

University of Alberta

Structural and Functional Studies of the Core Splicing Factor Prp8

by

Tao Wu

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

©Tao Wu
Spring 2012
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.

Dedicated to my wife Na Wei, who has suffered a lot
and
will continue to suffer during my career.

Abstract:

More than 90% of human genes undergo a processing step called splicing, whereby noncoding introns are removed from initial transcripts and coding exons are ligated together to yield mature messenger RNA. Splicing involves two sequential transesterification reactions catalyzed by a large RNA/protein complex called the spliceosome. RNA components of the spliceosome are at least partly responsible for splicing catalysis. However, the precise architecture of the spliceosome active site and whether it includes protein components remain unknown.

Numerous genetic and biochemical analyses have placed one of the largest and most highly conserved of nuclear proteins, Prp8, at the heart of the catalytic core of the splicing machinery during spliceosome assembly and through catalysis. Here we provide structural and functional evidence that the RNase H domain of Prp8 undergoes a conformational change during splicing which unmask a metal-binding site required for the second step of splicing. We are able to demonstrate that a magnesium ion essential for the catalysis binds to the RNase H domain of Prp8 and a metal specificity switch which abrogates metal binding severely inhibits the second step of splicing. We also show that yeast *prp8* first- and second-step alleles correspond to Prp8 mutants that favour one of the two distinct Prp8 conformations observed in the crystal structure of Prp8 RNase H domain. Together these data support the model of rearrangements within the spliceosome at the time of transition between the first and second step of splicing.

Our findings also establish that Prp8 is a metalloprotein which promotes exon ligation and are consistent with the designation of the spliceosome as a ribonucleoprotein enzyme.

Table of contents:

Dedication	<i>i</i>
Abstract	<i>ii</i>
Table of Contents	<i>iv</i>
List of Figures	<i>viii</i>
List of Tables	<i>xi</i>
List of Abbreviations	<i>xii</i>
Introduction. Pre-mRNA splicing and structures of spliceosome components.	1
I-1. Pre-mRNA splicing.	2
I-2. Spliceosomes and their assembly.	5
I-3. Structures of higher order complexes within spliceosomes.	10
I-3.1. U1 snRNP and U4 snRNP structures.	11
I-3.2. U4/U6•U5 tri-snRNP structure.	16
I-3.3. Structures of A, B and C complex.	18
I-4. Structures of protein components of spliceosomes.	20
I-4.1. SF1 and branch adenosine binding.	20
I-4.2. U2AF: PPT recognition and protein-protein interaction.	23
I-4.3. p14 SF3b155 complex.	26
I-4.4. Other spliceosome-related structures.	28
I-5. Structure of RNA components of spliceosome.	31
I-6. Summary.	36
I-7. References.	36
Chapter 1. Prp8: the pivotal spliceosomal protein at the catalytic center.	48
1-1. Introduction.	49
1-2. Prp8 domains.	51
1-2.1. N-terminal region.	51
1-2.2. Central region.	53
1-2.3. C-terminal region.	56
1-3. Prp8 and its interactions.	59
1-3.1. Prp8-RNA interactions.	59
1-3.2. Prp8 protein interactions.	63
1-4. Crystal structures of Prp8 domains.	65
1-4.1. Structure of the MPN domain.	65
1-4.2. Structure of the RNase H domain.	69

1-5. References.	74
Chapter 2. Structural and functional studies of the Prp8 active site.	83
2-1. Introduction.	84
2-1.1. Two monomers of Prp8 RNase H domain.	84
2-1.2. One metal vs. two metal mechanisms.	85
2-1.3. Metals in the spliceosome.	88
2-2. Results.	92
2-2.1. Mg ²⁺ binds to the Prp8 active site within monomer b.	92
2-2.2. Mutations of metal coordinating residues cause a splicing defect.	95
2-2.3. The impairment of D1853C function is not due to any change in structure.	98
2-2.4. The impairment of D1853C function is not due to defects in complex formation, non-snRNP functions or the first step.	99
2-2.5. The exon ligation step in yeast is ATP-independent but concentration-dependent.	106
2-3. Discussion.	109
2-4. Materials and Methods.	111
2-4.1. Protein expression, purification, and crystallization.	111
2-4.2. Data collection and processing.	112
2-4.3. Model building and refinement.	113
2-4.4. Construction of RNA substrates.	113
2-4.5. Creation of mutant <i>S. cerevisiae</i> strains and splicing extracts.	114
2-4.6. Splicing assays.	115
2-4.7. Native gel analysis.	116
2-4.8. Exon-ligation assay and chase experiment.	116
2-5. References.	117
Chapter 3. The conformational change of Prp8 is coupled to the transition between the two steps of splicing.	123
3-1. Introduction.	124
3-1.1. Transition between the first and the second step of splicing.	124
3-1.2. PRP8 alleles affecting the first and second step of splicing.	127
3-2. Results.	131
3-2.1. Structures of first-step alleles.	131
3-2.2. Structures of second-step alleles.	133
3-2.3. In vivo splicing of the first- and second-step alleles.	139
3-2.4. In vitro splicing of the first- and second-step alleles.	142
3-2.5. D1853C is not a first-step allele but can be rescued by second-step alleles.	145

3-3. Discussion.	147
3-4. Materials and Methods.	151
3-4.1. Protein expression, purification, and crystallization.	151
3-4.2. Construction of RNA substrates.	152
3-4.3. Creation of mutant <i>S. cerevisiae</i> strains and splicing extracts.	152
3-4.4. Temperature and copper growth tests.	153
3-4.5. Primer extension.	153
3-4.6. Splicing assays.	154
3-5. References.	155
 Chapter 4. Conclusions and future directions.	 158
4-1. Conclusions.	159
4-2. Future directions.	162
4-2.1. Designing new mutant alleles for the RNase H domain.	162
4-2.2. Mapping of protein–RNA crosslinking sites within the RNase H region.	164
4-2.3. Investigating the role of other Prp8 domains in different steps of splicing.	166
4-3. Materials and Methods.	169
4-3.1. Protein expression, purification, and crystallization.	169
4-3.2. Limited proteolysis.	170
4-4. References.	170
 Appendix I. Characterization of in vivo dimerization of ADAR proteins.	 173
I-1. Introduction.	174
I-2. Results.	178
I-2.1. Cellular localization of ADAR1 and ADAR2.	178
I-2.2. Homodimerization of ADAR1 and ADAR2.	179
I-2.3. Heterodimerization of ADAR1 and ADAR2.	182
I-2.4. RNA-independent ADAR dimerization.	183
I-3. Discussion.	187
I-4. Materials and Methods.	189
I-4.1. Mammalian Expression Constructs.	189
I-4.2. Cell Culture and Transfection.	190
I-4.3. Fluorescence Microscopy.	191
I-4.4. Fluorescence Resonance Energy Transfer.	191
I-5. References.	192

Appendix <i>II</i> . Characterization of ADAR1-containing extra-nucleolar bodies.	196
<i>II-1</i> . Introduction.	197
<i>II-2</i> . Results.	200
<i>II-2.1</i> . ADAR1 specific extra-nucleolar bodies.	200
<i>II-2.2</i> . FRAP of AdB.	202
<i>II-2.3</i> . AdBs are not known major nuclear bodies.	204
<i>II-2.4</i> . Isolation of ADAR-containing complexes.	206
<i>II-3</i> . Discussion.	213
<i>II-4</i> . Materials and Methods.	215
<i>II-4.1</i> . Mammalian Expression Constructs.	215
<i>II-4.2</i> . Cell Culture and Transfection.	216
<i>II-4.3</i> . Fluorescence Microscopy.	216
<i>II-4.4</i> . Fluorescence Recovery After Photobleaching.	216
<i>II-4.5</i> . Affinity purification of FLAG-ADAR1 and FLAG-ADAR2.	217
<i>II-4.6</i> . Editing reactions.	218
<i>II-5</i> . References.	219
Appendix <i>III</i> . Sequences of oligonucleotides used in this thesis.	223
<i>III-1</i> . Primer sequences 5' → 3'.	224
<i>III-1.1</i> . Human Prp8.	224
<i>III-1.2</i> . Yeast Prp8.	225
<i>III-1.3</i> . ACT1 transcription template.	228
<i>III-1.4</i> . Primer extension.	229
<i>III-1.5</i> . ADAR proteins.	229
<i>III-1.6</i> . Other proteins.	230
<i>III-2</i> . RNA sequences 5' → 3'.	231

List of Figures:

Introduction.

I-1. pre-mRNA splicing.	3
I-2. Assembly and disassembly pathway of spliceosomes.	7
I-3. U1 snRNP and U4 snRNP core crystal structure.	13
I-4. EM structures and models of higher order complexes.	17
I-5. SF1 and U2AF structures.	22
I-6. p14 SF3b155 complex structures.	27
I-7. Group I and group II intron structures.	32

Chapter 1.

1-1. Diagrams of conserved domains in yeast Prp8.	52
1-2. Prp8-RNA interactions.	60
1-3. Crystal structures of Prp8 MPN domain.	66
1-4. Crystal structures of Prp8 RNase H domain.	70

Chapter 2.

2-1. Examples of general metal ion driven mechanisms in protein metallonucleases.	87
2-2. Proposed two metal mechanism for splicing.	90
2-3. Conformational plasticity in the Prp8 RNase H domain.	93
2-4. PRP8 D1853 mutations impair the second step of splicing.	97
2-5. PRP8 D1853C does not bind metal.	100
2-6. The second step defect of D1853C is not due to defects in complex formation, non-snRNP protein functions or the first step.	102
2-7. Bimolecular exon-ligation assay with WT and D1853C.	104
2-8. Kinetics of the first step of splicing.	105
2-9. The exon ligation step does not depend on Prp16 or Prp22.	107
2-10. The exon ligation step is not ATP-dependent but concentration-dependent.	108

Chapter 3.

3-1. Schematic representation of splicing pathway and assignment of mutant alleles of <i>prp8</i> , <i>prp16</i> , <i>prp22</i> and U6 snRNA that modulate spliceosomal transitions.	125
3-2. Diagram of Prp8 mapping the known yeast alleles.	128
3-3. Structure of the PRP8 first-step alleles.	132
3-4. Structure of the PRP8 second-step alleles.	135
3-5. Structure of the second-step allele T1789P.	138
3-6. Growth phenotypes of strains containing PRP8 first- and second-step alleles.	139
3-7. In vivo splicing with first- and second-step PRP8 alleles.	141
3-8. PRP8 first-step alleles improve the first step splicing of the BSC substrate.	143
3-9. PRP8 second-step alleles improve the second step splicing of the UuG substrate.	144
3-10. D1853C is not a first-step allele but can be rescued by second-step alleles.	146
Chapter 4.	
4-1. Model of the role of Prp8 RNase H domain in splicing.	160
4-2. Structure of the Prp8 I1790C/T1800C mutant.	164
4-3. Limited proteolysis of yPrp8 965-1189.	168
Appendix I.	
I-1. Chemistry of deamination reaction	174
I-2. Domain structure and localization of human ADAR proteins	177
I-3. Localization of human ADAR proteins	179
I-4. Homodimerization of ADAR1 and ADAR2	181
I-5. Heterodimerization of ADAR1 and ADAR2	182
I-6. RNA-independent homodimerization and heterodimerization of ADAR1 and ADAR2	184
I-7. ADAR2 constructs and mutations for FRET	186
Appendix II.	
II-1. ADAR1-specific extra-nucleolar bodies	201

<i>II-2. FRAP of AdBs</i>	203
<i>II-3. AdBs do not correspond to any known structure</i>	205
<i>II-4. Purification of ADAR1- and ADAR2-containing complexes</i>	207
<i>II-5. Proteins pulled down by ADAR1 do not co-localize with ADAR1</i>	210
<i>II-6. Association of ADAR1 with ILF2</i>	211
<i>II-7. Glycerol gradient separation of ADAR1-containing complexes</i>	212

List of Tables:

Introduction.

