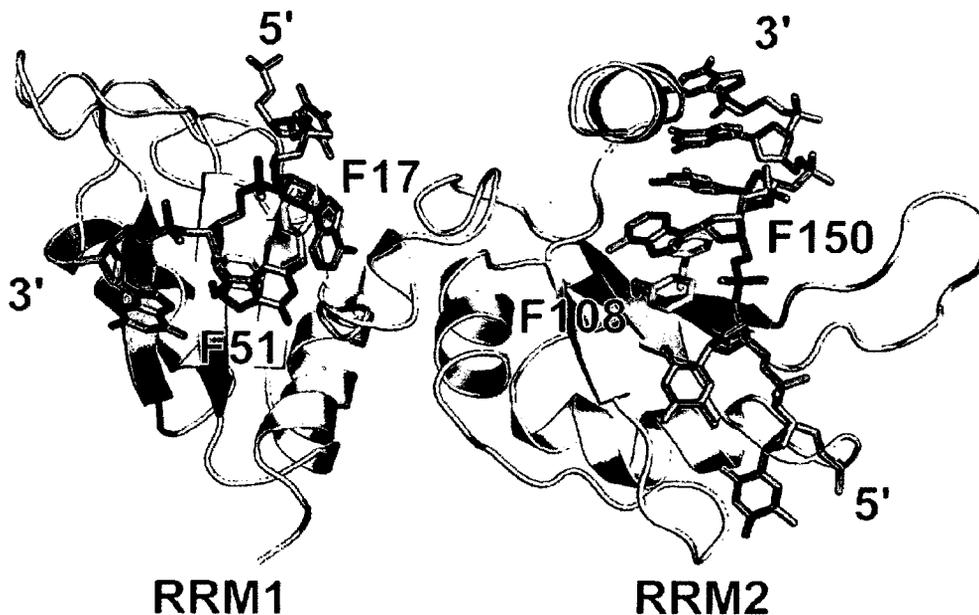


Figure 1-2. Structure of hnRNP A1 bound to DNA. RRM 1 and 2 of hnRNP A1 (grey) bind single stranded DNA with sequence TTAGGG (cyan) in an antiparallel fashion. Aromatic amino acids of the RNP1 and RNP2 motifs (yellow) form π -stacking interactions with nucleotides in the single-stranded nucleic acid substrate.

tandem repeats of the consensus binding site for hnRNP A1 (TTAGGG), but the majority of the interaction is with the central TAGG of each repeat. Each RRM binds the DNA across the face of the β -sheet, forming a π -stacking interaction with the two aromatic residues of the RNP1 and RNP2 motifs, while the nucleobases form hydrogen bonds with the protein. Interestingly, none of the structural interactions are specific for DNA or RNA, which implies that the RRM domain is not solely for RNA recognition, but rather RRMs recognize single stranded nucleic acid in general, showing how hnRNP A1 can function in both alternative splicing and telomere maintenance. The two RRMs do not act independently, but are bound together in an anti-parallel fashion such that two TAGG sequences (or UAGG in RNA) separated by a variable distance can function as an hnRNP A1 binding site, and this may cause localized RNA bending which could be important for hnRNP A1's ability to modulate alternative splicing (Ding et al., 1999). Using two short (3-4 nucleotides) binding sites which can be separated by a variable distance allows hnRNP A1 binding sites to be located within exons without limiting the protein coding ability of these regions. This recognition of a short sequence is echoed by the NMR structure of the second RRM of hnRNP D bound to single stranded DNA, where the protein-DNA interaction is essentially limited to the central TAG trinucleotide (Enokizono et al., 2005). The structure of the RRM of another protein, SRp20 also makes very few sequence-specific hydrogen bonds (Hargous et al., 2006), implying that recognition of short sequences is the method by which splicing factors recognize splicing signals which may reside within exons.

1-2.3. pre-mRNA recognition by Fox-1

Fox-1 is an alternative splicing factor with orthologues throughout the animal kingdom. Its expression has been shown to vary in different tissues, affecting splice site selection by the spliceosome (Nakahata and Kawamoto, 2005). Its RNA binding activity resides in a single RRM domain, and it recognizes a UGCAUG element which is typically found in introns of pre-mRNAs (Brudno et al., 2001). In contrast to the previous examples of RRMs, this single RRM binds a specific sequence with a very strong affinity, and is able to distinguish its binding site from sequences with a single nucleotide difference. A high resolution structure containing the Fox-1 RRM bound to the heptanucleotide RNA UGCAUGU demonstrates how Fox-1 is able to accomplish this specificity (Auweter 2006). Two nucleotides (U5, G6) bind across the face of the β -sheet, interacting with the RNP1 and RNP2 motifs in the canonical fashion for an RRM domain, while the first four nucleotides (U1-A4) are wrapped around a conserved phenylalanine (F126) at a site distinct from the other three nucleotides (Figure 1-3). In this structure, F126 of Fox-1 stacks between U1 and G2 while C3 forms a hydrophobic interaction with it, and A4 forms a shallow groove AG base pair with G2. The protein also makes extensive hydrogen bonds with the bases, including some that form between nucleotides due to the conformation imposed on it by Fox-1 binding. This induced RNA conformation enhances the protein's ability to discriminate between an RNA sequence with a single mismatch, as a difference in one nucleotide would propagate the effect of the mismatch,

disrupting RNA-protein interaction between multiple nucleotides. As the authors demonstrate in their paper, a single nucleotide change can cause up to a 1800 fold decrease in Fox-1 binding.

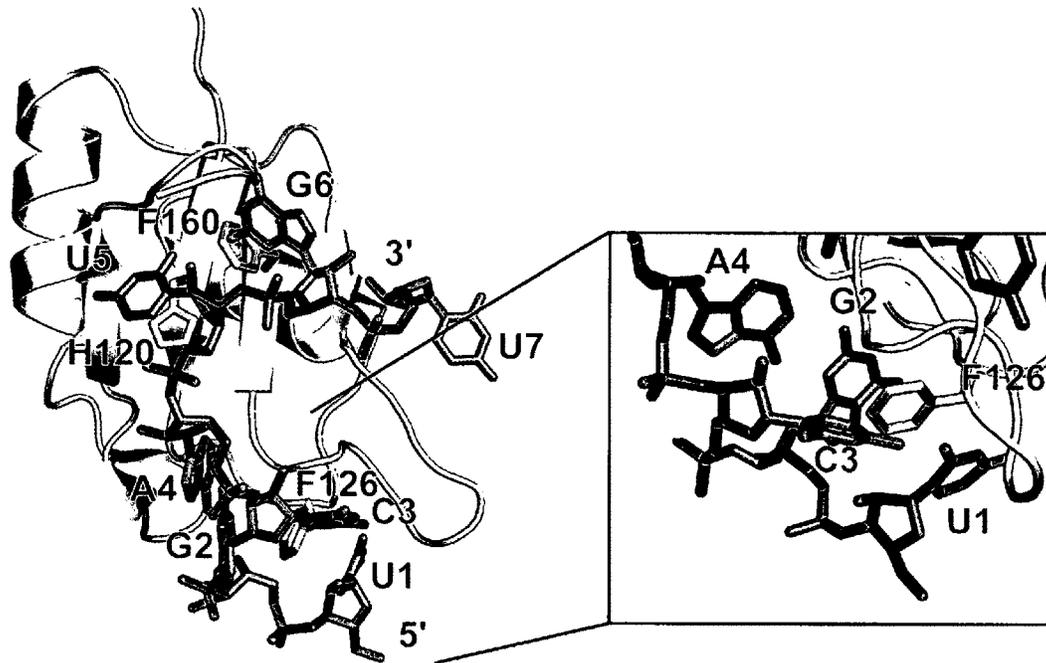


Figure 1-3. Structure of Fox-1 bound to RNA. (Left) The RRM domain of Fox-1 (grey) binds single stranded RNA with sequence UGCAUGU (cyan). Aromatic amino acids of the RNP1 and RNP2 motifs (H120, F160, yellow) form pi stacking interactions with nucleotides in the single-stranded nucleic acid substrate. (Right) Fox-1 RRM forms additional RNA interactions. F126 intercalates between U1 and G2 and forms a hydrophobic interaction with C3. Fox-1 also induces a shallow-groove basepair between G2 and A4

1-2.4. Polypyrimidine tract recognition by U2AF in E complex

The U2 auxiliary factor (U2AF) is a heterodimer composed of 65 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 domains originally

described as RRM. However, only the first two of these RRM domains are involved in RNA binding; the third represents a general protein interaction domain referred to as a UHM (U2AF Homology Motif; see section 1-2.5).

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 RRM 1 and 2 bound to a poly-uridine 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 allows hydrogen-bonding interactions with the occasional non-uridine nucleotide in the 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 proposed a model for the U2AF65 pyrimidine tract interaction in which the RRMs are in close proximity in a relatively condensed structure (Figure 1-4A). Solution data of the protein alone obtained by small angle X-ray scattering (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 1-4B). 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. 3'SS 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.

1-2.5. 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. 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 2 and β 4, and a conserved glutamate on α 1 which forms a salt bridge with the arginine side chain. The UHM consensus

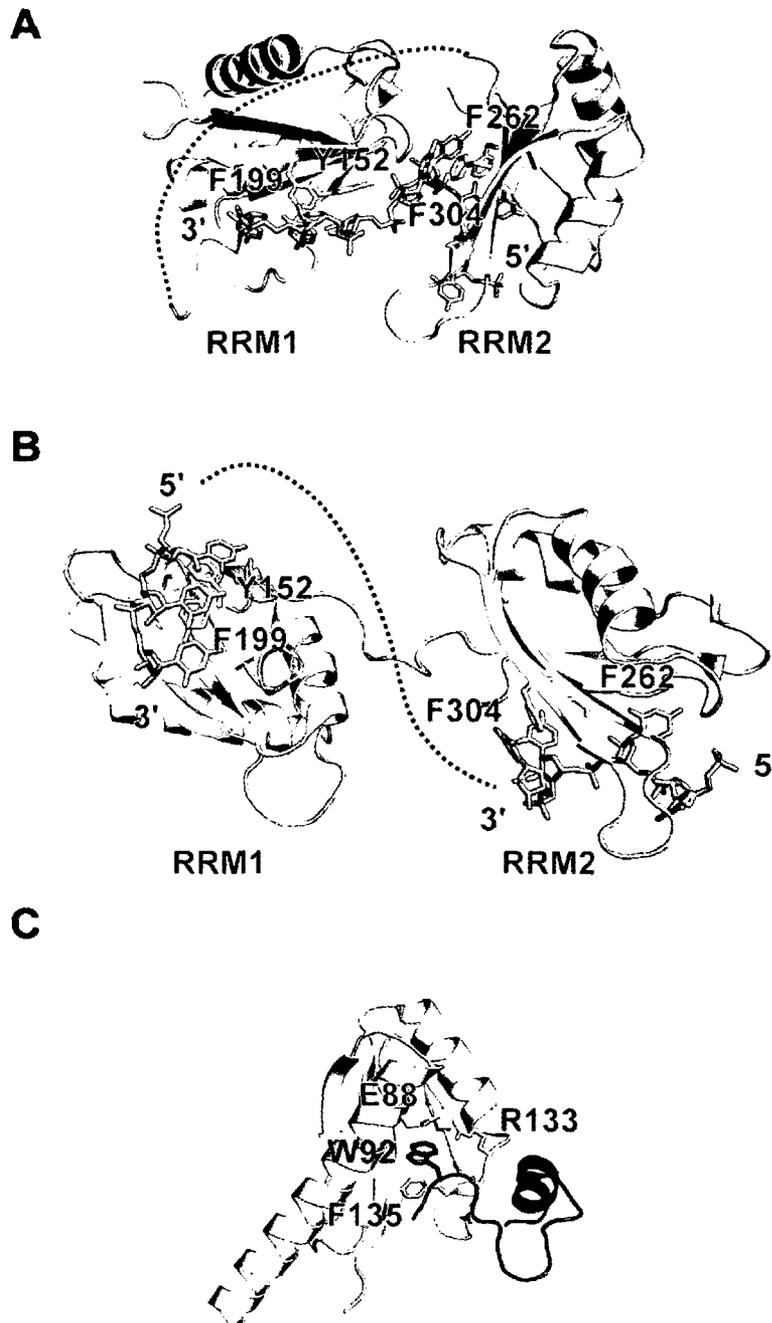


Figure 1-4. Early recognition of the polypyrimidine tract by U2AF. Two possible modes of interaction based on the X-ray structure of U2AF65 RRM1 and RRM2 (grey) bound to a seven nucleotide polyuridine RNA (cyan) (Sickmier et al., 2006). A) Model based on interaction with a single RNA with RRM1 and RRM2 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 (W92, green) is sandwiched between F135 and a salt bridge formed between E88 and R133 (yellow) of the U2AF35 UHM domain.

ligand is a peptide containing the [RK]-X-[RK]-W sequence with the tryptophan side-chain buried in a binding pocket formed between the UHM phenylalanine and the salt bridge (Kielkopf et al., 2004). This interaction mode was first observed in the X-ray structure of the RRM of U2AF35 bound to a short peptide from the N-terminal region of U2AF65 (Kielkopf et al., 2001). This UHM sandwiches W92 from the peptide ligand between F135 and the salt bridge formed between E88 and R133 (Figure 1-4C). 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 spliceosome 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 that have been shown to be bound preferentially by different UHM proteins (Corsini et al., 2007; Thickman et al., 2006).

1-2.6. 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 E complex, the association of SF1 with the branch sequence is mediated by the protein's hnRNP K homology (KH) domain.

The 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 I 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 1-5A). The branch adenosine is specifically recognized on the Watson-Crick face by hydrogen bonding with the peptide backbone of I177 (Figure 1-5B). The branch point

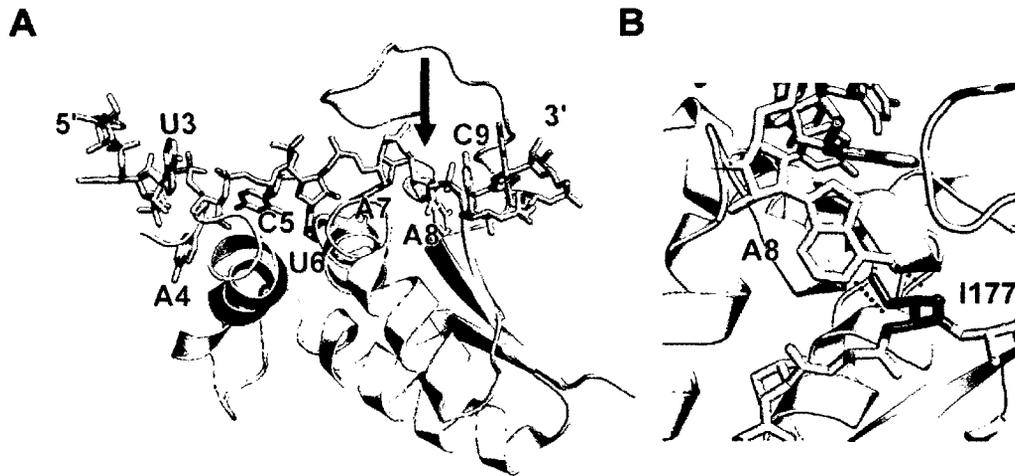


Figure 1-5. Early branch sequence recognition by SF1. A) NMR structure of the SF1 KH domain (grey) (Liu et al., 2001) bound to the branch point sequence RNA (cyan) with branch adenosine (yellow) buried within a pocket (arrow). B) The branch adenosine A8 forms Watson-Crick type hydrogen bonds with the peptide backbone of I177 (green) of SF1.

sequence is subsequently recognized by U2 snRNA which forms an imperfect duplex that extrudes the branch adenosine (Query et al., 1994); the conformation in which SF1 holds the branch point sequence likely templates the formation of this duplex.

1-2.7. U1 snRNP: a high resolution view

Initial identification of the 5'SS sequence by the spliceosomal machinery 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 U1 snRNP is composed of U1 snRNA, the proteins U1A, U1-70K, U1C, and the 7 Sm core proteins. 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 SL I-III and helix H of U1 snRNA, the threading of the RNA through the Sm core, and helix IV are all apparent in the structure (Figure 1-6A). 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, forming 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 (Figure 1-6B).

The Sm core proteins form a heptameric ring structure and the seven nucleotides which constitute a 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 1-6B). Thus, the

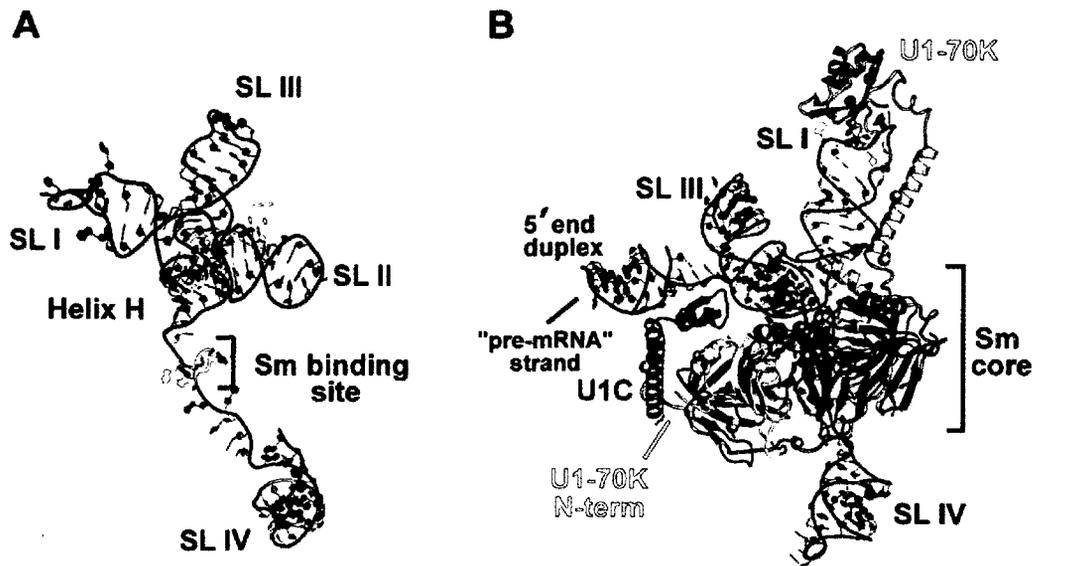
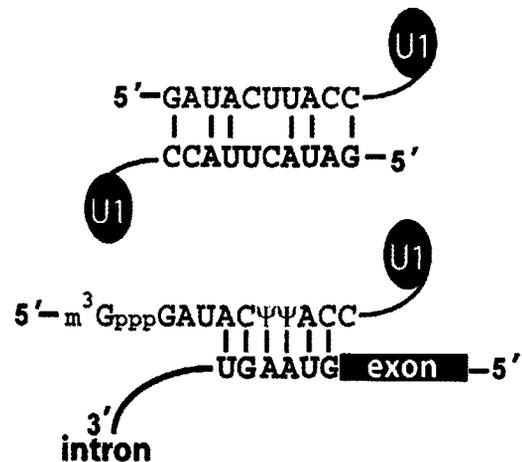


Figure 1-6. Crystal structure of U1 snRNP. A) Conformation of U1 snRNA within the snRNP. The RNA forms a four-helix junction between SLI-III and Helix H. The Sm binding site is 3' to this junction followed by SL IV. B) (upper) The Sm protein core forms a platform upon which the four helix junction is 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).



structure suggests that the Sm core acts as a platform for complex RNA structures; because complex RNA structures are found at the 5' side of the Sm binding site, Sm scaffolding of RNA structures is likely to be a common feature of the other spliceosomal snRNPs (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 1-6B). As described above, U1 snRNA•5'SS pairing is mimicked in the crystal lattice and the zinc finger of U1C interacts 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 identify and report formation of the correct 5'SS interaction to the rest of the U1 snRNP.

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 have 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 1-6B).

The U1 snRNP structure resolves a number of features left undefined in structural studies of individual components. 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-

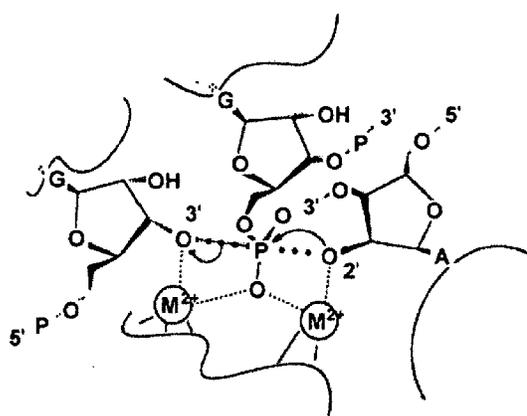
terminus of U1-70K could be attained because its conformation is dependent on its context within the larger particle.

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).

1-2.8. Structure of RNA at the spliceosome active site

Once assembled the spliceosome active site is believed to be composed of a U2/U6 snRNA structure. This model is supported by experiments where a protein-free preparation of U2/U6 snRNA based on the sequences of human U2 and U6 snRNAs 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 self splicing group II introns (Valadkhan et al., 2009). This RNA-catalyzed ligation reaction is very sensitive to changes in the sequence of two conserved elements of U6 snRNA, the AGC triad and the ACAGAGA box, similar to the 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).



1st step

Figure 1-7. 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).

A two-metal ion mechanism has been proposed for both the group II intron system as well as the spliceosomal reaction on the basis of the mechanism of DNA and RNA polymerases in phosphoryl-transfer reactions (Figure 1-7) (Steitz and Steitz, 1993). Domain 5 of group II introns is a metal-binding platform that coordinates the Mg^{2+} ions that are likely responsible for activating the first step nucleophile and 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 (Figure 1-8A; 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).

Domain 5 of group II introns and the spliceosomal U6 ISL are believed to be structurally and mechanistically analogous (Sigel et al., 2000). The NMR data with respect to the yeast U2/U6 RNA structure, 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 1-8B) (Rupert and Ferre-D'Amare, 2001; Sashital et al., 2004). 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. 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 (Sashital et al., 2004) analogous to the situation observed in the crystal structure of the *O. theyensis* group II intron (Toor et al., 2008). An unpaired residue, U80 in the U6 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).

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 thought to make base-pairing interactions with U2 snRNA residues to form U2/U6 helix Ib. 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). However, this model has been challenged recently by data showing that the AGC triad is involved in base-pairing to U2 snRNA (Mefford and Staley, 2009). Furthermore, it has been shown that a protein free U2/U6 complex can change between 3 different conformations in solution (Guo et al., 2009) which emphasizes that spliceosomal snRNAs are likely dynamic in their structure, and one snapshot will not tell the entire story.

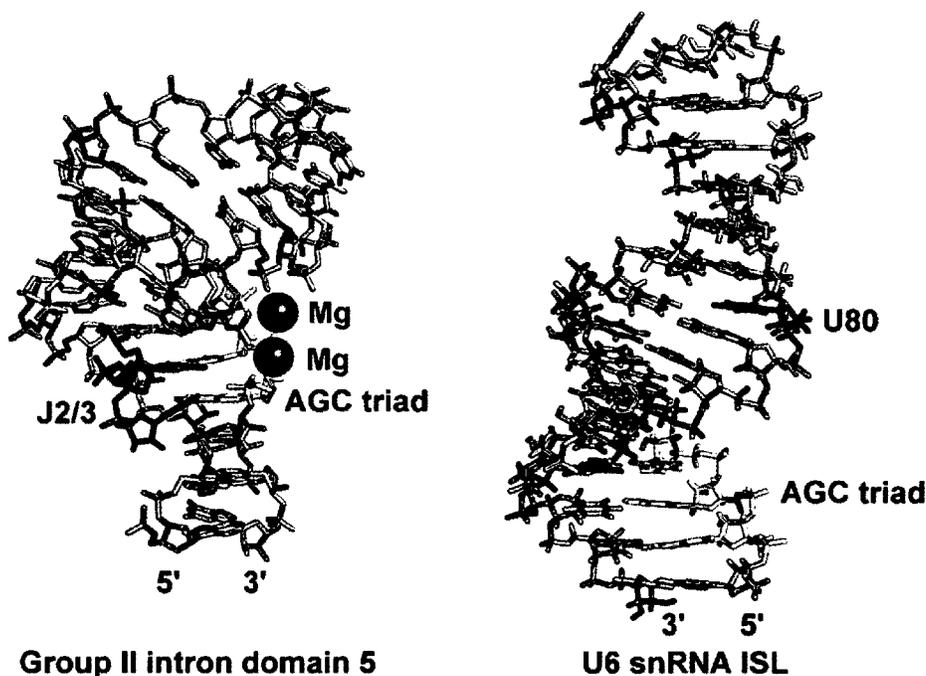


Figure 1-8. Comparison of group II intron and spliceosomal active site RNA structures. The crystal structure of domain 5 from the *Oceanobacillus theyensis* group II intron (Toor et al., 2008) (left) alongside the NMR structure of yeast U6 snRNA ISL (right) (Sashital et al., 2004). The location of two 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).

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. In addition, 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 1-8; 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).

While high resolution structures of spliceosomal complexes have been elusive, the landmark X-ray analysis of U1 snRNP represents the culmination of current 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 between the RNA and protein catalyzed worlds.

1-3. Crosslinking to spliceosomal proteins

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). A synthetic pre-mRNA substrate containing the photo-activatable crosslinker benzophenone tethered to the branch adenosine shows several proteins are within 15Å of the branch adenosine, including a 14 and a 220 kDa protein (Figure 1-9A; Macmillan et al., 1994). Furthermore, a substrate containing a short range 2,6-diaminopurine nucleotide crosslinker in the branch position crosslinks strongly to the 14 kDa protein, showing it is within ~3Å of the branch nucleotide (Query et al., 1996). Because these two proteins are near the reactive nucleotides of the spliceosome, a definition of their function will greatly enhance our understanding of pre-mRNA splicing. This thesis will describe the results such investigations.

1-4. 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.

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.

Brudno M., Gelfand M., Spengler S., Zorn M., Dubchak I., Conboy J. (2001) Computational analysis of candidate intron regulatory elements for tissue-specific alternative pre-mRNA splicing. *Nucleic Acids Res* **29**, 2338–2348

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).

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.

Ding, J., Hayashi, M.K., Zhang, Y., Manche, L., Krainer, A.R., Xu, R.M. (1999) Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. *Genes Dev.* **13**, 1102-15.

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.

Enokizono, Y., Konishi, Y., Nagata, K., Ouhashi, K., Uesugi, S., Ishikawa, F., Katahira, M. (2005) Structure of hnRNP D complexed with single-stranded telomere DNA and unfolding of the quadruplex by heterogeneous nuclear ribonucleoprotein D. *J Biol Chem.* **280**, 18862-70.

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.

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.

Guo Z., Karunatilaka K.S., Rueda D. (2009) Single-molecule analysis of protein-free U2-U6 snRNAs. *Nat. Struct. Mol. Biol.* **11**, 1154-9.

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.

Hargous, Y., Hautbergue, G.M., Tintaru, A.M., Skrisovska, L., Golovanov, A.P., Stevenin, J., Lian, L.Y., Wilson, S.A., Allain, F.H. (2006) Molecular basis of RNA recognition and TAP binding by the SR proteins SRp20 and 9G8. *EMBO J.* **25**, 5126-5137.

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. Mol. 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., and Moore, M.J. (2003). Pre-mRNA splicing: awash in a sea of proteins. *Mol. Cell* **12**, 5-14.

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.

Koodathingal, P., Novak, T., Piccirilli, J.A., Staley, J.P. (2010) The DEAH Box ATPases Prp16 and Prp43 Cooperate to Proofread 5' Splice Site Cleavage during Pre-mRNA Splicing. *Mol Cell.* **39**, 385-395.

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.

Lardelli R.M., Thompson J.X., Yates J.R. 3rd, Stevens S.W. (2010) Release of SF3 from the intron branchpoint activates the first step of pre-mRNA splicing. *RNA*, **16**, 516-28.

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.

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.

Long J.C., Caceres J.F. (2009) The SR protein family of splicing factors: master regulators of gene expression. *Biochem J.* **417**, 15-27.

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.

Luco R.F., Pan Q., Tominaga K., Blencowe B.J., Pereira-Smith O.M., Misteli T. (2010). Regulation of alternative splicing by histone modifications. *Science* **327**, 996-1000.

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.

Mayeda, A., Munroe, S.H., Xu R.M., Krainer, A.R. (1998) Distinct functions of the closely related tandem RNA-recognition motifs of hnRNP A1. *RNA*, **4**, 1111-1123.

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.

Nakahata S, Kawamoto S (2005) Tissue-dependent isoforms of mammalian Fox-1 homologs are associated with tissue-specific splicing activities. *Nucleic Acids Res* **33**, 2078–2089

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.

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.

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.

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

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., 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.

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.

Query C.C., Strobel S.A., Sharp P.A. (1996) Three recognition events at the branch-site adenine. *EMBO J.* **15**, 1392-402.

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.

Schwer B., Guthrie C. (1992) A conformational rearrangement in the spliceosome is dependent on PRP16 and ATP hydrolysis. *EMBO J* **11**, 5033–5039.

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.

Small E.C., Leggett S.R., Winans A.A., Staley J.P. (2006) The EF-G-like GTPase Snul14p regulates spliceosome dynamics mediated by Brr2p, a DExD/H box ATPase. *Mol. Cell.* **23**, 389-99.

Smith, D.J., Konarska M.M. (2008) Mechanistic insights from reversible splicing catalysis. *RNA* **14**, 1975-1978.

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.

Xu, Y.Z., Query, C.C. (2007) Competition between the ATPase Prp5 and branch region-U2 snRNA pairing modulates the fidelity of spliceosome assembly. *Mol Cell*. **28**, 838-849.

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, 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.

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 2⁽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, and Schellenberg et al, (2010). *RNA*, *in press*.

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. R.E., M.S., and M.G solved the crystal structure, and M.M.G, H.S., and R.L. performed the fitting of the p14•SF3b155 structure into the cryo-EM maps.

2-1. Introduction

A consideration of available data describing the structure of the active sites of other ribozymes provides a useful starting point to probe the architecture of the spliceosome active site. The ribosome is another ribozyme similar in size to the spliceosome. X-ray crystal structures of the ribosome show that the closest protein is about 18 Å from the active site, and therefore the active site is composed of RNA (Yusupov et al., 2001). pre-mRNA splicing in eukaryotes and most group II introns begins with the nucleophilic attack of the branch nucleotide on the 5' splice site (5'SS; Figure 1-1B); examining structures around this nucleophile provides a starting point to explore the active site catalyzing this reaction. Crosslinking data in a group II intron shows that domain 6 which contains the branch adenosine docks into a RNA structure contained in domain I (Hamil and Pyle, 2006). To date, no comparable structure of the spliceosome active site is available, but in vitro crosslinking data shows a situation different from the aforementioned ribozymes. A synthetic pre-mRNA substrate containing the photo-activatable crosslinker benzophenone tethered to the branch adenosine shows several proteins are within 15Å of the branch adenosine, including p14 and p220 (the human homologue of PRP8), which will be discussed in later chapters of this thesis (Figure 2-1A; MacMillan et al., 1994). Furthermore, a substrate containing a short range 2,6-diaminopurine nucleotide crosslinker in the branch position crosslinks strongly to a 14 kDa protein, called p14, showing it is within ~3Å of the branch nucleotide (Figure 2-1B; Query et al., 1996). Both crosslinking experiments show p14 associates with the branch adenosine in A through C

complex, at the same stages that the branch adenosine is bulged from a duplex with U2 snRNA. Furthermore, the presence of the crosslink in C complex shows that p14 is in direct contact with the branch adenosine up to the time splicing occurs, although it was recently shown that SF3b is displaced from the spliceosome just prior to splicing catalysis (Lardelli et al., 2010).

Crosslinked p14 protein was initially isolated from purified mammalian spliceosomes and subsequently identified as a constituent of both the U2 and U12 snRNPs of the major and minor spliceosomes (a variant of the spliceosome

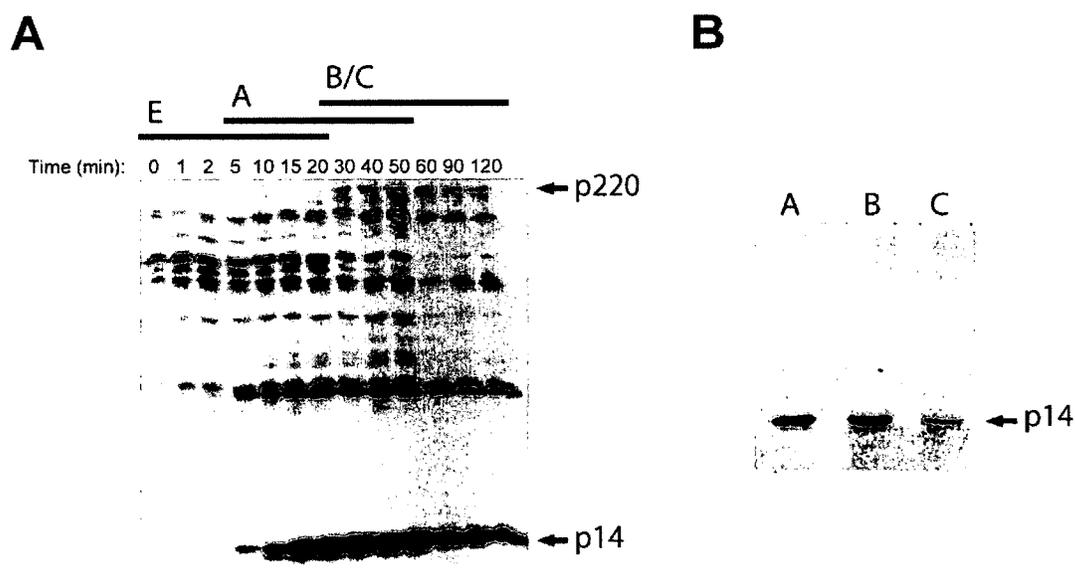


Figure 2-1. Crosslinking of Proteins to the Branch Nucleotide. A). Proteins cross-linked to the benzophenone-modified branch adenosine. Spliceosomal complexes were immunoprecipitated with anti-Sm mAb Y12 and mAb B1C8, which immunoprecipitate spliceosomal complexes during a time course of spliceosomal assembly and subjected to RNase A degradation of pre-mRNA before analysis by SDS-PAGE. The crosslinks corresponding to p14 and p220 (human Prp8) are indicated. The approximate timeframe of spliceosomal complex formation is indicated above the gel by black bars. B) Proteins cross-linked to a pre-mRNA substrate containing a 2,6 diaminopurine branch nucleotide. Spliceosomal A, B, and C complexes were separated on native polyacrylamide gels, isolated, and subsequently treated with RNase and crosslinked proteins are detected by SDS-PAGE. The crosslink corresponding to p14 is indicated. (Adapted from MacMillan et al., 1994 and Query et al., 1996).

responsible for splicing a subset of pre-mRNAs) respectively. More specifically, p14 is the 14 kDa component of SF3b, a salt-dissociable multiprotein complex conserved between the U2 and U12 snRNPs. A strong interaction between SF3b14 (referred to as p14 throughout this thesis) and the SF3b protein SF3b155 has been demonstrated (Will et al. 2001).