Table I-1. Major protein components of snRNPs.	6
--	---

Chapter 3.

Table 3-1. Summary of the Prp8 alleles mapping to the RNase H region.	129
---	-----

Appendix I.

Table I-1. FRET efficiency of ADARs in HeLa cells.	180
--	-----

Table I-2. FRET efficiency of ADAR2 constructs.	187
---	-----

Appendix II.

Table II-1. Proteins pulled down by ADAR1 and ADAR2.	208
--	-----

List of Abbreviations:

Δ	delta, depleted
3D	three dimensional
5-FOA	5-fluoroorotic acid
5-HT _{2C} R	2C subtype of serotonin receptor
A	adenosine
aa	amino acid
ADAR	adenosine deaminase that acts on RNA
AdB	ADAR1 specific nucleolar body
ALS	amyotrophic lateral sclerosis
ATP	adenosine triphosphate
BME	β -mercaptoethanol
BPS	branch point sequence
Br domain	bromodomain
BSC	ACT-1 pre-mRNA branch site A to C mutation
BSG	ACT-1 pre-mRNA branch site A to G mutation
C	cytosine
CB	Cajal body
CC	commitment complex
CFP	cyan fluorescence protein
cs	cold-sensitive
CTD	C-terminal domain
CTP	cytidine triphosphate
CypH	Cyclophilin
DAPI	4',6-diamidino-2-phenylindole
DMEM	Dulbecco's modified Eagle's medium
dsRBD	double-stranded RNA binding domain
DGCR8	DiGeorge syndrome critical region gene 8
DTT	dithiothreitol
DV	domain V
E complex	early complex
EDTA	ethylene diamine tetraacetic acid
eIF3	eukaryotic initiation factor 3
EM	electron microscopy
FRAP	fluorescence recovery after photobleaching
FRET	fluorescence resonance energy transfer
G	guanidine
GTP	guanosine triphosphate
hnRNP	heterogeneous nuclear ribonucleoprotein
HDAC4	histone deacetylase 4
HuR	Hu-antigen R
I-RNA	inosine-rich RNA
IP6	inositol hexakisphosphate
ISL	internal stem loop

JAMM motif	JAB1/MPN/Mov34 metalloenzyme motif
K_d	dissociation constant
KH	hnRNP K homology
LC-MALDI	Liquid chromatography/Matrix-assisted laser desorption/ ionization
MBP	maltose-binding protein
MEF2	myocyte enhancer binding factor 2
MPN	Mpr-1, Pad-1, N-terminal domain
mRNA	messenger RNA
NFAT	nuclear factor of activated T-cells
NLS	nuclear localization signal
NMR	nuclear magnetic resonance
nSG	nuclear stress granules
nt	nucleotide
NTX	nineteen complex
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PDB	protein databank
PEG	polyethylene glycol
PEPCK	phosphoenolpyruvate carboxykinase
PML	promyelocytic leukaemia
poly-Q	poly-glutamine
PPT	poly-pyrimidine tract
PROCN	Prp8 central domain
PRP	pre-mRNA processing
pre-mRNA	precursor messenger-RNA
PSP	paraspeckle protein
QUA2	Quaking homology 2
RBD	RNA binding domain
RISC	RNA-induced silencing complex
RP	retinitis pigmentosa
RRM	RNA recognition motif
rRNA	ribosomal RNA
RS	arginine-serine rich domain
RT	reverse transcriptase
RT-PCR	reverse transcription polymerase chain reaction
SAXS	small angle X-ray scattering
SDS	sodium dodecyl sulfate
SF1	splicing factor 1
siRNA	small interfering RNA
SL	stem loop
SMN	survival of motor neuron
snRNA	small nuclear RNA
snRNP	small nuclear ribonucleoprotein particle

snoRNA	small nucleolar RNA
SRP9	signal recognition particle 9
SS	splice site
TCEP	tris(2-carboxyethyl)phosphine
TERT	telomerase reverse transcriptase
TEV	tobacco etch virus
Th/X	thumb/X/maturase domain
TRBD	telomerase RNA binding domain
ts	temperature-sensitive
tRNA	transfer RNA
Tudor-SN	Tudor staphylococcal nuclease
U	uridine
U2AF	U2 auxiliary factor
U6i	U6 interaction domain
UBD	ubiquitin-binding domain
UTP	uridine triphosphate
UuG	ACT-1 pre-mRNA 3' splice site A to U mutation
VDR	vitamin D receptor
XPO5	Exportin 5
YFP	yellow fluorescence protein
YTH	yeast two-hybrid

Introduction

Pre-mRNA splicing and structures of spliceosome components

I-1. Pre-mRNA splicing

Ever since the discovery of the double-stranded structure of DNA and the proposal of the central dogma of molecular biology in the 1950s, numerous advances have occurred regarding the mechanism and regulation of gene expression. Many of these significant findings were made in the field of RNA splicing.

In the majority of eukaryotic genes, protein-coding sequences are interrupted at intervals by non-coding sequences. The coding regions are known as exons and the non-coding regions as introns. During gene expression, the DNA sequence is first transcribed by RNA polymerase to produce precursor messenger RNA (pre-mRNA). The transcripts then undergo RNA splicing to remove the intron sequence and ligate the flanking exons to produce the mature mRNA (Figure I-1 A). This “split gene” structure was first discovered during the mapping of adenoviral gene structure, which showed the formation of single stranded DNA loop within the hybrid between the 5' terminal sequences of adenovirus 2 (Ad2) mRNA and its encoding genomic DNA (Berget et al., 1977; Chow et al., 1977). Later it was discovered that more than 90% of eukaryotic genes contain intron sequences.

There have been four different types of RNA splicing identified, three of which are carried out by sequential phosphotransesterification reactions (Figure I-1 B). The first and second transesterification reactions will be hereafter referred to

as the first step and second step of splicing for convenience. Both group I and group II intron splicing is self-catalyzed. Group I introns are found in some

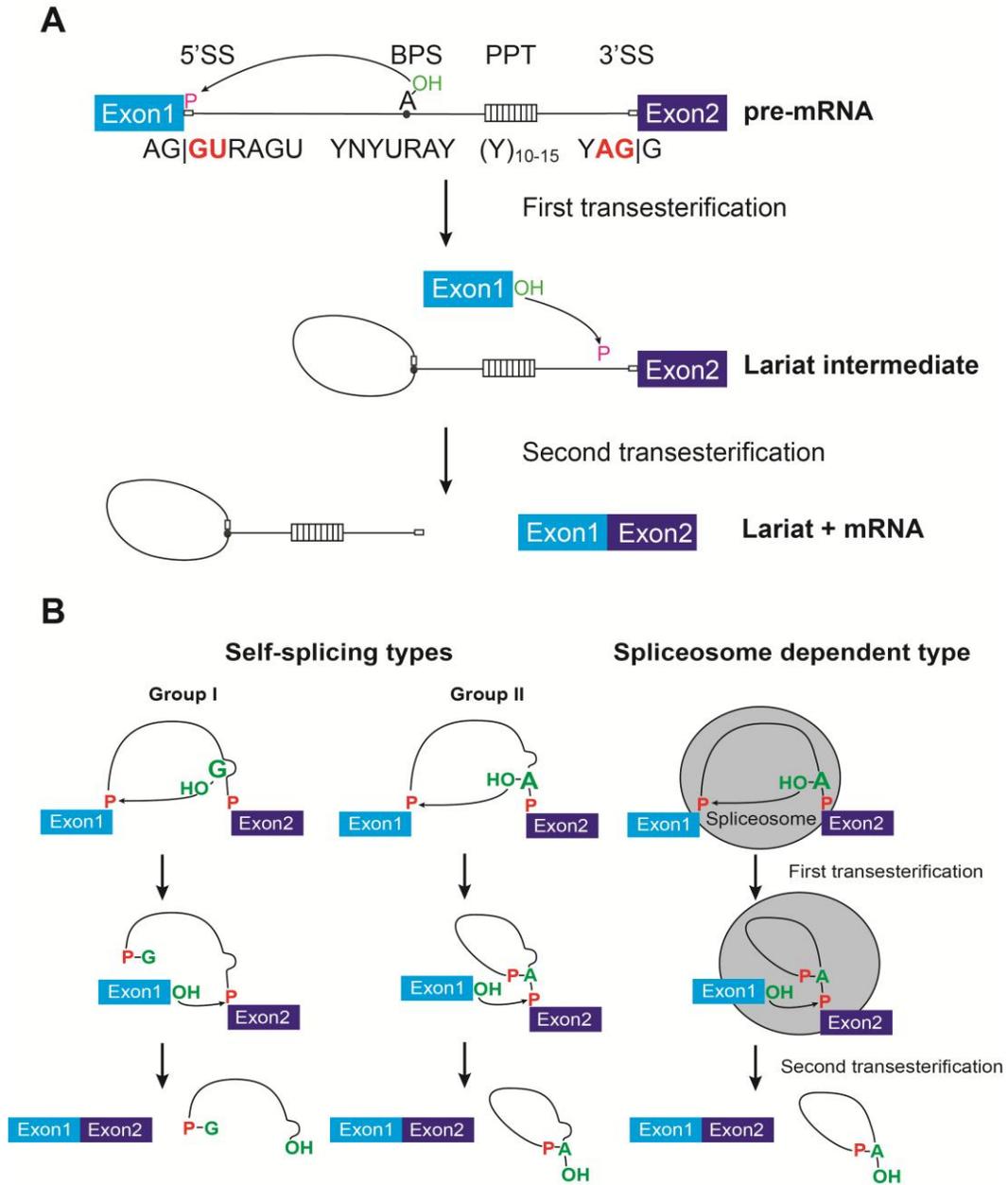


Figure I-1. pre-mRNA splicing. A) Representation of the two-step splicing mechanism. Critical sequence elements are also shown. Y = pyrimidine, R=purine, N=any nucleotide. B) Comparison of self-splicing and spliceosome-catalyzed splicing. The nucleophilic hydroxyl groups are shown in green, and the scissile phosphates are shown in red.

nuclear, mitochondrial and chloroplast genes coding for ribosomal RNAs (rRNAs), messenger RNAs (mRNAs) and transfer RNAs (tRNAs). Their splicing requires the binding of a guanine (G) nucleotide to the intron sequence, which is used as a nucleophile to attack the 5' splice site (5'SS). The newly formed 3' end of the mRNA chain then attacks the other side of the intron at the 3'SS to complete the reaction (reviewed in: Nielsen and Johansen, 2009). In the group II introns, the mechanism of splicing is similar, but the nucleophile in the first step is an adenosine (A) residue within the intron (reviewed in: Ritchie et al., 2009). The third and the largest class of introns are those found in nuclear mRNA primary transcripts. These undergo splicing by a mechanism similar to that of the group II introns. However, specialized RNA-protein complexes called small nuclear ribonucleoproteins (snRNPs) are required in this case (Rio 1993; Valadkhan and Jaladat, 2010). The fourth class of introns, found in certain tRNAs, is distinctive from the above three groups in that the splicing reaction requires ATP and a protein endonuclease. The splicing endonuclease utilizes the energy of ATP hydrolysis to catalyze intron removal and ligation of the two exons by a mechanism similar to the DNA ligase reaction (reviewed in: Calvin and Li, 2008).

Although group II introns and nuclear mRNA primary transcripts share similar transesterification steps, they differ significantly in consensus sequence and snRNP requirement. In group II introns, six typical stem-loop structures, called domains I to VI, are formed within the intron sequence. Group II introns possess very few conserved nucleotides, and those important for catalysis are

spread over the complete intron (Bonen and Vogel, 2001). Spliceosomal introns are defined by conserved sequence elements. In eukaryotic genes, the 5'SS contains an almost invariant GU dinucleotide which is part of a larger, less highly conserved consensus region. The 3' splice site (3'SS) has a seven nucleotide branch point sequence (BPS) with a conserved branched adenosine and a downstream poly-pyrimidine (C and U) tract (PPT), followed by a conserved YAG trinucleotide which terminates the intron sequence. There are also many splicing enhancer and repressor sequences in the vicinity of splice sites (Barash et al., 2010). Once the intron sequence is recognized by the spliceosome, the 2' hydroxyl (2'OH) of the branch adenosine in BPS attacks the 5'SS, thereby forming the lariat and a free 5' exon. The 3'OH of the 5' exon then attacks the 3'SS, releasing the lariat-structure intron and completing the two transesterification reactions (Figure I-1 A).

I-2. Spliceosomes and their assembly

The spliceosome is a 60S biochemical machine with protein and RNA components. The major spliceosome contains around two hundred proteins and five small nuclear RNAs (snRNAs), named U1, U2, U4, U5 and U6 (U3 is involved in ribosomal RNA processing; Table I-1). The RNA-protein complexes are also called small nuclear ribonucleoproteins (snRNPs), named after their RNA component. The RNA components play critical roles in splicing and they closely

resemble a number of critical RNA domains of group II introns. All except U4 have functional counterparts in group II introns, implying that snRNAs may have evolved from group II intron sequences (Cech, 1986; Valadkhan, 2010). Each snRNP contains seven common core proteins called Smith (Sm) proteins (SmB/SmB', SmD1, SmD2, SmD3, SmE, SmF, and SmG), named after the patient whose serum contained antibodies specific for the Sm complex (Kattah et al., 2010). snRNPs also contain snRNP-specific proteins, such as U1A, U1C and U1-70K in the U1 snRNP.

Table I-1. Major protein components of snRNPs

U1	U2	U4	U5	U6
Sm	Sm	Sm	Sm	Lsm
U1A	U2A'	Prp3	Prp8	Prp24
U1C	U2B''	Prp4	Brr2	
U170K	SF3a	Prp31	Snu114	
Prp40	SF3b	CypH	Prp6	
	Prp9	Snu13	Prp28	
		15.5K	Snu40	
			Dib1	

The canonical spliceosome assembly model involves an ordered, stepwise assembly of snRNPs on the pre-mRNA substrate (Figure I-2). The assembly is initiated by U1 snRNP binding to and base-pairing with the 5'SS to form the commitment complex, or early (E) complex. Other non-snRNP associated factors also contribute to spliceosomal assembly. For example, the U2 snRNP auxiliary

factor (U2AF) recognizes the PPT/3'SS, and the splicing factor 1 (SF1) binds to the BPS (Wu et al., 1999; Zorio and Blumenthal, 1999; Berglund et al., 1997). The formation of the E complex is ATP-independent and commits the substrate to the splicing pathway (Legrain et al., 1988; Seraphin and Rosbash 1989; Jamison et al., 1992).

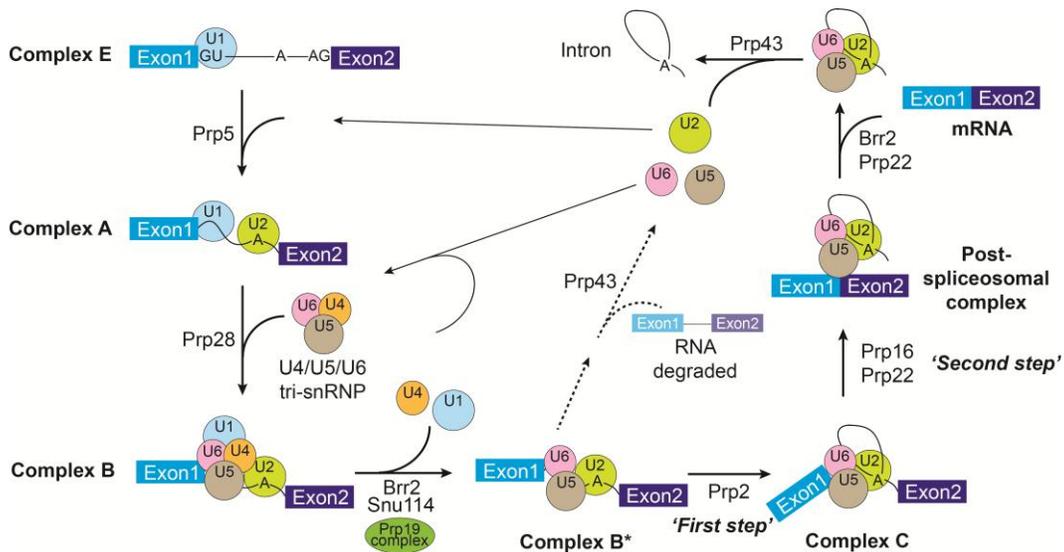


Figure I-2. Assembly and disassembly pathway of spliceosomes. The details of the process are discussed in the text. A Prp16/Prp43 dependent discard pathway is also shown. The first and second transesterification steps are shown in italics (Adapted from Will et al., 2010).

Next, U2 snRNP is recruited to the branch region through interactions with U2AF and possibly U1 snRNP in an ATP-dependent manner, resulting in the formation of the prespliceosome or A complex. This recruitment also depends on the binding of U2 snRNP-associated proteins SF3a and SF3b to flanking sequences upstream of the branchpoint, and one human SF3b subunit, p14, was

shown to contact the BPS directly (Query et al., 1994). SF3b155, a protein component of U2 snRNP, was found to bind to U2AF to recruit U2 snRNP to the BPS (Gozani et al., 1996; Will et al., 2001; Gozani et al., 1998). Base-pairing between U2 snRNA and the branch region bulges out the branch adenosine for the first transesterification reaction. A DExD/H-box helicase Prp5, acting with SF3a and SF3b, seems to be a proof-reader and ensures the fidelity of U2 snRNA/BPS association by preventing U2 snRNA from binding sub-optimal branch point sequence (Xu and Query, 2007; Perriman and Ares, 2010).

Subsequently, the interactions between U1 snRNP and U4/U6•U5 complex recruit the tri-snRNP to the 5'SS to form the precatalytic spliceosome or B complex (reviewed in: Brow, 2002). Within the tri-snRNP, U4 and U6 snRNA form an extensive base-pairing interface, which must be unwound for splicing to occur. U6 snRNA also forms base-pairs with 5'SS (Bindereif et al., 1990).

The tri-snRNP completely displaces U1 snRNA from the 5'SS, forming a new base-pairing between the ACAGAGA box of U6 snRNA and the 5'SS. U4/U6 base-pairs are unwound; the new U2/U6 base-pairs are formed and the U1 and U4 snRNPs dissociate from the spliceosome. The DExD/H-box helicase Prp28 was shown to destabilize the U1/5'SS helix (Ares and Weiser, 1995; Madhani and Guthrie, 1994; Staley and Guthrie, 1998). These transitions move B complex into the activated spliceosome or B* complex and form the active site of spliceosome. The importance of replacing U4/U6 base-pairing with U2/U6 has been shown by the U4-cs1 mutation, where the U4/U6 base-pairing is extended

and the U6/5'SS interactions are masked; a cold-sensitive block is conferred to splicing (Li and Brow, 1996). The DExD/H-box protein Prp44, also called Brr2, has been implicated in unwinding U4/U6 (Staley and Guthrie, 1999; Noble and Guthrie, 1996; Xu et al., 1996). After disassociation of U1 and U4 snRNP the nineteen complex (NTC) containing Prp19 helps to stabilize association of U5 and U6 with the spliceosome (Chan et al., 2003).

After the spliceosome is fully assembled, the DExD/H box protein Prp2 displaces the SF3a/b complex from the spliceosome. This activates the first step transesterification reaction, moving the splicing machine into the catalytic step 1 spliceosome or C complex (reviewed in Lardelli 2010). Additional rearrangements occur in C complex before the second catalytic step. After the second step occurs, the spliceosome dissociates and the mature mRNA is released.

Two additional RNA helicases are required during the splicing process. The first helicase Prp16 functions in proofreading the first step of splicing by discarding sub-optimal substrates and causing a conformational change required for the second step (reviewed in: Wahl et al., 2009; Koodathingal et al., 2010). During the second step of splicing, the helicase Prp22 proofreads exon ligation by sensing aberrant substrates. Along with the activity of Brr2, Prp22 also helps to release the mature mRNA (Schwer, 2008). In both proofreading pathways, another helicase Prp43 stimulates the discard of defective substrates and dissociation of the spliceosome. Prp43 also functions in the intron release and

disassembly of the spliceosome following splicing (Pandit et al., 2006; Koodathingal et al., 2010).

I-3. Structures of higher order complexes within spliceosomes

The complicated and dynamic process of spliceosome assembly is reminiscent of the ribosome, another ribonucleoprotein nanomachine. Mature ribosomes contain four ribosomal RNAs (rRNAs) and over 90 different proteins. The assembly process involves the coordinated function of over 200 proteins in the synthesis and processing of the four rRNAs, as well as assembly of those rRNAs with the ribosomal proteins (reviewed in: Staley and Woolford, 2009). High-resolution structures have been achieved for both the ribosome alone and the ribosome complexed with tRNA and mRNA (Schuwirth et al., 2005; Selmer et al., 2006). No such structure is available for the spliceosome due to the sheer number of factors involved and their dynamic interactions with the mRNA, snRNA and with each other. However, structures of individual components and complexes within the spliceosome have been obtained at different resolution levels. The following sections of this introduction will highlight the recent progress in this field.

I-3.1. U1 snRNP and U4 snRNP structures

U1 snRNP is composed of U1 snRNA, the seven common core Sm proteins, and three U1-specific proteins (U1-70K, U1A, and U1C). The crystal structure of the U1 snRNP complex (Krummel et al., 2009), along with previous structural and biochemical data, provide information about how this complex is assembled.

Human U1 snRNA forms four stem-loops (SL1-4) (Krol et al., 1990). Seven core Sm proteins assemble around the Sm site heptad (AUUUUUG) between SL3 and SL4. U1-70K and U1A bind to SL1 and SL2, respectively, through their RNA binding domains (RBDs) (Query et al., 1989; Scherly et al., 1990; Oubridge et al., 1994). U1C is recruited to the U1 snRNA by association with the N-terminus of U1-70K and the Sm core (Nelissen et al., 1994). The crystal structure of the N-terminal RBD of U1A bound to SL2 has been reported, as well as the NMR structure of the zinc finger-region of U1C (Oubridge et al., 1994; Muto et al., 2004). Electron microscopy (EM) studies and a later cryo-electron microscopy (cryo-EM) single-particle analysis of U1 snRNP revealed the overall globular core domain with two protruding regions corresponding to U1-70K and U1A. The U1 snRNP shows a ring-shaped core domain interpreted as the Sm ring with a funnel-shaped hole passing through it. (Kastner et al., 1992; Stark et al., 2001). However, due to the low resolution (~ 10 Å), the secondary structure elements cannot be identified in the complex.

In a more recent study, the X-ray crystal structure of the functional core of human U1 snRNP at 5.5 Å resolution was obtained (Krummel et al., 2009; Figure I-3 A). In this structure, the Sm core proteins form a heptameric ring structure. In contrast to the EM result, the four-helix junction lies directly over the centre of the Sm ring and no funnel is observed. Thus, the interpretation of the EM envelope is not consistent with the crystal structure (Krummel et al., 2010). An N-terminal extension of SmD2 forms an extra α -helix that points into the minor groove of helix H. The N-terminal helix of SmB also interacts with the phosphate backbone of SL2. These interactions likely stabilize the four-way junction of U1 snRNA. The RBD of U1-70K binds to SL1 and the region between residues 60 and 90 forms an α -helix running along SL1. The N-terminus of U1-70k and the C-terminus of SmD3 form short α -helices and contact the long α -helix of U1C, which explains the requirement of U1-70K to recruit U1C to the core domain (Nelissen et al., 1994).