Little genetic information is available about p14. No mutations or their associated phenotypes have been identified. Although most species have a p14 ortholog, the model organism *S. cerevisiae*, which is the organism most often used to identify mutations in spliceosomal proteins does not have a p14 homologue, and it is not known if a domain of any other protein replaces p14. Another model organism, *S. pombe* has a p14 gene, but a knockout of this gene has only a mild effect on growth and no known defect in pre-mRNA splicing. p14 seems to be essential for viability in HeLa cells, as attempts to knock-out the p14 gene have been unsuccessful (C. Query, pers. comm. to A. MacMillan, also see chapter 3).

p14 interacts directly with the pre-mRNA branch adenosine within the spliceosome (Query *et al.* 1996), which implies that an understanding of the function of p14 will help us understand the functioning of the spliceosome near the nucleophile for the first step of the splicing reaction. As an initial step towards this understanding, an analysis of the amino acid sequence of human p14 and a comparison amongst the sequences of putative p14 orthologs from different species reveals some general clues (Figure 2-2). A p14 gene is found in nearly all eukaryotic species with sequenced genomes, with the exception of some of the hemiascomycetes. They contain a conserved RNA recognition motif (RRM)

resolution structure of p14 bound to a peptide from another spliceosomal protein (SF3b155). From this structure and some biochemical evidence, we propose a mechanism by which p14 binds the branch point adenosine within the spliceosome.

2-2.1. Structure of the p14•SF3b155 complex

Initial investigations of p14 suggested that in isolation, the protein was partially unfolded as evidenced by poor solubility, aggregation, and ¹H NMR (M.J.S., L. Spyrocolous, and A.M.M., unpublished data). We therefore focused further studies on the complex formed between p14 and SF3b155 (Will et al., 2001). 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 2-3A). The sequence of this portion of SF3b155 is conserved amongst eukaryotes, with the exception of *S. cerevisiae*, which is consistent with the absence of a p14 homologue in this species. 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. The K_d of this interaction has been reported by another group as ~10nM, which indicates it is highly unlikely this interaction is disrupted in the spliceosome at any point (Spadaccini et al. 2006). We crystallized a selenomethionine-substituted complex of p14 bound to SF3b155 aa 373-415 (herein referred to as the SF3b155 peptide) and determined its structure to 2.5 Å resolution by using multiwavelength anomalous diffraction methods (Figure 2-3B, table 2-1).

Table 2-1. Crystallographic data collection and refinement statistics.

	SeMet Derivative ¹		Y22M Mutant ²	Adenine complex ³
Data collection				
Space group	C 2 2 2 ₁		P 4 ₁	C 2 2 2 ₁
Cell dimensions				
<i>a</i> , <i>b</i> , <i>c</i> (Å)	101.87, 115.23, 82.54		71.27, 71.27, 102.28	101.85, 112.7, 82.2
α , β , γ (°)	90, 90, 90		90, 90, 90	90, 90, 90
	<i>Remote</i>	<i>Peak</i>	<i>Native</i>	<i>Native</i>
Wavelength (Å)	1.019859	0.979547	1.115869	1.1158
Resolution (Å)	2.5	3.1	3	2.4
<i>R</i> _{sym} or <i>R</i> _{merge}	0.076 (0.399)	0.090 (0.317)	0.059 (0.460)	0.055 (0.311)
<i>I</i> / σ <i>I</i>	25.5 (2.4)	17.2 (5.0)	23.3 (3.3)	21.8 (2.8)
Completeness (%)	99.5 (95.6)	100	99.8 (100)	95 (84.0)
Redundancy	7.6	4.5	4.2	3.8
Refinement				
Resolution (Å)	76-2.50		71-3.0	74-2.4
No. reflections	16225		10288	17384
<i>R</i> _{work} / <i>R</i> _{free}	0.221/0.278		0.227/0.290	0.214/0.258
No. atoms				
Protein	2560		2416	2545
Ligand/ion	0		0	20
Water	67		14	80
<i>B</i> -factors (Å ²)				
Protein	45.3		76.4	49.2
Ligand/ion	na		na	na
Water	42.5		53.5	46.6
R.m.s deviations				
Bond lengths (Å)	0.014		0.011	0.016
Bond angles (°)	1.56		1.35	1.55
PDB accession number	2F9D		2F9J	3LQV

[1] Peak and remote data were collected from a single crystal. Refinement was performed against the remote data. Highest resolution shell is shown in parenthesis (2.59-2.50 for remote and 3.21-3.10 for peak).

[2] Highest resolution shell is shown in parenthesis (3.11-3.00).

[3] Highest resolution shell is shown in parenthesis (2.40-2.49).

As predicted, p14 contains a central RRM domain spanning residues 20–91 (Figure 2-3C). In addition, the C terminus of p14 contains two additional α -helices: a short 5-residue 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 (α 1, aa

380–396) and a second shorter helix (α_2 , aa 401–407). Interestingly, the C terminus of the SF3b155 fragment contains a short β -strand (which interacts with β -3 of the p14 RRM) connected to α_2 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 ≈ 1 kDa, we speculate that spliceosomal p14 lacks ≈ 10 aa of the predicted full-length protein. Furthermore, these missing residues are poorly conserved among the p14 orthologs (Figure 2-2); therefore, they are probably not relevant to the structure of p14 within the spliceosome.

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 $\approx 1,300$ Å². The interface between the two proteins is extensive and includes a hydrophobic core (Figure 2-4A) surrounded by a set of hydrogen bonds and salt bridges (Figure 2-4B). 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·SF3b155 peptide surface (Figure 2-4C). The surface surrounding this pocket includes four basic residues: R24, R57, R96, and K100 (Figure 2-4C). 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

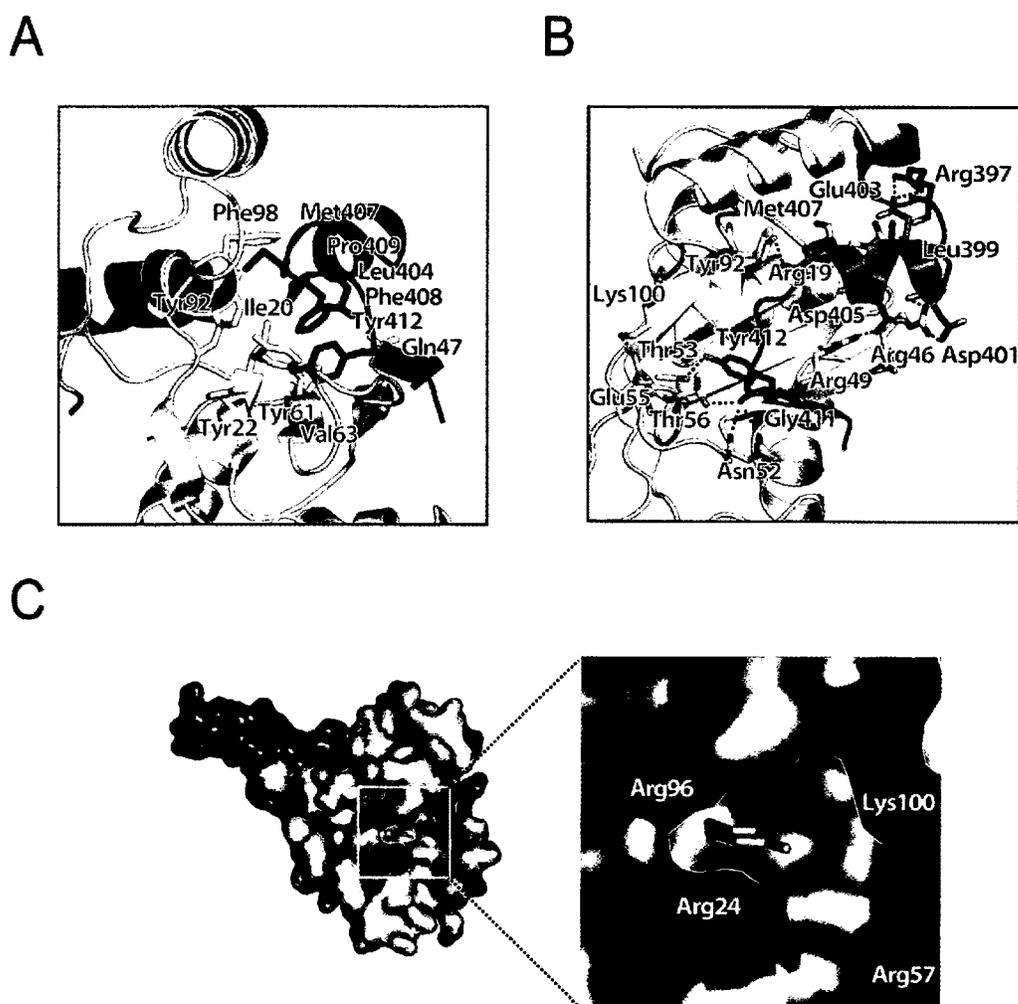


Figure 2-4. Details of p14·SF3b155 peptide interface. A) Hydrophobic core of the p14·SF3b155 interface. p14 is coloured yellow, and SF3b155 peptide is coloured blue. B) Hydrogen-bonding network and salt bridges surrounding the hydrophobic core. p14 is coloured yellow, and SF3b155 peptide is coloured 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 grey, SF3b is shaded dark grey, Y22 is coloured yellow, and R24, R57, R96, and K100 are coloured blue.

R96 and K100 are conserved among p14 orthologs but are not part of a canonical RRM domain.

2-2.2. 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 2-5A). In one orientation, the β -sheet of the RRM is masked (Avis et al., 1996); an $\sim 135^\circ$ 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 intermolecular between p14 and the SF3b155 peptide (Figures 2-3A,B and 2-5A). It is unlikely that a rearrangement of structure similar to that proposed for U1A would expose the p14 RNP motifs for branch duplex recognition. A number of proteins contain atypical RRM's called UHM's (see section 1-2.5) that differ from canonical RRM's 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.

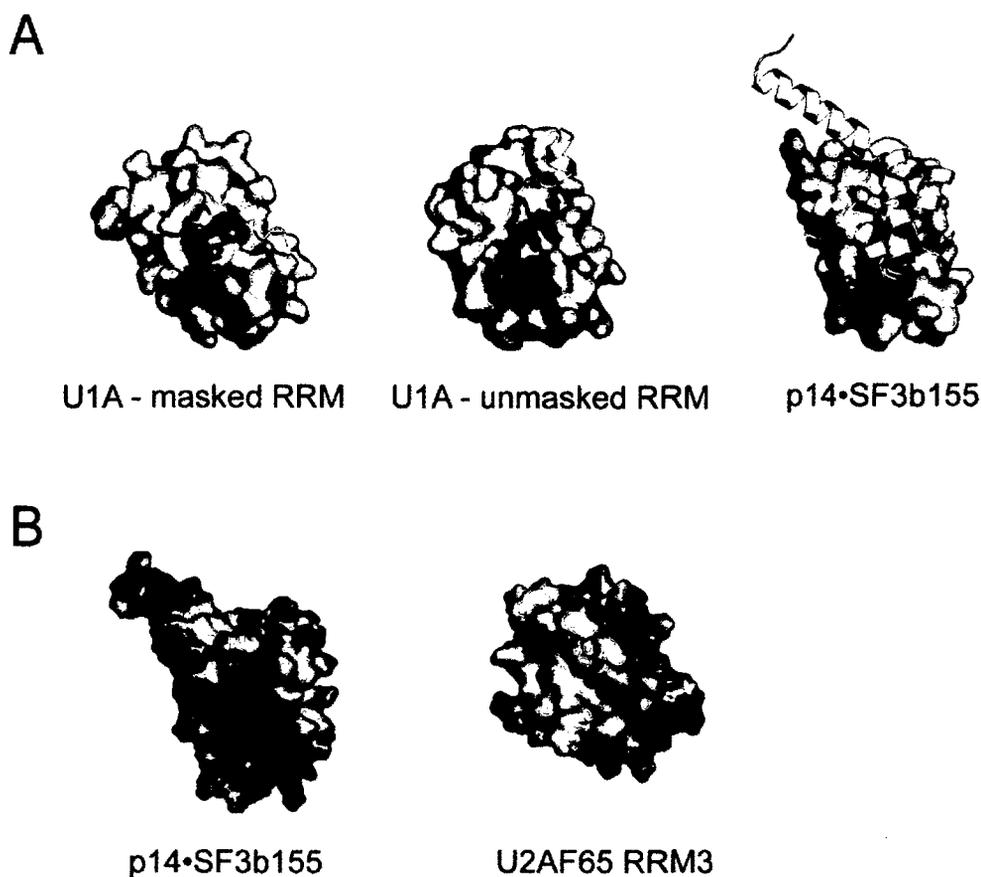


Figure 2-5. Comparison between p14•SF3b155 peptide surface and structures of U1A and U2AF65 RRM3. A) (Left) NMR structure of free U1A with C-terminal helix (in yellow) positioned across RNA-binding surface (Avis et al., 1996). (Middle) X-ray structure of free U1A showing C-terminal helix rotated to unmask the RNA binding surface (Rupert et al., 2003). (Right) Structure of p14•SF3b155 peptide complex showing C-terminal helices of p14 and Sf3b155 peptide (in yellow). RNP1 and RNP2 are coloured red and blue, respectively. B) Representation of surface charges of p14•SF3b155 and U2AF homology motif (RRM3) of U2AF65. The colour scheme is as follows: blue, positively charged; red, negatively charged; white, neutral.

Recent high-resolution structural studies of atypical RRM 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). A comparison of both the p14 RNP sequences

and the structure of the p14 RRM with standard RRMs and UHMs suggests that p14 is best classified as containing a canonical RRM. In addition, the overall positive charge of the p14 RRM in 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 2-5B).

2-2.3. 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 the p14•peptide complex show only weak RNA•protein association ($>100 \mu\text{M}$ 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).

To determine which amino acid residues of p14 interact 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 2-6A and data not shown).

After nuclease P1 treatment, p14 was crosslinked to a single 32 P labelled adenosine. We used a variety of site-specific proteolytic reagents to determine the portion of p14 which crosslinked to the adenosine. Treatment with increasing amounts of cyanogen bromide (CNBr) cleaved the crosslinked protein to a smaller peptide which is present both in the lowest concentrations of CNBr and as a final cleavage product (Figure 2-6A), indicating it represents an N- or C-terminal fragment. Cleavage with endoprotease Asp-N yields a similar size fragment (Figure 2-6A), and further cleavage of the CNBr fragment with Asp-N produces no further cleavage which indicates the radiolabel is contained between amino acids 4 to 35 (the C-terminal CNBr fragment contains an aspartate). Further treatment of the CNBr fragment with endoprotease Lys-C or Glu-C produces a slightly smaller fragment (Figure 2-6B) indicating the crosslink is located

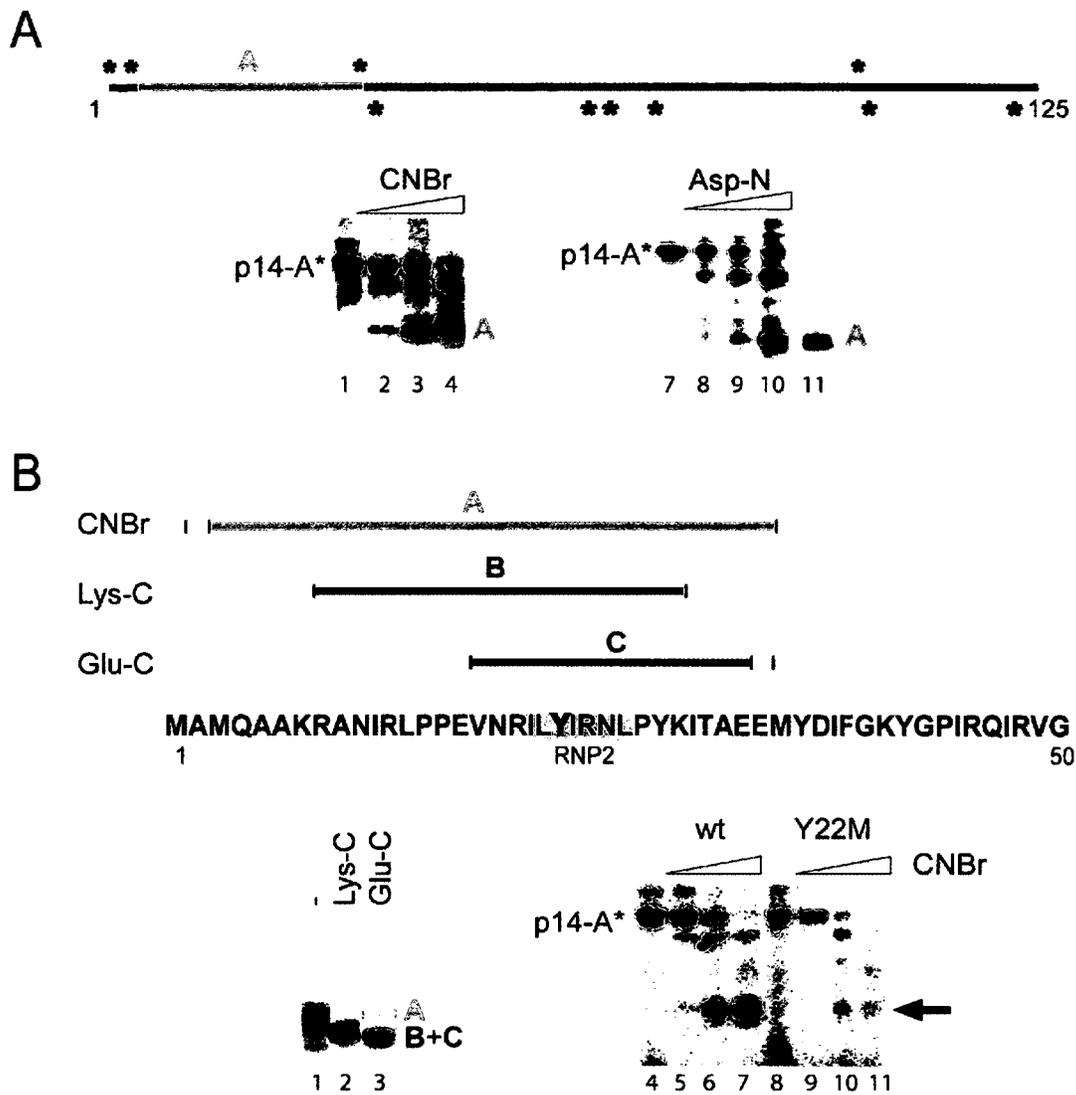


Figure 2-6. Mapping of the p14•branch nucleotide interaction. A) Treatment of p14•SF3b155 peptide complex–RNA crosslink with nuclease P1 yields a single radiolabeled band on SDS/PAGE corresponding to p14 (lane 1). Cleavage with CNBr produces fragment A, which appears immediately (lane 2) and persists as a final digestion product (lanes 3 and 4). Cleavage of p14–RNA crosslink (lane 7) with increasing amounts of endoprotease Asp-N (lanes 8–10) results in a fragment with the same mobility as fragment A (lane 11). B) (lower left) Further digestion of purified CNBr fragment A (lane 1) with endoproteinase Lys-C produces fragment B (lane 2). Alternatively, cleavage of fragment A with endoproteinase Glu-C produces fragment C (lane 3). 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). (lower right) CNBr treatment of p14•RNA (lanes 4–7) and Y22M p14•RNA (lanes 8–11) crosslinks shows disappearance of crosslinked Y22M protein (arrow) upon CNBr treatment.

between amino acids 8 to 29 or 17 to 33 respectively. These mapping results were supported by performing the identical experiments with an N-terminal truncation of p14 (Δ 1-11) in which the mobility of the CNBr and Lys-C, but not the Glu-C fragment is affected (data not shown).

These experiments demonstrate that the branch nucleotide crosslinks to p14 between amino acids 17 and 29 (Figure 2-6B); this region includes the RNP2 consensus motif. Thus, the pre-mRNA branch interacts with a portion of the RRM of p14. We hypothesized that the branch nucleotide may be interacting with the conserved aromatic amino acid Y22, which is accessible at the base of a pocket on the surface of p14. We were able to express and purify a mutant of p14, Y22M, and we examined the interaction of this mutant with RNA. 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 (Table 2-1) 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•SF3b155 peptide complex and the bulged duplex RNA and observed the formation of a crosslink between the branch nucleotide and Y22M p14 (Figure 2-6B). 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 labelled protein upon treatment with the reagent (Figure 2-6B). This result is consistent with the crosslink occurring between the branch nucleotide and the terminal methyl group of the M22 side chain (Clement et al., 2005), which suggests that

the conserved aromatic Y22 of RNP2 interacts directly with the bulged nucleotide. 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 Y22 with an unpaired nucleotide at the branch position.

2-2.4. Adenine recognition by the p14/SF3b155 peptide complex

In parallel with the mapping experiments described above, we attempted co-crystallization and soaks of p14•SF3b155 peptide crystals with a variety of nucleobases, nucleosides, and mononucleotides. Following a crystal soak with adenine alone, we solved the 2.4 Å resolution structure of an adenine•p14/SF3b155 peptide complex (Table 2-1). We used crystals of a cysless mutant of p14 (C74V, C83S) which produced crystals that diffracted to a higher resolution than those of wild type p14. These mutations were chosen based on a phylogenetic comparison, and the structure of this mutant shows these mutations do not perturb the structure of p14 (data not shown).

Electron density corresponding to the adenine is clearly visible in a 2Fo-Fc map within the pocket on the p14 surface as predicted on the basis of crosslinking to the Y22M mutant. The resolution of the adenine•p14/SF3b155 peptide structure allows us to unambiguously model the specific disposition of adenine within the pocket (Figure 2-7A). Consistent with the crosslinking results, the base stacks on the conserved Y22 of RNP2. Specific recognition of the base within the

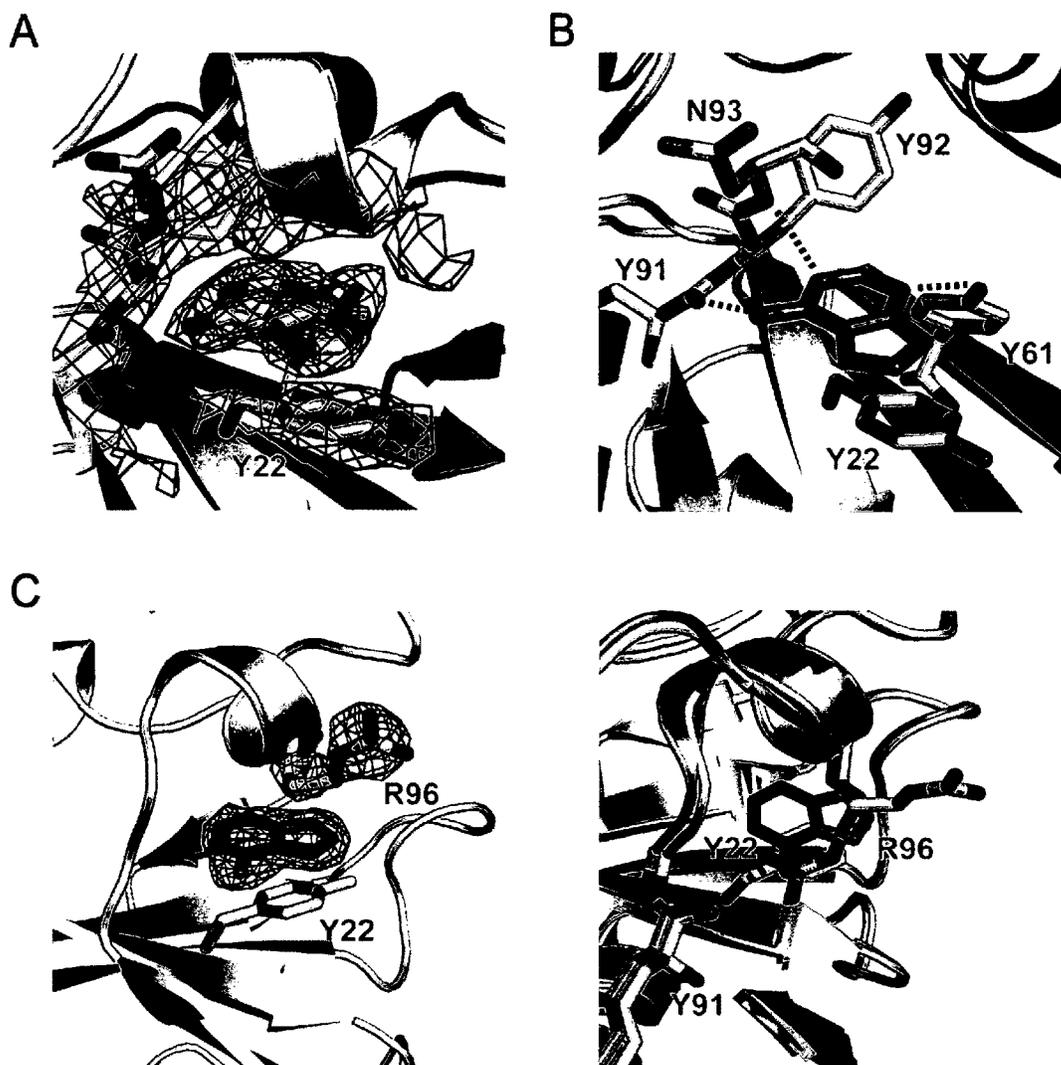


Figure 2-7. Recognition of branch nucleotide by p14. A) 2Fo-Fc map at 2.4 Å resolution using phases calculated from the final, refined model and contoured at 1 σ (blue) and Fo-Fc map using phases calculated from a model lacking adenine contoured at +5 σ (green) shows clearly the location and orientation of bound adenine (green) which stacks on Y22. p14 is coloured yellow and the SF3b155 peptide is coloured grey. B) Interactions between adenine and the SF3b14 pocket. The purine stacks on Y22 of RNP2, features Watson-Crick-like hydrogen-bonding interactions between the N-6 exocyclic amine and main chain carbonyl of Y91 and between N-1 and the main chain N-H of N93, as well as hydrogen-bonding between N-3 and the hydroxyl of otherwise buried Y61 of RNP1. C) R96 in p14/SF3b155/adenine structure. (left) A 2Fo-Fc map contoured at 1 σ shows density corresponding to adenine and R96. (right) In the absence of adenine R96 of p14 (cyan) partially overlaps with the adenine binding site and hydrogen bonds to the carbonyl oxygen of Y91. R96 is displaced by adenine (green) in the adenine•p14•SF3b155 peptide structure (yellow). The SF3b155 peptide is coloured grey for both structures.

pocket is mediated by a Watson-Crick-like interaction with main-chain functionalities (the N-6 exocyclic amine with the carbonyl of Y91 and N-1 with

the amide N-H of N93) as well as hydrogen-bonding of N-3 to the phenolic hydroxyl of the otherwise buried Y61 of RNP1 (Figure 2-7B). Adenine was also observed in the same position of wild type p14•SF3b155 peptide crystals which were soaked in adenine. However, because these crystals diffracted to a lower resolution, assignment of the adenine•p14 interactions was more difficult.

Overall, upon adenine binding the structure of the p14•SF3b155 peptide complex is unperturbed, with the exception of the side chain of R96. In the absence of adenine, this lies along the side of the pocket, forming a hydrogen bond with the main chain carbonyl of Y91. As noted above, adenine binding replaces this with an N6 interaction; the R96 side chain is swung out from the surface of the complex, into a conformation which would be suitable to interact with the phosphate backbone of a bound RNA duplex (Figure 2-7C,D).

2-2.4. 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 which performs the homologous function that U1 and U2 snRNPs perform for the major 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).

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 2-8A). The co-crystal 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 co-crystal structure of the p14•SF3b155 peptide fits well into the outer globular domain of the U11/U12 di-snRNP previously suggested (Golas et

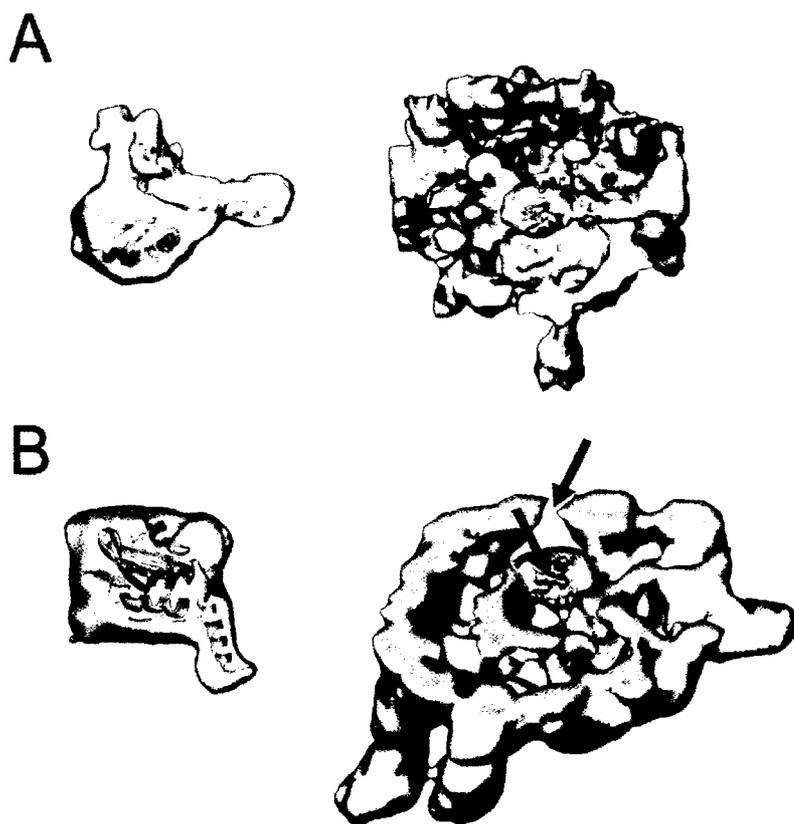


Figure 2-8. p14•SF3b155 peptide structure within SF3b and U11/U12 snRNP. A) Fitting of p14•SF3b155 peptide density into the SF3b cryo-EM structure (Golas et al., 2003). (Left) cryo-EM density corresponding to p14•SF3b155 peptide (yellow), with the p14•SF3b155 x-ray structure represented as a ribbon diagram. (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.

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 2-8B). The C-terminal α -helix of p14 is located on the outside, consistent with the 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 2-8B). The space between the outer wall of the U11/U12 di-snRNP and the p14 RRM is >2 nm wide indicating this cleft is large enough to accommodate even double-stranded RNA. It is possible, however, that pre-mRNA association involves conformational changes around 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.

2-3. Discussion

The results of the crosslinking experiments and the adenine complex x-ray structure reported here suggest that the pre-mRNA branch adenosine interacts directly with the conserved aromatic Y22 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, interacting 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, the branch duplex must interact with p14•SF3b155 in a noncanonical fashion. Although other RRMs (Oubridge et al. 1994) have been shown to undergo a conformational change to “un-block” their RRM upon RNA binding, the extensive interface between the β -sheet of p14 and the C-terminal α -helices of p14 and the SF3b155 peptide makes such a change unlikely for p14. Modeling of the p14•SF3b155 peptide structure into the cryo-EM structures also suggests that no rearrangement occurs upon incorporation of SF3b into U12 (or U2) snRNP, and provides a landmark by which to identify other components within these cryo-EM structures.

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.

Specific recognition of adenine, revealed by the X-ray structure reported here, supports the suggestion that the unpaired adenosine of the pre-mRNA•U2 snRNA bulged duplex is bound within a pocket on the p14 surface. This interaction is specific and stabilized by three hydrogen-bonding interactions. This includes the interaction of the adenine N6 exocyclic amine with the main chain of Y91 that is dependent on the displacement of the R96 side-chain. The hydrogen-bond between Y61 and N3 of adenine may only be representative of a subset of branch nucleotide p14 interactions since phenylalanine frequently occupies this position in p14 orthologs (Figure 2-2). The NMR structure of a single-stranded RNA, representing the human branch sequence, bound to the SF1 KH domain has also been described (Liu et al. 2001). Interestingly, in this structure recognition of the nucleotide corresponding to the branch adenosine is also specified by a Watson-Crick-like interaction, in this case with the backbone amide and carbonyl of I177 (Figure 1-5). Together, the SF1•RNA and p14•adenine structures reveal specific recognition of the branch adenosine at multiple stages of spliceosome assembly from E complex (SF1) through the A, B, and C complexes (p14). The observed specific interactions with adenine provide a molecular explanation for the pronounced preference for that nucleotide at the branch position and may function along with a spliceosomal proof-reading mechanism (Xu and Query 2007) to help ensure fidelity of branch selection.

Further experiments will use the location of the adenine in the p14•SF3b155 structure as a landmark to dock a model of the pre-mRNA•U2 structure that features an extruded adenosine at the branch position (Berglund et al., 2001). The next chapter of this thesis will investigate the complex formed between p14 and such an RNA duplex.

2-4. Methods

2-4.1. Identification and alignment of p14 ortholog sequences

The amino acid sequence of human p14 (NP_057131) was used to perform a BLAST search of the genomes of other eukaryotes. Sequences of p14 orthologs from diverse organisms including other animals (*T. nigroviridis*, *D. melanogaster*, *C. elegans*), plants (*A. thaliana*, *V. vinifera*, *Z. mays*, *P. patens*), ciliates (*P. tetraurelia*), and fungi (*S. pombe*, *Y. lipolytica*, *U. maydis*, and *D. hansenii*) were identified and subsequently aligned using CLUSTALW.

2-4.2. Determination of SF3b155 p14 interaction domain

Maltose binding protein (MBP)-p14 fusion protein (300 pmol) was incubated with 300 pmol of GST-SF3b155 constructs and 10 μ l of amylose-sepharose 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 \times 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-glutathione-S-transferase (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.

2-4.3. Protein expression and purification

DNA encoding full-length human p14 was cloned into the EcoRI and PstI sites of pMAL-C2x (NEB) 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 (Macrogen). The resulting MBP-p14 fusion proteins were expressed in *E. 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 pGEX6P-1 (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 Source 15S HR10/10

column (GE Healthcare) followed by gel filtration with a Superdex-75 26/60 column (GE Healthcare).

2-4.4. 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 μ L of 10 mg/ml protein solution (10 mM Tris, pH 8.0/60 mM KCl/1 mM EDTA/5 mM 2-mercaptoethanol/0.02% NaN_3) with 2 μ L of precipitant (14–18% polyethylene glycol 3350/100 mM MOPS, pH 6.0/200 mM NaHCO_2). Y22M-containing crystals were grown by mixing 2 μ L of protein solution with 2 μ L of precipitant (1.8 M 60%/40% $\text{NaH}_2\text{PO}_4/\text{K}_2\text{HPO}_4$ and 500 mM β -alanine). Crystals of the Cys-less p14•SF3b155 peptide complex were grown by mixing 1 μ L of protein solution with 1 μ L of precipitant (12-14% PEG 3350, 100 mM Tris pH 7, and 200 mM NaCl). For adenine soaks, a 100 mM solution of adenine hemisulfate (Sigma) was heated to dissolve the adenine, quickly cooled to room temperature, and 0.2 μ L was added to a 2 μ L drop containing p14•SF3b155 crystals. Crystals were cryoprotected in reservoir solution containing 15% glycerol, and flash-cooled in liquid nitrogen for data collection.

2-4.5. 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 for the Y22M and adenine complex crystals were collected at a single wavelength. Data were processed and scaled with the HKL package (Otwinowski and Minor, 1997).

2-4.6. 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 REFMAC (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 and p14•SF3b155•adenine complex was solved by molecular replacement by refinement using PDB entry 2F9D with REFMAC.

2-4.7. 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 template, followed by purification on a 15% denaturing PAGE gel. Recombinant p14•SF3b155 373–415 (20 μ M) 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 μ g 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/ μ l, 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 NH₄HCO₃, 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.

2-4.8. 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.

2-5. 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.

Hamill S., Pyle A.M. (2006) The receptor for branch-site docking within a group II intron active site. *Mol. Cell* **23**, 831–840.

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.

Lardelli R.M., Thompson J.X., Yates J.R. 3rd, Stevens S.W. (2010) Release of SF3 from the intron branchpoint activates the first step of pre-mRNA splicing. *RNA*, **16**, 516-28.

Liu Z., Luyten I., Bottomley M.J., Messias A.C., Houngninou-Molango S., Sprangers R., Zanier K., Kramer A., Sattler M. (2001). Structural basis for recognition of the intron branch site RNA by splicing factor 1. *Science* **294**, 1098-1102.

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.

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.

Spadaccini R., Reidt U., Dybkov O., Will C., Frank R., Stier G., Corsini L., Wahl M.C., Lührmann R., Sattler M. (2006). Biochemical and NMR analyses of an SF3b155-p14-U2AF-RNA interaction network involved in branch point definition during pre-mRNA splicing. *RNA* **12**, 410-425.

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.

Xu Y.Z., Query C.C. (2007). Competition between the ATPase Prp5 and branch region-U2 snRNA pairing modulates the fidelity of spliceosome assembly. *Mol. Cell* **28**, 838-849.

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.

Chapter 3⁽¹⁾
RNA binding by the spliceosomal protein p14

¹ Adapted from Schellenberg et al, (2010). *RNA, in press*. This work represents a collaboration between authors. M.S. and E.D. cloned and prepared mutant proteins, M.S. prepared protein crystals, collected X-ray scattering data, and performed X-ray and SAXS analysis. M.S. performed the disulfide tethering and reduction experiments. M.S. and A.M.M. performed the data analysis.

3-1. Introduction

One of the key steps in spliceosome assembly is the recognition of the branch region. In contrast to the yeast branch sequence, 5'-UACU AAC-3' which is very highly conserved, the human branch region sequence is variable (Burge et al., 1999; Gao et al., 2008); branch sequence plasticity in humans is important in splicing regulation and mutations are associated with a variety of disease states resulting from aberrant splicing (reviewed in Hartmann, 2009). In humans, initial recognition of the single-stranded branch sequence within the pre-mRNA is mediated by the branch-binding protein SF1 which is part of a network of factors that define the E or commitment complex (Reed, 1990). The stable association of U2 snRNP with the pre-mRNA in A complex includes the formation of a bulged duplex structure between the pre-mRNA branch sequence and U2 snRNA that specifies the bulged nucleotide as the nucleophile for the first transesterification step of splicing (Query et al., 1994).

Sequential association of proteins with the pre-mRNA branch region has been demonstrated by site-specific modification of the branch adenosine with photocrosslinkers followed by analysis by SDS-PAGE (Figure 2-1; MacMillan et al. 1994; Query et al., 1996)). These studies detect the early association of SF1 with the branch region, followed subsequently by a short-range crosslink to p14 which appears in A complex and persists within the fully assembled spliceosome. The kinetics of the appearance of this crosslink suggest that p14 is associated with the bulged branch region•U2 snRNA duplex (Figure 3-1A; Query et al., 1994).

An intriguing feature of the p14•SF3b155 (amino acids 373-415 of SF3b155, herein referred to as the SF3b155 peptide) complex, as revealed by the X-ray structure, is that the canonical β -sheet of the RRM is occluded by one of the C-terminal helices of p14, the central helix of the SF3b155 peptide, and the loop connecting the peptide to the C-terminal β -strand resulting in a large buried surface. The interface between the two proteins is extensive including a hydrophobic core surrounded by a set of hydrogen bonds and salt bridges.

The blocking of one face of p14 is significant because the β -sheet of the RRM represents the canonical RNA binding surface of the domain including the highly conserved RNP1 and RNP2 motifs (Oubridge et al. 1994; Deo et al. 1999; Handa et al. 1999; Allain et al. 2000; Wang and Hall 2001; Clery et al. 2008). In the p14•SF3b155 peptide complex, residues of RNP1 and RNP2 are largely buried with the exception of a portion of RNP2, which 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 3-1B). In the previous chapter, we were able to characterize a crosslink between the bulged adenosine of a model branch region duplex and the mutant Y22M p14•SF3b155 peptide complex where the introduced methionine forms the floor of the pocket on the p14 surface. Mapping the RNA crosslink to M22 in the mutant allowed us to infer that the bulged nucleotide is bound within the p14 pocket and that the surrounding surface represents the RNA-binding face of p14. This result was supported by the observed adenine in the crystal structure,

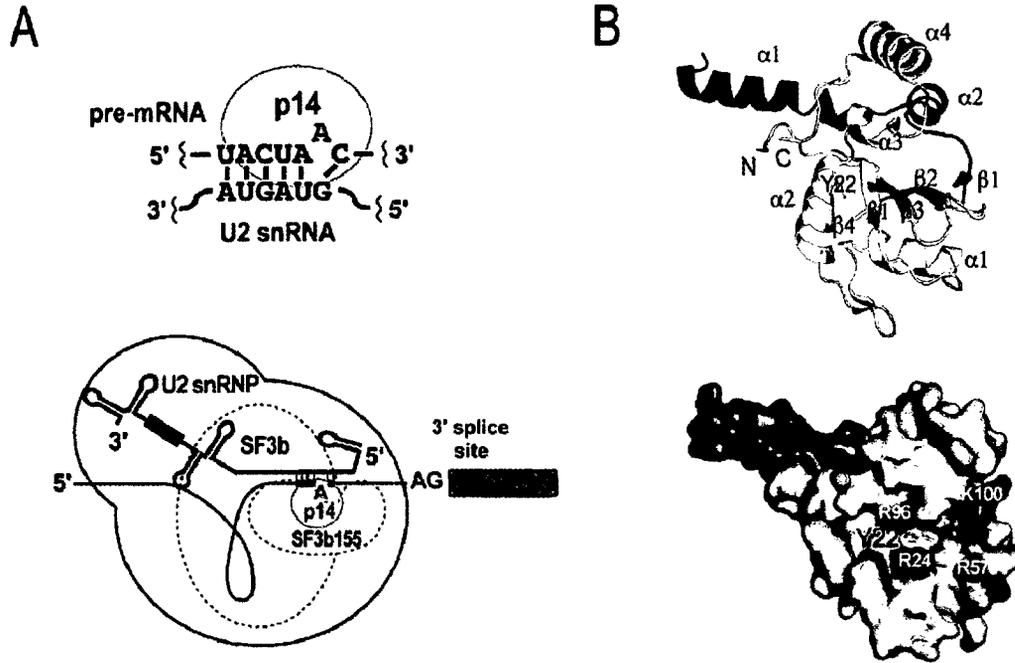


Figure 3-1. Protein-protein and protein-RNA interactions of p14 in the human spliceosome. (A) (upper) p14 directly contacts the bulged duplex formed between the pre-mRNA branch sequence and U2 snRNA; (lower) p14 as a constituent of the U2 snRNP-associated SF3b particle. (B) (upper) Ribbon diagram showing the p14•SF3b155 peptide complex based on the X-ray structure (Schellenberg et al. 2006). p14 and the SF3b155 peptide are yellow and blue respectively; Y22 is shown in green. (lower) Surface representation of the p14•SF3b155 peptide complex showing Y22 exposed within the surface pocket surrounded by conserved basic residues; p14 (light grey), SF3b155 peptide (dark grey), Y22 (yellow), and R24, R57, R96, and K100 (blue).

which forms hydrogen bonds between the Watson-Crick face of adenine and p14, with a third hydrogen bond that may be less conserved amongst p14 orthologues (see chapter 2 for discussion).

In order to further characterize the interaction of p14 with the pre-mRNA branch duplex we have used a disulfide tethering approach to compare the stabilities of different RNA•protein complexes. The results of these studies, combined with solution-scattering studies and *in vivo* knockdown of p14 allow us to propose a model of branch duplex•p14 interaction with significant implications with respect to the role of p14 in the assembled spliceosome.

3-2. Results and Discussion

3-2.1 Interrogating the p14•RNA complex using a disulfide tethering strategy

The affinity of the isolated p14•SF3b155 peptide complex for short model RNA duplexes is not strong ($>100 \mu\text{M}$); neither is there a marked specificity for a specific bulged duplex over double-stranded or single-stranded RNA (Spadaccini et al., 2006; Schellenberg and MacMillan, unpublished). These observations are not surprising and may reflect several different features of branch duplex recognition within the spliceosome. These include the cooperative nature of RNA•protein interactions — 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) — as well as the fact that p14 likely interacts with multiple bulged duplex structures due to the variable nature of the branch sequence in humans (Gao et al., 2008).

To address problems of affinity/specificity and to facilitate biochemical and structural characterization of p14•RNA interaction, we chose to use a disulfide tethering approach to stabilize a complex between thiol-derivatized RNA and p14 containing a single exposed Cys residue (Figure 3-2). This approach, first developed by Verdine and coworkers, has been very successfully used in the characterization of protein•DNA complexes (Lee et al., 2008; Komazin-Meredith et al., 2008; Huang et al., 1998; Huang et al., 2000; He and Verdine, 2002; Johnson et al., 2006; Fromme et al., 2004; Bannerjee and Verdine, 2006; Zhao et al., 2008; Corn and Berger, 2007) and has been adopted by others to accelerate

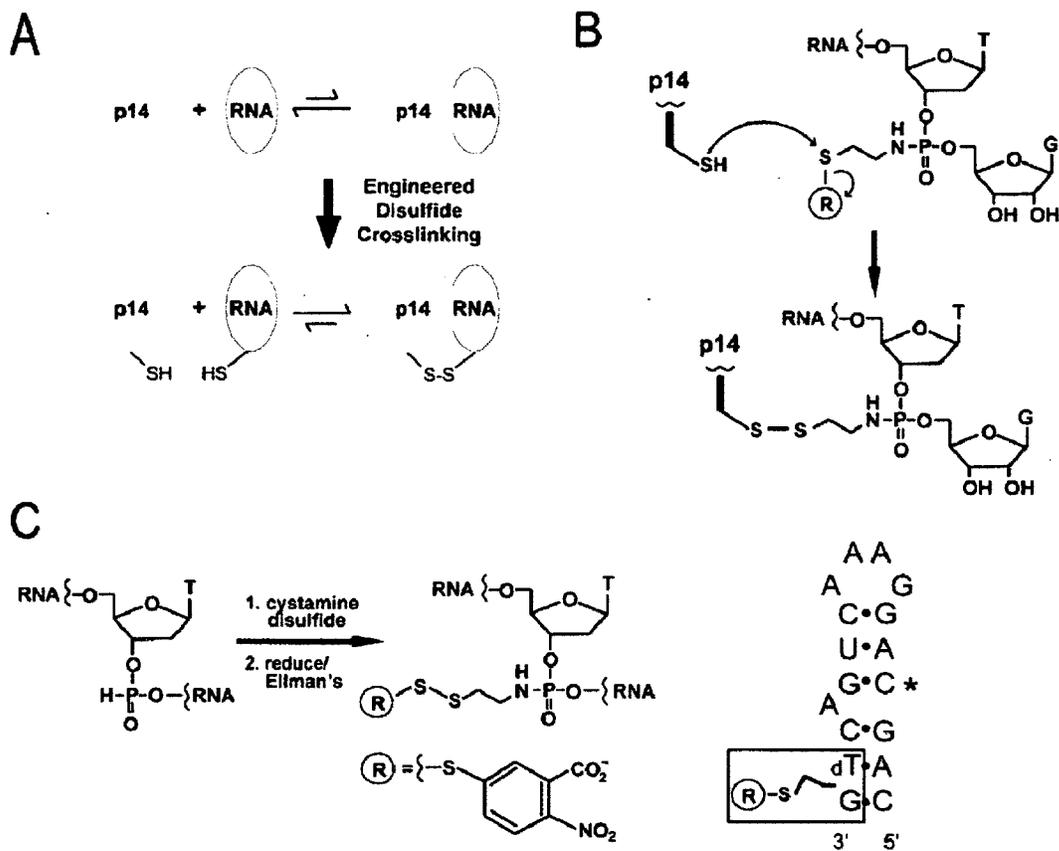


Figure 3-2. Disulfide Crosslinking Strategy for interrogating the p14•SF3b155 peptide•RNA Complex. (A) Crosslinking of engineered Cys containing p14•SF3b155 peptide to thiol-derivatized RNA generates a tethered RNA-protein complex in which cognate association is predicted to stabilize the disulfide bond to reduction. (B) p14•RNA crosslinking reaction. Chemistry of disulfide bond formation between single Cys containing p14•SF3b155 peptide and an N-thioalkyl modified RNA where attachment to the backbone is mediated by single phosphoramidate substitution for a backbone phosphate. (C) Synthesis of thiol-modified RNA. (left) Oxidation of an H-phosphonate monomer with cystamine disulfide is followed by reduction and trapping with DTNB during the course of chemical synthesis. (right) sequence of thiol-modified RNA synthesized as a mimic of the pre-mRNA•U2 snRNA bulged duplex containing C in place of pseudouridine at position 4 (*).

screening of drug-target interactions (Cancilla et al., 2008; Erlanson et al., 2000; Erlanson et al., 2003a; Erlanson et al., 2003b). Sequence-specific and non-specific complexes of the *E. coli* Ada protein with DNA have been trapped via intramolecular disulfides (He and Verdine, 2002). X-ray structures of the DNA repair protein MutY bound to a DNA lesion (Fromme et al., 2004) and a complex

of HIV reverse transcriptase with a DNA template primer were also based on a disulfide trapping strategy (Huang et al., 1998).

The positioning of the adenine base within the pocket on p14 allows us to crudely model the interaction of the protein complex with a cognate RNA based on X-ray and NMR structures of bulged duplex RNAs (Berglund et al., 2001; Lin and Kielkopf, 2008; Newby and Greenbaum, 2002). Given the positioning of adenine within the pocket, and minimizing steric clashes with the p14•SF3b155 peptide surface, two possible orientations of a bulged-adenosine RNA duplex, related by a 180° rotation, may be modeled (1 and 2 in Fig. 3-3A). We created a Cys-less version of p14 (C83S; C74V) based on phylogenetic comparisons (Figure 2-1) and determined the structure of the mutant p14•SF3b155 peptide complex to show that it is not perturbed by these changes (r.m.s.d. 0.2 Å with wildtype p14). From Cys-less p14, we then created a panel of single Cys containing p14 mutants where the introduced residues were designed, based on our X-ray structure and crude model, to sample the phosphodiester backbone with respect to both possible RNA orientations on the proposed RNA binding surface (Fig. 3-3A). The SF3b155 peptide used in these experiments does not contain any cysteine residues. One p14 mutant, D37C is located on the face of the p14•SF3b155 peptide complex opposite to the proposed RNA binding surface, to serve as a negative control. This combination of modified proteins, especially the D37C mutant, is important with respect to addressing any concern that the tethering strategy might force a non-biological interaction; screening and comparison of a panel of disulfide linked complexes serves as a control for the

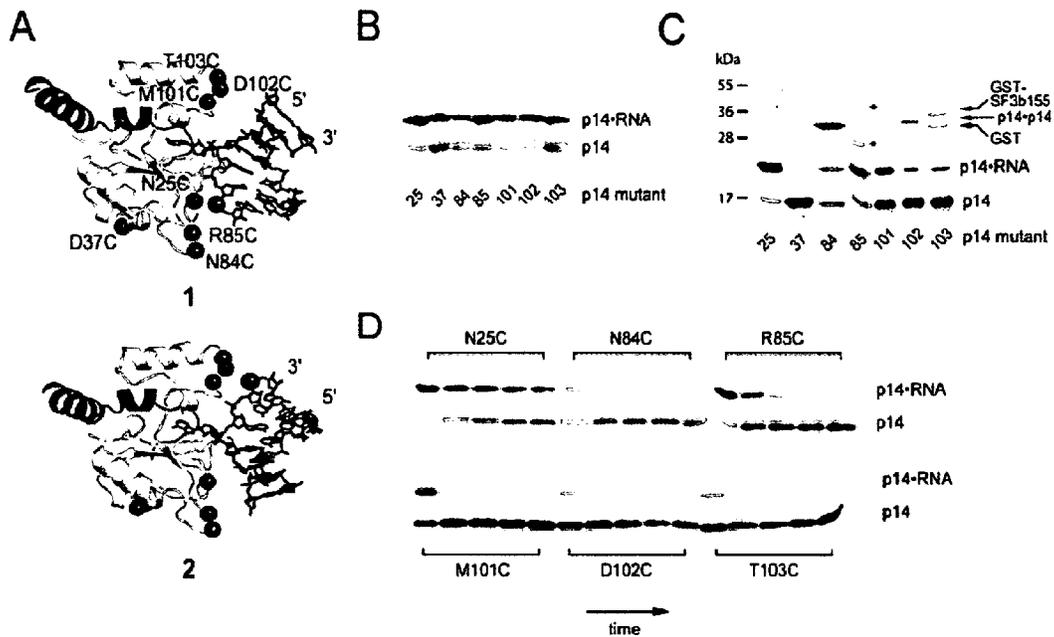


Figure 3-3. Formation and Reduction of p14•SF3b155 peptide•RNA disulfides. (A) Two models of p14•SF3b155 peptide•RNA interaction (1 and 2), related by a 180° rotation, based on crosslinking of a bulged nucleotide to the pocket on the p14 surface (Schellenberg et al. 2006). p14 is shown in yellow, the SF3b155 peptide in blue, and RNA in green. Shown are sites of single-Cys mutations introduced into Cys-less p14 (red balls) and site of RNA backbone modification with the C2 thiol functionality (green ball). Modeling based on the p14•SF3b155 peptide X-ray structure (PDB: 2F9D) and a bulged RNA X-ray structure (Berglund et al. 2001; (PDB: 1I9X)). (B) Coomassie-stained SDS-PAGE gel showing initial formation of p14•RNA disulfide crosslinks. (C) Coomassie-stained SDS-PAGE gel showing p14•RNA disulfide crosslinks under mild equilibrating conditions with panel of single Cys mutants (p14•p14 disulfide is formed upon dialysis of p14 samples into non-reducing buffer before RNA addition; * indicates uncharacterized contaminant). (D) Coomassie-stained SDS-PAGE gel showing kinetics of p14•RNA reduction (0, 10, 30, 60, 120 min).

validity of the approach (see below).

We chemically synthesized model bulged RNA hairpins, containing a single, commercially available 2'-deoxy-H-phosphonate precursor (Froehler 1986), to mimic the pre-mRNA•U2 snRNA duplex (Fig. 3-2C). These were derivitized with a thiol by oxidation of the H-phosphonate with cystamine disulfide during the synthesis, which we then reduced and protected as a mixed disulfide with 5,5'-Dithio-Bis(2-Nitrobenzoic Acid) (DTNB). The position of RNA modification was based on an initial model of RNA•protein interaction

where placement of the bulged nucleotide within the p14 pocket sets the register of duplex association with that face of the protein. The length of RNA was similarly based on the initial crude model. A highly conserved pseudouridine in U2 snRNA base-pairs with the nucleotide 5' to the branch position and has been proposed to stabilize an extrahelical disposition of the branch adenosine (Newby and Greenbaum 2002). Therefore, two RNAs were initially synthesized, one containing a single pseudouridine to match the position of this modified base within the U2 snRNA and another containing cytidine at the same position (Fig. 3-2C). Disulfide bond formations between these RNAs and p14 cysteine mutants were essentially the same suggesting that the replacement of pseudouridine with cytidine is a valid model of the bulged duplex (data not shown).

Mutant p14•SF3b155 peptide complexes were dialyzed overnight into a low concentration of reductant (0.1mM β -mercaptoethanol). These were mixed with an equal molar amount of DTNB-protected RNA (Figure 3-2), and incubated in a final concentration of 60 μ M β -mercaptoethanol. Initial formation of disulfide linkages (which is thermodynamically driven by reduction of the mixed RNA-DTNB disulfide) was assayed by quenching an aliquot of these reactions after twenty minutes with excess iodoacetamide and analysis by non-reducing SDS PAGE (Fig. 3-3B). The location of the protein-RNA band on the SDS-PAGE was confirmed by RNase treatment. This produced a roughly equivalent yield of almost all of the RNA-protein disulfides showing that no single Cys mutant formed the initial disulfide with significantly greater or less (due to steric constraints for example) efficiency.

Allowing Cys mutant protein-RNA disulfides to equilibrate under mild reducing conditions (60 μ M β -mercaptoethanol) overnight reduced the RNA-protein disulfides to a variable extent as assayed by SDS-PAGE (Figure 3-3C). The highest yields of complex were those involving formation of RNA disulfides with the N25C and R85C proteins suggesting that these are the most thermodynamically stable linkages. A roughly 50% yield of the M101C and N84C disulfides were obtained; the yields of both the D102C and T103C complexes were much lower. Significantly, almost no D37C disulfide was observed. In several cases, intermolecular p14 disulfide formation, upon dialysis into non-reducing buffer and before RNA addition, was observed; this was significant in the case of N84C likely reflecting the location of the introduced cysteine on a flexible loop of p14. The relative stabilities of the protein-RNA stabilities are very informative immediately suggesting that the bulged duplex interacts with one face of the p14•SF3b155 peptide complex and furthermore indicating the preferred binding orientation of the bulged duplex on this face. The lack of D37C disulfide under mild reducing conditions is a control for non-specific protein-RNA disulfide formation on the rear face of the complex and the varying yields of the other disulfides support one of the two proposed orientations on the front face (1 in Figure 3-3A). The thermodynamic stability of the N25C and R85C disulfide with respect to all of the others suggests that these disulfides most represent a cognate RNA-protein association. The relatively high amount of M101C disulfide formed under mild equilibrating conditions (Figure 3-3C) shows that an ensemble of p14•RNA complexes (specific and non-specific) is likely

present under the conditions of the experiment; however this disulfide and those to D102C and T103C are less stable to reduction suggesting they are unstable/strained and do not represent cognate complexes (see below).

We were also able to show that the N25C protein-RNA disulfide is the most stable to reduction with respect to the other disulfides by treatment with a high concentration of β -mercaptoethanol followed by SDS-PAGE analysis (Figure 3-3D). At 5 mM β -mercaptoethanol, all the disulfides except N25C and R85C are reduced within 10 minutes, with N25C being exceptionally stable to this concentration of β -mercaptoethanol. Furthermore, this experiment distinguished the N25C mutant as being the most stable, suggesting that this disulfide most represents the interactions within the cognate p14-RNA interaction.

The disulfide tethering approach described here confirms a model of RNA interaction with one face of p14 and allows us to distinguish between two proposed orientations of RNA on the p14 surface. Enhanced stability of disulfides to reduction has been clearly shown to reflect the specificity of protein•ligand interaction in tethered complexes. Relative disulfide stability is a measure of relative binding affinity (Gilbert 1995; Stanojevic and Verdine 1995; O'Shea et al. 1998) and reflects the fact the disulfide bond formation/reformation in specific complexes is essentially an intramolecular process that is entropically favoured. The high efficiency of disulfide linked N25C and R85C formation under equilibrating conditions suggests the specific orientation of RNA on the p14 surface (1 in Figure 3-3A). Furthermore, the kinetic stability of the N25C complex to reduction suggests that tethering at this position was optimal in terms

of representing a cognate complex. The different efficiencies of disulfide formation strongly support this model and serve as a control for the approach.

3-2.2. SAXS Analysis of tethered p14•SF3b155 peptide•RNA complex

The results of the tethering experiments supported by the X-ray structure of p14 containing adenine (Figure 2-7) suggest that the N25C complex represents a cognate model of a bulged RNA duplex•p14•SF3b155 interaction. We therefore decided to characterize this specific complex further using small angle X-ray scattering (SAXS). This method has several advantages over others in the structural characterization of macromolecular structures. As a solution technique, it does not require crystallization, relatively small absolute amounts of sample are required, and finally, a significant amount of structural information is contained within the X-ray scattering data.

We performed SAXS on samples of both the N25C p14•SF3b155 peptide complex and purified N25C p14•SF3b155 peptide•RNA disulfide (Table 3-1). On scale-up, in the final purification of the N25C p14•SF3b155•RNA complex by anion exchange chromatography, we were able to separate two species present in a roughly 20:1 ratio. The minor species showed a scattering curve that was relatively featureless and did not flatten at low angles, consistent with a disordered or unfolded structure (data not shown). We interpret this minor crosslinked species, which we did not characterize further, to represent a strained disulfide formed with one of the two possible diastereomers at phosphorous that result from the oxidation of the H-phosphonate precursor during the chemical

synthesis. We examined both the p14•SF3b155 peptide and p14•SF3b155 peptide•RNA samples by dynamic light scattering and detected no evidence of higher order multimers or aggregates (data not shown).

Table 3-1. Parameters and quality indicators derived from scattering data.

	p14/SF3b155	p14/SF3b155•RNA
R _g (Guinier)	19.7 ± 0.02	22.7 ± 0.09
R _g (P(r))	19.73 ± 0.04	22.84 ± 0.03
D _{max} (P(r))	68	74
NSD	1.006	1.117
Chi ² (<i>ab initio</i> model)	1.174	1.068
Chi ² (PDB 2F9D) ¹	2.7	
Chi ² (PDB 2F9D and RNA) ¹		2.0

¹ PDB model modified using CHADD (Petoukhov et al. 2002) to compensate for residues disordered in the structure.

With respect to the N25C p14•SF3b155 peptide complex and the major purified N25C p14•SF3b155 peptide•RNA disulfide, the SAXS scattering plot suggests the presence of ordered, folded structures (Figure 3-4A). The flattening of the raw scattering plots at low angles shows that there is minimal aggregation in these samples. The bell-shaped distribution function, P(r), for both is representative of a folded, essentially globular structure (Figure 3-4B); the asymmetry represented by a shoulder at ~50 Å is consistent with the protrusion of the long helix, α 1, of the SF3b155 peptide observed in the X-ray structure (Figure 3-1B). Linear Guinier plots (Figure 3-4C) indicate the absence of aggregation in both samples. The bell-shaped structure of the Kratky plot (Figure 3-4C; inset) is parabolic in shape, which is also consistent with an essentially globular structure.

We next used the *ab initio* modeling program GASBOR (Svergun et al., 2001) in real space mode to generate models refined against the P(r) function. Ten rounds of model-building from different and random starting positions for dummy

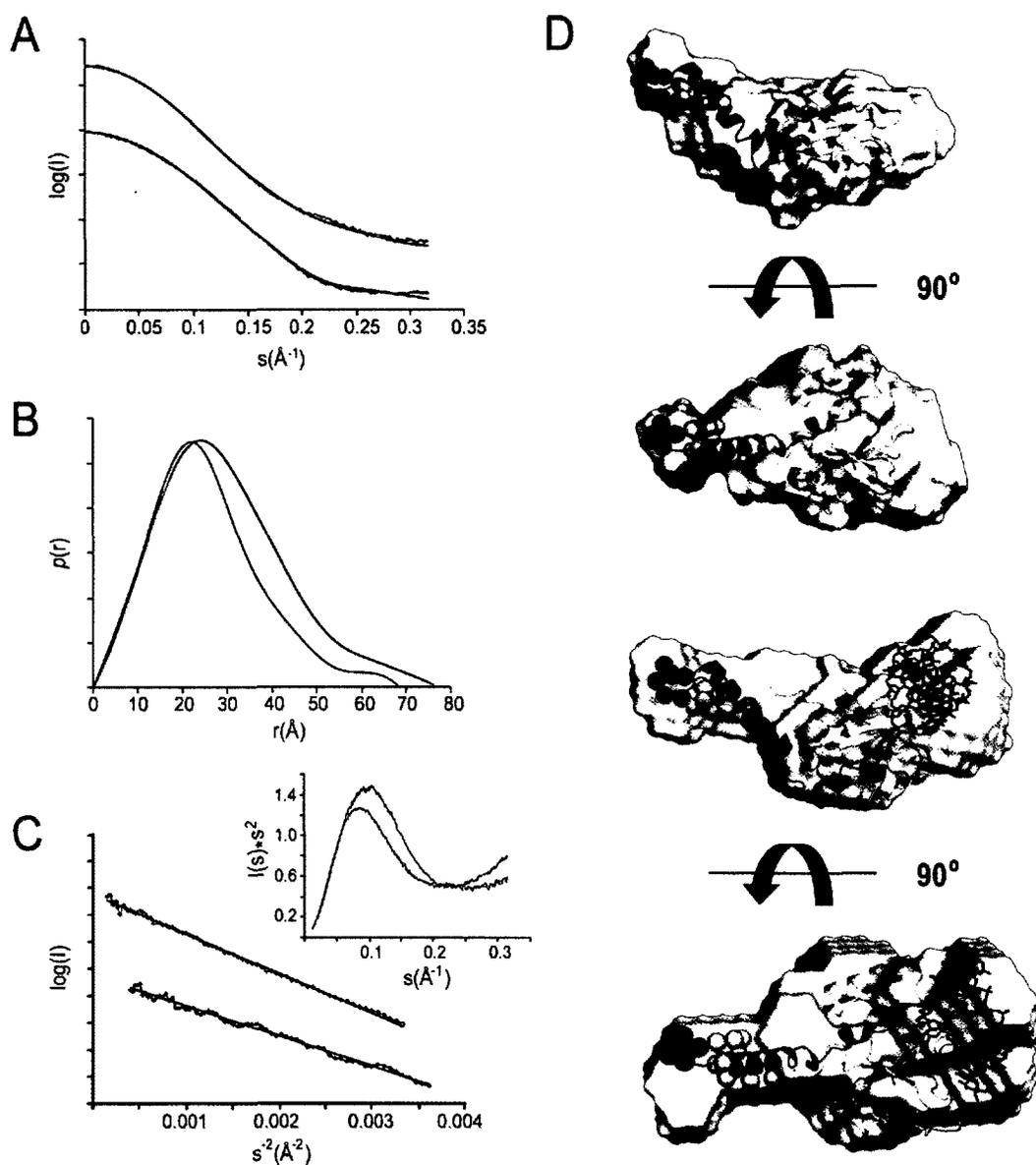


Figure 3-4. Small angle X-ray scattering (SAXS) analysis of p14•SF3b155 peptide and tethered p14•SF3b155 peptide•RNA complexes. (A) Experimental SAXS curves for p14•SF3b155 peptide (blue) and p14•SF3b155 peptide•RNA (red); also shown in black are scattering curves calculated from the CHADD (p14•SF3b155 peptide) or SASREF models (p14•SF3b155 peptide•RNA). (B) Distance distribution functions, $p(r)$, for p14•SF3b155 peptide (blue) and p14•SF3b155 peptide•RNA (red) complexes calculated from the experimental scattering data using GNOM. (C) Guinier plots of p14•SF3b155 peptide (blue) and p14•SF3b155 peptide•RNA data (red); inset: Kratky plots. (D) (upper) SAXS envelope for the p14•SF3b155 peptide complex derived from DAMMIN refined *ab initio* model superimposed on the model from CHADD. (lower) SAXS envelope for the p14•SF3b155 peptide•RNA complex derived from DAMMIN refined *ab initio* model superimposed on the model from SASREF. Dummy residues added for p14 and SF3b155 using CHADD are shown in yellow and blue respectively.

atoms were averaged using the program DAMAVER (Volkova and Svergun, 2003). The unfiltered model from DAMAVER was further refined by using it as a starting model for the program DAMMIN (Svergun, 1999) to yield the final calculated envelope. The *ab initio* envelope for the p14•SF3b155 peptide complex is bipartite featuring a globular density and long extension which corresponds to the shape expected for the p14/SF3b peptide crystal structure. In comparing the *ab initio* model to the p14/SF3b peptide X-ray structure, we first used the program CHADD (Petoukhov et al., 2002) to add dummy residues to that structure representing the poorly ordered/disordered N-terminal residues (eleven and seven residues of p14 and the SF3b155 peptide respectively). The radius of gyration, R_g , derived from this model, 19.3Å, is close to that derived from the experimental scattering data (19.7 Å); the theoretical scattering curve calculated from this model using CRY SOL (Svergun et al., 1995) fit the experimental with a χ^2 of 2.7 (Figure 3-4A). We compared the fit of the p14•SF3b155 peptide model created by the program CHADD to the *ab initio* DAMMIN model using the program SUPCOMB (Kozin and Svergun, 2001; Figure 3-4D; upper). The globular portion of the crystallographic X-ray density fits well and was positioned by virtue of the close correspondence of the helical extension of SF3b155 $\alpha 1$ with the protrusion from the calculated envelope which leaves room for the five amino acids N-terminal to $\alpha 1$ that are disordered in the crystal structure.