In the asymmetric unit of the U1 snRNP crystal structure, the 5' end of U1 snRNA base-pairs with its counterpart from an adjacent U1 snRNP, which mimics the base-pairing between U1 snRNA and 5'SS. Helix A of the U1C zinc-finger region binds across the minor groove of this duplex where C8 and A7 of the U1 snRNA base-pair with nucleotides corresponding to the invariant GU dinucleotide of the 5'SS. The polypeptide loop between the two Zn-coordinating His residues is in proximity to the pre-mRNA strand and the conserved Arg21-Lys22 on helix A and Arg28-Lys29 in the loop may be involved in the interaction with the

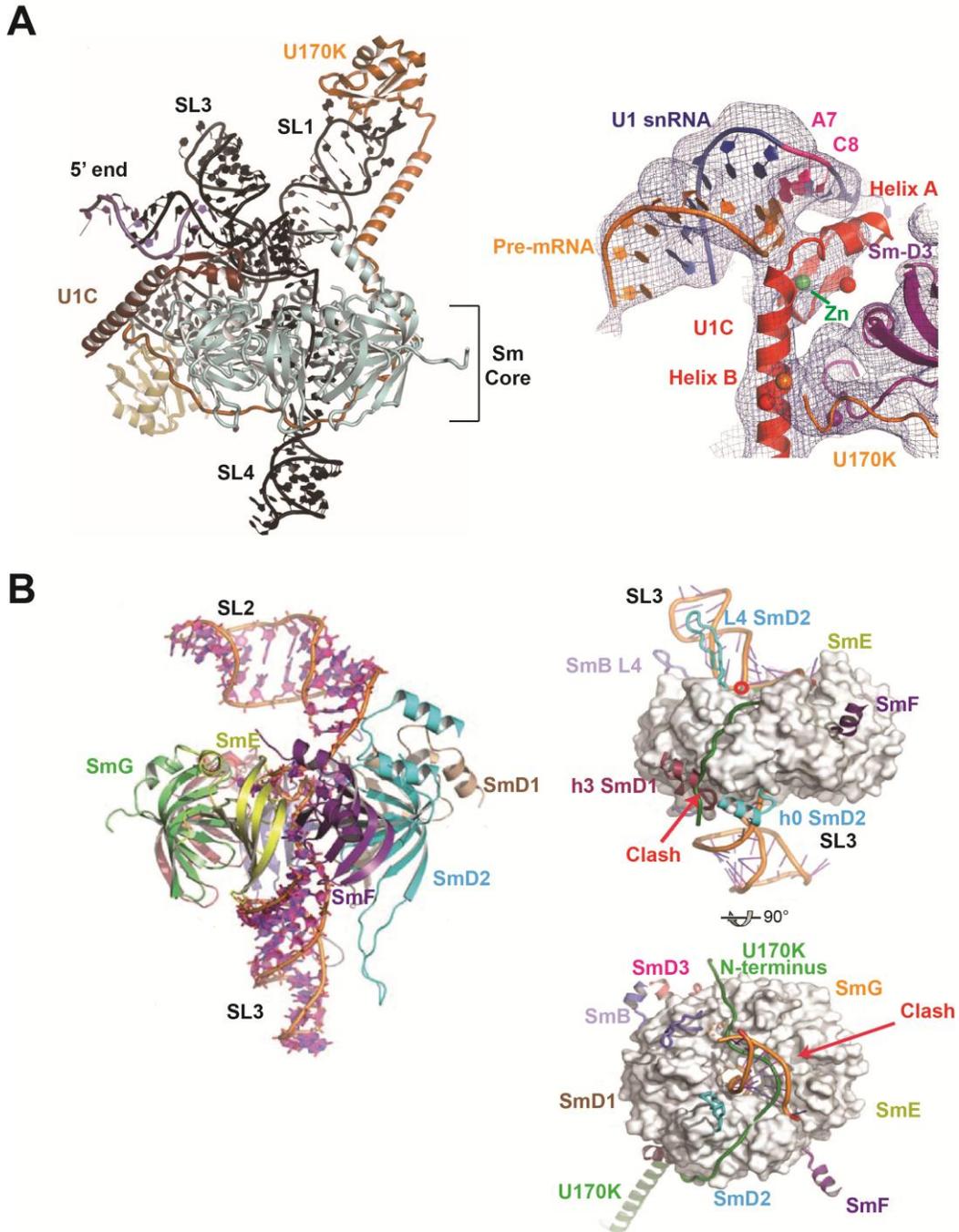


Figure I-3. U1 snRNP and U4 snRNP core crystal structure. A) Crystal structure of U1 snRNP (*left*) and the 5'SS recognition (*right*) (Figure adapted from Krummel et al., 2009). A zinc ion is shown in green sphere. B) Crystal structure of U4 snRNP (*left*) and mapping the U1-70K N-terminus onto the U4 core domain (*right*) (Figure adapted from Leung et al., 2010). The arrows indicate the steric clash between U1-70K and SmD1 (*upper*) and between U1-70K and SL3 of U4 snRNA (*lower*).

phosphate backbone (Newman et al., 2010). A double mutant (R28G, K29S) fails to promote E complex formation (Will et al., 1996). Therefore the zinc finger domain of U1C is proposed to play a significant role in stabilizing this base-pairing, although the specific details of interaction are not available due to the limited resolution. This is consistent with the previous crosslinking between U1C and 5'SS (Du and Rosbash, 2002).

The U1 snRNP crystal structure is the first successful attempt to reveal the architecture of a spliceosomal snRNP at high resolution and provided important insights into the mechanism of the assembly of U1 snRNP as well as the recognition of the 5'SS. Some previously unknown features, such as the N-terminal extension of SmD2 and the structure of the N-terminus of U1-70K, which are both dependent on their context within the entire snRNP complex, can now be visualized. Recently, another U1 snRNA crystal structure has been solved at a higher resolution (Weber et al., 2010). However, both the 5.5 Å and the 4.4 Å structures are still limited, leaving the side chain interactions between the Sm proteins and Sm site unresolved.

Recently a 3.6 Å crystal structure of U4 snRNP core domain was published by the same group as 5.5 Å U1 snRNP (Leung et al., 2011). At this resolution, the detailed interactions between Sm site and Sm proteins were depicted. In this study, a truncated U4 snRNA containing SL2 and SL3 was mixed with Sm core proteins and crystallized. The U4 Sm site heptad interacts with the Sm proteins asymmetrically in the central hole of the heptamer ring through a series of

hydrogen bonds and ring stackings (Figure I-3 B). Interestingly, in mammalian U4 and U5 snRNAs the Sm site and the 3' SL are linked by a single nucleotide, while in U1 and U2 snRNA there are five nucleotides in the linker region (Guthrie and Patterson, 1988). When the structures of U1 snRNP and U4 core domain are aligned, the snRNAs thread the central Sm hole in a similar way but with different stem orientations so that their 3' termini fall on opposite sides (Leung et al., 2011). When trying to map the U1-70K N-terminus onto the U4 core domain, the 3' strand of U4 SL3 would clash with U1-70K and obstruct its pathway. This is consistent with the result that the N-terminus of U1-70K failed to bind to the U5 core domain as U4 and U5 adopt the same single nucleotide separation between Sm site and the 3' end (Nilissen et al., 1994). In addition, the N-terminal extension of SmD2 and C-terminal extension of Sm D1 are disordered in the absence of RNA, whereas they adopt different conformations in U1 snRNP and U4 snRNP (Kambach et al., 1999). Therefore, although the same Sm core proteins bind to all five snRNAs, their associations and conformations depend on the RNA sequence, which provides selectivity for the complex-specific proteins and plays a critical role in snRNP assembly. In the future, effort should be exerted to elucidate the disposition of U4 snRNP-specific proteins and the 5'SL of U4 snRNA which base-pairs with U6 snRNA since this region is absent from the U4 core domain structure.

I-3.2. U4/U6•U5 tri-snRNP structure

The 3D structure of the U4/U6•U5 tri-snRNP has been studied by performing 3D cryo-EM on the human tri-snRNP and its two major subunits, the U5 snRNP and the U4/U6 di-snRNP (Sander et al., 2006; Figure I-4 A). The tri-snRNP shows a tetrahedral shape with a broad upper portion and a slim lower part. The U5 snRNA has an 11 nucleotide loop I, which is located very close to the catalytic centre of the spliceosome (Ségault et al., 1999). This region has been proposed to tether the two exon ends in the orientation needed for the second catalytic step of splicing by base-pairing with both exons (McConnell and Steitz, 2001). In the EM structure the loop I was located near the centre of the tri-snRNP, in close proximity to the U4/U6 di-snRNP portion.

More recently, the Luhrmann group solved the EM structure of the yeast tri-snRNP (Häcker et al., 2008; Figure I-4 B). The particles appear to adopt a Y-shaped structure containing head, foot and arm domains. Also, in this study it was possible to localize seven key proteins (Brr2, Prp8, Snu114, Prp6, Prp31, Prp3 and Lsm8) within the EM structure. With respect to U5 snRNP components, Brr2 is found in the head domain; Prp8 and Snu114 are found between the foot and head domains. Lsm8, a U6 snRNP protein, and Prp3, a U4 snRNP protein, are located in the arm region and are connected to the main body through the linker region containing Prp6 and Prp31, which are both components of U4/U6 di-snRNP. The most striking feature of the yeast tri-snRNP is that two alternative conformations - closed and open – are indicated. In the closed form Brr2 may be

in contact with U4/U6 snRNP in the arm domain so that Brr2 can carry out the U4/U6 unwinding helicase activity. Snu114, located in the central hinge region, may act as the switch between open and closed forms, depending on GTP/GDP-binding state (Fabrizio et al., 1997).