We next generated an *ab initio* envelope for the tethered N25C p14•SF3b155•RNA complex using the protocol described above for the p14•SF3b155 peptide complex alone. The resulting envelope is similar to that

derived for the protein complex alone featuring, again, a globular region and a narrow protusion which we assign to $\alpha 1$ of the SF3b155 peptide. Fixing the p14•SF3b155 peptide crystal structure within the envelope using this constraint reveals significant additional density in the *ab initio* envelope corresponding to the proposed RNA-binding face of the protein complex (Figure 3-4D; lower).

We used the program SASREF (Petoukhov and Svergun 2005). to generate a model of the p14•SF3b155 peptide•RNA complex. The protein model was derived from the crystal structure with the missing residues added using CHADD, and the RNA model was a six base-pair bulged duplex based on the larger X-ray structure of Schultz and colleagues capped with a canonical GAAA tetraloop. Ten rounds of SASREF docking using restraints corresponding to the N25C disulfide and bulged-adenosine binding pocket were performed. The models were quite similar, and the one with the best fit to the experimental data (Chi^2 of 2.0) is shown in Figure 3-4A. Using SUPCOMB, we were able to fit this model within the *ab initio* envelope which once again positions the helical extension of the SF3b155 peptide within the protrusion from the globular main body of the calculated mass.

The envelopes generated from the SAXS data for both the p14•SF3b155 peptide and the major RNA•p14•SF3b155 peptide complex correspond with the density expected on the basis of the p14•SF3b155 peptide X-ray structure, in the latter case with the positioning of bound RNA on one face of the protein complex. The p14•SF3b155 peptide•RNA model formed on the basis of this data is of low resolution but is fully consistent with our previous photo-crosslinking results

(Schellenberg et al. 2006), the disulfide crosslinking reduction studies and the pocket-bound adenine X-ray structure.

Chemical shift mapping has been used to examine the interaction of the p14•SF3b155 peptide complex with single-stranded and bulged duplex RNAs representing U2 snRNA and the pre-mRNA•U2 snRNA duplex respectively (Spadaccini et al. 2006; Kuwasako et al. 2008). Overall, the NMR results are consistent with the essentials of the structural model proposed here, specifically with respect to the identification of the face of p14 involved in RNA binding and the occlusion of the canonical RNA binding surface of the RRM. They suggest limited interaction with the RNP motifs (with the exception of Y22 and K24), consistent with their being largely buried and inaccessible; the principal chemical shift perturbations suggest RNA binding involves areas flanking the central β -sheet including the loop between β 2 and β 4 and the C-terminal portion of p14 proximal to β 3. Although it is possible that association of p14 with single-stranded RNA occurs prior to the recruitment of U2 snRNP to the branch region, the p14•SF3b155•RNA interaction that occurs in the A and higher order complexes must involve recognition of a bulged duplex structure (Query et al. 1994).

3-2.3. Implications for role of p14 and SF3b in spliceosome assembly and activation

The model of p14•duplex interaction described here suggests that an intimate complex is formed between p14 and the pre-mRNA•U2 snRNA duplex.

One face of the RNA and more specifically the bulged branch nucleotide is bound on the protein surface consistent with previous crosslinking results showing direct contact between the branch residue and p14 in the A through B and C complexes during spliceosome assembly.

In the model described here, the branch nucleotide, and more specifically the 2' hydroxyl which is the nucleophile for the first step of splicing, is buried at the interface between the bulged duplex and p14 protein complex (Figure 3-3A); this is true even in the mature spliceosome, as represented by the C complex, since photo-crosslinking shows a direct interaction between p14 and the bulged branch adenosine (Query et al. 1996). Thus although the p14•RNA complex persists into the mature spliceosome, it must be disrupted to allow the first step of splicing to occur. Consistent with this, it has been demonstrated that the entire SF3b particle is displaced from the spliceosome just prior to catalysis (Bessonov et al., 2008; Lardelli et al., 2010). It has been proposed that the SF3 association with the pre-mRNA prevents premature nucleophilic attack by the branch adenosine until the correct orientation of the proper substrate within the spliceosome (Lardelli et al., 2010). The structural model reported here, and the resulting requirement for disruption of the bulged duplex p14 interaction, are consistent with this proposal. Interestingly, in *S. cerevisiae*, which lacks a p14 homolog, the branch nucleophile can attack U2 snRNA in the presence of a weak 5' splice site where presumably the positioning of the proper substrate has been decoupled from spliceosome activation (Smith et al., 2007). Furthermore, bulged RNA duplexes have also been shown to be susceptible to hydrolysis 3' to the

bulged residue by virtue of the backbone geometry (Portmann et al., 1996). Thus the p14•branch duplex association may serve as a regulatory step preventing aberrant chemistry during the assembly of the correct spliceosome active site.

In addition to preventing aberrant chemistry, the high degree of conservation of p14 implies it may have a function more essential for the spliceosome. The branch point sequence in most eukaryotes in general is quite degenerate, and therefore the BPS•U2 snRNA duplex in spliceosomes would involve several mismatches (Irimia and Roy, 2008). An examination of the genomes of sequenced hemiacomycetous yeast shows that in general, the BPS of their introns are more highly conserved with respect to U2 snRNA base-pairing (Bon et al., 2003). Furthermore, several of these yeasts (which include *S. cerevisiae*) lack p14 homologues, as determined by BLAST searches. The conservation of p14 in species which have more degenerate branch point sequences suggests that p14 may function to stabilize weak base-pairing of a BPS•U2 snRNA duplex which contains mismatches. The structural model described in the previous section shows that p14 could serve such a function, as it could stabilize the protruded conformation of the branch adenosine while providing a positively charged surface which binds the backbone of the BPS-U2 duplex. This would likely serve to stabilize the formation of this duplex by reducing the charge repulsion between the negatively charged RNA strands, as well as reducing the entropic penalty of a bulged nucleotide.

3-2-4. *in vivo* depletion of p14

To help understand the function of p14 in the spliceosome, we have investigated pre-mRNA splicing in a p14-knockout strain of the yeast *S. pombe*, in which the branch-point sequence is moderately conserved. This yeast grew only ~7% slower in liquid culture than a wild type strain, and an RT-PCR assay of splicing of 10 genes containing introns showed no detectable defect in pre-mRNA splicing (Schellenberg and MacMillan, unpublished), suggesting that the growth defect caused by a loss of the p14 gene only results from defects in splicing of a subset of pre-mRNAs.

Human branch point sequences are much more degenerate than those of *S. pombe*, and often contain multiple mismatches between the BPS and U2 snRNA. We have developed a methodology to deplete p14 protein from HeLa cells using RNAi. Although we were able to detect an ~90% knockdown of p14 pre-mRNA levels, 66 hours after transfection of siRNA we detected only a slight decrease in p14 protein levels (Figure 3-5A,B). In contrast, we could detect similar levels of knockdown of the SF3b155 mRNA by targeting it with siRNAs, but a more significant decrease in protein expression. To assay splicing in these HeLa cells we transfected a reporter plasmid in which the PIP85b intron (Query et al., 1994) was cloned into the 5' end of a green fluorescent protein open-reading frame, and splicing was assayed by RT-PCR using primers which annealed to the exons. When levels of SF3b155 protein decreased, an increase in the level of pre-mRNA relative to spliced mRNA was observed indicating a deficiency in splicing in the absence of this protein (Figure 3-5C), consistent with published reports (Kaida et

al., 2005). This effect was significant at 66 hours, and was followed by cell death. For the p14 knockdown, it was found that a second transfection of siRNA 48 hours after the first was necessary to deplete p14 protein levels, and at 96 hours a significant decrease in p14 protein levels was observed (Figure 3-5D,E). This caused a defect in splicing of the pre-mRNA from the reporter, although to a lesser extent than depletion of the SF3b155 protein (Figure 3-5F). This was followed by a cell death by a minority of cells (~25%), with the remaining cells

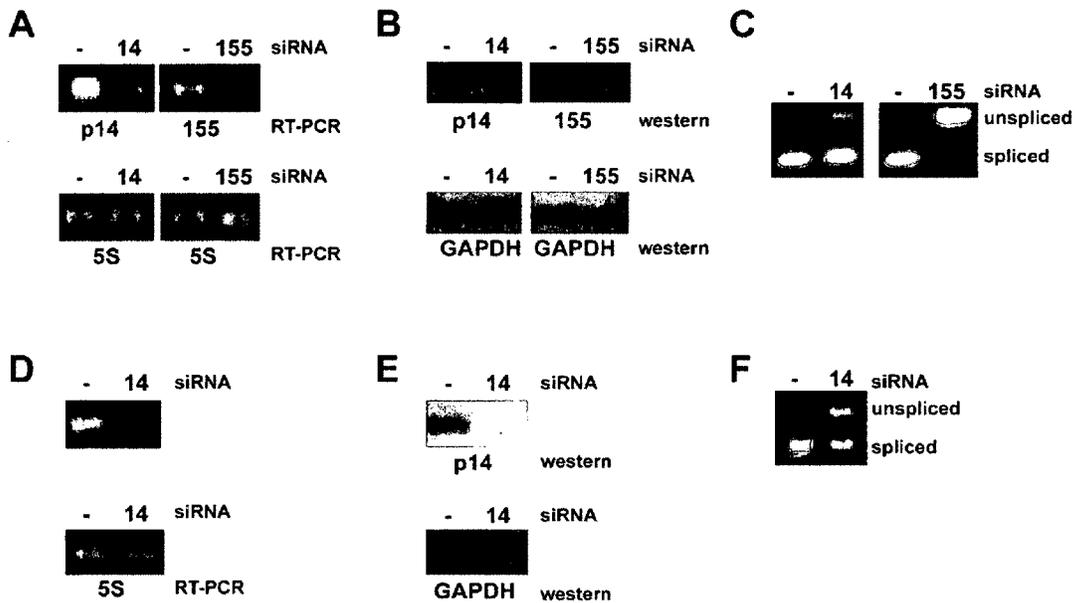


Figure 3-5. *in vivo* knockdown of p14 and SF3b155 proteins using RNAi at 66 or 96 hours. (A) 5S rRNA and p14 and SF3b155 mRNA levels measured using low-copy RT-PCR 66 hours after transfection of scrambled negative control siRNA (-), p14 (14), or SF3b155 (155) siRNAs, using primers annealing to indicated RNAs. (B) Protein levels measured by western blot 66 hours after transfection of scrambled negative control siRNA (-), p14 (14), or SF3b155 (155) siRNAs, using indicated antibodies. (C) splicing of pre-mRNA reporter assayed using RT-PCR 66 hours after transfection of scrambled negative control siRNA (-), p14 (14), or SF3b155 (155). (D) 5S rRNA and p14 mRNA levels measured using low-copy RT-PCR 96 hours after transfection of scrambled negative control siRNA (-) or p14 (14) siRNAs, using primers annealing to indicated RNAs. (E) Protein levels measured by western blot 96 hours after transfection of scrambled negative control siRNA (-) or p14 (14) siRNAs, using indicated antibodies. (F) splicing of pre-mRNA reporter assayed using RT-PCR 96 hours after transfection of scrambled negative control siRNA (-) or p14 (14) siRNAs.

recovering and returning to growth after a few more days. The larger splicing defect caused by SF3b155 knockdown compared to p14 knockdown is likely due to the proposed role of SF3b155 as a scaffolding protein for the SF3b complex (Golas et al., 2003); depletion of SF3b155 likely causes the failure of the entire SF3b particle to function during spliceosome assembly. Nonetheless, these results indicate that both SF3b155 and p14 are required for pre-mRNA splicing in HeLa cells.

3-3. Methods

3-3.1. Protein Expression and Purification

DNA fragments encoding mutant p14 were generated by overlapping PCR and cloned into pMALC2-X as described (section 2-4.3). All mutants were verified by sequencing. Protein expression and purification of mutant p14•SF3b155 aa 373-415 were carried out as previously described (section 2-4.3).

3-3.2 RNA Synthesis

RNAs were synthesized on an ABI 394 synthesizer using 2'-ACE chemistry (Dharmacon) modified to incorporate the thiol tether during synthesis. After the first detritylation step, the RNA synthesis column was removed from the machine, and the following steps were performed manually with a syringe: (1) wash with 1 mL of 50:50 acetonitrile:pyridine, (2) coupling with 28 mg of 5'-trityl 3'-H phosphonate thymidine monomer (ChemGenes) dissolved in 1 ml of 50:50 acetonitrile:pyridine and 40 μ L of pivaloyl chloride (Sigma) for 60 s, (3) wash

with 1 mL of 50:50 acetonitrile:pyridine, (4) wash with 5 ml of acetonitrile, (5) oxidation for 1 hr with 50:25:25 carbon tetrachloride:pyridine:cystamine disulfide (freshly dried over MgSO_4 in dichloromethane), (6) wash with 5 mL of pyridine, and finally (7) wash with 5 mL of acetonitrile. The column was returned to the ABI 394 synthesizer, and capped with acetic anhydride followed by detritylation, and the synthesis was continued using standard 2'-ACE chemistry. The oligonucleotides were deprotected according to the manufacturers instructions with the added steps of addition of 0.1 M β -mercaptoethanol during base deprotection and 20 mM tris(2-carboxyethyl)phosphine (TCEP) during the 2'-ACE removal step. Crude oligonucleotides were then modified with 5,5'-dithiobis-(2-nitrobenzoic) acid (DTNB) prior to separation by denaturing PAGE (20%, 19:1 acrylamide:bis acrylamide, 8M urea) in TBE running buffer. The band corresponding to the full length oligonucleotide containing the thiol modification was identified, excised, and extracted from the gel slice.

3-3.3. Protein-RNA Disulfide Bond Formation

Mutant p14•SF3b155 peptide complex was dialyzed overnight into buffer (10 mM Tris pH 8.0, 60 mM KCl, 0.1 M EDTA) containing 1 mM β -mercaptoethanol, followed by dialysis in the same buffer containing 0.1 mM β -mercaptoethanol for 1 hour. Protein complex was mixed with RNA (50 μM) in the same buffer such that the final concentration of β -mercaptoethanol was 60 μM , and disulfide bond formation was allowed to proceed overnight at 4°C. For analysis of the kinetic products of crosslinking, samples were removed after 20

minutes and treated as described below. All samples were analyzed by 16% (w/v) SDS-PAGE under non-reducing conditions after first quenching all free thiols with 2 mM iodoacetamide for 10 min and using a loading dye lacking β -mercaptoethanol. For SAXS analysis the p14•SF3b155 peptide RNA complex was purified on a Superdex 75 column (GE Healthcare Life Sciences) followed by ion exchange on a mono-Q HR5/5 column (GE Healthcare Life Sciences).

3-3.4. Protein-RNA Disulfide Reduction

p14•SF3b155 peptide RNA complex was added to reactions containing 5 mM β -mercaptoethanol and incubated at room temperature for between 10 and 120 min. At the end of each time-point, free thiols were quenched with a two-fold excess of iodoacetamide for 10 min followed by non-reducing 16% (w/v) SDS-PAGE using loading dye lacking β -mercaptoethanol. SDS-PAGE gels were stained with coomassie blue.

3-3.5. Small Angle X-ray Scattering

Small Angle X-ray Scattering (SAXS) data were collected at beamline 12.3.1 of the Advanced Light Source, Lawrence Berkeley National Laboratory, at 12 keV on a MAR165 detector. Samples were maintained at 10°C during data collection using a thermostatically controlled cuvette. p14•SF3b155 peptide samples at 2.5 and 10 mg/ml and a p14•SF3b155 peptide-RNA sample at a concentration of 10 mg/mL were buffer-exchanged overnight at 4°C in 25 μ L dialysis buttons, utilizing 6-8 kDa nominal molecular weight cut-off regenerated

cellulose dialysis tubing, against a 50 mL volume of the SAXS buffer (10 mM Tris pH 8.0, 60 mM NaCl, 0.1mM EDTA, and 5% glycerol). A series of exposures (6, 60, and 6 s) were measured to assess for radiation sensitivity. Scattering of the dialysis medium alone from 6 or 60 s exposures was subtracted from the scattering of the equivalently exposed samples to yield scattering curves for the macromolecules alone. Data were analyzed using the programs PRIMUS43 (Konarev et al., 2003), GNOM44 (Svergun, 1992), and AUTORG45 (Konarev et al., 2003). No radiation damage was apparent as the first and last 6 s exposures produced curves which were superimposable, no aggregation was visible to the program AUTORG45, and a linear dependence of $\log(I(s))$ vs s^2 in the range $sRG < 1.3$ was observed. The scattering curves for the 6 s exposure (data points 1 to 250) and the 60 s exposure (data points 100 to 504) of the 10mg/ml p14•SF3b155 peptide RNA sample were merged. Data points 1 to 250 for the 60 s exposure of the 2.5 mg/ml p14•SF3b155 peptide and data points 100 to 504 for the 60 s exposure of the 10mg/ml sample were merged using PRIMUS43. The merged scattering curves were used for analysis with the program GNOM44 which generated a $P(r)$ curve. Ten rounds of model building using the program GASBOR (Svergun et al., 2001) were averaged using DAMAVER (Volkova et al., 2003). The starting model from DAMAVER was used as an initial model for the program DAMMIN (Svergun, 1999), which produced the final refined *ab initio* model.

The program CHADD (Petoukhov et al., 2002) was used to add dummy residues to the crystal structure of p14•SF3b155 peptide (PDB entry 2F9D, chains

A and P) to model amino acids which were disordered in the crystal structure. The program SASREF (Petoukhov and Svergun, 2005) was used to model the complex between the p14•SF3b155 peptide model generated by CHADD and a model RNA duplex (PDB entry 1I9X chain A, nt 2-8 and chain B, nt 7-12 (Berglund et al. 2001), with a GAAA tetraloop at the end of the helix from PDB entry 1TLR, nt 5-8 (Butcher et al., 1997)), using restraints corresponding to the N25C-RNA disulfide tether and the bulged adenosine-Y22 interaction. Theoretical scattering from models was compared to the experimental scattering curves using the program CRY SOL27 (Svergun et al., 1995). The *ab initio* models were superimposed with the CHADD and SASREF models using the program SUPCOMB20 (Kozin and Svergun, 2001).

3-3.6 *in vivo* depletion of p14 and SF3b155 using siRNAs.

HeLa cells were grown in Dulbecco's Modified Eagle Medium (DMEM, Invitrogen) supplemented with 6% Fetal Bovine Serum (FBS) to 60-70% confluency in a 24-well plate (BD) in an incubator at 37°C and 5% CO₂. Cells were washed twice with phosphate buffered saline (PBS) and transfected with 500 µL transfection mix (30 nM siRNA and 2 µM Transductin reagent (IDT) dissolved in Opti-MEM media (Invitrogen). After a 4 hour incubation, cells were washed once with PBS, removed from the plates using 0.02% trypsin solution (Invitrogen), and plated at a 1:5 fold dilution in a 24-well plate in DMEM + 6% FBS. For assay of pre-mRNA splicing, after 44 hours 200 ng of reporter plasmid was transfected using 1 µL Lipofectamine 2000 reagent (Invitrogen) according to

the manufacturer's instructions. For a second transfection of p14 siRNA, 20 nm siRNA (final concentration) was included with the plasmid DNA. 4 hours after this transfection, cells were removed from the plates using 0.02% trypsin solution and plated at a 1:4 fold dilution in a 6 well plate. At the indicated times protein and RNA were extracted from cells using Trizol (Invitrogen) according to the manufacturer's instructions.

3-3.7 Western blotting and RT-PCR.

For western blots, 8µg of total protein, as assayed using Bradford reagent (BioRad) was run on a 16% 29:1 acrylamide SDS-PAGE. Protein was transferred to a nitrocellulose membrane (BioRad) in transfer buffer (20mM glycine, 10mM ethanolamine, 20% methanol) for 2 hours at 250mA in a Mini-Trans Blot system (BioRad), and blocked using Western Block Solution (Qiagen) for 1 hour. Membranes were probed for 1 hour with primary antibody, washed three times, and probed with the appropriate secondary antibody for 1 hours followed by washing 3 times. Antibodies were visualized using Immobilon reagent (Millipore) and visualized using ECL Plus film (GE Healthcare). Anti-p14 antibody was purchased from Protein Biotech, anti SF3b155 from AbCam, anti-GAPDH from AbCam, anti-rabbit-HRP from NEB, and anti-mouse-HRP from Bio-Rad.

For low-copy RT-PCR analysis of 5s rRNA, p14 mRNA, and SF3b155 mRNA, 10 ng of total RNA was used as the template in a 10 uL RT-PCR reaction using the Superscript II one-step RT-PCR kit (Invitrogen) with 1.25µL of each

primer. The RT-PCR cycle was as follows: 30 mins at 50°C followed by N x cycles (30 sec at 92°C, 30 sec at 50°C, and 90 sec at 72°C), where N=6 for 5S rRNA, 15 for SF3b155, and 19 for p14. RT-PCR products were run on a 2% agarose gel stained with ethidium bromide.

3-4. References

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

Banerjee, A., Verdine, G.L. (2006). A nucleobase lesion remodels the interaction of its normal neighbor in a DNA glycosylase complex. *Proc. Natl. Acad. Sci. USA* **103**:15020-15025.

Berglund, J.A., Rosbash, M., 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., Luhrmann, R. (2008) Isolation of an active site step I spliceosome and composition of its RNP core. *Nature* **452**, 846-850.

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

Burge, C.B., Tuschl, T., Sharp, P.A. (1999) The RNA World 2nd edn. (eds. Gesteland RF, Cech TR, Atkins JF), pp 525-560, *Cold Spring Harbor Laboratory Press*, Cold Spring Harbor, NY.

Cancilla, M.T., He, M.M., Viswanathan, N., Simmons, R.L., Taylor, M., Fung, A.D., Cao, K., Erlanson, D.A. (2008) Discovery of an Aurora kinase inhibitor through site-specific dynamic combinatorial chemistry. *Bioorg. Med. Chem. Lett.* **18**, 3978-3981.

Cléry, A., Blatter, M., Allain, F.H. (2008) RNA recognition motifs: boring? Not quite. *Curr. Opin. Struct. Biol.* **18**, 290-298.

Corn, J.E., Berger, J.M. (2007) FASTDXL: a generalized screen to trap disulfide stabilized complexes for use in structural studies. *Structure* **15**, 773-780.

Deo, R.C., Bonanno, J.B., Sonenberg, N., Burley, S.K. (1999) Recognition of polyadenylate RNA by the poly(A)-binding protein. *Cell* **98**, 835-845.

Erlanson, D.A., Lam, J.W., Wiesmann, C., Luong, T.N., Simmons, R.L., DeLano, W.L., Choong, I.C., Burdett, M.T., Flanagan, W.M., Lee, D., Gordon, E.M., O'Brien, T. (2003a) In situ assembly of enzyme inhibitors using extended tethering. *Nat. Biotechnol.* **21**, 308-314.

Erlanson, D.A., McDowell, R.S., He, M.M., Randal, M., Simmons, R.L., Kung, J., Waight, A., Hansen, S.K. (2003b). Discovery of a new phosphotyrosine mimetic for PTP1B using breakaway tethering. *J. Am. Chem. Soc.* **125**, 5602-5603.

Erlanson, D.A., Braisted, A.C., Raphael, D.R., Randal, M., Stroud, R.M., Gordon, E.M., Wells, J.A. (2000) Site-directed ligand discovery. *Proc. Natl. Acad. Sci. USA* **97**, 9367-9372.

Froehler, B.C. (1986) Deoxynucleoside H-Phosphonate diester intermediates in the synthesis of internucleotide phosphate analogues. *Tett. Lett.* **27**, 5575-5578

Fromme, J.C., Banerjee, A., Huang, S.J., Verdine, G.L. (2004) Structural basis for removal of adenine mispaired with 8-oxoguanine by MutY adenine DNA glycosylase. *Nature* **427**, 652-656.

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

Gilbert, H.F. (1995) Thiol/disulfide exchange equilibria and disulfide bond stability. *Met. Enzymol.* **251**, 8–28.

Gozani, O., Potashkin, J., 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., Yokoyama, S. (1999) Structural basis for recognition of the tra mRNA precursor by the Sex-lethal protein. *Nature* **398**, 579-585.

Hartmann, B., Valcárcel, J. (2009) Decrypting the genome's alternative messages. *Curr. Opin. Cell. Biol.* **21**, 377-386.

He, C., Verdine, G.L., (2002) Trapping distinct structural states of a protein/DNA interaction through disulfide crosslinking. *Chem. Biol.* **9**, 1297-1303.

Huang, H., Chopra, R., Verdine, G.L., Harrison, S.C. (1998) Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science* **282**:1669-1675.

Huang, H., Harrison, S.C., Verdine, G.L. (2000) Trapping of a catalytic HIV reverse transcriptase template:primer complex through a disulfide bond. *Chem. Biol.* **7**:355-364.

Johnson, A.A., Santos, W., Pais, G.C., Marchand, C., Amin, R., Burke, T.R., Verdine, G.L., Pommier, Y. (2006) Integration requires a specific interaction of the donor DNA terminal 5'-cytosine with glutamine 148 of the HIV-1 integrase flexible loop. *J. Biol. Chem.* **281**, 461-467.

Komazin-Meredith, G., Santos, W.L., Filman, D.J., Hogle, J.M., Verdine, G.L., Coen, D.M. (2008) The positively charged surface of herpes simplex virus UL42 mediates DNA binding. *J. Biol. Chem.* **283**, 6154-6161.

Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J., Svergun, D.I. (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Cryst.* **36**, 1277-1282.

Kozin, M.B., Svergun, D.I. (2001) Automated matching of high- and low-resolution structural models. *J. Appl. Crystallogr.* **34**, 33-41.

Kuwasako, K., Dohmae, N., Inoue, M., Shirouzu, M., Taguchi, S., Güntert, P., Séraphin, B., Muto, Y., Yokoyama, S. (2008) Complex assembly mechanism and an RNA-binding mode of the human p14-SF3b155 spliceosomal protein complex identified by NMR solution structure and functional analyses. *Proteins* **71**, 1617-1636.

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

Lee, S., Radom, C.T., Verdine, G.L. (2008) Trapping and structural elucidation of a very advanced intermediate in the lesion-extrusion pathway of hOGG1. *J. Am. Chem. Soc.* **130**, 7784-7785.

Lin, Y., 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., Houngrinou-Molango, S., Sprangers, R., Zanier, K., Kramer, A., Sattler, M. (2001) Structural basis for recognition of the intron branch site RNA by splicing factor 1. *Science* **294**, 1098-1102.

MacMillan AM, Query CC, Allerson CA, Chen S, Verdine GL, Sharp PA. 1994. Dynamic association of proteins with the pre-mRNA branch region. *Genes Dev* **8**, 3008-3020.

McRee DE. 1999. XtalView/Xfit--A versatile program for manipulating atomic coordinates and electron density. *J Struct Biol* **125**, 156-165.

Murshudov GN, Vagin AA, Dodson EJ. 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Cryst D* **53**, 240-255.

Newby MI, Greenbaum NL. 2002. Sculpting of the spliceosomal branch site recognition motif by a conserved pseudouridine. *Nature Struct Biol* **9**, 958-965.

O'Shea EK, Rutkowski R, Stafford WF, Kim PS. 1989. Preferential heterodimer formation by isolated leucine zippers from fos and jun. *Science* **245**, 646-648.

Otwinowski Z, Minor W. 1997. Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Methods Enzymol* **276**, 307-326.

Oubridge C, Ito N, Evans PR, Teo CH, 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.

Petoukhov MV, Eady NA, Brown KA, Svergun DI. 2002. Addition of missing loops and domains to protein models by x-ray solution scattering. *Biophys. J.* **83**, 3113-3125.

Petoukhov MV, Svergun DI. 2005. Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys. J.* **89**, 1237-1250.

Portmann S, Susan Grimm, S, Workman, C, Usman, N, Egli M. 1996. Crystal structures of an A-form duplex with single-adenosine bulges and a conformational basis for site-specific RNA self-cleavage. *Chem Biol.* **3**, 173-184.

Query CC, Moore MJ, Sharp PA. 1994. Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model. *Genes Dev* **8**, 587-597.

Query CC, Strobel SA, Sharp PA. 1996. Three recognition events at the branch-site adenine. *EMBO J.* **15**, 1392-13402.

Reed R. 1990. Protein composition of mammalian spliceosomes assembled in vitro. *Proc Natl Acad Sci USA* **87**, 8031-8035.

Schellenberg MJ, Edwards RA, Ritchie DB, Kent OA, Golas MM, Stark H, Luhrmann R, Glover JN, MacMillan AM. Crystal structure of a core spliceosomal protein interface. *Proc Natl Acad Sci USA* **103**, 1266-1271.

Smith DJ, Query CC, Konarska MM. 2007. Trans-splicing to spliceosomal U2 snRNA suggests disruption of branch site-U2 pairing during pre-mRNA splicing. *Mol Cell* **26**, 883-890.

Spadaccini R, Reidt U, Dybkov O, Will C, Frank R, Stier G, Corsini L, Wahl MC, Lührmann R, Sattler M. 2006. Biochemical and NMR analyses of an SF3b155-

p14-U2AF-RNA interaction network involved in branch point definition during pre-mRNA splicing. *RNA* **12**, 410-425

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

Stanojevic D, Verdine GL. 1995. Deconstruction of GCN4/GCRE into a monomeric peptide-DNA complex. *Nat Struct Biol* **2**, 450 – 457.

Svergun DI. 1992. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J Appl Cryst* **25**, 495-503.

Svergun DI. 1999. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **76**, 2879-2886.

Svergun DI, Barberato C, Koch MHJ. 1995. CRY SOL a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. App. Crystallogr.* **28**, 768-773.

Svergun DI, Petoukhov MV, Koch MHJ. 2001. Determination of domain structure of proteins from X-ray solution scattering. *Biophys J* **80**, 2946-2953.

Volkova VV, Svergun DI. 2003. Uniqueness of *ab initio* shape determination in small-angle scattering. *J Appl Cryst* **36**, 860-864.

Wang X, Hall TMT. 2001. Structural basis for recognition of AU-rich element RNA by the HuD protein. *Nat Struct Biol* **8**, 141-145.

Will CL, Schneider C, MacMillan AM, Katopodis NF, Neubauer G, Wilm M, Lührmann R Query CC. 2001. A novel U2 and U11/U12 snRNP protein that associates with the pre-mRNA branch site. *EMBO J* **20**, 4536-4546.

Xu YZ, Query CC. 2007. Competition between the ATPase Prp5 and branch region-U2 snRNA pairing modulates the fidelity of spliceosome assembly. *Mol. Cell* **28**, 838-849.

Zhao Z, McKee CJ, Kessl JJ, Santos WL, Daigle JE, Engelman A, Verdine GL, Kvaratskhelia M. 2008. Subunit-specific protein footprinting reveals significant structural rearrangements and a role for N-terminal Lys-14 of HIV-1 Integrase during viral DNA binding. *J Biol Chem* **283**, 5632-5641.

Chapter 4⁽¹⁾
The Structure 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. M.S. performed the X-ray data collection, solved the crystal structure, and performed the limited proteolysis RNA footprinting. D.R. crystallized the domain IV protein and performed the EMSA experiments. E.G., S.R., and D.S. assisted in experiments not described in the chapter and provided helpful discussions.

4-1. Introduction

4-1.1. The highly conserved spliceosomal protein Prp8

Assembly of the spliceosome active site on the pre-mRNA proceeds through distinct stages and involves the sequential association of spliceosomal and non-spliceosomal factors with conserved sequences at the splice sites, within the intron, and with other sequences in the pre-mRNA. The assembled spliceosome contains a U2/U6 snRNA structure, which has been proposed to form the active site for catalysis of the two transesterifications (Brow, 2002; Valadkhan et al., 2007). 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 separate RNA-protein cross-linking studies in humans and yeast, Prp8 has been demonstrated to crosslink to important nucleotides in the spliceosome (Reyes et al., 1999; Vidal et al., 1999). Furthermore, mutant PRP8 alleles in yeast strongly suggest that within the spliceosome this factor interacts with both substrate and snRNA catalytic structures in the spliceosome (Grainger and Beggs, 2005). Thus, the catalytic heart of the spliceosome may include protein as well as RNA components.

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). Prp8 is a component of the U5 snRNP (Lossky et al. 1987; Stevens et al. 2001) and U5•U4/U6 tri-snRNP (Teigelkamp et al. 1997; Gottschalk et al. 1999; Stevens and Abelson 1999). With 61% sequence identity between *S. cerevisiae* 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

For the remainder of this thesis, the human protein will be distinguished from the yeast (*S. cerevisiae*) protein by hPrp8 or PRP8 respectively. 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.

4-1.2. Prp8 crosslinks to important nucleotides in the spliceosome

Cross-linking studies in yeast and HeLa nuclear extract revealed direct contact between Prp8 and the 5' splice site (5'SS), 3' splice site (3'SS), and branch point sequence (BPS) (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, its interaction with the branch nucleotide and 3'SS occurs concomitant with or subsequent to step 1 (Figure 2-1A; MacMillan et al., 1994; 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 to amino acids 1894-1898 of hPrp8 (Reyes et al., 1999). This 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.

Interestingly, more recent cross-linking experiments carried out in yeast followed by proteolytic mapping suggest regions of the protein different from those previously identified are responsible for interaction with the 5'SS. A screen where tobacco etch virus (TEV) protease sites were randomly inserted throughout the PRP8 gene was used to map the crosslinks between PRP8 and spliceosomal RNA (Turner et al., 2006). The 5' splice site (5'SS)- and branch point sequence (BPS)-interacting regions of PRP8 were identified by this method 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 determined to crosslink to the 5'SS in the human system (Turner et al., 2006). The difference may be explained by differences in the stage of assembly of spliceosomal complexes at the time of crosslinking and by different crosslinking reagents. Konarska's cross-link (equivalent to aa 1946-1950 of yeast PRP8) likely reflects

an interaction present in B complex (Reyes et al., 1999) while Newman's cross-links (aa 871-970, 1281-1413, and 1503-1673 of PRP8) is more indicative of interactions occurring in the catalytic C complex (Turner et al., 2006). This crosslinking system also 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, implying they directly affect 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. This TEV insertion system was used to map cross-links between PRP8 and both the U6 snRNA and U5 snRNA in the context of purified U5•U4/U6 tri-snRNPs (Turner et al., 2006). Cross-links from the invariant U-rich loop of U5 snRNA were mapped to PRP8 amino acids 770-871 and 1281-1413; the latter range overlaps with the region crosslinking to the 5'SS and BPS. The cross-link to the ACAGA region of U6 snRNA was similarly mapped to amino acids 1503-1673, which also overlaps with a region that crosslinks to the 5'SS and BPS. These results suggest that Prp8 directly contacts the complexes formed between the spliceosomal RNAs and the reactive sites in the pre-mRNA, supporting the idea that Prp8 is directly involved in catalyzing the splicing reaction.

4-1.3. Prp8 protein domain structure

Prp8 is one of the largest (>220 kDa) and most highly conserved nuclear proteins known to date, with 61% sequence identity between *S. cerevisiae* and

humans. Despite the size and high sequence identity and conservation of this protein between species, little is known with respect to Prp8 structure. Inspection of the primary sequence suggests the presence of a central RNA recognition motif (RRM) and a C-terminal Jab1/MPN domain, within which mutations associated with retinitis pigmentosa in humans have been characterized (Tiegelkamp et al., 1995; Grainger and Beggs, 2005; Figure 4-1A); the presence of an MPN domain has been verified by high-resolution structural analysis (Pena et al., 2007). However, there is no genetic or biochemical evidence to implicate either of these domains in direct interactions with the spliceosomal RNA core. A clustering of mutations specifies another two domains, domain 3 and IV, for which an analysis of the amino acid sequence provides no clues as to the function of these domains (Grainger and Beggs, 2005).

4-1.4. A putative RRM in Prp8

Grainger and Beggs (2005) offered the first report of an RRM conserved among all Prp8 orthologs. Genetic interactions between the RRM of PRP8 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 three amino acids, KDM, which is the location of three mutations that suppress U4-*cs1*. Interestingly, the KDM tripeptide is located in a loop between β -strands

2 and 3, a region important for determining specificity of RNA-binding by RRM. It is not known whether the RRM in Prp8 binds RNA directly, or whether it is a protein-interaction motif, similar to a UHM (section 1-2.5).

4-1.5. The MPN domain of Prp8

The amino acid sequence of the C-terminus of Prp8 indicates the presence of an MPN domain (also known as Mov34, JAB, MPN+, PAD-1, or JAMM domain). 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 (Maytal-Kivity et al., 2002; Grainger and Beggs 2005). The typical model of an MPN domain is the proteasome component Rpn11, which has been proposed to be the proteasomal subunit responsible for recognizing and cleaving ubiquitin from its substrates (Verma et al., 2002). Crystal structures of MPN domains have been solved and reveal a Zn^{2+} binding site consisting of 4 conserved coordinating ligands (Tran et al., 2003; Ambroggio et al., 2004). The Zn^{2+} coordinating residues are found in 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 a histidine which gives the sequence H-x-Q-x[7]-S-xx-D. This change indicates that Prp8's MPN domain would likely not coordinate a Zn^{2+} ion.

High resolution structure analysis of the C-terminus of Prp8 shows it folds into an MPN domain with N- and C-terminal extensions which are part of the folded domain (Pena et al., 2007; Zhang et al., 2007). As predicted by the lack of

a histidine in the JAMM motif, no zinc ion is observed in this structure. Nonetheless, this domain appears to be capable of binding ubiquitin (Bellare et al., 2006); this binding has been shown to be important for U5 assembly into the U5•U4/U6 tri-snRNP (Bellare et al., 2008). The MPN domain has also been shown to stimulate the ATPase activity of Brr2, which is an important helicase responsible for unwinding the U4/U6 snRNA structure during spliceosome activation, as well as the U2/U6 structure after splicing catalysis (Maeder et al., 2009; Madhani and Guthrie 1992;).

4-1.6. Prp8 Domain 3

Domain 3 (aa 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 PRP8 alleles in domain 3 that suppress not only mutations at the 3'SS but also mutations at both the 5'SS and BPS (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 do not 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, domain 3.1 and the more C-terminal domain 3.2 (1547–1660 in yeast), which is more highly conserved (72% identity between yeast and human) and contains 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 three-dimensional structure of domain 3.

4-1.7. Prp8 Domain IV

Similar to domain 3, domain IV was originally defined by the observed clustering of PRP8 alleles that suppress pre-mRNA mutations at the 5'SS, 3'SS, branch sequence, poly-pyrimidine tract, and mutations in U4 and U6 snRNA (Liu et al., 2007; Grainger and Beggs, 2005; Umen and Guthrie, 1995a; Umen and Guthrie, 1995b; Kuhn et al., 1999). 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; the stretch of five amino acids QACLK (h1894–1898) in hPrp8 that cross-link to the 5'SS are not conserved in yeast, the corresponding 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, suggesting that the 5 amino acids identified are not essential for RNA binding, 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).

4-1.8. Domain IV as a target for structural studies

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. Alleles mapping to domain IV have been found which suppress mutations at the 5'SS, 3'SS, BPS, and poly-pyrimidine tract (PPT) within the

Table 4-1. Mutant PRP8 alleles mapping to domain IV core and corresponding phenotypes.

Yeast Allele (Human)	5'SS	3'SS	BPS	PPT	U6	U4cs
F1834S/L (Y1762)				X		
F1851L (F1779)						X
V1860N (V1788)						X
V1860D (V1788)						X
T1861P (T1789)	X	X				X
V1862D (I1790)						X
V1862A/Y (I1790)						X
H1863E (H1791)		X	X			
K1864E (K1792)	X	X				
N1869D (N1797)	X	X	X		X	
V1870N (L1798)	X	X	X			
I1875T (I1803)						X
E1960K/G (E1888)				X		
T1982A (T1910)	X	X	X			
V1987A (T1915)	X	X	X			

Amino acid numbering is for the *S. cerevisiae* protein (human amino acid number shown in brackets). 5'SS = 5' splice site mutation suppression, 3'SS = 3' splice site mutation suppression, PPT = loss of preference for a 3' splice site with a polypyrimidine tract, BPS = suppression of branch point sequence mutation, U4cs = suppression of mutations which hyperstabilize the U4/U6 snRNA structure, U6 = suppression of a mutation which hyperstabilizes the ISL of U6 snRNA. Based on Grainger and Beggs, 2005, Liu et al., 2007, and Yang et al., 2008.

pre-mRNA, as well as mutations in U4 and U6 snRNAs (Table 4-1). Furthermore, PRP8 domain IV alleles have been found which either suppress or exacerbate mutations in the important DEA/H box helicases PRP16, PRP22, and PRP28 (Grainger and Beggs 2005; Liu et al., 2007). The diverse phenotypes and biochemical data with respect to domain IV both emphasize its importance in spliceosomal function, as well as pose the question as to how so many functions can reside in a single region of Prp8.

This chapter describes the high resolution X-ray structure of a stable core of domain IV. Analysis of this structure, combined with existing 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.

4-2. Results and Discussion

4-2.1. 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 4-1B; see Appendix 1). However the structure was a dimer, and it is known that Prp8 exists as only one copy per U5 snRNP. Nevertheless, analysis of this 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 4-1C; Table 4-2).

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 1918-1990). The RNase H fold exhibits a characteristic five-stranded mixed parallel/anti-parallel β -sheet, buttressed by two α -helices. This structural homology was not predicted by analysis of the 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 forms a linker between the two sub-domains of the core.

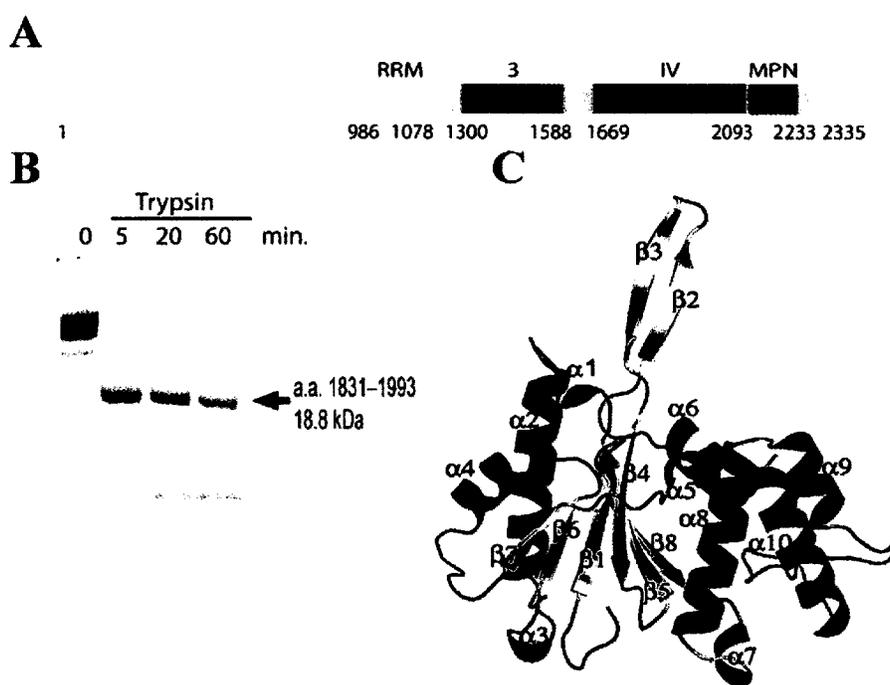


Figure 4-1. Structure of hPrp8 domain IV. A) Primary structure of hPrp8. Predicted domains include those inferred from primary-sequence homologies (a central RRM and C-terminal Jab1/MPN domain) and from the identification of mutant alleles in yeast, coupled with tolerance of amino acid insertions and 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.

The very high sequence similarity of Prp8 domain IV core between human 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 4-2;

Table 4-1). The majority of U4/U6 cold-sensitive suppressor alleles correspond to amino acids in the β -hairpin of the RNase H domain. The same is true for alleles which suppress mutations at either the splice sites or the BPS (Figure 4-3). One

Table 4-2. 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				
a, b, c (Å)	75.397, 78.063, 93.575		76.047, 78.003, 93.008	
α, β, γ (°)	90, 90, 90		90, 90, 90	
	<i>Native</i>	<i>peak</i>	<i>inflection</i>	<i>remote</i>
Wavelength (Å)	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			
B-factors (Å ²)				
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).

of these alleles, T1789 (T1861P in yeast), is located in the β -hairpin, and mutation to proline would be expected to disrupt the β -sheet hinting at conformational flexibility or change within the context of the spliceosome. An allele causing a loss of preference for a PPT (E1888G/K) is located on the surface of the RNase H domain, and a suppressor of U4/U6-*cs* mutation is located in the core of the protein (I1875T), indicating that regions outside the β -hairpin are either involved in any conformational change or can affect the function of the spliceosome.



Figure 4-2. 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 ($\alpha 5$) that forms the base of the channel separating the RNase H homology domain and the C-terminal helical cluster (Figures 4-1C, 4-3).

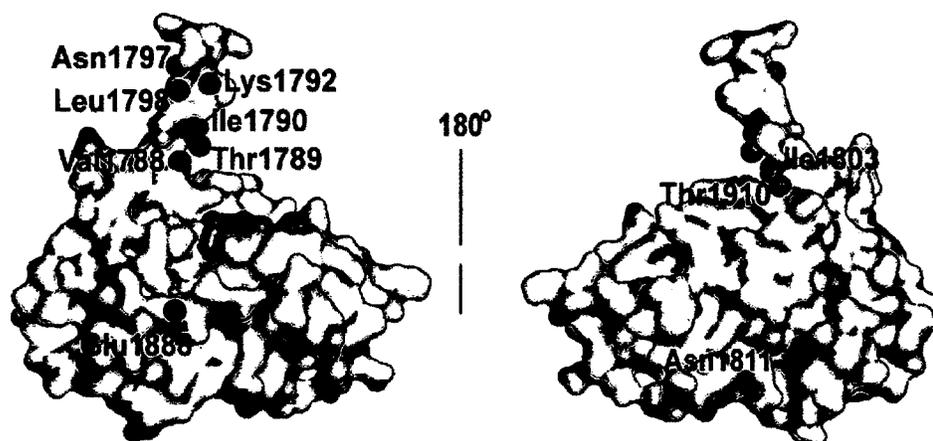


Figure 4-3. Mapping of yeast mutant alleles and a 5'SS cross-link in the context of the Prp8 domain IV core. Thirteen mutant alleles, corresponding to mutation at ten positions in yeast Prp8, map to the surface of the Prp8 domain IV core. All of the wild type residues are identical or conservative substitutions in humans, and the vast majority 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 4-1) 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 aa1755-2016 (PDB ID code 3E9L) and two yeast constructs consisting of aa1827-2092 (PDB ID code 3E9P) and aa1836-2092 (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 C α atoms, indicating that the structures are very similar. The only differences between the yeast and

human structures are the relative orientations of the β -hairpin to the RNase H β -sheet as well as a slightly longer N-terminus, such that the yeast amino acid F1834 can be seen located adjacent to E1960, consistent with the similar phenotype caused by mutation of these amino acids (Table 4-1).

4-2.2. 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 4-4A). 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 4-4A) (Song et al., 2004).

The RNase H homology within the domain IV core structure includes conserved secondary and tertiary structure but little relationship at the 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 through a combination of water nucleophile activation 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 4-4B).

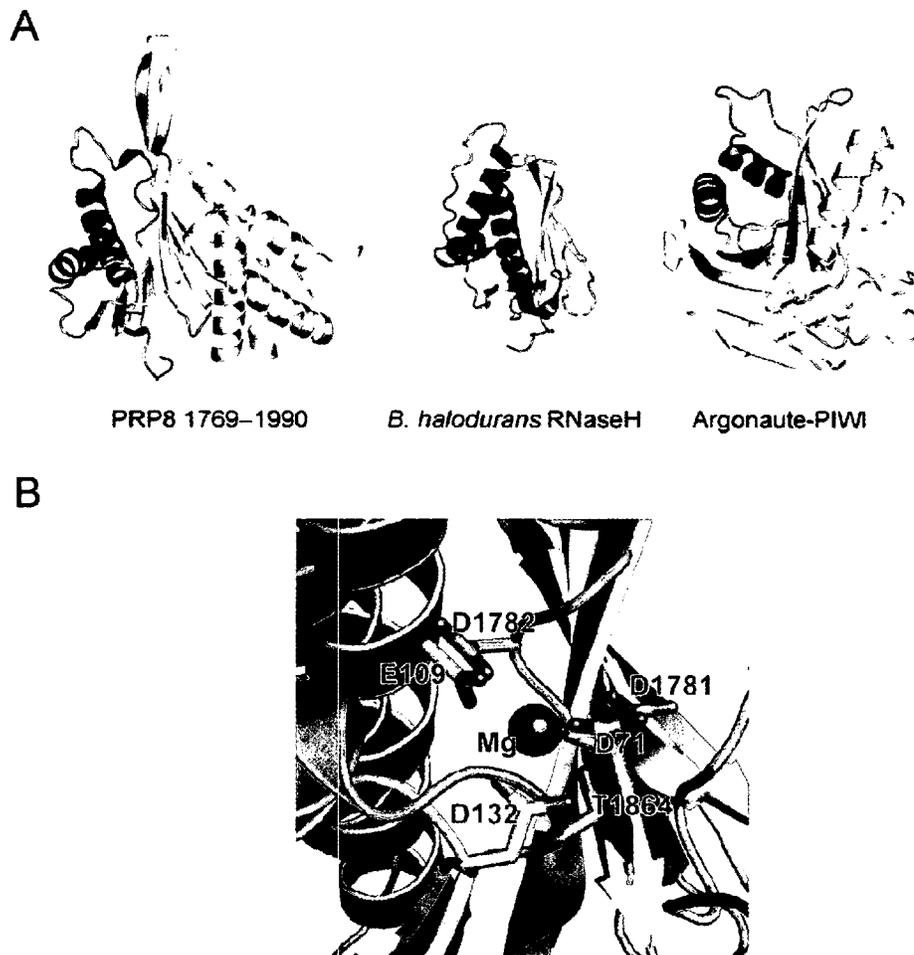


Figure 4-4. Comparison of the Prp8 domain IV core with RNase H folds. A) Structure of the hPrp8 domain IV core alongside *Bacillus halodurans* RNase H (Nowotny et al., 2005) and the PIWI domain of *P. furiosus* Argonaute (Song et al., 2004), with the RNase H fold indicated in red and yellow. B) Alignment 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 a Mg^{2+} ion in the core domain site despite the fact that crystals were grown in 200 mM $MgCl_2$, and Mg^{2+} is observed bound at another site bridging a symmetry related glutamate between neighbouring molecules in the crystal lattice. Inspection of the Prp8 Domain IV site shows that

in place of bound metal, the spatially conserved aspartates are stabilized by a network of water molecules (Figure 4-5A). In addition, the side chain of Arg1865 - part of the loop between $\beta 6$ and $\alpha 4$ - is hydrogen-bonded to Asp1782 effectively blocking the potential metal binding site. This blockage 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 4-5A) (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 Prp8 function has been investigated by mutating the corresponding residues in yeast and testing their effects on viability (Figure 4-5B; 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 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 to be seen whether these residues are coordinating an essential metal ion.

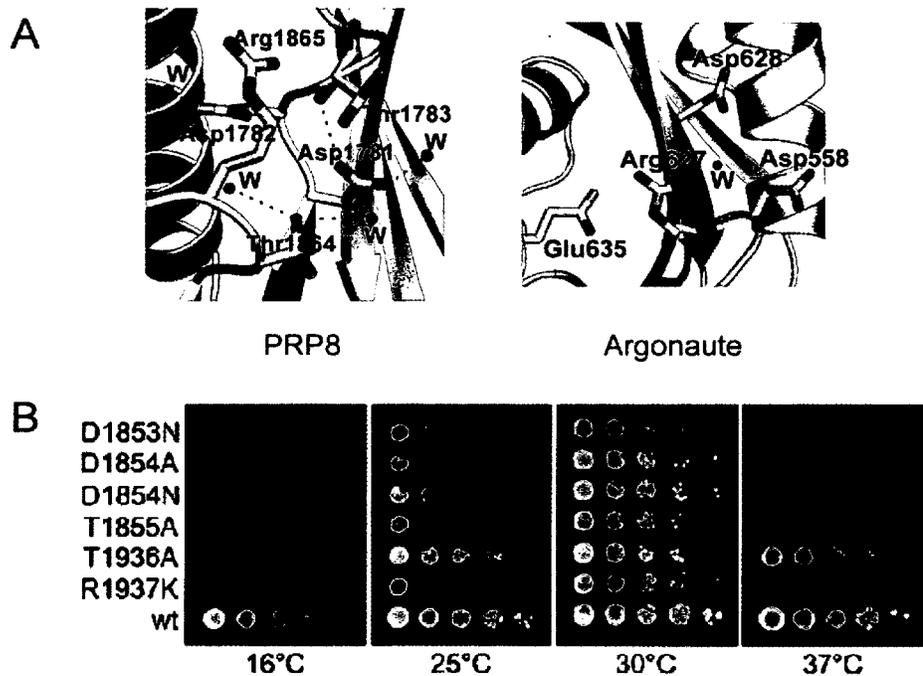


Figure 4-5. Analysis of the putative conserved binding site in Prp8 domain IV core. A) The potential metal binding site in the core domain (left) includes four hydrogen-bonding water molecules and is blocked by the side chain of Arg1865 hydrogen bonded to Asp1782 in a similar fashion to the interaction 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).

4-2.3 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 suggest a direct interaction in the spliceosome active site. There are numerous examples of cold-sensitive 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 aa 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 -hairpin, 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 the formation of a proper ACAGAGA-box/5'SS duplex, and 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 5, 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 ion, 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. This 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 plays 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 interacting with multiple strands of RNA, consistent with Prp8's multiple roles in mediating rearrangements in RNA structure on the way to spliceosome activation.

4-2.4. 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 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. 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 BPS•U2 snRNA duplex, and several more complex RNAs based on proposed structures in the mature spliceosome (Figure 4-6 and data not shown). Electrophoretic mobility shift assays with this panel of RNAs assayed against wild type hPrp8 domain IV core and several mutants revealed dissociation constants in the range of 20 to >300 μ M. 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 \sim 100 μ M vs >300 μ M).

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

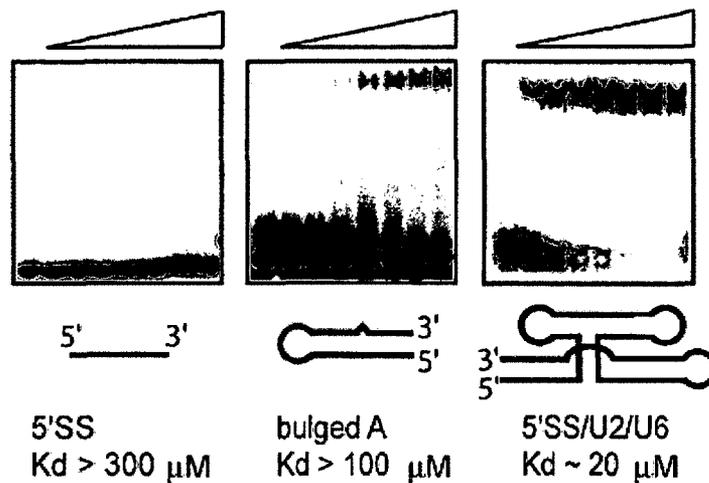


Figure 4-6. RNA binding by hPrp8 domain IV core. Gel mobility-shift assays for RNA binding to the core domain showing the affinity of three RNAs incubated with increasing concentrations (2–100 μM) of protein. Left to right: an 11-nucleotide RNA representing a pre-mRNA 5'SS, a 14-base-pair, 32-nucleotide 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).

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 Ferred'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 μM) was 5-10 fold greater than observed for simple single- or double-stranded RNAs. We also synthesized a catalytically compromised 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 (data not shown).

4-2.5. 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 Domain IV core using partial proteolysis footprinting (Figure 4-7). 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 R1832 within the loop connecting $\alpha 1$ and $\alpha 2$ (Figure 4-7), as determined by mass spectrometry. Similarly, treatment of the protein with chymotrypsin yielded a slightly larger ~21 kDa fragment representing the C-terminal product of cleavage at W1827 just C-terminal to $\alpha 1$ (Figure 4-7). When these proteolytic treatments were repeated in the presence of saturating four-helix-junction RNA, cleavage at R1832 was reduced three-fold while that at W1827 was not affected (Figure 4-7). Despite their proximity, the two residues in question are oriented towards opposite surfaces of the core domain, with R1832 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 R1832 but not W1827).

The overall electrostatics of the core domain surface matching the RNA binding face of RNase H are consistent with RNA-binding (Figure 4-8A). The observed protections from proteolysis combined with existing crosslinking data (Reyes et al., 1999) strongly implicate a surface comprising the RNase H fold as

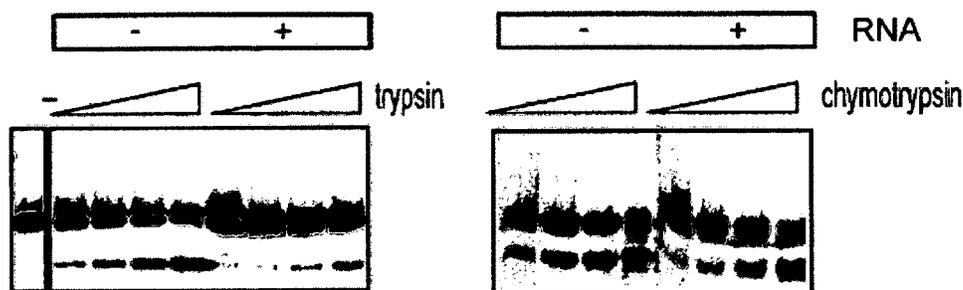


Figure 4-7. Footprinting analysis of a Prp8/RNA complex. Partial proteolysis of the core domain alone and in the presence of saturating four-helix-junction RNA showing protection against cleavage by trypsin at Arg1832 upon binding but not against cleavage by chymotrypsin at Trp1827.

well as the narrow channel across the core domain as being involved in RNA binding (Figure 4-8B). 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 4-3).

4-2.6. Implications of the core structure for 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 5'SS 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

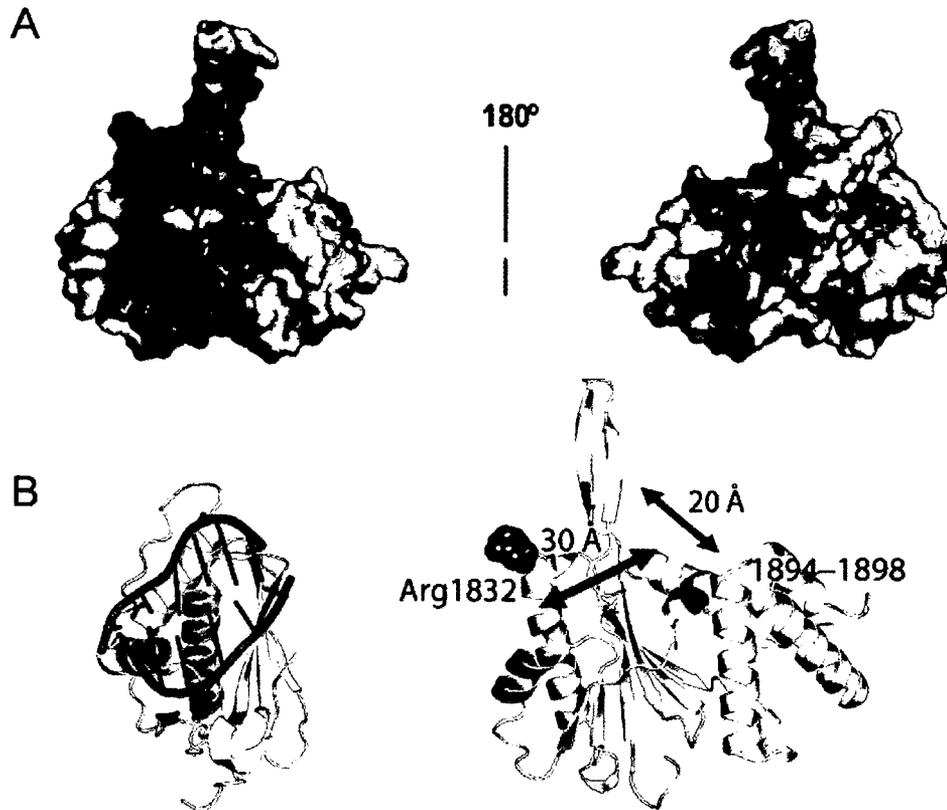


Figure 4-8. Representations of the proposed Prp8 domain IV core RNA binding surface. A) Surface electrostatics for two surfaces of the Prp8 domain IV core. Blue, basic; red, acidic; white, neutral. B) Comparison of RNA binding by RNase H and proposed core binding surface. Left, structure of an A-form 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.

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). However, in marked contrast to the ribosome, 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 spliceosome assembly.

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. Despite the absence of a bound metal in this structure and the fact that the putative metal binding site is blocked by a basic side chain, the absolute conservation of the putative metal binding ligands, as well as the deleterious effect of their mutation suggests that metal binding at this site is regulated or blocked. Therefore, 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. The multiple phenotypes of mutant alleles characterized in yeast suggest the possibility of distinct RNA-binding functions at various stages of spliceosome assembly and 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). Distinct alleles mapping to PRP8 domain 3 have been characterized that correspond to phenotypes associated with Domain IV mutants, suggesting 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.

4-3. Methods

4-3.1. 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 pMAL-C2x (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.

4-3.2. Crystallization

Crystals of hPrp8 aa1769-1990 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 μ L of 10 mg/mL protein solution (10 mM Tris, pH 8.0, 0.1 mM EDTA, 5 mM DTT, 0.02% NaN₃) with one μ L of precipitant (2.5 M NaCl, 100 mM Tris, pH 7, 100 mM MgCl₂). Crystals were

transferred to precipitant containing 20% glycerol and flash-cooled in liquid nitrogen for data collection.

4-3.3. Data collection and processing

Data were collected at beam line 8.3.1 of the Advanced Light Source, Lawrence Berkeley National Laboratory. Anomalous diffraction data were 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 4-2) (Otwinowski and Minor, 1997).

4-3.4. 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 4-2.

4-3.5. 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.

4-3.6. 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'

4-3.7. Limited proteolysis

hPrp8 domain IV core (100 pmol) was incubated in 4 µl 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 µg of chymotrypsin. Samples were incubated at 22°C for 30 min, quenched by addition of 6 µL 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.

4-4. 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., Auchynnika 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.

Burge, C.B., Tuschl, T. & Sharp, P.A. (1999) 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, 1999).

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.

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

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.

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.

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.

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

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.

Raghunathan, 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.

Rupert, P.B., Ferre-D'Amare, A.R. (20010) Crystal structure of a hairpin ribozyme-inhibitor complex with implications for catalysis. *Nature* **410**, 780–786.

Sashital, D.G., Cornilescu, G., 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.

Small, E.C., Leggett, S.R., Winans, A.A., and Staley, J.P. (2006). The EF-G-like GTPase Snul14p 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.

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

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.

Verma, R., Aravind, L., Oania, R., McDonald, W.H., Yates, J.R. 3rd, Koonin, E.V., Deshaies, R.J. (2002) Role of Rpn11 metalloprotease in deubiquitination and degradation by the 26S proteasome. *Science* **298**, 611-615.

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

Yang, K., Zhang, L., Xu, T., Heroux, A. Zhao, R. (2008) Crystal structure of the β -finger domain of Prp8 reveals analogy to ribosomal proteins. *Proc. Natl. Acad. Sci. USA* **105**, 13817–13822.

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 5

Metal-ion coordination mediated by PRP8 conformational change is coupled to pre-mRNA exon ligation

The work presented in this chapter represents a collaboration between authors. M.S. performed the X-ray data collection and solved the crystal structures. M.S., D.R, T.W., and K.A. prepared proteins for crystallization, grew the crystals, prepared mutant yeast strains, and performed splicing assays

5-1. Introduction

5-1.1. Prp8 alleles affecting the 1st and 2nd step of splicing

The two steps of pre-mRNA splicing are not chemically identical; therefore the spliceosome has either two active sites which the substrate must shuffle between, or a single active site which is reorganized between the two steps of splicing (Moore and Sharp, 1993). This requirement has also been demonstrated for group II introns (Podar et al., 1998) which suggests that the rearrangement likely involves the spliceosomal snRNAs. In support of this idea, Konarska and co-workers have demonstrated that the pairing between U6 snRNA and the 5'SS sequence needs to be disrupted after the first step of splicing in order for the second step to occur. Hyperstabilization of the U6•5'SS pairing inhibits the second step of splicing, but will recover splicing of introns impaired at the first step of splicing by a sub-optimal branch sequence. The reverse has also been demonstrated where splicing of an intron impaired at the second step by a mutant 3'SS can be recovered by a mutation at the 5'SS which reduces U6•5'SS base-pairing (Konarska et al., 2006).

Experiments from the Cheng lab showed that changes in spliceosome active site conformation between the two steps of splicing are not necessarily complex and are thermodynamically close in free energy. To show this, they isolated spliceosomes which had carried out both of the chemical steps of splicing, but were stalled before mRNA release by a mutation in the Prp22 protein (Tseng and Cheng, 2008). By adjusting the cation concentrations of the buffer, they were able to reverse both chemical steps of splicing. Incubation of purified

post-catalytic spliceosomes in the absence of KCl caused them to reverse the first step of splicing. Subsequent addition of KCl caused a fraction of spliceosomes to catalyze the forward second step of splicing, while some catalyzed the reverse of the first step of splicing. This result implies that there are two different active site conformations in competition with each other. The splicing equilibrium was also sensitive to divalent metal ions (Mg^{2+} and Mn^{2+}). Interestingly, Mn^{2+} also had the effect of increasing the error rate in the forward reaction of the second step of splicing.

This proposed mechanism provides a rationale for the evolution of proteins that would help the spliceosome proceed through the reaction coordinate of the splicing reaction (Smith and Konarska, 2008). Indeed several conserved, essential ATPases are required for spliceosome function: notable examples are Prp2, required before the first step of splicing, Prp16 which is important for the first and second step of splicing, and Prp22 which is required for the second step and for spliceosome disassembly after the splicing reaction (Koodathingal et al., 2010; Lardelli et al., 2010).

Further experiments have shown that mutations in these spliceosomal ATPases as well as in PRP8 can affect the equilibrium between the first and second steps of splicing (Liu et al., 2007 and references therein). Mutations in PRP8 can be generally classed as first- or second-step alleles (Table 5-1), where a first-step allele will recover splicing for introns impaired at the first step of splicing; conversely, second-step alleles will recover splicing of introns which are impaired at the second step (Figure 5-1). The effect of combining a first- and

Table 5-1. Mutant PRP8 alleles mapping to domain IV core and corresponding phenotypes.