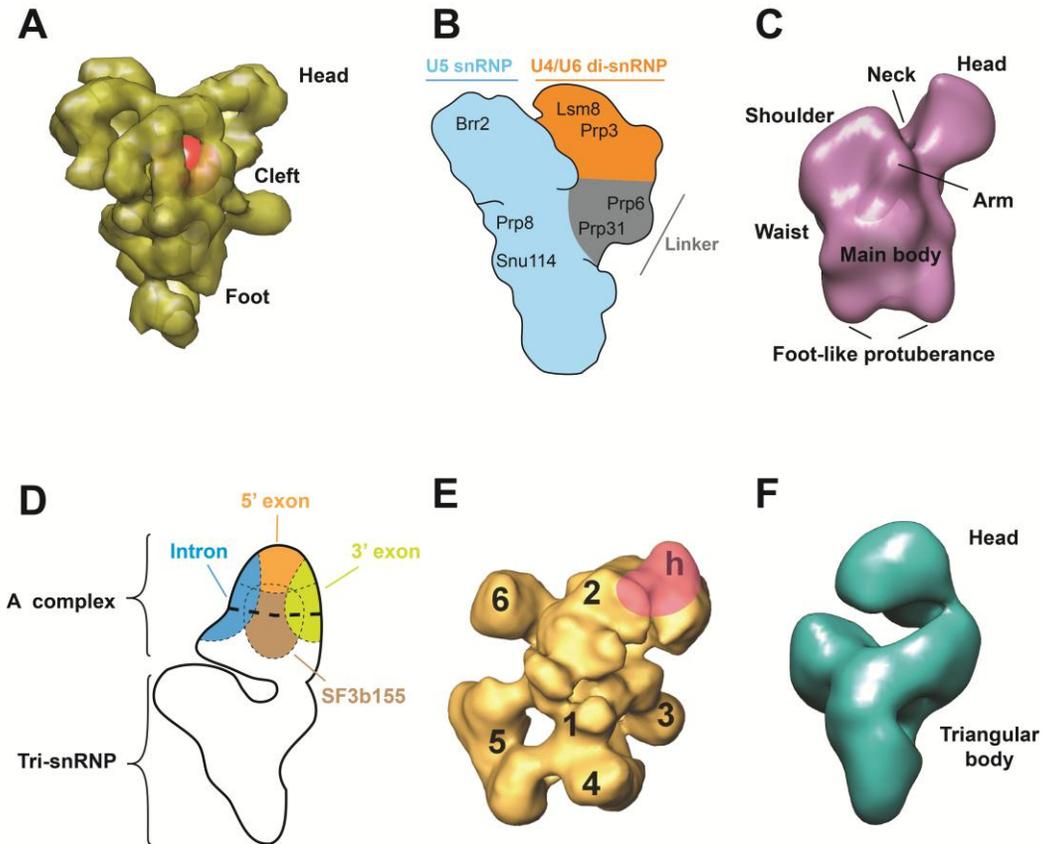


Figure I-4. EM structures and models of higher order complexes. A) 3D structure of human U4/U6 U5 tri-snRNP (Sander et al., 2006). The location of U5 snRNA loop I is shown by red sphere. B) Diagram of locations of important proteins in the ‘closed’ state of yeast tri-snRNP (Adapted from H äcker et al., 2008). C) 3D structure of human A complex (Bahzadnia et al., 2007). D) Diagram of locations of mRNA and SF3b155 within human B complex (Adapted from Wolf et al., 2009). E) 3D structures of human C complex (Golas et al., 2010). The position of 5' exon is shown in pink. F) 3D structure of human BAU1 complex (Boehringer et al., 2004). All EM 3D structures are downloaded from the EM database and built with chimera.

I-3.3. Structures of A, B and C complex

The 3D-EM structure of human A complex has been reported at a very low resolution (~40 Å) (Bahzadnia et al., 2007; Figure I-4 C). Very limited information can be obtained due to the low resolution. This includes the observation of the overall 260 Å length, two foot-like regions at the bottom which are connected to a globular main body, and a shoulder region at the top linked to a protruding head region via the neck domain. It is impossible to fit SF3b or other important components into the EM map. The A complex structure here is similar to the reported EM structure of U11/U12 di-snRNP, functional analogues of U1 and U2 (Golas et al., 2005).

The A complex is converted into the B complex with the addition of the U4/U6•U5 tri-snRNP. Recently, the EM structure of the human B complex was obtained (Wolf et al., 2009; Figure I-4 D). The B complex appears rhombic with a triangular body and a globular head region. The body can be divided into four subdomains: a central mass, a lower foot branch, an upper stump branch and a thinner neck branch between the body and the head. The head domain can also be divided into two parts: a globular head top and an elongated head base. The authors were also able to map the positions of pre-mRNA sequences and SF3b155 within the EM model by binding protein A-coated colloidal gold bound to the targets using specific antibodies. In summary, the 5' exon was mapped to the top of the globular head region, the intron was located more in the middle towards the stump side of the head, and the 3' exon was just on the opposite side to the 5' exon.

SF3b155, which binds U2 snRNP and the pre-mRNA to stabilize the association of U2 with the BPS, was found to span from the lower part of the globular domain to the base of the head. Since these components are at the interface between the head and the body, the catalytic centre of the spliceosome may be found in this region.

The C complex contains the active splicing centre and differs from the B complex in the formation of the new U2/U6 base-pair and disassociation of U1 and U4 snRNP from the spliceosome. Recently, a C complex with which U2, U5 and U6 snRNAs are stably associated was purified and 3D-EM structure was determined to a resolution of 20-29 Å (Golas et al., 2010; Figure I-4 E). The catalytically active C complex contains a central domain (domain 1) surrounded by several connected domains (domains 2-6). Domain 2 notably forms the top one-third of the complex and there is deep cleft between domains 1 and 2. The author also tried to map the position of the 5' exon into the C complex, and this corresponds to domain h as well as a small region in the upper part of domain 2 (Figure I-4 E).

So far no human B* complex has been isolated, but the EM structure of a similar BΔU1 in which the U1 snRNP and several non-snRNP proteins were absent has been reported (Boehringer et al., 2004; Figure I-4 F). The BΔU1 complex has a rhombic shape and is composed of a triangular main body and a globular head domain. When compared, the BΔU1 structure appears very similar to B complex, only smaller in the head domain. But the BΔU1 and C complex

have very different conformations, B Δ U1 being compact and the C complex adopting a multimodular appearance. Therefore, B* complex probably has the same overall architecture as B complex, and a structural rearrangement occurs during the B* to C transition. A similar result was observed in yeast B and C complexes (Fabrizio et al., 2009).

I-4. Structures of protein components of spliceosomes

To this point the majority of higher order spliceosomal complex structure studies depend on EM with the only exceptions of very recently successful U1 and U4 snRNP crystal structures. In order to fit secondary structure elements of proteins into the cryo-EM 3D density maps, a <10 Å resolution is required. Unfortunately this is beyond the resolution level that has been so far obtained for spliceosomes and snRNP complexes. However, high resolution crystal or NMR structures of individual spliceosome components have shed light on the dynamic process of spliceosome assembly. Some well-studied models will be outlined here.

I-4.1. SF1 and branch adenosine binding

Splicing factor SF1 (SF1) is involved in the ATP-dependent formation of the E complex. It binds specifically to the BPS as well as U2AF⁶⁵, the larger subunit of the U2AF dimer, and the latter association dramatically increases the SF1-BPS interaction cooperatively (Berglund et al., 1998). When the A complex

is formed, SF1 is thought to be replaced by the U2 snRNP protein SF3b155, the biggest subunit of SF3b (Gozani et al., 1998).

The NMR structure of SF1 bound to its target RNA shows that the protein recognizes pre-mRNA via a KH (hnRNP K homology)-QUA2 (Quaking homology 2) domain. (Liu et al., 2001; Figure 1-5 A) The KH domain is a ~70 amino acid single-stranded nucleic acid binding domain. The type I KH domain is found in eukaryotes and it adopts a conserved three anti-parallel stranded β -sheet abutted by three α -helices (Valverde et al., 2008). The QUA2 domain contains ~20 amino acids and it adds another α -helix to KH domain. The BPS RNA in the structure adopts an extended single-stranded conformation. The 3' part of the BPS (UAAC) is specifically recognized in a hydrophobic cleft formed by a Gly-Pro-Arg-Gly motif and the variable loop of the KH domain, and the QUA2 region binds the 5' nucleotides of the BPS (ACU). This KH-QUA2/BPS RNA complex is stabilized by a network of hydrophobic interactions, hydrogen bonding, and electrostatic contacts. Most remarkably, the branch adenosine (A8) forms hydrogen bonds with I177, which mimic the Watson-Crick A-U base-pairing. This interaction buries the branch adenosine in an orientation that may facilitate the following formation of the U2-BPS duplex with the adenosine excluded from the complex.

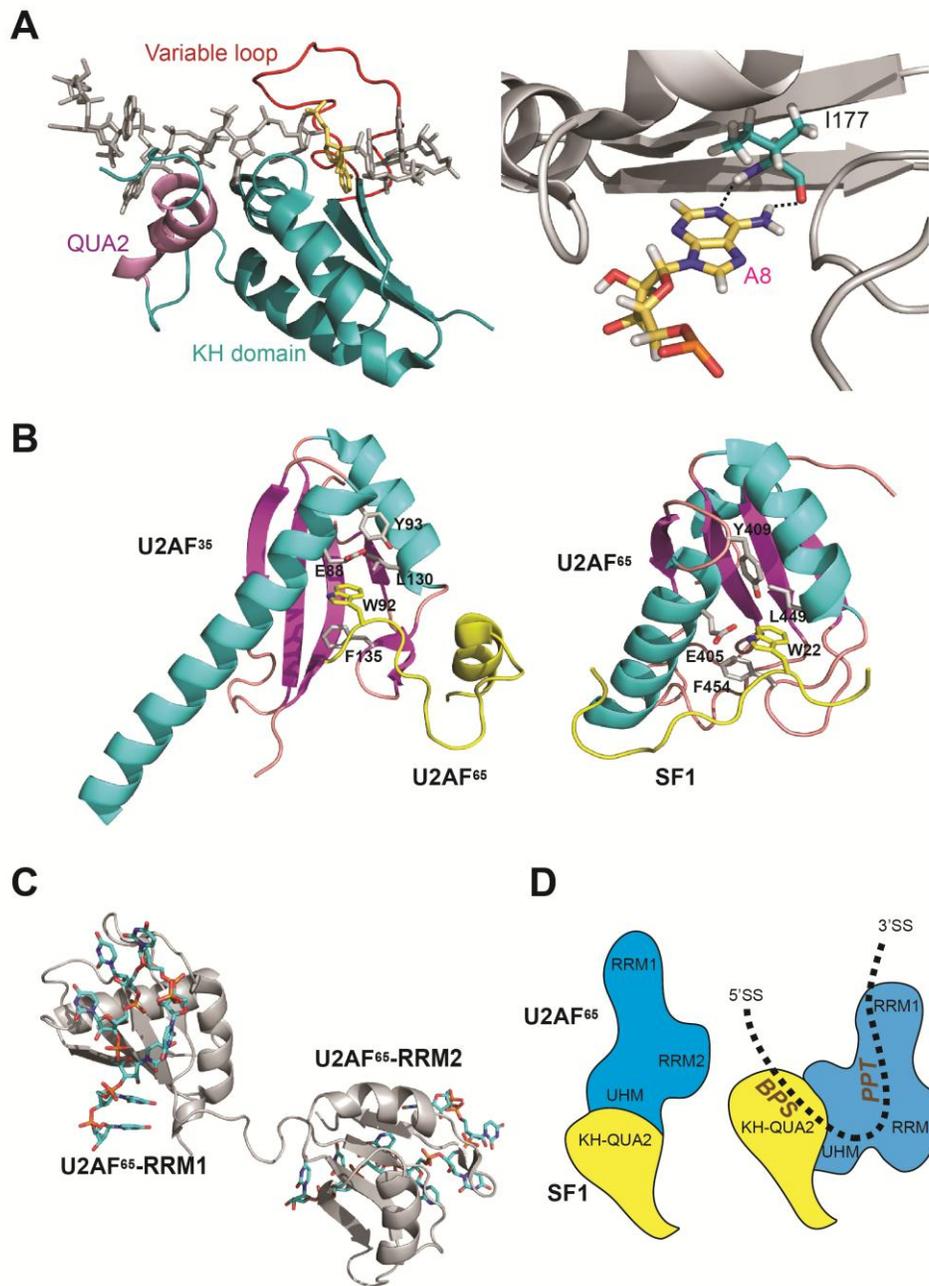


Figure I-5. SF1 and U2AF structures. A) Structure of SF1 complexed with the BPS (Liu et al., 2001). The KH (cyan), QUA2 (violet) and variable loop involved in RNA recognition are indicated (*left*). The hydrogen bond between A8 and I177 mimicking the Watson-Crick A-U base-pairing is detailed (*right*). B) Tryptophan recognition by UHM (Selenko et al., 2003). In both the U2AF³⁵ UHM in complex with the proline-rich region of the N terminus of U2AF⁶⁵ (*left*) and the U2AF⁶⁵-UHM/SF1 complex (*right*), the tryptophan makes hydrophobic interactions with Glu, Tyr and Leu residues and stacks with a conserved Phe. C) The model of U2AF⁶⁵ binding to poly (U) RNA. D) A suggested model for the changes observed following binding of the splice site RNA in the SF1/U2AF⁶⁵/3'SS complex (Adapted from Gupta et al., 2011).