Yeast Allele (Human)	Suppress Mutations in						Allele Class	
	5'SS	3'SS	BPS	PPT	U6	U4cs	1 st Step	2 nd Step
F1834S/L (Y1762)				X				
F1851L (F1779)						X		
V1860N (V1788)						X		
V1860D (V1788)						X	c	
T1861P (T1789)	X	X				X		X
V1862D (I1790)						X	-	-
V1862A/Y (I1790)						X		
H1863E (H1791)		X	X					X
K1864E (K1792)	X	X						
N1869D (N1797)	X	X	X		X			X
V1870N (L1798)	X	X	X					X
I1875T (I1803)						X		
E1960K/G (E1888)				X			X	
T1982A (T1910)	X	X	X					X
V1987A (T1915)	X	X	X					X

Amino acid numbering is for the *S. cerevisiae* protein (human amino acid number shown in brackets). 5'SS = 5' splice site mutation suppression, 3'SS = 3' splice site mutation suppression, PPT = loss of preference for a 3' splice site with a polypyrimidine tract, BPS = suppression of branch point sequence mutation, U4cs = suppression of mutations which hyperstabilize the U4/U6 snRNA structure, U6 = suppression of a mutation which hyperstabilizes U6 snRNA. "X" = displays this phenotype "-" = demonstrated to lack this phenotype, "c" = confirmed in this thesis (chapter 5), Based on Grainger and Beggs, 2005, Liu et al., 2007, and Yang et al., 2008.

second-step mutation in PRP8 serves to neutralize the effect of each individual mutation, producing a wild type phenotype, suggesting that they affect a single transition between the first-step and second-step equilibrium. Mutations in PRP16 inhibit the transition from the first to the second step of splicing and are therefore classed as first-step alleles, while mutations in PRP22 inhibit the transition out of the second step, thus suppressing defects of the second step. Although a double mutation of PRP8 with both a first- and second-step allele leads to a phenotype which is essentially wild type, the combination of a first-step PRP16 allele with a PRP8 second-step allele serves to suppress defects of both

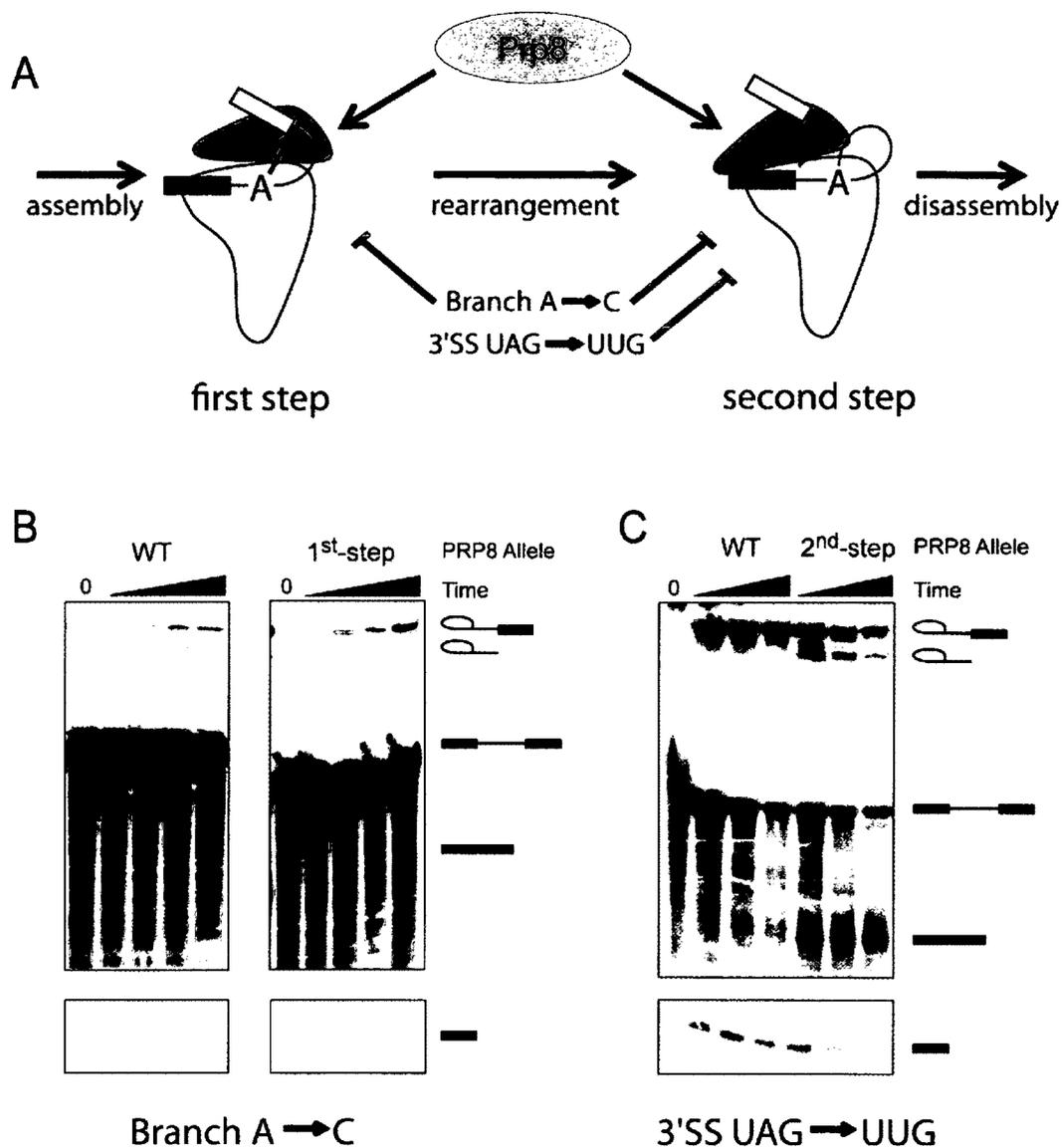


Figure 5-1. First- and second-step alleles of PRP8. A) Two-step model of spliceosomal catalysis. Two distinct conformations of the spliceosome catalyze the two steps of the splicing reaction. Mutations in a pre-mRNA will impair either or both of these steps. The branch-A to C mutation interferes with both steps of splicing and the 3'SS UAG to UUG mutation only interferes with the second step. Defects in the first step of splicing can be suppressed by first-step alleles of PRP8 and defects in the second step are suppressed by second-step PRP8 alleles (adapted from Liu *et al.*, 2007). B) *in vitro* splicing of ^{32}P -labelled ACT1 pre-mRNA containing a branch A to C mutation, which impairs both steps of splicing in yeast whole cell extract containing wild type PRP8. RNA products were separated on a denaturing PAGE and analyzed by autoradiography. RNA products are indicated diagrammatically on the right side of the gel, from top to bottom: lariat-intermediate (1st step), lariat product (2nd step), pre-mRNA (substrate), mRNA (2nd step), and 5' exon (1st step). Splicing in extracts containing a 1st-step allele of PRP8 (E1960K) show an improvement in the efficiency of the first step. C) *in vitro* splicing of ^{32}P -labelled ACT1 pre-mRNA containing a 3'SS UAG to UUG mutation, which impairs the second step of splicing in yeast whole cell extract containing wild type PRP8. Splicing in extracts containing a 2nd-step allele of PRP8 (V1870N) show an improvement in the efficiency of the second step.

steps of splicing, suggesting that these two proteins function differently between the two steps of splicing (Liu et al. 2007).

This proposal of two active site conformations explains why mutations of so many different amino acids of PRP8 are able to suppress mutations at multiple positions in a pre-mRNA, as after their initial identification, many alleles were shown to be able to suppress mutations at several sites in the spliceosome, including those within the snRNAs themselves (Table 5-1). By affecting the equilibrium between the two active site conformations corresponding to the two steps of splicing, PRP8 suppresses the effect of mutations that cause a defect in either step of splicing. The next two chapters of this thesis will investigate how domain IV of Prp8 is capable of performing this function.

5-1.2. C-terminal region of Prp8 at the spliceosome active site

The previous chapter of this thesis reported the X-ray structure of human domain IV of Prp8 (herein referred to as the PRP8 RNase H domain). The original crystals of the human Prp8 domain IV (amino acids 1769-1990) featured two monomers in the asymmetric unit, with the prominent features being the presence of a C-terminal cluster of five helices and an N-terminal RNase H fold (Figure 4-1B). This latter structural homology was not predicted by primary sequence analysis due to a seventeen amino acid insertion (amino acids 1787-1803) between two β -strands of the RNase H-like domain. In "monomer a" of the unit cell, this insertion was well structured forming a two-stranded anti-parallel β -hairpin while in "monomer b" this region was disordered as a displaced loop. The

translation of this loop back $\sim 45^\circ$ pulls aa 1782-1784 to extend the $\beta 1$ strand of the RNase H fold. This movement is probably related to the disruption/rearrangement of $\alpha 1$ and part of $\alpha 2$; these residues now partially fill the space formerly occupied by the base of the finger (Figure 5-2A).

Intriguingly, the largest proportion of suppressor alleles associated with the PRP8 RNase H domain map to the two-stranded β -hairpin (Figure 4-3; Grainger and Beggs, 2005). This suggests that the disposition of this structure is critical to the function of this domain within the spliceosome.

5-2. Results and Discussion

5-2.1. A conformational change unmasks a Mg^{2+} binding site

The presence of an RNase H fold within a domain of PRP8 associated with the individual catalytic steps of splicing is provocative since RNase H is a metalloenzyme promoting chemistry on an RNA substrate. Catalysis of RNA cleavage by RNase H like enzymes involves a two metal mechanism in which divalent magnesium ions, bound at adjacent sites separated by $\sim 4 \text{ \AA}$, promote hydrolysis by activation of a water nucleophile combined with transition state stabilization (Nowotny et al., 2005; Steitz and Steitz, 1993). Inspection of the domain IV Prp8 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. Despite the fact that our original crystals were grown in 200 mM $MgCl_2$, we did not detect a metal ion in the Prp8 RNase H domain site. Instead, the

spatially conserved aspartates are stabilized by a network of water molecules and the side chain of Arg1865 is hydrogen-bonded to Asp1782 possibly blocking the metal binding site (Figure 4-5A). This has previously been observed in the RNase H fold of the Argonaute PIWI domain and also in the Tn5 transposase (Figure 4-5B).

If R1865 was blocking a canonical RNase H-like metal-binding site of Prp8, this site might be unmasked by displacement of this side chain, for example by RNA binding during spliceosome assembly/activation. In order to model this, we expressed, crystallized, and solved the 1.15Å X-ray structure of the R1865A hPrp8 mutant which mimics a displacement of R1865 (Table 5-2). This structure is remarkably similar to that of the wild type protein including two monomers in the asymmetric unit but with the disrupted β -hairpin of monomer b significantly more ordered. One consequence of the structural rearrangement involved in the conformation a to b transition is that D1782 moves closer to D1781 and T1783 is displaced ~ 4 Å upwards (Figure 5-2B). Significantly, within monomer b of this structure, a single Mg^{2+} ion bound at high occupancy is observed in the canonical RNase H site. Coordination of this ion includes inner sphere contact with the side chain of D1781 and five waters with bond lengths of ~ 2.1 Å in an octahedral arrangement. The waters are hydrogen bonded to the carboxylate of D1782, carbonyls of D1782 and L1891, the amide carbonyl of Q1894, and the hydroxyl of T1864 (Figure 5-2C,D). We confirmed assignment of the Mg^{2+} ion by soaking the R1865A crystals in $CoCl_2$ and observing strong anomalous diffraction from Co^{2+} bound in the RNase H site (Figure 5-2D, table 5-2). These crystals were

Table 5-2. Crystallographic data collection and refinement statistics

Prp8 Mutant	R1865A	R1865A	WT	V1788D	T1789P/R1865A	T1789P	T1789P + WT mix	N1797D	L1798N	D1781C	D1781C/R1865A
Divalent Cation	Mg ²⁺ (300 mM)	Co ²⁺ (100mM)	Co ²⁺ (100mM)	Mg ²⁺ (200 mM)	Mg ²⁺ (300 mM)						
Data collection											
X-ray Source	CMCF-1	CMCF-1	CMCF-1	ALS 8.3.1	ALS 8.3.1	ALS 8.3.1	ALS 8.3.1	CMCF-1	Cu anode	CMCF-1	ALS 8.3.1
Space group	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁
Cell dimensions											
a, b, c (Å)	76.20, 78.04, 94.06	76.25, 78.22, 94.37	76.23, 78.29, 94.09	75.07, 78.24, 93.57	76.17, 77.25, 93.60	76.18, 77.08, 93.04	76.30, 77.95, 94.19	76.01, 78.08, 93.86	76.19, 77.78, 93.48	75.90, 77.87, 93.97	75.90, 77.87, 93.97
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Wavelength	0.97949	0.97949	0.97949	1.115879	1.1158	1.1158	1.1158	0.97949	1.5418	1.89223	1.28282
Resolution (Å)	1.15	1.32	1.5	2	1.95	1.65	1.8	1.45	2.2	2.3	1.7
R_{sym} or R_{merge}^1	0.038 (0.482)	0.045 (0.390)	0.067 (0.447)	0.074 (0.290)	0.054 (0.528)	0.049 (0.607)	0.053 (0.472)	0.049 (0.261)	0.054 (0.441)	0.046 (0.080)	0.040 (0.535)
$I / \sigma I^2$	26.7 (2.5)	23.4 (2.1)	16.4 (2.9)	14.6 (5.1)	24.6 (2.6)	25.3 (2.2)	23.4 (2.7)	28.7 (3.8)	25.2 (3.3)	21.1 (10.7)	41.1 (3.0)
Completeness (%) ¹	98.9 (96.6)	95.2 (85.3)	88.1 (96.3)	99.9 (100)	94.3 (99.8)	95.0 (97.8)	94.9 (96.8)	98.6 (90.4)	97.5 (99.3)	96.7 (81.1)	99.7 (100)
Redundancy	4.3	4.3	3.6	4	4	3.9	3.9	5	4.2	3.4	7
Refinement											
Resolution (Å)	60-1.15	60-1.32	60-1.5	47-2.0	60-1.95	59-1.65	60-1.8	60-1.45	60-2.2	60-2.3	60-1.7
No. reflections	186162	119605	75859	35887	36295	60115	46079	93120	26765	22868	58928
$R_{\text{work}} / R_{\text{free}}$	0.145/0.164	0.148/0.174	0.164/0.205	0.230/0.281	0.196/0.244	0.178/0.227	0.201/0.243	0.153/0.175	0.173/0.223	0.197/0.251	0.166/0.206
No. atoms											
Protein	3777	3783	3717	3468	3459	3509	3500	3645	3557	3489	3489
Ligand/ion	11	10	16	0	1	1	1	11	0	1	7
Water	707	661	535	191	478	451	505	614	319	190	580
B -factors											
Protein	19.1	18.2	18.3	21.8	29.5	23.8	23.1	21.1	36.6	30.1	26.9
Ligand/ion	30.9	26.2	24.6	na	32.7	38.9	23.1	31.8	na	37.4	43.7
Water	22.2	21.4	24.9	23.2	33.9	24.9	25.9	25.1	39.6	35.2	31.2
R.m.s deviations											
Bond lengths (Å)	0.025	0.22	0.022	0.015	0.018	0.022	0.022	0.018	0.02	0.015	0.015
Bond angles (°)	1.99	1.84	1.81	1.34	1.51	1.97	1.97	1.63	1.68	1.48	1.48
	¹ Values in parenthesis correspond to highest resolution shell										

grown in 300 mM MgCl₂, which is higher than the conditions under which wild type protein crystals were grown. Therefore, we soaked R1865A crystals in decreasing concentrations of MgCl₂ before data collection. We observed the Mg²⁺ ion in concentrations of MgCl₂ as low as 50 mM, but not in 20 mM MgCl₂. This apparent K_d is higher than the physiological concentration under which pre-mRNA splicing occurs (~2mM), but additional Mg²⁺ ligands (ie. phosphodiester RNA backbone) in the spliceosomal context likely serve to augment Mg²⁺ ion occupancy at this site.

Interestingly, under new crystallization conditions, we were ultimately able to observe partial occupancy of this site by Mg²⁺, Mn²⁺, or Co²⁺ exclusively in monomer b of wild type protein (Table 5-2 for Co²⁺). Repeating the same Mg²⁺ titration for crystals before data collection shows no Mg²⁺ ion is bound at 50mM MgCl₂, and even as high as 1 M MgCl₂, less than a full occupancy of the Mg²⁺ site is observed. The weaker Mg²⁺ association may be explained by the proximity of the R1865 positive charge. It thus appears that it is the larger conformational change observed between a and b that is required for binding of divalent metal to the Prp8 RNase H site, although rearrangement of R1865 likely contributes to enhanced metal- binding within the context of the spliceosome.

5-2.2. Disruption of Mg²⁺ ion binding by Prp8 Domain IV impairs splicing

Luhrmann, Zhao, and coworkers have carried out extensive mutagenesis of the RNase H domain metal-binding site and shown that mutation of the residues shown here to be metal coordinating is deleterious in yeast (Pena et al.,

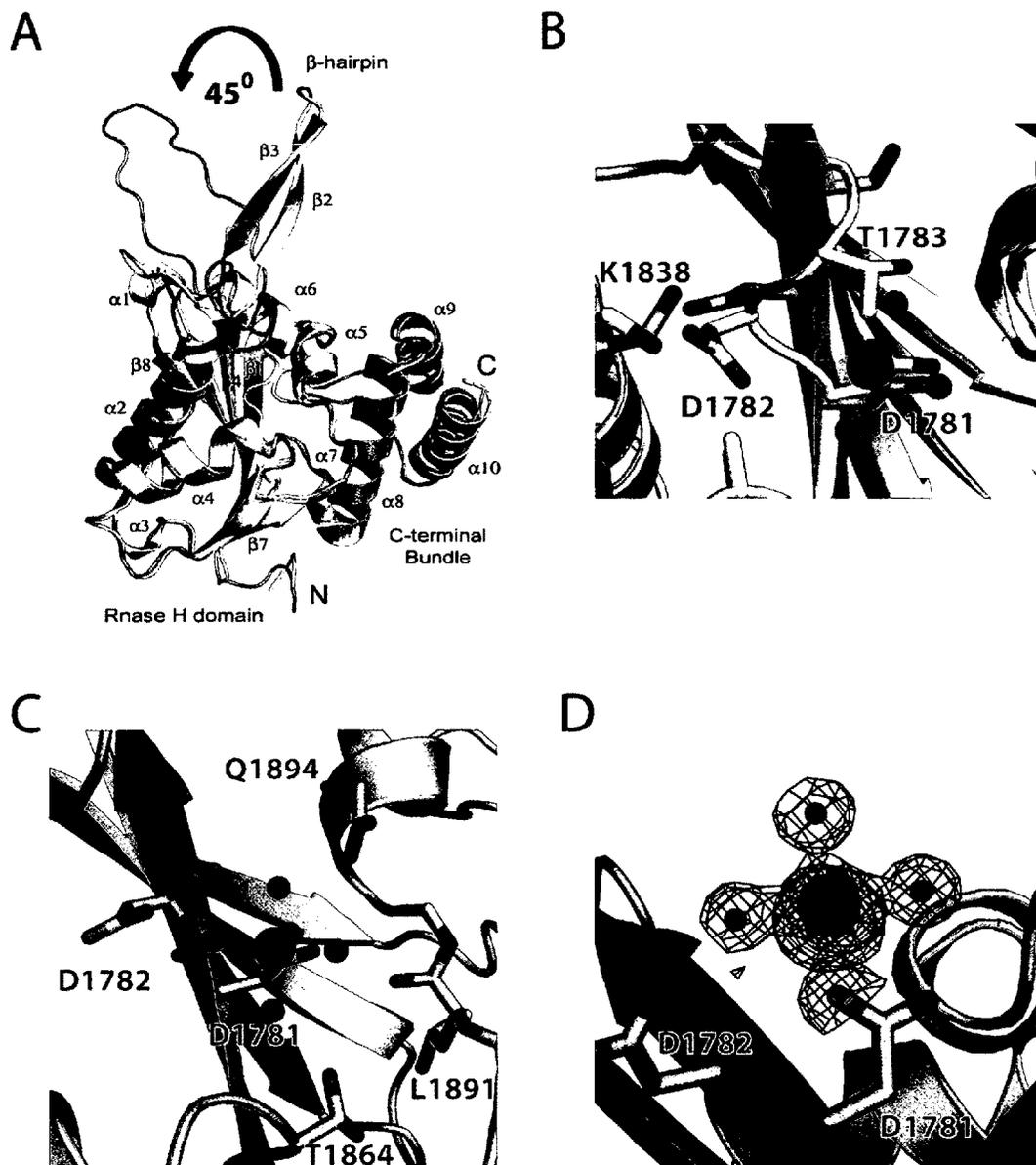


Figure 5-2. Conformational plasticity in the PRP8 RNase H domain. A) Overlay of two monomers from the asymmetric unit of the PRP8 RNase H domain crystal structure. In monomer b (blue), the β -hairpin is disrupted and translated $\sim 45^\circ$, the $\alpha 2$ helix shortened and $\beta 1$ strand extended. B) Displacement of T1783 upon conformational rearrangement facilitates binding of Mg^{2+} (purple) and its inner-sphere waters (green) at the canonical site. Shown is overlay of monomer a (yellow) and monomer b (cyan). C) Detail of coordination of Mg^{2+} ion (purple) and its inner-sphere waters (green) in monomer b of Prp8 Domain IV. D) Detail of 1.15 Å resolution 2Fo-Fc map contoured at 1.5 σ (blue) showing octahedral coordination of Mg^{2+} bound in monomer b. Map of anomalous scattering from Co^{2+} soak contoured at 8 σ is superimposed in red.

2008; Yang et al., 2008). For example, mutation of the outer sphere ligand, D1854A (human D1782A), results in severe growth defects at both 16° and 37° while the mutation of the inner sphere D1853 to alanine is lethal (Figure 4-5B). In order to further investigate the role of metal-binding in Prp8 function, we mutated the inner sphere ligand D1853 to cysteine in yeast PRP8 based on the expectation that this mutation should abrogate Mg²⁺ coordination because the softer ligand strongly disfavours binding a hard metal ion.

The D1853C yeast strain grows at ~half the rate of wild type at 30°C and manifests a severe growth defect at 37°C. We examined *in vitro* splicing of an ACT1 pre-mRNA comparing extracts prepared from wild type and PRP8 D1853C mutant yeast strains. The D1853C mutant extract forms spliceosomal complexes at a comparable rate to wild type. However, it is mildly impaired for the first step of splicing and exhibits a strong defect in the second step (Figure 5-3A). Attempted rescue of the second step in the presence of the thiophilic divalent manganese ion was unsuccessful, in contrast to Mn²⁺ rescue of splicing in phosphorthioate-modified U6 snRNA where the substitution disrupts the critical U80 Mg²⁺ binding site. However, it has been shown that recovery of enzymatic activity where Mg²⁺ protein ligands (aspartate or glutamate) have been mutated to cysteine is not always possible (Gao et al., 2004).

The impairment in function of PRP8 by D1853C is not due to any gross structural change. We solved the structure of the corresponding mutation in the human protein, D1871C, and a D1871C/R1865A double mutant Prp8 RNase H domain and the architecture of the metal binding site is preserved (Table 5-2). As

expected, Mg^{2+} does not bind to this site; however in contrast to the wild type protein, we were unable to observe bound Mn^{2+} following crystal soaks.

It has long been believed that the spliceosome is in essence a ribozyme based, in part, on the mechanistic similarity between self-splicing group II introns and the processing of nuclear pre-mRNAs. There is also considerable evidence for the role of the spliceosomal U2/U6 snRNA structure and particularly U6 snRNA in catalysis of splicing (Yean et al., 2000). High resolution structural analysis of the group I and group II introns has supported a general two-metal ion mechanism proposed for catalytic RNA, including the spliceosome, based on analysis of phosphoryl transfer by protein enzymes (Lipchick et al., 2008; Toor et al., 2008; Steitz and Steitz, 1993). There is nevertheless scope for the involvement of more or fewer metal ions in similar enzymatic transformations. For example, Berger and coworkers have recently proposed a mechanism for type II and IA topoisomerases where the transition state is stabilized by a single bound metal and conserved arginine (Schmidt et al., 2010).

5-2.3. PRP8 D1853C is not a first-step allele.

The model proposed by Query and Konarska has suggested that Prp8 regulates the equilibrium between two distinct spliceosomal conformations associated with the first and second transesterification steps respectively (Liu et al., 2007). Two sets of PRP8 alleles, designated as first- or second-step, are proposed to act by shifting this equilibrium to favour one step over the other. In general, these alleles do not affect splicing of a wild type pre-mRNA substrate,

however, we wished to determine if the D1853C PRP8 allele was causing a defect in splicing by acting as a strong first-step allele. To this end, we compared *in vitro* splicing of a pre-mRNA where the branch adenosine of the intron was mutated to a cytosine (BS-C). This mutation has been shown to cause a severe

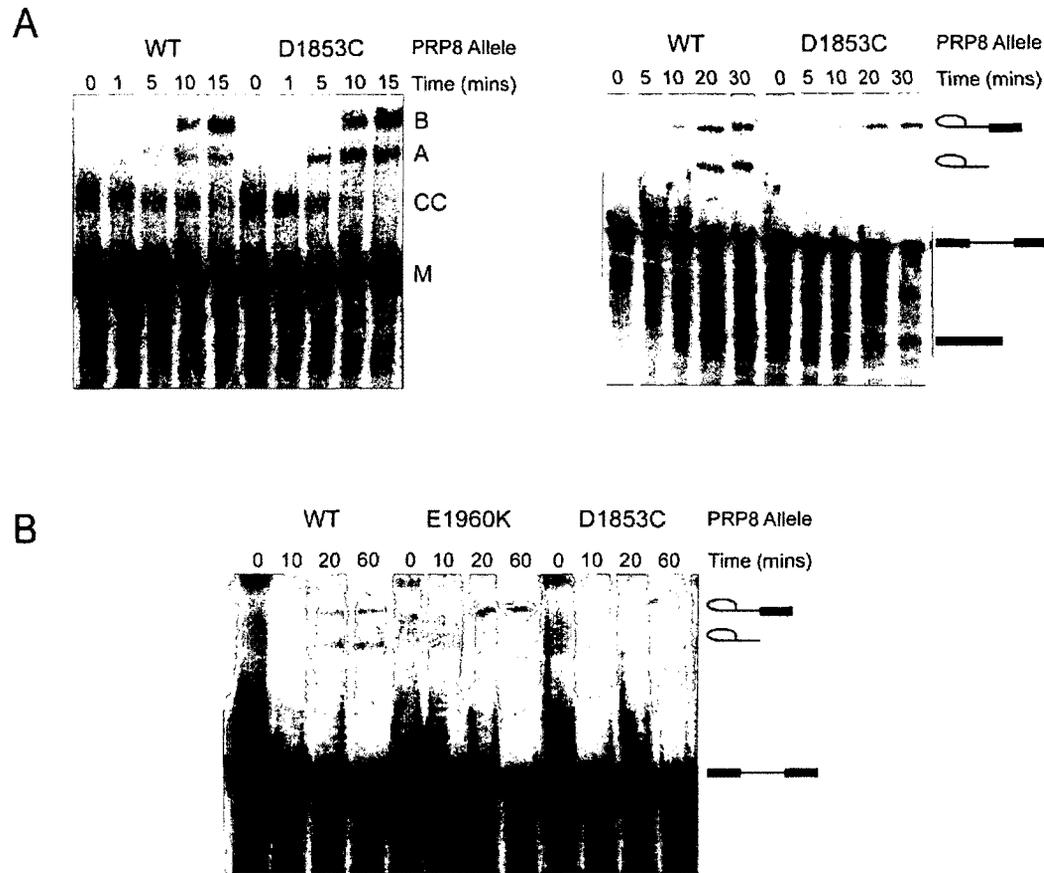


Figure 5-3. The PRP8 D1853C mutation impairs the second step of splicing. A) (left) spliceosome complex formation – D1853C mutant splicing extracts form spliceosomal complexes on the ACT1 pre-mRNA at a comparable rate to splicing extract from wild type yeast. On the right side of the gel, from bottom to top are indicated are the M – pre-mRNA, CC – commitment complex, A – A complex, and B- B complex. (right) Pre-mRNA splicing – D1853C mutant extract is mildly impaired for the first step of splicing, but severely impaired for the second step. On the right side of the gel, from bottom to top are indicated the mRNA, pre-mRNA, intron-lariat, and intron-lariat intermediate, by RNA diagrams as described in figure 1-1B. (B) Splicing of ACT1 pre-mRNA with a Cytosine branch nucleotide (BS-C) is impaired in D1853C splicing extracts. E1960K splicing extract (a first-step allele) is able to partially suppress the defect in the first step of splicing for this sub-optimal substrate. In contrast, D1853C further impairs the first step of splicing for this substrate, indicating it is not a first-step allele.

defect in both the first and second step of splicing, and it has been shown that first-step alleles of PRP8 will partially alleviate the defect in the first step of splicing. We compared splicing of this substrate in extracts from yeast containing wild type, D1853C, and E1960K (a known first-step) alleles of PRP8 (Figure 5-3B). In agreement with previous reports, E1960K improves the efficiency of the first step of splicing of this substrate, however, D1853C impairs the first step of splicing of this substrate. This indicates that the severe second-step defect caused by D1853C is not caused by stabilizing the first-step conformation of the spliceosome.

5-2.4. First- and second-step PRP8 alleles favour one of two conformations observed in the crystal structure of Prp8 RNase H domain.

Because the largest group of suppressor alleles in the RNase H domain, including those designated as first- or second-step map to the β -hairpin (Figure 4-3, Table 5-1; Grainger and Beggs, 2005; Yang et al., 2008), we endeavoured to determine whether these are specifically associated with either of the two conformations observed crystallographically. We investigated mutants of the human protein corresponding to both first- and second-step suppressor alleles that crystallized as described above with one monomer of the asymmetric unit containing the β -hairpin and the second divalent metal-binding monomer featuring the displaced loop structure.

We solved the 2.0 Å crystal structure of the human V1788D mutant corresponding to yeast V1860D (Table 5-2). In this case, compared to the wild

type (or R1865A which has a more ordered monomer b), monomer a is stabilized by an additional hydrogen-bonding interaction between the side chains of D1788 and Y1786. In monomer b, these two side chains are distant from each other, and D1788 mostly contacts the solvent and would not make any extra intramolecular hydrogen bonds (Figure 5-4A). This argues that the yeast allele favours the β -hairpin/non-metal binding conformation. The D1788 and Y1786 interaction

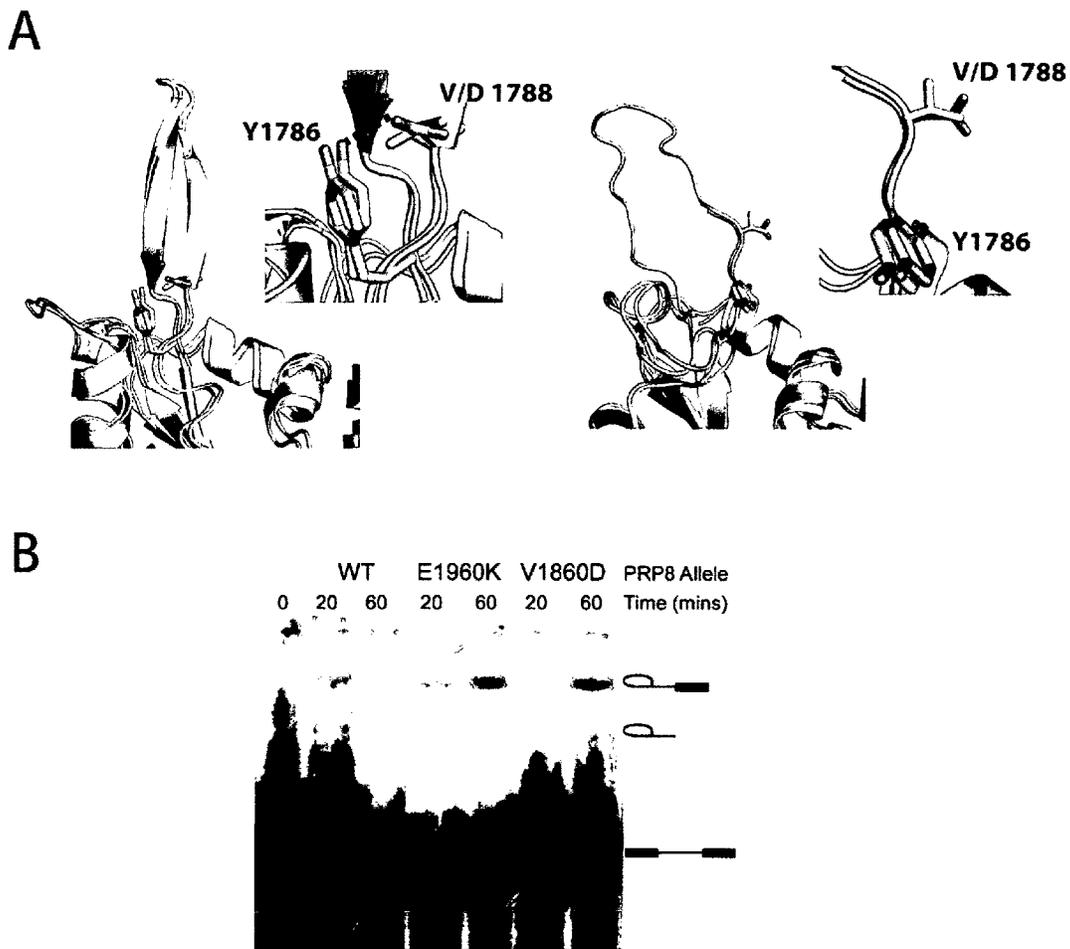


Figure 5-4. Structure and phenotype of V1860D/V1788D Prp8. A) Details of the 2 Å X-ray structure of V1788D Prp8 RNase H domain corresponding to the yeast first-step allele V1860D; shown are superimposed mutant (yellow and blue) and wild type (grey) for monomers a and b (left and right). The non-metal binding β -finger conformation of monomer a is stabilized by a hydrogen-bond formed between D1788 and Y1786. B) The first step of splicing of the BS-C reporter is enhanced by the first-step PRP8 allele E1960K, as well as the V1860D allele. On the right side of the gel, from bottom to top are indicated the pre-mRNA, intron-lariat, and intron-lariat intermediate, by RNA diagrams as described in figure 1-1B.

effectively anchors the β -strand in position, stabilizing it relative to the more extended and translated loop structure.

We examined *in vitro* splicing of extracts from yeast containing the corresponding PRP8 allele, V1860D with the mutant BS-C pre-mRNA substrate (Figure 5-4B). We observed a phenotype similar to that of E1960K, where the rate of the first step of splicing is increased. In agreement with previously published results, this clarifies that V1860D is a first-step allele, and implies that the monomer a conformation with the β -hairpin corresponds to the conformation of PRP8 in the spliceosome during the first step of splicing (Yang et al., 2008).

An examination of the sites of mutation of previously reported second-step alleles supports the proposal that monomer b represents the conformation of this domain in the spliceosome during the second step of splicing. The PRP8-201 allele contains a T1861P mutation, which corresponds to T1789 in the human Prp8 structure. In order to examine this second-step allele (Collins and Guthrie, 1999), we crystallized and solved a 1.65 Å structure of the human T1789P mutant and a 1.95 Å structure of the T1789P/R1865A double mutant (Table 5-2). Mg^{2+} ion binding is unchanged by this mutant, with a partial or full occupancy Mg^{2+} ion bound to D1781 of the respective structures. The T1789P mutation lies in the β -hairpin, and as expected, the β -hairpin is disrupted in monomer a, but the proline substitution is accommodated in metal-binding monomer b (Figure 5-5A). We performed a crystallographic mixing experiment where equal concentrations of R1865A and T1789P/R1865A double mutant protein were mixed in the

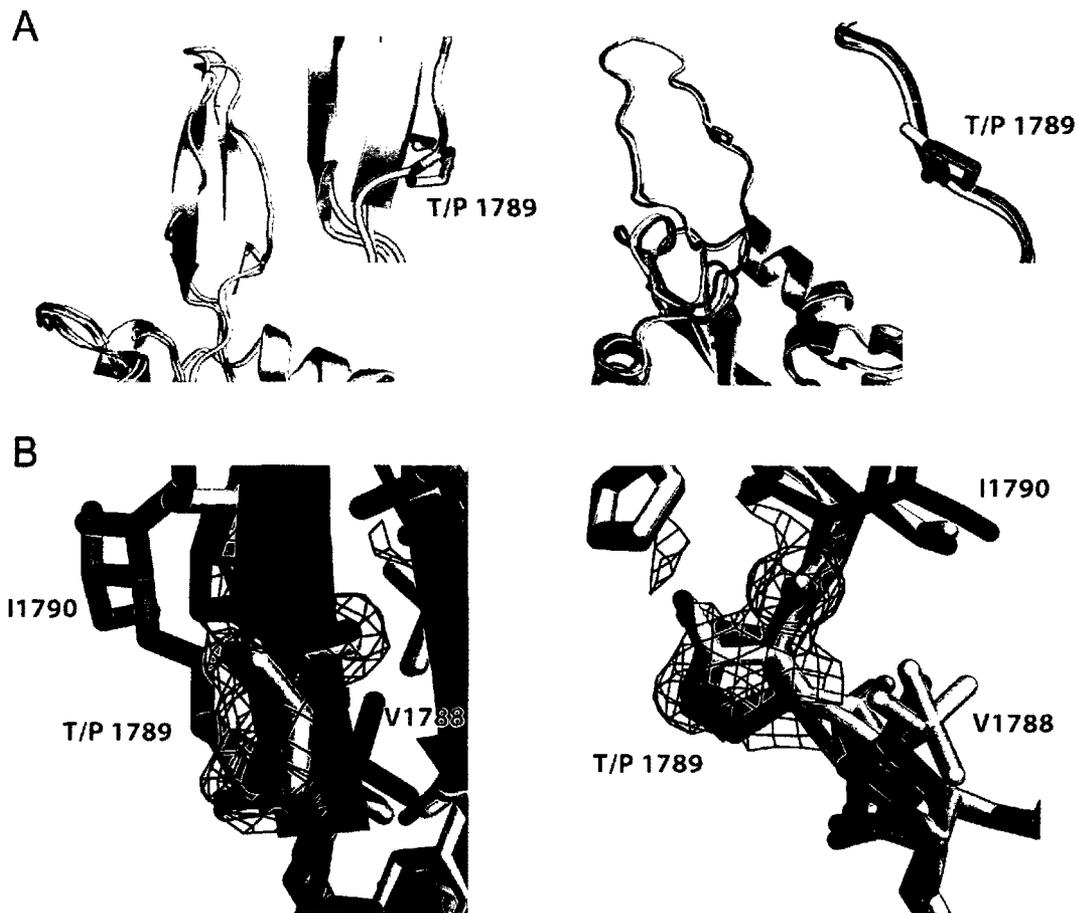


Figure 5-5. Structure of the T1789P Prp8 Domain IV. A) Details of the 1.65 Å X-ray structure of the human T1789P mutant corresponding to the yeast T1861P second-step allele. (left) The non-metal binding β -hairpin conformation of monomer a is destabilized by the introduction of proline. (right) the T1789P mutation is accommodated in the disrupted β -hairpin of monomer b. Shown are superimposed mutant (yellow and blue) and wild type (grey) for monomers a and b. B) Shown are the models for the T1789P/R1865A double mutant (cyan), R1865A (yellow) and T1789P/R1865A + R1865A mixture (green) for monomer a (left) and monomer b (right). A 2Fo-Fc map contoured at 1σ calculated from the dataset of the T1789P/R1865A + R1865A mixture crystal is shown in blue. In monomer a, the density map indicates only a threonine residue is present at position 1789, and no density is observed for the conformation of I1790 caused by P1789 mutation. In monomer b, the density map shows a mixture of threonine and proline residues at position 1789, indicating that both proteins are in this conformation in the crystal.

crystallization conditions. We chose to include the R1865A mutation in the proteins for this experiment (rather than compare wild type and T1789P single mutant proteins) because the R1865A mutation increases the order in the disrupted β -hairpin conformation of monomer b. This makes the resulting

electron density maps much clearer at the position of amino acid 1789. The 1.8 Å structure of the resulting crystal shows that only wild type protein which contains a threonine at position 1789 is visible in the β-hairpin of monomer a. In monomer b, a mixture of threonine and proline is observed for this residue (Figure 5-5B). This demonstrates the pronounced preference of this amino acid for this conformation, implies that monomer b conformation corresponds to the second-step conformation of the spliceosome.

The PRP8-162 allele contains a V1870N mutation, and displays a strong second-step phenotype (Liu et al., 2007). An analysis of the structure of the R1865A protein (where the loop corresponding to the disrupted β-hairpin is visible in monomer b) shows that the Cδ of the corresponding amino acid (L1798) is 3.1 Å away from the peptide carbonyl of a nearby amino acid (Figure 5-6A). This suggests that mutation of this amino acid to asparagine (which has a similar overall geometry) would create an additional hydrogen bond in monomer b. In monomer a, the sidechain of L1798 is not positioned near any hydrogen-bond donors or acceptors. Unfortunately, in the 1.75 Å structure of the L1798N crystal the disrupted β-hairpin conformation of monomer b is poorly ordered, making assignment of structure in this region somewhat uncertain. Density is visible for the N1798 sidechain, showing that it lies in the equivalent position to L1798, and therefore likely forms an extra hydrogen bond to the carbonyl oxygen of G1796 in this conformation. In monomer a, no extra hydrogen bonds are observed in the L1798N mutant, indicating that this mutation stabilizes the monomer b conformation (Figure 5-6A).

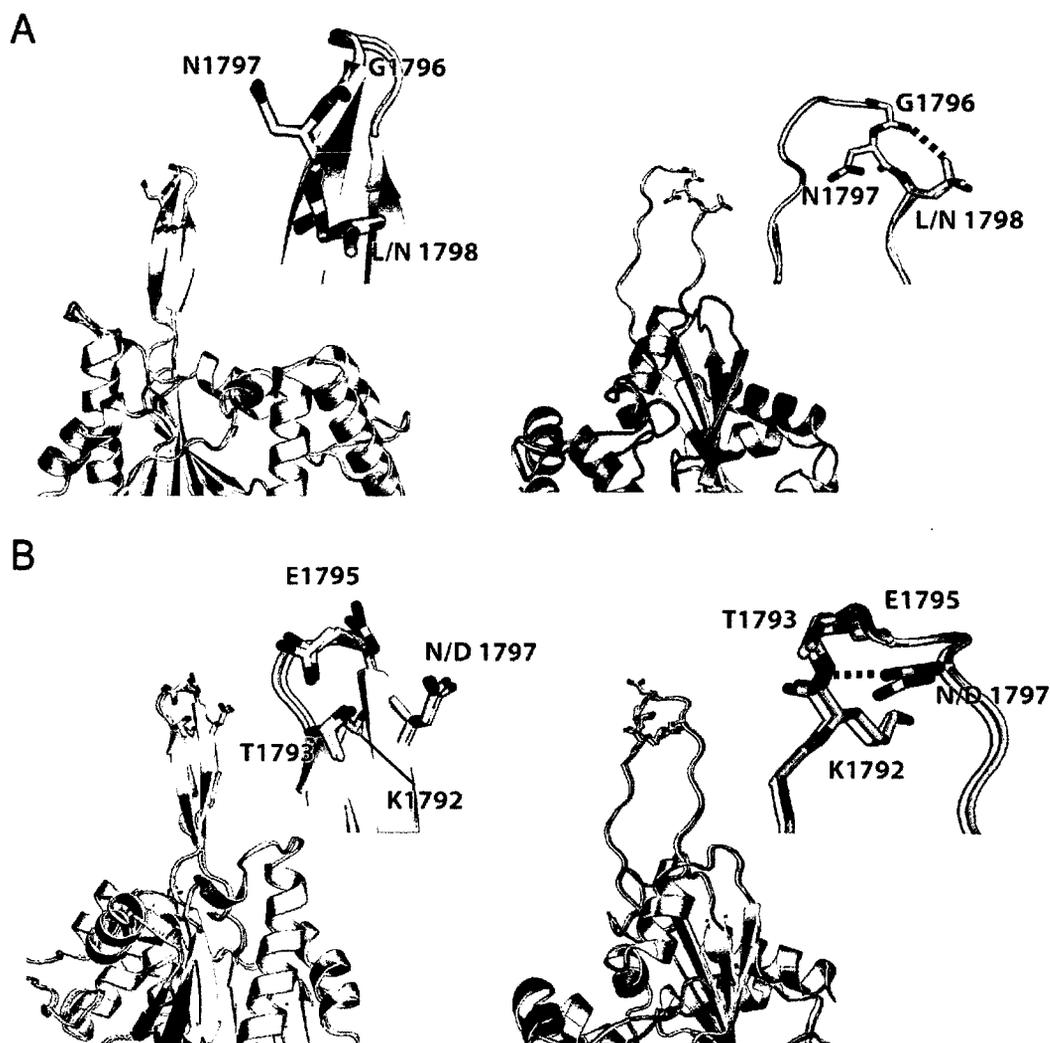


Figure 5-6. Structure of the L1798N and N1797D Prp8 Domain IV mutants. A) Details of the 2.2 Å X-ray structure of the human L1798N mutant corresponding to the yeast V1870N second-step allele. (left) The non-metal binding β -hairpin conformation of monomer a accommodates the mutation with no changes in structure. (right) The L1798N mutation forms an additional hydrogen bond to the carbonyl oxygen of G1796 in monomer b. Shown are superimposed mutant (yellow and blue) and wild type (grey) for monomers a and b (left and right). B) Details of the 1.75 Å X-ray structure of the human N1797D mutant corresponding to the yeast N1869D second-step allele. (left) The non-metal binding β -hairpin conformation of monomer a accommodates the mutation, an alternate conformation of E1795 which is closer to D1797 is no longer observed. (right) The N1797D mutation forms a hydrogen bond to the peptide nitrogen of T1793 in monomer b. This hydrogen bond is likely stronger than is observed for N1797. Consistent with this, D1797 is better positioned for this hydrogen bond than N1797. Shown are superimposed mutant (yellow and blue) and wild type (grey) for monomers a and b (left and right).

Another known PRP8 second-step allele, PRP8-151 contains a N1869D mutation (Siatecka et al., 1999; Liu et al., 2007). An analysis of the R1865A protein structure shows that the corresponding amino acid (N1797) forms a

hydrogen bond between the carbonyl of its amide group and the peptide amide of a nearby amino acid in monomer b. Mutation of this amino acid to aspartate would be expected to strengthen this hydrogen bond, by increasing the negative charge on the oxygen. Consistent with this effect, this portion of the crystal structure is much more ordered than in the wild type protein structure. In monomer a, N1797 is mostly exposed to the solvent, but is located 4.4Å from one conformation of a flexible glutamate (E1795), suggesting there may be additional charge-charge repulsion in this conformation. In the N1797D structure, E1795 is more ordered, and populates only the conformation which is further from N1797D (Figure 5-6B). The combination of these two effects would be expected to stabilize the monomer b conformation relative to monomer a.

5-2.5. The Prp8 Domain IV conformation change in the spliceosome

In the studies described here, we have observed crystallographically two distinct conformations of the Prp8 RNase H domain, one of which permits coordination of Mg²⁺ at the conserved RNase H metal-binding site. We have shown that several first- or second-step alleles form extra hydrogen bonds or disrupting one conformation of the Prp8 RNase H domain, and therefore we propose that this domain undergoes a similar conformational change between the two steps of splicing. The structures of proteins corresponding to first- and second-step alleles argue that the non-metal binding conformation corresponds to a first-step conformation and link divalent metal coordination by Prp8 to the

second step of splicing. The latter hypothesis is supported by the results of mutagenesis of the inner-sphere ligand D1853 in yeast.

The transition between first- and second-step states in the spliceosome is more complex than that described here as evidenced by domain IV alleles located outside the region involved with the conformational change, as well as those associated with domain 3 of PRP8 (Grainger and Beggs, 2005; Liu et al., 2007). Nevertheless, the conformational change described here is most likely a representation of the switch between the two steps of splicing. The fact that it is coupled to binding of a required metal ion implicates it as a key regulatory mechanism in catalysis of the second step of splicing.

The conformation change of Prp8 shown in this structure allows the assessment of additional mutants which have yet to be classified as first- or second-step alleles. It also provides a starting point for the design of additional mutants which will not only allow further interrogation of the spliceosome, but be useful for stalling spliceosomes at certain stages of the splicing reaction for other biochemical experiments. It also provides a starting point from which to investigate the conformational change of the spliceosome between the two steps of splicing. Furthermore, because many of the U4/U6 *cs* alleles are not the same as first- or second-step alleles, this implies the Prp8 RNase H domain has multiple functions in the spliceosome. It remains to be shown how Prp8 can accomplish this other function, and it will be exciting to see whether the same or perhaps another conformational change is involved in U4/U6 unwinding.

5-3. Materials and Methods

5-3.1. Protein expression, purification, and crystallization.

All mutants were generated using the overlapping PCR method. Cloning, expression, and purification of all mutant proteins was done as described in section 4-3.1. Crystals were grown using the procedure described in section 4-3.2, but with a different precipitant solution (10-14% PEG 4000, 100 mM Tris pH 8.0, and 300 mM MgCl₂), and flash-cooled in the same solution with 16% glycerol for data collection. For soaks with other metals, the crystals were soaked in the same precipitant, with the MgCl₂ replaced with 100 mM of the indicated metal chloride salt for 1 hour prior to soaking and flash-cooling in the same solution with 16% added glycerol.

5-3.2. Data collection and processing

Data were collected at beam line 8.3.1 of the Advanced Light Source, Lawrence Berkeley National Laboratory, beamline CMCF-1 at the Canadian Light Source, University of Saskatchewan, Saskatoon, and on an R-Axis rotating Cu anode X-ray source. For each dataset, a single wavelength experiment was performed at, or near the Se K edge or Cu K edge. Data were processed and scaled using the HKL package (Otwinowski and Minor, 1997); Friedel pairs were not merged when anomalous scattering maps were required. Anomalous scattering maps were calculated using the CCP4 program suite (Potterton et al., 2003).

5-3.3. Model building and refinement

The structures were solved using molecular replacement using the program REFMAC (Murshudov et al., 1997) to refine using PDB id 3ENB as a starting model. Water molecules were built using the program ARPWARP (Terwilliger, 2002). Iterative cycles of refinement in REFMAC against the merged dataset and manual model building using COOT was used to complete and refine the models (Emsley and Cowtan, 2004). Refinement statistics for all structure from this chapter are summarized in Table 5-2.

5-3.4. Creation of mutant *S. cerevisiae* strains splicing extracts, and splicing

Mutant PRP8 genes were created using gap repair of plasmid pJU186 (Umen and Guthrie, 1995), which contains a HIS selectable marker and PRP8 gene. Mutants plasmids were transformed into strain JDY8.06 (*ura3-52, leu2-3,-112, ade2, his3-A1, trp1-289, prp8::LEU2*, pJU169(*PRP8, URA3, CEN, ARS*), kindly provided by Richard Grainger and Jean Beggs, University of Edinburgh, UK), containing wild type PRP8 on a counter-selectable URA3-marked plasmid (Brown and Beggs, 1992). Before plasmid shuffling (Boeke et al, 1987), cells were selected by growth at 30°C in a medium lacking histidine and leucine for 16 hours. Transformants were streaked once on medium which lacked histidine and leucine, but contained 5-fluoroorotic acid (5-FOA) to select for cells lacking the URA3 plasmid. Cells that survived on 5-FOA plates were grown in media lacking histidine and total DNA was extracted using a DNeasy kit (Qiagen), and all mutant PRP8 strains were verified by sequencing. Splicing extract was prepared

from wild type and mutant yeasts as described (Umen and Guthrie, 1995). Splicing reactions were done as described (Lin et al., 1985), and separated on a 7%, 19:1 Acrylamide:Bis-acrylamide, 8 M Urea PAGE, exposed to a phosphor storage screen, and scanned using a phosphorimager (Molecular Dynamics).

5-4. References

Boeke, J.D., Trueheart, J., Natsoulis, G., Fink, G.R. (1987) 5-Fluoroorotic acid as a selective agent in yeast molecular genetics. *Met. Enzymol.* **154**, 164-175.

Brown, J.D., Beggs, J.D. (1992) Roles of PRP8 protein in the assembly of splicing complexes. *EMBO J.* **11**, 3721-3729.

Collins, C.A., 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.

Emsley, P., Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Cryst. .D* **60**, 2126-2132.

Gao, K., Wong, S., Bushman, F. (2004) Metal Binding by the D,DX35E Motif of Human Immunodeficiency Virus Type 1 Integrase: Selective Rescue of Cys Substitutions by Mn²⁺ In Vitro. *J Virol.* **78**, 6715-6722.

Grainger, R.J., and Beggs, J.D. (2005). Prp8 protein: at the heart of the spliceosome. *RNA* **11**, 533-557.

Koodathingal, P., Novak, T., Piccirilli, J.A., Staley, J.P. (2010) The DEAH Box ATPases Prp16 and Prp43 Cooperate to Proofread 5' Splice Site Cleavage during Pre-mRNA Splicing. *Mol Cell.* **39**, 385-395.

Konarska M.M., Vilardell J., Query C.C. (2006) Repositioning of the Reaction Intermediate within the Catalytic Center of the Spliceosome. *Mol. Cell.* **21**, 543-553.

Lardelli R.M., Thompson J.X., Yates J.R. 3rd, Stevens S.W. (2010) Release of SF3 from the intron branchpoint activates the first step of pre-mRNA splicing. *RNA*, **16**, 516-28.

- Lin, R.J., Newman, A.J., Cheng, S.C., Abelson, J. (1985) Yeast mRNA splicing in vitro. *J Biol. Chem.* **260**, 14780-14792.
- Lipchock, S.V., Strobel, S.A. (2008) A relaxed active site after exon ligation by the group I intron. *Proc. Natl. Acad. Sci.* 2008 **105**, 5699-5704.
- 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.
- Moore M.J., Sharp P.A. (1993) Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. *Nature.* **365**, 364-368.
- 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., 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.
- Podar M, Perlman PS, Padgett RA. (1998) The two steps of group II intron self-splicing are mechanistically distinguishable. *RNA* **4**, 890-900.
- Potterton, E., Briggs, P., Turkenburg, M., Dodson E. (2003) A graphical user interface to the CCP4 program suite *Acta. Cryst. D* **59**, 1131-1137.
- Schmidt, B.H., Burgin, A.B., Dewese, J.E., Osheroff, N., Berger, J.M. (2010) A novel and unified two-metal mechanism for DNA cleavage by type II and IA topoisomerases. *Nature* **465**, 641-644.
- Siatecka, M., Reyes, J., Konarska, M.M. (1999) Functional interactions of Prp8 with both splice sites at the spliceosomal catalytic center. *Genes Dev.* **13**, 1983-1993.
- Smith, D.J., Konarska M.M. (2008) Mechanistic insights from reversible splicing catalysis. *RNA* **14**, 1975-1978.

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.

Terwilliger, T.C. (2002). Automated main-chain model-building by template-matching and iterative fragment extension. *Acta Crystallogr. D Biol. Crystallogr.* **59**, 38-44.

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.

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.

Yang, K., Zhang, L., Xu, T., Heroux, A. Zhao, R. (2008) Crystal structure of the β -finger domain of Prp8 reveals analogy to ribosomal proteins. *Proc. Natl. Acad. Sci. USA* **105**, 13817-13822.

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

Chapter 6

Summary and Conclusions

6-1. Summary and conclusions

6-1.1. The spliceosomal protein p14 plays an essential role in the spliceosome

Chapters 2 and 3 of this thesis describe a high resolution structure of the complex formed between p14 and a peptide from the SF3b155 protein. The strength of the interaction between these proteins can be explained by the large interface between these two polypeptides (Figure 2-3A, 2-4A,B). Therefore, this structure likely represents a stable structure which is unchanged during spliceosome assembly and activation.

The evidence presented in chapters 2 and 3 suggests p14 serves two functions in the spliceosome. The X-ray crystal structure as well as the p14•RNA model shows that p14 occludes the 2'OH of the branch adenosine (Figure 6-1), therefore necessitating the removal of the SF3b complex before the first step of splicing, which is consistent with other experiments. This suggests that p14 binds the BPS•U2 snRNA duplex to prevent any splicing chemistry until the appropriate time when the spliceosome is fully assembled. Furthermore, given that depletion of p14 protein disrupts pre-mRNA splicing, this supports the idea that p14 performs an additional, essential function such as stabilizing base-pairing between the BPS and U2 snRNA. The p14•RNA model proposed in this chapter proposes that p14 binds and stabilizes the bulged conformation of the branch adenosine. This interaction surrounds the bulged adenosine with a positively charged surface which would stabilize the BPS•U2 snRNA duplex by reducing the charge repulsion between the negatively charged RNA strands (Figure 6-1). This

function is consistent with the absence of a p14 homologue in *S. cerevisiae*, where base-pairing between the BPS and U2 snRNA is much stronger due to the high sequence conservation of the BPS. Future experiments that compare the defect of pre-mRNA splicing caused by p14 knockdown with the amount of mismatches between the BPS and U2 snRNA will demonstrate whether this is the case.

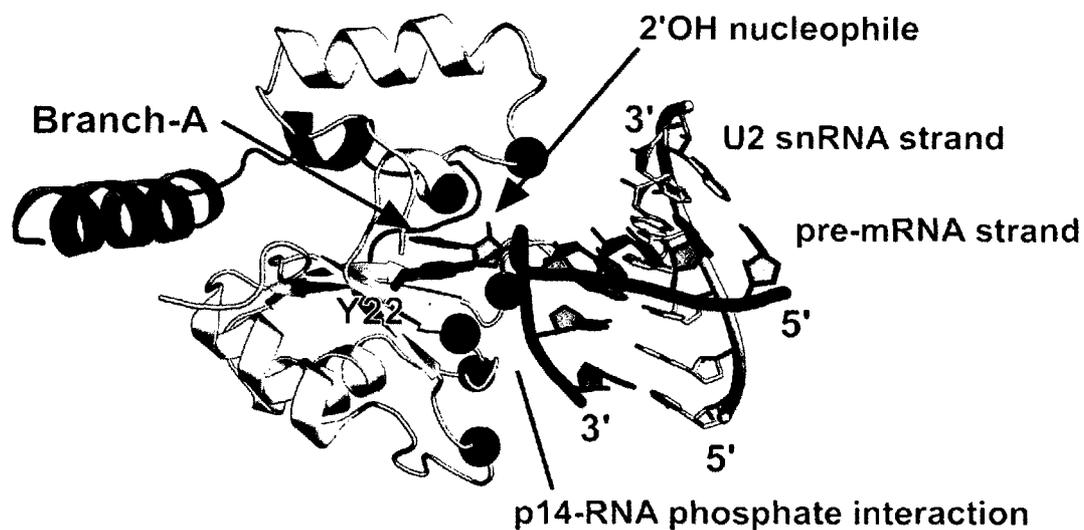


Figure 6-1. Model of p14•RNA complex. The X-ray structure of the p14•SF3b 155 peptide complex is shown bound to an RNA duplex as described in chapter 3. p14 is coloured yellow, SF3b155 in grey, the pre-mRNA RNA strand in dark green and U2 snRNA strand in light green. The disulfide tethering and SAXS data from chapter 3 describes a model where the bulged adenosine of the BPS•U2 snRNA duplex stacks on Y22 (red) of the p14•SF3b155 peptide complex. Accommodating the RNA with minimal steric clashes positions the negatively charged phosphate-backbone of the pre-mRNA strand adjacent to several conserved positively charged amino acids (C α atoms shown as blue spheres). This complex stabilizes the bulged conformation of the branch adenosine while obstructing access to the 2'OH which is the nucleophile for the first step of splicing.

6-1.2. Domain IV of Prp8

The structure of the domain IV core of Prp8 revealed an unexpected RNase H domain, which was not predicted by primary sequence analysis due to a 17 aa insertion which forms a β -hairpin that protrudes from the RNase H fold (Figure 4-1). This finding is quite provocative, as RNase H catalyzes the nucleophilic attack of a water molecule on the backbone of an RNA, resulting in strand cleavage. The spliceosome catalyzes a nucleophilic attack of a hydroxyl group on the backbone of an RNA, which also results in strand cleavage. Furthermore, the presence of an RNase H domain in Prp8 provides an excellent platform with which to interact with structures containing double-stranded RNA, and we have demonstrated that this domain preferentially interacts with complex RNA structures that contain double-stranded regions (Figure 4-7).

By overlaying the Prp8 domain IV structure and a structure of RNase H, we found that two of the acidic metal-coordinating residues of RNase H are spatially conserved in Prp8 domain IV (Figure 4-4B), and mutation of either of these to alanine in Prp8 is either lethal or severely impairs growth (Figure 4-5B), demonstrating that these metal ligands are required for splicing in yeast. The structure of the R1865A mutant Prp8 domain IV shows that a Mg^{2+} ion is bound by the two conserved aspartates, and that Mg^{2+} binding is regulated by a conformational change in Prp8. The crystal lattice of Prp8 contains two monomers which differ by a conformational change involving a separation of the β -hairpin with an $\sim 45^\circ$ tilting back of the separated strands along with a

rearrangement of $\alpha 1$ and part of $\alpha 2$ and the connecting loop region. Importantly, this change causes the sidechain of a conserved threonine to shift by $\sim 4\text{\AA}$ which unblocks the Mg^{2+} ion binding site (Figure 5-2).

The structure of several domain IV Prp8 mutants corresponding to first- or second-step alleles show that these amino acid substitutions stabilize or destabilize one of the conformations observed in the crystal lattice, and allow us to propose that the two conformations observed in our crystal lattice correspond to the conformations of this domain during the two steps of splicing (Figure 6-2).

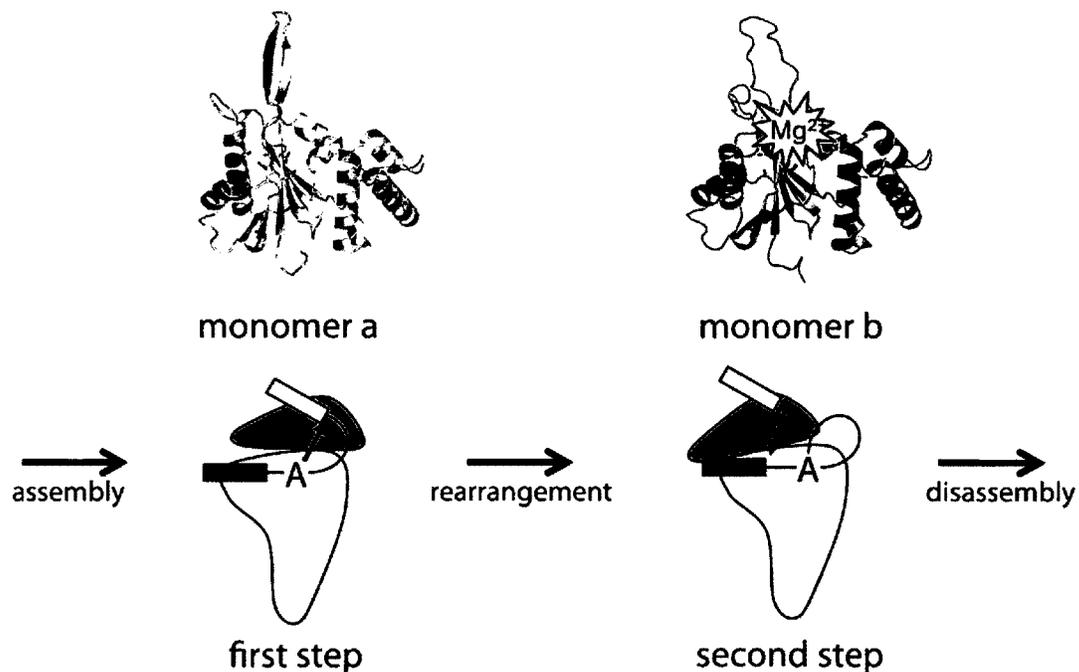


Figure 6-2. The two-step model of the spliceosome active site. The spliceosome rearranges between two conformations to catalyze the two steps of the splicing reaction. The “monomer a” conformation of Prp8 domain IV corresponds to the first-step conformation of the spliceosome, and the “monomer b” conformation of Prp8 domain IV corresponds to the second-step conformation of the spliceosome, and binds a Mg^{2+} ion which is required for the second step of splicing.

It is very interesting that disrupting the putative Mg^{2+} binding site inhibits the second step of splicing, and the structures presented in chapter 5 suggest that the Mg^{2+} binding site is only present during the second-step conformation of the

spliceosome. This is the strongest evidence presented to date that part of the active site of the spliceosome is composed of protein, and highlights the importance of additional experiments to probe the function of this domain in the spliceosome.

Context-dependent remodeling of structure in two large protein fragments⁽¹⁾

⁽¹⁾ Adapted from Schellenberg et al., (2010) *J. Mol. Biol.* **402**, 720-730. This work represents a collaboration between authors. M.S. and D.R. grew the crystals and M.S. solved the structure. M.S., T.W., K.A., and C.M. performed biochemical experiments, and M.S., C.M., L.S., and A.M.M performed the data analysis.

I-1. Introduction

The relationship between protein primary, secondary, and tertiary structural elements is a central concern in the description of protein structure. An axiom in this respect is that “the native conformation of proteins is determined by...the amino acid sequence [of a polypeptide] in a given environment” (Anfinsen, 1973). The notion that sequence determines structure (and hence function) has proven immensely useful in, for example, facilitating homology models between proteins of closely related sequence and justifying structural study or other characterization of individual domains extracted from large proteins or more complex assemblies. This has proven true even within the context of substantial amino acid substitutions or deletions in a specific primary sequence particularly with respect to non-conserved regions that may form loops or unstructured connective elements in a folded protein.

The dependence of protein structure on sequence is a function of the combination of individual chemical properties and conformational preferences over a given sequence; it is precisely this combination that represents the "environment" which is difficult to describe especially over extended sequences (Anfinsen, 1973). An analysis of these properties and preferences can result in a remarkably accurate prediction of secondary structure but the next level of structural prediction is more difficult and homology with known structures is a standard method of structure prediction. As a general rule proteins that share at least 40% sequence identity may be considered homologs that adopt the same fold in their native state (Davidson 2008). This is indicative of the robustness of a

specific protein fold even within the context of considerable sequence divergence. As an extreme case, mutation of 22 of 60 amino acids of the Msx-1 homeodomain to alanine does not compromise the fold or DNA-binding specificity of the protein (Shang et al., 1994).

A variety of investigations of both model and biological peptides/proteins have highlighted the extent to which protein structure derived from primary sequence is context dependent. "Chameleon" sequences have been described where the adoption of a specific secondary structure is highly dependent on local context. For example, Minor and Kim characterized an 11 amino acid sequence that folded as either an α -helix or β -strand of a sheet depending upon its position in the IgG-binding domain of protein G (Minor and Kim, 1996). Within a biological context, dramatic secondary structural change within a specific sequence occurs as a function of pH with hemagglutinin in a loop-helix transition that induces viral fusion with the cellular membrane (Bullough et al., 1994). The inhibitory interaction of serpins with target proteases results in significant structural rearrangement including a coil to strand transition coupled with insertion into a β -sheet (Cabrita et al., 2004).

Complicating our understanding of tertiary structure conservation based on sequence identity/homology are recent examples of proteins which share a high degree of sequence similarity and yet possess significantly different structures (Roessler et al., 2008; Alexander et al., 2007; Alexander et al., 2009). In an extreme example, two strikingly different conformations of a designed protein, a three helix bundle versus a four-stranded β -sheet with one helix, could

be interconverted by a single amino acid change (Alexander et al., 2009). Substantial misfolding or rearrangement of protein structure is characteristic of a variety of disease states and may be favoured by specific mutations in primary sequence (Ono et al., 2009; Knaus et al., 2001). There is also evidence of dynamic equilibria between radically different folded structures. Lymphotactin adopts two distinct structures, characterized by a rearrangement of all of the core hydrophobic contacts as well as a monomer to dimer conversion, each of which is reasonably populated under physiological conditions with an interconversion on the millisecond time scale (Tuinstra et al., 2008).

A conclusion that emerges from the studies described above is that protein structure is remarkably plastic even within a given sequence context and that moderate changes can result in substantially different folded structures. One aspect of protein folding that has not been extensively documented at the structural level is the extent to which substantial insertions or deletions either outside or within a particular primary sequence can alter the fold of a protein. There are several notable exceptions. For example, in an examination of insertion and deletion mutants of T4 lysozyme, Matthews and coworkers observed that the insertion of three alanines within the interdomain helix not only extended it but caused the reorganization of two orthogonal neighbouring helices into a single helix (Vetter et al., 1996). Insertion of five residues into the λ Cro sequence resulted in the formation of a β -hairpin replacing the anti-parallel β -ribbon structure that forms the dimer interface in the native protein (Mossing et al., 1990). Finally, the NFP protein is a homolog of the luciferase monomers which

feature a classic $(\beta/\alpha)_8$ barrel fold but lacking ~100 amino acids found in the latter, folds into a seven-stranded β -barrel with a novel dimerization interface (Moore et al., 1994; Fisher et al., 1996).

Here we present the crystal structure of a stable fragment of the human splicing protein Prp8, Δ D4, and compare it to our previously reported structure of the full Prp8 D4 domain that contains an additional 62 N-terminal amino acids (Ritchie et al., 2008). While most of the secondary structural elements of D4 are retained in the shorter protein, their arrangement differs strikingly at the tertiary structural level due to the absence of these N-terminal residues. The fact that such considerable differences are represented in these two fragments demonstrates the plasticity of protein structure even with respect to exon-size insertions or deletions.

I-2. Results

I-2.1. Identification, purification, and characterization of Prp8 Δ D4 and D4

As part of our efforts to characterize the structure and function of the human splicing protein Prp8, we set out to identify stable domain fragments including one representing Prp8 Domain IV (D4). We initially cloned and expressed a sub-domain, Δ D4, corresponding to amino acids 1831-1990 of human Prp8, that we identified by partial proteolysis of extended Prp8 domain fragments expressed in *E. coli*. Several interesting features of the protein were observed during its purification. Δ D4 eluted from a sizing column as a dimer, although

available evidence, including the results of cryo-electron microscopy studies of the Prp8-containing U4/U6•U5 tri-snRNP, indicates that the native protein is monomeric (Stevens et al., 2001; Hacker et al., 2008). As well, purified Δ D4 in low salt buffer precipitated on warming to room temperature but resolubilized on cooling to 4°C suggesting that it might represent a meta-stable structure. We were able to crystallize Δ D4 and solve its structure using X-ray methods (see below).