I-4.2. U2AF: PPT recognition and protein-protein interaction

U2 auxiliary factor (U2AF) is composed of a large and a small subunit, U2AF⁶⁵ and U2AF³⁵. It is a non-snRNP protein required for the binding of U2 snRNP to the pre-mRNA branch site. U2AF⁶⁵ contains an N-terminal arginine-serine rich (RS) domain followed by two RNA recognition motif (RRM) domains and a U2AF homology motif (UHM) domain. Crystal structures have been reported for both the RRM1, 2/poly (U) complex and RRM1 in the absence of RNA substrate (Sickmier et al., 2006; Thickman et al., 2007).

The RRM domain is one of the most abundant RNA binding domains in eukaryotes. A typical RRM motif has a ~80 residue core composed of four β -strands packed against two α -helices. Two consensus sequences on the β_3 and β_1 , called RNP1 (conserved K/R-G-F/Y-G/A-F/Y-V/I/L-X-F/Y) and RNP2 (less conserved I/V/L-F/Y-I/V/L-X-N-L) respectively, are important for RNA binding (reviewed in: Cassola et al., 2010). The U2AF⁶⁵ RRM1, 2 is shown to recognize the poly (U) strand by forming unique hydrogen bonds with the uracils. Many of these hydrogen bonds are formed with side chains, which provide the flexibility to tolerate various PPT sequences within metazoans. In addition, two water molecules mediating these hydrogen bonds may also contribute to accommodate the PPT sequence diversity. When compared to the RRM1 alone structure, alternative conformations of three side chains are stabilized (R150, K225, R227) and a conformational change occurs within a flexible loop connecting the β_2/β_3 strands when bound to RNA.

In the RRM1,2/poly (U) complex structure, the RNA is bound by both RRMs but with each domain provided by different protein monomers. Since there is a 20-residue deletion within the inter-RRM linker, the arrangement of U2AF⁶⁵ on the PPT remains unclear (Sickmier et al., 2006). Thus the same group used small angle X-ray scattering (SAXS) technique to address this question (Jenkins et al., 2008). The result favours the proposed model that the RRM1 and RRM2 form an extended structure and act independently to bind the PPT, rather than binding in a tightly coupled manner (Figure I-5 C).

The UHM domain of U2AF⁶⁵ was originally identified as a third RRM motif within U2AF (reviewed in: Kielkopf et al., 2004). However, the RNP1 and RNP2 consensus sequence are not conserved in this domain and UHM domains do not bind RNA. Instead, the U2AF⁶⁵ UHM was later shown to mediate protein-protein interactions by binding UHM ligand motifs (ULMs), which contain a (R/K)_nXRW(DE) consensus sequence. The first example of this interaction comes from the X-ray structure of U2AF⁶⁵/U2AF³⁵ heterodimer (Kielkopf et al., 2001; Figure I-5 B). The UHM domain in the middle U2AF³⁵ was co-crystalized with a proline-rich peptide from U2AF⁶⁵. The structure revealed that subunits of the heterodimer are closely associated by a network of interactions around two tryptophans, each from a distinct subunit and recognized by the other subunit. These interactions include aromatic stacking with each of the tryptophans, as well as hydrogen bonds and salt bridges. This tryptophan recognition is found again in the following NMR structure of the U2AF⁶⁵/SF1 complex, although not in the

same reciprocal manner as in U2AF⁶⁵/U2AF³⁵ (Selenko, 2003). In this study, an N-terminal peptide of SF1 binds to the UHM of U2AF⁶⁵ by inserting the tryptophan into the hydrophobic pocket in a strikingly similar way to that observed in the U2AF heterodimer, but with a much lower affinity. Therefore, the association via UHM ULM interaction helps to recruit the U2AF dimer to the BPS-bound SF1 in such a way that SF1-U2AF interaction is only transient, while U2AF⁶⁵/U2AF³⁵ dimer is rather tight. Since the original studies, UHM-mediated interactions have been reported in several other splicing factors (Corsini et al., 2007; Manceau 2008; Corsini et al., 2008). The most notable one involves SF3b155, a component of U2 snRNP. SF3b155 has several ULM domains in the N-terminus region and it potentially act as platform for A complex assembly by binding multiple UHM-containing proteins (Thickman et al., 2006).

Recently, the SAXS study of SF1/U2AF⁶⁵ and SF1/U2AF⁶⁵/RNA complexes has yielded some interesting results (Gupta et al., 2011). While no significant changes occurred when SF1 and U2AF⁶⁵ form complexes with the RNA substrate individually, the molecular dimensions of SF1/U2AF⁶⁵/RNA contracted greatly, compared to the SF1/U2AF⁶⁵ complex alone (Figure I-5 D). These changes probably correspond to a bend in the 3'SS conformation and help to position the RNA substrate in the optimal position for the subsequent stages of splicing to occur. This is consistent with a model of substrate pre-organization originally proposed on the basis of footprinting experiments (Kent and MacMillan, 2002).

I-4.3. p14 SF3b155 complex

After U2 snRNP is recruited to the E complex, U2 snRNA forms an imperfect duplex with the BPS and bulges out the branch adenosine. SF1 is excluded and p14, also termed SF3b14, which is a subunit of SF3b complex, comes into contact with the excluded adenosine (reviewed in: Kuwasako et al., 2008). An RNA p14 cross-link was observed in A complex, and p14 was the only protein shown to cross-link to the branch adenosine through A, B and C complex (MacMillan et al., 1994). Also, p14 was shown to associate tightly with SF3b155 (Gozani et al., 1998).

Initial trials to crystalize p14 yielded poor results, as p14 needs to bind SF3b155 to fold properly. Ultimately, a p14/SF3b155 peptide complex was purified and the X-ray structure was determined (Schellenberg et al., 2006; Figure I-6 A). p14 adopts a canonical RNA RRM fold ($\beta\alpha\beta\beta\alpha\beta$), with an extensive network of hydrophobic and hydrophilic interactions formed. In addition, the N-terminal part of the SF3b155 fragment and the C-terminal extended helix of p14 ($\alpha 3$) block the β -sheet surface of the RRM fold. Considering p14 interacts with a bulged duplex instead of a single-stranded RNA, the p14 RNA complex probably represents a non-canonical RRM-RNA interaction pattern. Indeed, four basic residues (R24, R57, R96 and K100) form a surface pocket and a tyrosine residue (Y22) resides at the base of this pocket. When this Tyr was mutated, its association with branch A was dramatically reduced, suggesting the potential aromatic stacking between Y22 and the branch A. The NMR structure of a

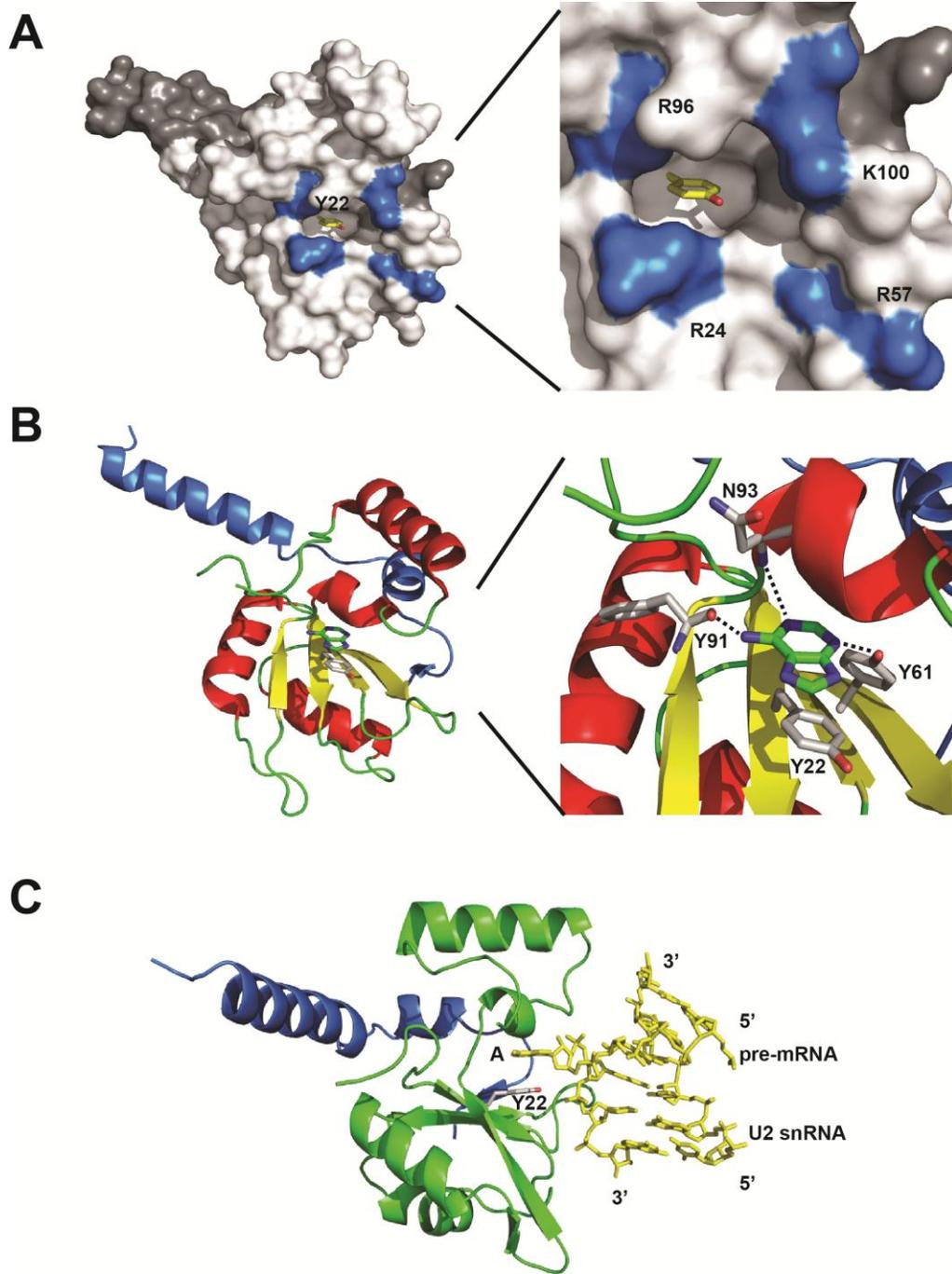


Figure I-6. p14 SF3b155 complex structures. A) Surface representation of p14 SF3b155 complex, showing Y22 within a surface pocket surrounded by basic residues (Schellenberg et al., 2006). B) Structure of p14 SF3b155 adenosine complex, showing the recognition of branch adenosine. The p14 structure is shown in red and yellow; the SF3b155 peptide is shown in blue (Schellenberg et al., 2011). C) Model of p14 SF3b155 RNA interaction suggested by disulphide tethering and SAXS.

p14/SF3b155 complex was later solved and elicited the same model (Kuwasado et al., 2008). Subsequent to the original structure, our lab made a series of biochemical and structural studies of RNA p14 complexes, which have yielded a more detailed model of the p14 RNA interaction (Schellenberg et al., 2011; Figure I-6 B and C).

I-4.4. Other spliceosome-related structures

Several structures of spliceosomal proteins containing RRM have been reported. Prp24 is the only protein component of the U6 snRNP other than the core Lsm proteins. It binds to the ISL region of U6 snRNA and helps to anneal the U4/U6 RNAs (Ghetti et al., 1995). Prp24 has four RRM domains. The crystal structure of the first three RRMs and the NMR structure of RRM2 bound to U6 snRNA have been solved (Rae et al., 2007; Martin-Tomasz et al., 2010). RRM1-3 were shown to form multiple intermolecular interactions, with RRM2 binding the AGAGAU sequence on U6 snRNA. RRM1 was in a position to interact with the internal helix of U6 snRNA and destabilize it.