The solution properties and high-resolution structure of Δ D4 suggested that it represented a non-native fold. The largest number, though not all, of PRP8 Domain IV mutant yeast alleles corresponds to amino acid sequence immediately N-terminal to the beginning of Δ D4 (Grainger and Beggs 2005); a structural association of this sequence with Δ D4 seemed reasonable. We therefore examined a series of N-terminal Δ D4 extensions ultimately determining that the addition of 62 amino acids resulted in a fragment, D4, that we characterized as the core of native Prp8 Domain IV. The D4 domain purified as a monomer and was structurally characterized using X-ray methods (Figure I-1A).

I-2.2. Overview of Δ D4 and D4 structures

We were able to crystallize and solve the X-ray structures of both Δ D4 and D4. The 1.85 Å X-ray structure of Δ D4 involves an extended dimer with approximately 2600 Å² buried surface area (Figure I-1A; Table I-1). Overall, the structure of Δ D4 is essentially two-fold symmetric with minor variations in

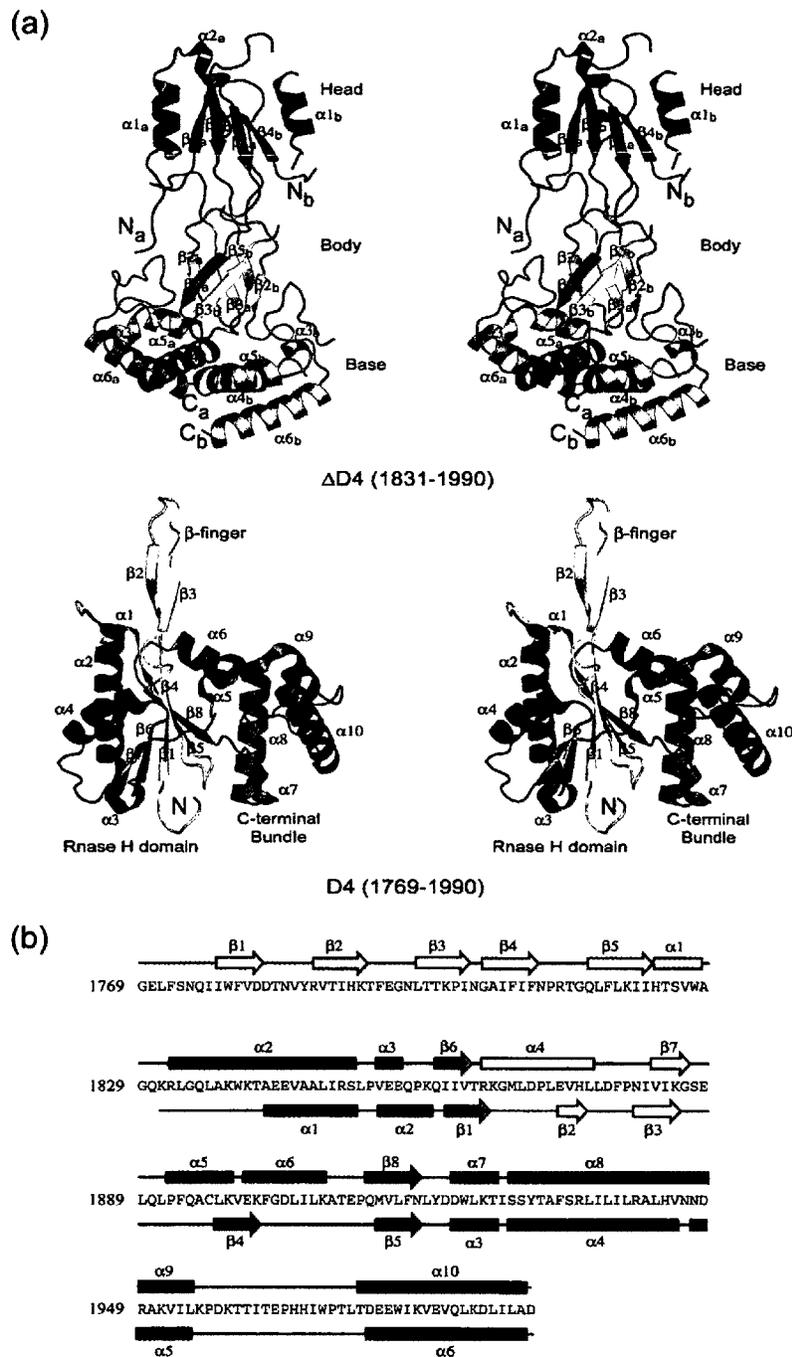


Figure I-1. Structural analysis of D4 and ΔD4. (a) Stereo views of ribbon diagrams of ΔD4 (top) and D4 (bottom) colour-coded by primary sequence. Grey represents the N-terminal 62 residues of D4 (amino acids 1769-1830) missing in ΔD4. (b) Secondary structure diagram depicting α -helices and β -strands of D4 (upper) and ΔD4 (lower). Secondary structural elements are coloured corresponding to their position in primary sequence. All numbering is based on full length human PRP8

no metal is observed in the D4 structure. This region of the protein is of great importance for Prp8 function *in vivo* as individual mutation of these amino acids to alanine is either lethal or produces a severe growth defect in yeast.

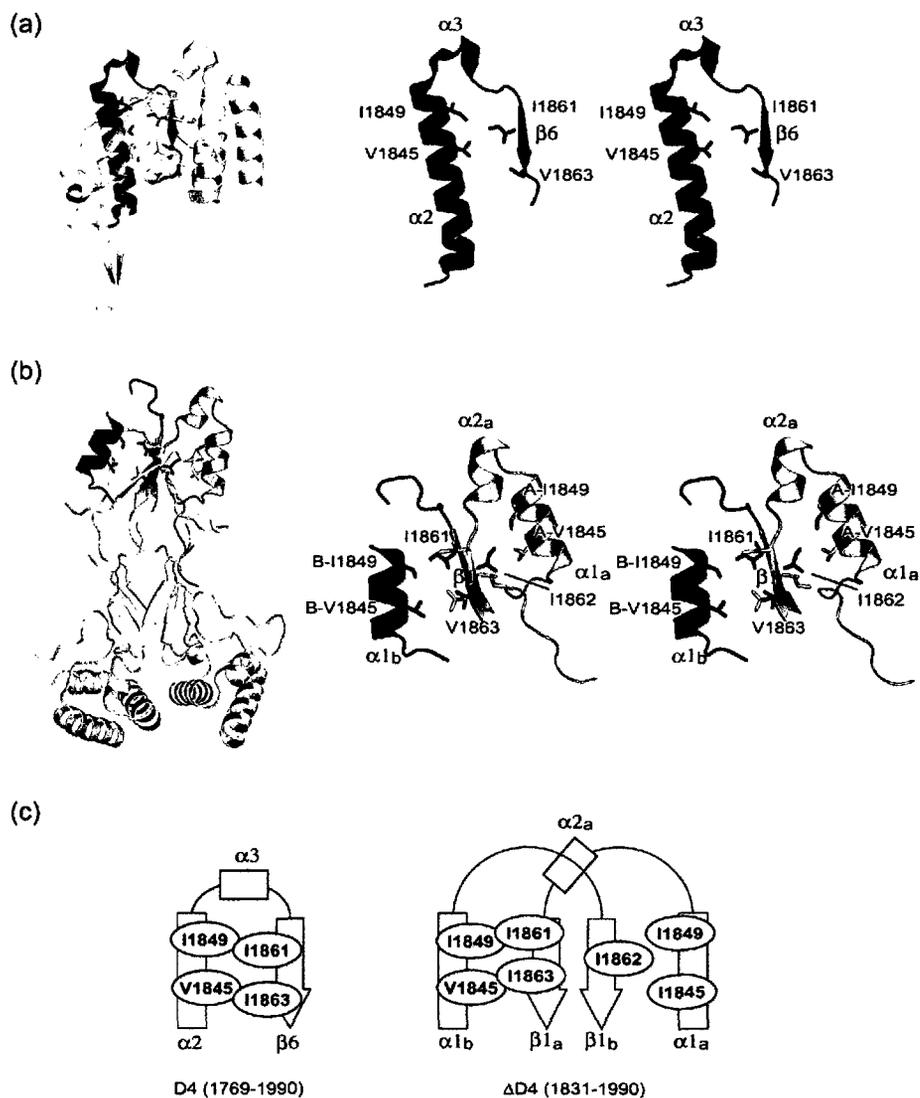


Figure I-2. Conservation of a helix-strand tertiary interaction in D4 and Δ D4. (a) (left) $\alpha 2/\alpha 3 \cdot \beta 6$ interaction within context of D4; (right) Stereo view of packing of $\alpha 2$ and $\beta 6$ in D4 involving hydrophobic interactions which include van der Waals interactions of V1845 and I1849 of $\alpha 2$ with I1861 and V1863 of $\beta 6$. (b) (left) Conserved $\alpha \beta$ packing interaction within context of Δ D4; (right) Stereo view of $\alpha \beta$ packing. The $\alpha 2/\beta 6$ interaction of D4 is conserved in one half of Δ D4 where V1845 and I1849 of $\alpha 1B$ (red) interact with I1861 and V1863 of $\beta 1A$ (pink) which involves a strand swap. In the other half of Δ D4, packing of $\alpha 1A$ V1845 and I1849 pack against I1862 of $\beta 1B$, which involves the opposite face of $\beta 1B$ because $\beta 1A$ and $\beta 1B$ are parallel. (c) Secondary structure diagram depicting partial conservation of D4 $\alpha \beta$ packing in Δ D4 through strand swap.

I-2.5. C-terminal conservation of tertiary structure in D4 and Δ D4

The structure of the C-terminal region of Δ D4 and D4 from amino acids 1912-1990 is remarkably conserved between the two structures (Figure I-3A); the regions spanning amino acids 1918-1990 are superimposable with an RMSD of 0.6 Å. Noteworthy is a strand forming edges of the β -sandwich and RNase H fold β -sheet in Δ D4 and D4 respectively which directly precedes the C-terminal bundle of four alpha helices. The disposition of this strand with respect to the helical bundle is similar in Δ D4 and D4; with amino acids 1918-1990 superimposed the angle between the two β -strands is $\sim 30^\circ$ (Figure I-3B). A hydrophobic face of D4 $\alpha 8$ mediates the packing of the helical bundle against the rest of the protein. In Δ D4, this same face is the basis of a two-fold symmetric dimerization interface with the inter-digitation of hydrophobic amino acids F1933 and H1944 mimicking the hydrophobic environment of these amino acids in the native structure (Figure I-3C).

I-2.6. Analysis of Δ D4/D4 structure and stability using CD spectroscopy

We examined both Δ D4 and D4 in solution using CD spectroscopy. This enables us to make a rough comparison of the overall secondary structural features of the two proteins; analysis of the changes in CD spectra during urea-induced denaturation allows a comparison of the relative stability of the two proteins.

The CD spectra of both proteins suggest a mixed α -helical/ β -sheet structure (Figure I-4A). The precipitation of Δ D4 in low salt buffer at room

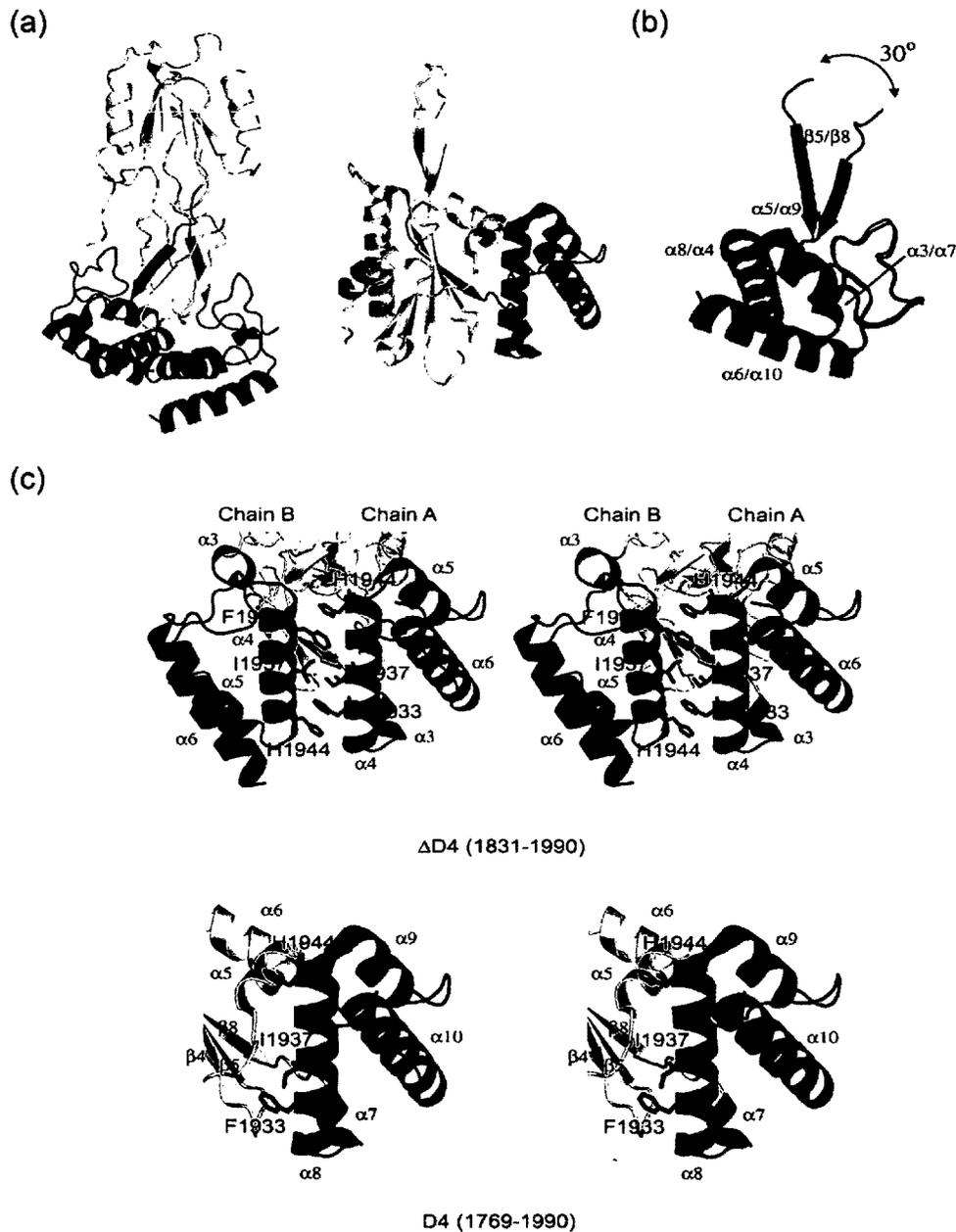


Figure I-3. Remodeling of the helical bundle protein interface. (a) Conserved C-terminal helical bundle within context of D4 and Δ D4; (b) Overlay of the C-terminal helical bundle of D4 (cyan) and Δ D4 chain A (blue) shows the similarity of this domain between the two structures, with the exception of the β -strand which differs $\sim 30^\circ$ in orientation. (c) Stereo views of packing of helical bundle protein interface in D4 and Δ D4 (upper). The α 4 helices of Δ D4 chain A and B pack together with the stacking of aromatic amino acids F1933 and H1944 to form a 2-fold symmetric dimerization interface. (lower) In D4, the equivalent α 8 helix packs against the RNase H domain.

temperature suggests that this protein is less stable than D4. Therefore, we studied the unfolding of both Δ D4 and D4 in the presence of increasing concentrations of urea using CD spectroscopy; these experiments suggest a gradual unfolding of Δ D4 with a midpoint corresponding to a urea concentration of ~ 5 M compared to a sharper transition in the CD spectrum of D4 at ~ 7 M urea and confirm that Δ D4 is less stable. The relative sharpness of the D4 transition compared to Δ D4 (Figure I-4B) suggests a relatively greater exposure of surface area upon unfolding of D4 which is the opposite from the expected result (Myers at al., 1995). This argues that unfolding of Δ D4 occurs by other than a two state process and would be consistent with the multi-domain structure of this fragment.

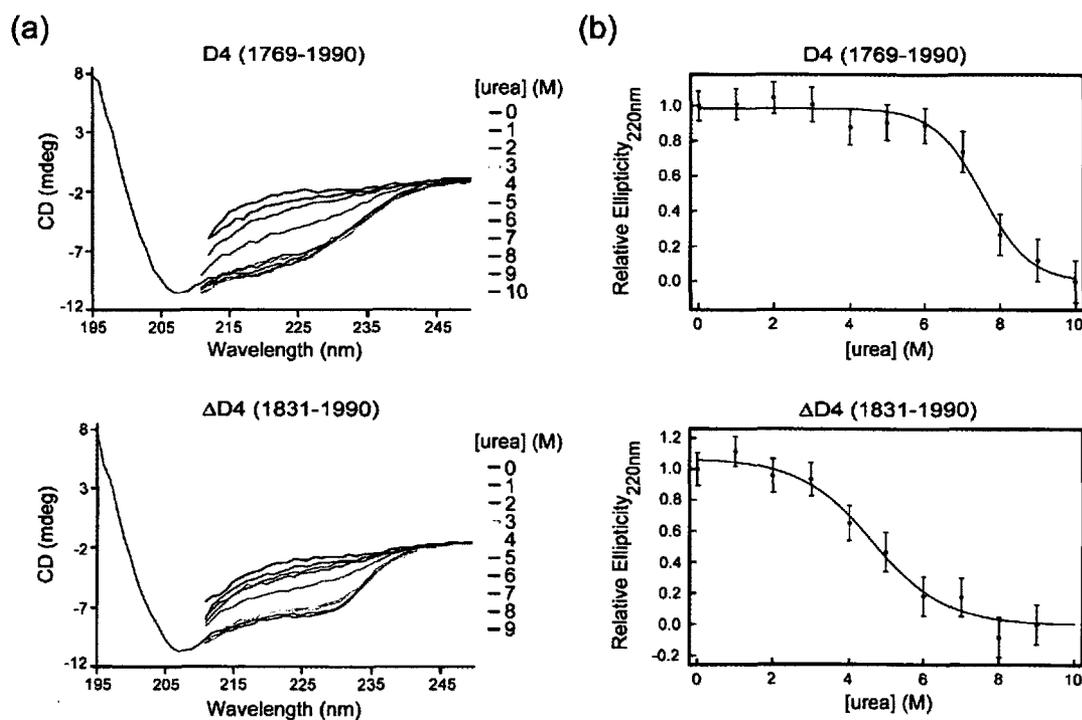


Figure I-4. Solution structure and stability of Δ D4 and D4. (a) CD denaturation curves of D4 and Δ D4; (b) Plots of molar ellipticity at 220 nm vs [urea] for D4 and Δ D4 fit to a two state transition model.

We also attempted to express and characterize sub-domains of $\Delta D4$ in order to assess their stability and structure. A fragment representing the Prp8 C-terminal amino acids 1918-1990 can be expressed at reasonable levels in *E. coli* and purifies as a dimer on a size exclusion column. CD analysis of this protein shows it to be essentially all α -helical; in the presence of increasing amounts of denaturant an unfolding transition was observed in the presence of $\sim 2M$ urea. We were unable to express the N-terminal portion of $\Delta D4$ suggesting that this protein is unstable. The N-terminal portion of D4, amino-acids 1769-1918 could be expressed in *E. coli* but was an aggregate in solution.

I-3. Discussion

The folds of $\Delta D4$ and D4 differ strikingly although their secondary structures are $\sim 75\%$ conserved. The N-terminal seven amino acids of $\Delta D4$ lose the helical structure of $\alpha 2$ in D4 but are disordered in one chain and likely do not contribute to the overall fold. As well, there are structural changes at the edges of secondary structural elements but the major differences involve the conformations of the 12 amino acids of D4 $\alpha 4$ and the 17 amino acids of D4 $\alpha 5/\alpha 6$ that correspond to $\sim 20\%$ of the shared $\Delta D4/D4$ sequence. These transitions involve a combination of helix to loop and helix to strand conversions.

Two elements of higher order structure, the N-terminal helix-strand interaction and the C-terminal strand and helical bundle are conserved between $\Delta D4$ and D4 suggesting that these are fundamentally stable units of structure in the two proteins. Of these, the latter, based on our ability to express and purify it

as a sub-domain, appears most stable in and of itself and likely stabilizes both the D4 and Δ D4 structures. Both the purification properties and CD spectroscopic analysis are consistent with D4 and Δ D4 adopting stable or meta-stable structures respectively in solution, which are likely reflected accurately in the crystallographic analysis.

Comparing interactions within the monomer to those observed in the dimer, we would note that the preservation of the D4 α 2/ α 3 \cdot β 6 packing in Δ D4 involves a strand swap that thus contributes to the dimer interface (Figure I-2). The same surface of α 8 that mediates packing of the D4 C-terminal bundle to the RNase H domain is present as the anti-parallel symmetric interface between the C-terminal bundles of the Δ D4 dimer (Figure I-3). Elsewhere in Δ D4, for example in the central body, hydrophobic surfaces observed within D4 are packed so as to contribute to the dimer interface. Thus, while the sense of interactions observed in D4 is preserved within the dimer, aside from the preservation of the C-terminal sub-domain fold, only one, the helix \cdot strand interaction, is preserved in the latter.

In contrast to many reported examples of context-dependent structure, Δ D4 and D4 share complete sequence identity across an extended length differing only by the N-terminal deletion of Δ D4. In this respect the two proteins might be viewed as a model of structural differences between two natural protein isoforms. A major source of proteome diversity in higher eukaryotes involves the generation of isoforms by alternative splicing of pre-mRNAs. It is estimated that \sim 95% of pre-mRNAs undergo some form of alternative splicing (Wang et al., 2008; Pan et

al., 2008) but the structural and functional consequences of such changes are not well understood at the structural level: the PDB contains only a small representation of isoform structures. Perhaps ~30% of alternative splicing events result in the introduction of a premature stop codon that targets the resulting message for nonsense mediated decay (Lewis et al., 2003). Several analyses of alternative splicing events within the context of known protein structures have been performed (Romero et al., 2006; Peneff et al., 2001; Birzele et al., 2008); for example, an examination of ~500 alternative splicing events based on an analysis of all Swissprot proteins with annotated alternative splicing events has suggested that proteins are remarkably tolerant to sequence insertions, deletions, or replacements (Birzele et al., 2008). While approximately 50% of the events in this survey correspond to entire domains or variable regions within a protein superfamily, many alternative splicing events map to conserved regions within proteins including those with well defined secondary structures. The inference of this work is that a large number of alternative splicing events have the potential to generate significant structural (and hence functional) diversity in proteins by altering the protein fold.

Alternative splicing events corresponding to the removal/addition of structural features both at the edges and within folds have been documented. With respect to $\Delta D4$, the removal of 62 amino acids from D4 essentially corresponds to the deletion of three strands $\beta 1$, $\beta 4$, and $\beta 5$ from the central portion of the RNase H fold sheet. A handful of alternative splicing events resulting in significant structural rearrangement have been characterized. Also, to date, well-

documented examples of insertion/deletion in isoforms involve relatively short sequences (Hymowitz et al., 2003). A splice variant of the p65 subunit of NF- κ B corresponds to the deletion of an internal β -strand; the resultant protein has altered properties that may function in NF- κ B regulation (Ruben et al., 1992). A nine residue insertion in the C-terminus of the pre-synaptic scaffolding protein Picollo induces a large conformational change that prevents calcium binding (Garcia et al., 2004). However, the prediction from sequence analysis of documented alternative splicing events within the context of known structures suggest that much more extensive insertions/deletions in protein structures occur. Our analysis of Δ D4/D4 demonstrates that an exon sized insertions/deletions can be accommodated in a fundamentally remodeled protein structure.

I-4. Materials and Methods

I-4.1. Identification and expression of a Prp8 domain IV core

A maltose binding protein (MBP)-Prp8 fusion protein encompassing domain IV residues 1669–2001 was expressed in *E. coli*, purified, and subjected to limited proteolysis with trypsin resulting in the identification of a stable ~18.8 kDa (Prp8 1831-1993) fragment by MALDI mass spectrometry. Based on our analysis of this sequence, a cDNA representing Prp8 1831-1990 was cloned into the EcoRI and HindIII sites of pMAL-c2x (NEB) using PCR primers to insert a TEV protease cleavage site between MBP and hPrp8 fragment 1831-1990. The resulting MBP–fusion protein was expressed in *E. coli* and purified by sequential

amylose resin, anion exchange, and size-exclusion chromatography. All other protein fragments were cloned, expressed, and purified in an analogous fashion.

I-4.2. Crystallization

Crystals of Prp8 1831–1990 were grown at 4 °C using the hanging drop vapor diffusion technique. Crystals of native protein and protein containing SeMet substitutions were grown by mixing 1 μL of 10 mg ml^{-1} protein solution (10 mM Tris, pH 8.0, 0.1 mM EDTA, 5 mM DTT, 0.02% (w/v) NaN_3) with 1 μL of precipitant (22% PEG 8000, 100 mM Tris, pH 8, 250 mM NaCl, 10 mM 1,3-diaminopropane HCl, 0.25% ethyl acetate, 100 mM NaSCN). Crystals were transferred to precipitant containing 20% (v/v) glycerol and flash-cooled in liquid nitrogen for data collection.

I-4.3. Data collection and processing

Data were collected at beamline 8.3.1 of the Advanced Light Source, Lawrence Berkeley National Laboratory. Anomalous diffraction data were collected from a single SeMet-derivatized crystal; a single-wavelength SAD experiment was performed collecting data in an inverse-beam mode at the experimentally determined Se/K edge. Data were processed and scaled using the HKL package (Table I-1; Otwinowski and Minor, 1997).

I-4.4. Model building and refinement

The program SOLVE was used to determine the positions of the four expected selenium atoms (Terwilliger and Berendzen, 1999). Initial phases to 2.5 Å were improved with density modification in RESOLVE (Terwilliger, 2000). An initial model automatically built by ARPWARP served as the basis for subsequent manual model building and refinement (Terwilliger, 2002). Iterative cycles of refinement in REFMAC against 1.85-Å data collected on the same crystal and manual model building using XFIT were used to complete and refine the model (Murshudov et al., 1997; McRee, 1999). Refinement statistics for the structures are summarized in Table I-1 and 4-2. Coordinates for ΔD4 have been deposited under accession code 3LRU.

I-4.5. CD spectroscopy

CD spectra were collected at room temperature on a J-720 (Jasco) spectrophotometer. Spectra were collected from protein samples (0.2 mg/mL) in PBS buffer or PBS with added urea in a 50 μm path-length cuvette. For each sample, the CD spectrum between 195 and 250 nm was measured 8 times and averaged, followed by a 2 minute long measurement at 220 nm. Ellipticity at 220 nm as a function of urea concentration was normalized to the value at 0 M urea and numerically fit to a function describing a two-state unfolding process between native and unfolded states, assuming a linear dependence of protein stability on denaturant concentration (Pace and Shaw 2000).

I-5. References

- Alexander, P. A., He, Y., Chen, Y., Orban, J. & Bryan, P. N. (2007). The design and characterization of two proteins with 88% sequence identity but different structure and function. *Proc. Natl Acad. Sci. USA* **104**, 11963-11968.
- Alexander, P. A., He, Y., Chen, Y., Orban, J. & Bryan, P. N. (2009). A minimal sequence code for switching protein structure and function. *Proc. Natl Acad. Sci. USA* **106**, 21149-21154.
- Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science* **181**, 223-230.
- Birzele, F., Csaba, G. & Zimmer, R. (2008). Alternative splicing and protein structure evolution. *Nucleic Acids Res.* **36**, 550-558.
- Bullough, P. A., Hughson, F. M., Skehel, J. J. & Wiley, D. C. (1994). Structure of influenza haemagglutinin at the pH of membrane fusion. *Nature* **371**, 37-43.
- Cabrita, L. D., Dai, W. & Bottomley, S. P. (2004). Different conformational changes within the F-helix occur during serpin folding, polymerization, and proteinase inhibition. *Biochemistry* **43**, 9834-9839.
- Davidson, A. R. (2008). A folding space odyssey. *Proc. Natl Acad. Sci. USA* **105**, 2759-2760.
- Fisher, A. J., Thompson, T. B., Thoden, J. B., Baldwin, T. O. & Rayment, I. (1996). The 1.5 Å resolution crystal structure of bacterial luciferase in low salt conditions. *J. Biol. Chem.* **271**, 21956-21968.
- Garcia, J., Gerber, S. H., Sugita, S., Sudhof, T. C. & Rizo, J. (2004). A conformational switch in the Piccolo C2A domain regulated by alternative splicing. *Nat. Struct. Mol. Biol.* **11**, 45-53.
- Grainger, R. J. & Beggs, J. D. (2005). Prp8 protein: At the heart of the spliceosome. *RNA* **11**, 533-557.
- Hacker, I., Sander, B., Golas, M. M., Wolf, E., Karagoz, E., Kastner, B. *et al.* (2008). Localization of Prp8, Brr2, Snu114 and U4/U6 proteins in the yeast tri-snRNP by electron microscopy. *Nat. Struct. Mol. Biol.* **15**, 1206-1212.
- Hymowitz, S. G., Compaan, D. M., Yan, M., Wallweber, H. J. A., Dixit, V. M., Starovasnik, M. A. *et al.* (2003). The crystal structures of EDA-A1 and EDA-A2: splice variants with distinct receptor specificity. *Structure* **11**, 1513-1520.

- Knaus, K. J., Morillas, M., Swietnicki, W., Malone, M., Surewicz, W. K. & Yee, V. C. (2001). Crystal structure of the human prion protein reveals a mechanism for oligomerization. *Nat. Struct. Biol.* **8**, 770-774.
- Lewis, B. P., Green, R. E. & Brenner, S. E. (2003). Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc. Natl Acad. Sci. USA* **100**, 189-192.
- McRee, D. E. (1999). XtalView/Xfit: a versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **125**, 156-165.
- Minor, D. L. & Kim, P. S. (1996). Context-dependent secondary structure formation of a designed protein sequence. *Nature* **380**, 730-734.
- Moore, S. A. & James, M. N. G. (1994). Common structural features of the ZuxF protein and the subunits of bacterial luciferase: evidence for a $(\beta\alpha)_8$ fold in luciferase. *Protein Sci.* **3**, 1914-1926.
- Mossing, M. C. & Sauer, R. T. (1990). Stable, monomeric variants of λ Cro obtained by insertion of a designed beta-hairpin Sequence. *Science* **250**, 1712-1715.
- Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240-255.
- Myers, J. K., Pace, C. N., Scholtz, J. M. (1995). Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci.* **4**, 2138-2148.
- Ono, K., Condrón, M. M. & Teplow, D.B. (2009). Structure-neurotoxicity relationships of amyloid beta-protein oligomers. *Proc. Natl Acad. Sci. USA* **106**, 14745-14750.
- Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307-326.
- Pace, C. N., and Shaw, K. L. (2000). Linear extrapolation method of analyzing solvent denaturation curves. *Proteins: Structure, Function and Genetics Supp* **4**, 1-7.
- Pan, Q., Shai, O., Lee, L. J., Frey, B. J. & Blencowe, B. J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* **40**, 1413-1415.

Pena, V., Rozov, A., Fabrizio, P., Luhrmann, R. & Wahl, M.C. (2008). Structure and function of an RNase H domain at the heart of the spliceosome. *EMBO J.* **27**, 2929-2940.

Peneff, C., Ferrari, P., Charrier, V., Taburet, Y., Monnier, C., Zamboni, V. *et al.* (2001). Crystal structures of two human pyrophosphorylase isoforms in complexes with UDPGlc(Gal)NAc: role of the alternatively spliced insert in the enzyme oligomeric assembly and active site architecture. *EMBO J.* **20**, 6191-6202.

Ritchie, D. B., Schellenberg, M. J., Gesner, E. M., Raithatha, S. A., Stuart, D. T. & MacMillan, A. M. (2008). Structural elucidation of a Prp8 core domain from the heart of the spliceosome. *Nat. Struct. Mol. Biol.* **15**, 1199-1205.

Romero, P. R., Zaidi, S., Fang, Y. Y., Uversky, V. N., Radivojac, P., Oldfield, C. J. *et al.* (2006). Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms. *Proc. Natl Acad. Sci. USA*, **103**, 8390-8395.

Roessler, C. G., Hall, B. M., Anderson, W. J., Ingram, W. M., Roberts, S. A., Montfort, W. R. *et al.* (2008). Transitive homology-guided structural studies lead to discovery of Cro proteins with 40% sequence identity but different folds. *Proc. Natl Acad. Sci. USA* **105**, 2343-2348.

Ruben, S. M., Narayanan, R., Klement, J. F., Chen, C. H. & Rosen, C. A. (1992). Functional characterization of the NF-kappa B p65 transcriptional activator and an alternatively spliced derivative. *Mol. Cell. Biol.* **12**, 444-454.

Shang, Z., Isaac, V. E., Li, H., Patel, L., Catron, K. M., Curran, T. *et al.* (1994). Design of a "minimal" homeodomain: the N-terminal arm modulates DNA binding affinity and stabilizes homeodomain structure. *Proc. Natl Acad. Sci. USA* **91**, 8373-8377.

Stevens, S. W., Barta, I., Ge, H. Y., Moore, R. E., Young, M. K., Lee, T. D. *et al.* (2001). Biochemical and genetic analyses of the U5, U6, and U4/U6U5 small nuclear ribonucleoproteins from *Saccharomyces cerevisiae*. *RNA* **7**, 1543-1553.

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

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.

Tuinstra, R. L., Peterson, F. C., Kutlesa, S., Sonay Elgin, E., Kron, M. A. & Volkman, B. F. (2008). Interconversion between two unrelated protein folds in the lymphotactin native state. *Proc. Natl Acad. Sci. USA* **105**, 5057-5062.

Vetter, I. R., Baase, W. A., Heinz, D. W., Xiong, J. P., Snow, S. & B. W. Matthews, B. W. (1996). Protein structural plasticity exemplified by insertion and deletion mutants in T4 lysozyme. *Protein Sci.* **5**, 2399–2415.

Wang, E. T., Sandberg, R., Luo, S., Khrebtkova, I., Zhang, L., Mayr, C. *et al.* (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470-476.

Yang, K., Zhang, L., Xu, T., Heroux, A. & Zhao, R. (2008). Crystal structure of the beta-finger domain of Prp8 reveals analogy to ribosomal proteins. *Proc. Natl Acad. Sci. USA* **105**, 13817-13822.