Another RRM-containing protein is hnRNP A1, which functions in splicing, transcription, and telomere maintenance (Ding et al., 1999). In the crystal structure of hnRNP A1 complexed with single-stranded DNA, the two RRMs bind the substrate in an anti-parallel fashion. The most striking feature is that no

interaction is specific to DNA or RNA, which explains the involvement of hnRNP A1 in both DNA and RNA recognition.

A third example of RRM structure is Fox-1, an alternative splicing factor (Auweter et al, 2006). The NMR structure of Fox-1 bound to the UGCAUGU sequence shows an unusual recognition mechanism. The last three nucleotides, UGU, bind to the RRM in the canonical fashion, while the first three nucleotides, UGC, are wrapped around a Phe residue with G2 and A4 forming a base-pair. These unique structural features confer a very high specificity to Fox-1/RNA recognition and a single nucleotide mutation can dramatically reduce their association. These proteins, along with SF1 and U2AF⁶⁵, clearly demonstrate both the similarity and polymorphism of RRM motifs. In general, they each contain conserved tyrosines or phenylalanines which form stacking interactions with nucleotides, while the substrate specificity is determined by hydrogen-bonding interactions (reviewed in: Ritchie et al., 2009).

As mentioned earlier in this chapter, spliceosomes contain several members of the RNA-dependent ATPases of the DExD/H family. Despite their functional importance, little is known about their structures. The only evidence comes from the Brr2 protein which is responsible for U4/U6 unwinding during spliceosomal activation (Zhang et al., 2009; Pena et al., 2009). Brr2 is comprised of an N-terminal domain and two consecutive Hel308-like modules (Hel308-I and Hel308-II). The putative β -hairpin region in Hel308-I was shown to be functionally important for helicase activity, and Brr2 possibly uses an unwinding

mechanism similar to Hel308. Brr2 can cooperatively bind ATP and the duplex region of substrate RNA. On binding, the protein induces a sharp bend in one strand disrupting some base-pairs nearby. Subsequently, ATP is hydrolyzed and the unwound RNA is released (Sengoku et al, 2006). The Hel308-II domain is also shown to facilitate the binding of Brr2 with Prp8 C-terminal domain (CTR) and this association is required to couple ATP hydrolysis by Brr2 to RNA unwinding.

So far only the EM structure is available for U4/U6•U5 tri-snRNP, but a high resolution of part of this complex, consisting of the U4 snRNA 5' stem-loop, part of the Prp31 protein and the 15.5K protein has been reported (Liu et al., 2007). This structure clearly showed how 15.5K binds to and stabilizes the 5'-SL of the U4 snRNA in a conformation to favor a composite RNP binding site for Prp31. The latter then stabilizes the U4/U6•U5 tri-snRNP by interacting with Prp6 via a separate coiled-coil domain. Prp6 links Prp31 to Brr2 and Snu114 that are crucial for both spliceosome activation and disassembly.

Similarly, no high resolution structure of U4/U6 di-snRNP is available yet, but a crystal structure of Cyclophilin (CypH) complexed with U4/U6-60K (a homologue of Prp4) has been solved (Reidt et al., 2003). This structure indicates that Prp4 binds to CypH opposite to the CypH active site, which is occupied by the N-terminus of a neighbouring CypH. This interaction then recruits Prp3 and a hetero-trimer forms in the pre-spliceosome. Later, Prp4 leaves with the U4

snRNA and CypH binds to Prp18, which also contains a Prp4-like sequence and is required for the second transesterification step (Horowitz et al., 2002).

I-5. Structure of RNA components of spliceosome

Divalent metal ions are required for a variety of protein phosphoryltransfer enzymes, as well as group I, group II and protein-dependent splicing. Different models have been proposed for the function of metals in the active site. The first complete structural evidence in this respect comes from the crystal structure of the group I intron (Adams et al., 2004; Figure I-7 A). This structure clearly supports the two-metal-ion model (Steitz and Steitz, 1993). The two metals are well positioned to promote catalysis of the exon ligation reaction, with M_1 activating the nucleophile and M_2 stabilizing the leaving group, and both metals coordinating the scissile phosphate (Figure I-7 B). This structure suggests how the second splicing reaction is carried out. The reversible nature of splicing suggests the roles of nucleophilic and leaving group activation are also likely to be reversed between the two steps of the splicing (Stahley and Strobel, 2005). This model has subsequently been supported by a number of structural studies (Stahley and Strobel, 2005; Lipchock and Strobel, 2008).

A similar two-metal-ion mechanism has been proposed for both the group II introns and the spliceosome-dependent splicing. In 2008, the 3.1 Å crystal structure of a group II intron was solved (Toor et al., 2008; Figure I-7 A). Indeed,

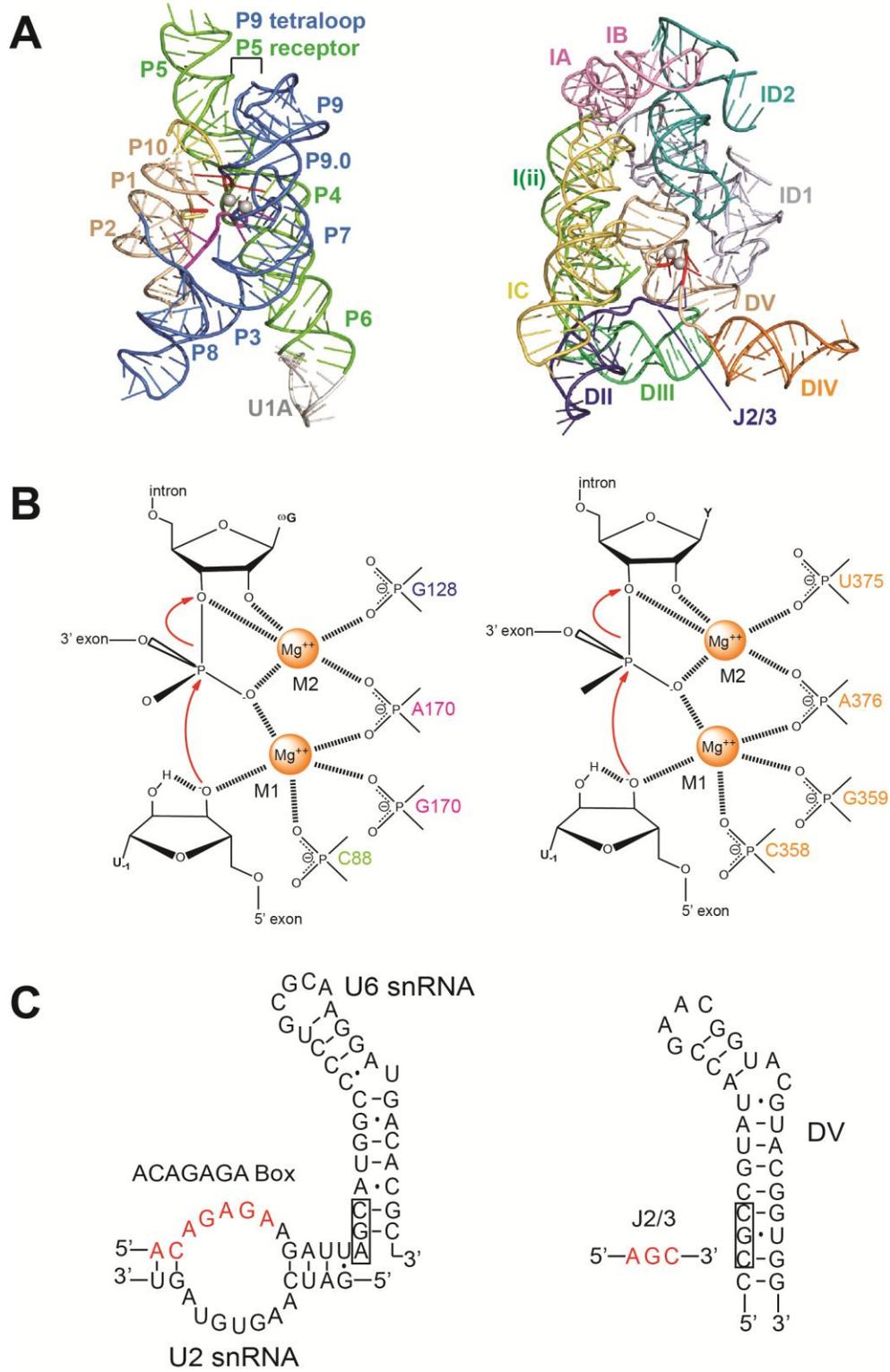


Figure I-7. (Legend on next page)

two metals were located on the surface of domain V (DV) and accessible to both the 5' and 3' splice sites. When compared to the group I active site, the coordination of the two metals reveals strong similarities, suggesting a convergent evolution (Figure I-7 B). As well, there are additional divalent metal ions within the core. Although not directly involved in catalysis, these may play important roles in stabilizing the structure and modulating the electrostatic environment of the core (review in: Toor et al., 2009).

The group II intron structure also provides information about the topology and tertiary RNA interaction network which have been suggested by previous biochemical studies (Costa et al., 1997; Boudvillain and Pyle, 1998; Gordon and Piccirilli, 2001). Domain I (DI) forms a scaffold and makes numerous interactions with other domains. The active site is docked within the large cavity of DV by a series of contacts. Among these are the base stacking interaction between ζ - ζ' , a conserved sequence on DI, with the GAAC tetraloop on DV, the docking of the bulge and lower helix of DV by associating with the Z-anchor motif and κ - κ' on DI, and the stacking between G5 at the 5' end of the intron with DV bulge base A376. Perhaps the most important feature of this structure is the triple helix

Figure I-7. Group I and group II intron structures. A) Crystal structures of group I intron (Adams et al., 2004) (*left*) and group II intron (Toor et al., 2008) (*right*). The two divalent metals at the active site are shown by grey spheres. B) A close look at the active sites of group I intron (*left*) and group II intron (*right*). C) Secondary structures of the U6 snRNA (*left*) and domain V of group II intron (*right*). The catalytic triads are indicated with boxes. The ACAGGA box (red) may form base triple interactions with the triad of the U6 snRNA, analogous to that observed between J2/3 and domain V (Adapted from Toor et al., 2009).

formed between J2/3, a highly conserved linker region between domains II and III, with the DV catalytic triad (usually AGC but often CGC). The first two nucleotides of the catalytic triad (5'-CGC) form a triple interaction with the corresponding J2/3 region (5'-AGC).

The J2/3 triad structure is reminiscent of a homolog within the U2/U6 snRNA (Sashital et al., 2004; Figure I-7 C). DV of the group II intron is a highly conserved hairpin loop structure that contains a small “elbow” or bulge of two bases (AC) which are both involved in metal binding (Schmidt et al., 1996), whereas U6 snRNA forms a stem-loop region (U6 ISL) with a single bulged U serving as a metal binding site as well. U6 snRNA also has the AGC catalytic triad, which was proposed to form an intramolecular base-pair, extend the U6 ISL and be juxtaposed with the conserved ACAGAGA box during the first step of splicing (reviewed in: Valadkhan, 2005) . In the second step, this triad has been proposed to form helix Ib with U2 snRNA, allowing an interaction between the U2 bulge and the ACAGAGA sequence. A recent study challenged this model by showing that helix Ib is important for both steps of splicing (Mefford and Staley, 2009). Also, the U2/U6 duplex was shown to adopt at least three distinct conformations in equilibrium (Guo et al., 2009). Therefore, more evidence is required in the context of the spliceosome before the detailed U2/U6 and active site picture can be drawn.

It has long been debated whether spliceosomes are ribozymes like the self-splicing introns or protei-catalyzed machines. Indeed, the existing data suggests

strong parallels between the arrangement of the spliceosomal active site and that of group II introns. Recently, a protein-free system containing only U2/U6 snRNA and two short RNA substrates was shown to carry out the second step of splicing (Valadkhan et al., 2009). This result strongly argues that the spliceosome is a ribozyme.

So what is the possible role of proteins within the spliceosome active site? Many ribozymes are associated with proteins *in vivo* which improve their catalytic activities by stabilizing the RNA structures as well as assisting the binding of metal ions (Hsieh et al., 2003). In addition, the spliceosomal snRNAs are unusually small for catalyzing the phosphodiester bond cleavage reactions, without enough tertiary structures formed between distinct regions. For example, group I and group II introns are at least two folds longer than the combined length of U2 and U6 snRNAs (reviewed in: Valadkhan, 2010). Indeed, the splicing reaction in the protein-free system is significantly slower than that in spliceosome-catalyzed systems (Valadkhan and Manley, 2001; Valadkhan et al., 2007). Therefore it is conceivable that the snRNAs can only form an inefficient ribozyme and other proteins factors are required for stable positioning of the active site elements. A leading candidate for this purpose is the U5 snRNP component Prp8 (Collins and Guthrie, 2000; Grainger and Beggs, 2005).

I-6. Summary

In the last two decades considerable progress has been achieved in understanding the structures of spliceosome components. X-ray, NMR and EM techniques have been employed to elucidate spliceosomal structures at different levels. However, considering the intricate and dynamic nature of spliceosome assembly, a tremendous amount of effort must be made before the delicate scenario can be elucidated piece by piece.

The following chapters of this thesis will focus on the biochemical and structural studies of a potential spliceosomal active site protein, the U5 snRNP component Prp8. The results presented here shed light on the nature of the spliceosomal active site as well as conformational changes between different steps along the splicing pathway.

I-7. References

Adams, P.L., Stahley, M.R., Kosek, A.B., Wang, J., and Strobel, S.A. (2004). Crystal structure of a self-splicing group I intron with both exons. *Nature* **430**, 45-50.

Ares, M. Jr., and Weiser, B. (1995). Rearrangement of snRNA structure during assembly and function of the spliceosome. *Prog. Nucleic Acid Res. Mol. Biol.* **50**, 131-159.

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

Bae, E., Reiter, N.J., Bingman, C.A., Kwan, S.S., Lee, D., Phillips, G.N., Jr.,

- Butcher, S.E., and Brow, D.A. (2007). Structure and interactions of the first three RNA recognition motifs of splicing factor prp24. *J. Mol. Biol.* **367**, 1447-1458.
- Barash, Y., Calarco, J.A., Gao, W., Pan, Q., Wang, X., Shai, O., Blencowe, B.J., and Frey, B.J. (2010). Deciphering the splicing code. *Nature* **465**, 53-59.
- Behzadnia, N., Golas, M.M., Hartmuth, K., Sander, B., Kastner, B., Deckert, J., Dube, P., Will, C.L., Urlaub, H., Stark, H., and Luhrmann, R. (2007). Composition and three-dimensional EM structure of double affinity-purified, human prespliceosomal A complexes. *EMBO J.* **26**, 1737-1748.
- Berget, S.M., Moore, C., and Sharp, P.A. (1977). Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc. Natl. Acad. Sci. U.S.A.* **74**, 3171-3175.
- Berglund, J.A., Abovich, N., and Rosbash, M. (1998). A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. *Genes Dev.* **12**, 858-867.
- Berglund, J.A., Chua, K., Abovich, N., Reed, R., and Rosbash, M. (1997). The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. *Cell* **89**, 781-787.
- Bindereif, A., Wolff, T., and Green, M.R. (1990). Discrete domains of human U6 snRNA required for the assembly of U4/U6 snRNP and splicing complexes. *EMBO J.* **9**, 251-255.
- Boehringer, D., Makarov, E.M., Sander, B., Makarova, O.V., Kastner, B., Luhrmann, R., and Stark, H. (2004). Three-dimensional structure of a pre-catalytic human spliceosomal complex B. *Nat. Struct. Mol. Biol.* **11**, 463-468.
- Bonen, L., and Vogel, J. (2001). The ins and outs of group II introns. *Trends Genet.* **17**, 322-331.
- Boudvillain, M., and Pyle, A.M. (1998). Defining functional groups, core structural features and inter-domain tertiary contacts essential for group II intron self-splicing: a NAIM analysis. *EMBO J* **17**, 7091-7104.
- Brow, D.A. (2002). Allosteric cascade of spliceosome activation. *Annu. Rev. Genet.* **36**, 333-360.
- Calvin, K., and Li, H. (2008). RNA-splicing endonuclease structure and function. *Cell. Mol. Life Sci.* **65**, 1176-1185.
- Cassola, A., Noe, G., and Frasch, A.C. (2010). RNA recognition motifs involved

- in nuclear import of RNA-binding proteins. *RNA Biol.* **7**, 339-344.
- Cech, T.R. (1986). The generality of self-splicing RNA: relationship to nuclear mRNA splicing. *Cell* **44**, 207-210.
- 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., Hothorn, M., Stier, G., Rybin, V., Scheffzek, K., Gibson, T.J., and Sattler, M. (2008). Dimerization and protein binding specificity of the U2AF homology motif of the splicing factor Puf60. *J. Biol. Chem.* **284**, 630-639.
- Costa, M., Deme, E., Jacquier, A., and Michel, F. (1997). Multiple tertiary interactions involving domain II of group II self-splicing introns. *J. Mol. Biol.* **267**, 520-536.
- Ding, J., Hayashi, M.K., Zhang, Y., Manche, L., Krainer, A.R., and Xu, R.M. (1999). Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. *Genes Dev.* **13**, 1102-1115.
- 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.
- Fabrizio, P., Dannenberg, J., Dube, P., Kastner, B., Stark, H., Urlaub, H., and Lührmann, R. (2009). The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. *Mol. Cell* **36**, 593-608.
- Fabrizio, P., Laggerbauer, B., Lauber, J., Lane, W.S., and Luhrmann, R. (1997). An evolutionarily conserved U5 snRNP-specific protein is a GTP-binding factor closely related to the ribosomal translocase EF-2. *EMBO J.* **16**, 4092-4106.
- Ghetti, A., Company, M., and Abelson, J. (1995). Specificity of Prp24 binding to RNA: a role for Prp24 in the dynamic interaction of U4 and U6 snRNAs. *RNA* **1**, 132-145.
- Golas, M.M., Sander, B., Bessonov, S., Grote, M., Wolf, E., Kastner, B., Stark, H., and Lührmann, R. (2010). 3D cryo-EM Structure of an active step I spliceosome and localization of its catalytic core. *Mol. Cell* **40**, 927-938.

- 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.
- Gordon, P.M., and Piccirilli, J.A. (2001). Metal ion coordination by the AGC triad in domain 5 contributes to group II intron catalysis. *Nat. Struct. Biol.* **8**, 893-898.
- Gozani, O., Feld, R., and Reed, R. (1996). Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A. *Genes Dev.* **10**, 233-243.
- 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.
- Grainger, R.J., and Beggs, J.D. (2005). Prp8 protein: At the heart of the spliceosome. *RNA* **11**, 533-557.
- Guo, Z., Karunatilaka, K.S., and Rueda, D. (2009). Single-molecule analysis of protein-free U2–U6 snRNAs. *Nat. Struct. Mol. Biol.* **16**, 1154-1159.
- Gupta, A., Jenkins, J.L., and Kielkopf, C.L. (2011). RNA induces conformational changes in the SF1/U2AF65 splicing factor complex. *J. Mol. Biol.* **405**, 1128-1138.
- Guthrie, C., and Patterson, B. (1988). Spliceosomal snRNAs. *Annu. Rev. Genet.* **22**, 387-419.
- Häcker, I., Sander, B., Golas, M.M., Wolf, E., Karagöz, E., Kastner, B., Stark, H., Fabrizio, P., and Lührmann, R. (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.
- Horowitz, D.S., Lee, E.J., Mabon, S.A., and Misteli, T. (2002). A cyclophilin functions in pre-mRNA splicing. *EMBO J.* **21**, 470-480.
- Hsieh, J., Andrews, A.J., and Fierke, C.A. (2004). Roles of protein subunits in RNA-protein complexes: lessons from ribonuclease P. *Biopolymers* **73**, 79-89.
- 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., Sousa, D., Moore, M.J., and Grigorieff, N. (2004). Three-dimensional structure of C complex spliceosomes by electron microscopy. *Nat. Struct. Mol. Biol.* **11**, 265-269.

Kambach, C., Walke, S., Young, R., Avis, J.M., de la Fortelle, E., Raker, V.A., Luhrmann, 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., Kornstadt, U., Bach, M., and Luhrmann, R. (1992). Structure of the small nuclear RNP particle U1: identification of the two structural protuberances with RNP-antigens A and 70K. *J. Cell Biol.* **116**, 839-849.

Kattah, N.H., Kattah, M.G., and Utz, P.J. (2010). The U1-snRNP complex: structural properties relating to autoimmune pathogenesis in rheumatic diseases. *Immunol. Rev.* **233**, 126-145.

Keating, K.S., Toor, N., Perlman, P.S., and Pyle, A.M. (2009). A structural analysis of the group II intron active site and implications for the spliceosome. *RNA* **16**, 1-9.

Kent, O.A., and MacMillan, A.M. (2002). Early organization of pre-mRNA during spliceosome assembly. *Nat. Struct. Biol.* **9**, 576-581.

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.

Koodathingal, P., Novak, T., Piccirilli, J.A., and 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.

Krol, A., Westhof, E., Bach, M., Luhrmann, R., Ebel, J.P., and Carbon, P. (1990). Solution structure of human U1 snRNA. Derivation of a possible three-dimensional model. *Nucleic Acids Res.* **18**, 3803-3811.

Krummel, D.A.P., Nagai, K., and Oubridge, C. (2010). Structure of spliceosomal ribonucleoproteins. *F1000 Biol. Rep.*

- Kuwasako, K., Dohmae, N., Inoue, M., Shirouzu, M., Taguchi, S., Güntert, P., S raphin, B., Muto, Y., and 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., and Stevens, S.W. (2010). Release of SF3 from the intron branchpoint activates the first step of pre-mRNA splicing. *RNA* **16**, 516-528.
- Legrain, P., Seraphin, B., and Rosbash, M. (1988). Early commitment of yeast pre-mRNA to the spliceosome pathway. *Mol. Cell. Biol.* **8**, 3755-3760.
- Leung, A.K.W., Nagai, K., and Li, J. (2011). Structure of the spliceosomal U4 snRNP core domain and its implication for snRNP biogenesis. *Nature* **473**, 536-539.
- Li, Z., and Brow, D.A. (1996). A spontaneous duplication in U6 spliceosomal RNA uncouples the early and late functions of the ACAGA element in vivo. *RNA* **2**, 879-894.
- Lipchock, S.V., and Strobel, S.A. (2008). A relaxed active site after exon ligation by the group I intron. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 5699-5704.
- Liu, S., Li, P., Dybkov, O., Nottrott, S., Hartmuth, K., Luhrmann, R., Carlomagno, T., and Wahl, M.C. (2007). Binding of the human Prp31 Nop domain to a composite RNA-protein platform in U4 snRNP. *Science* **316**, 115-120.
- Liu, Z., Luyten, I., Bottomley, M.J., Messias, A.C., Houngrinou-Molango, S., Sprangers, R., Zanier, K., Kr mer, A., and Sattler M. (2001). Structural basis for recognition of the intron branch site RNA by splicing factor 1. *Science* **294**, 1098-1102.
- L hrmann, R., and Stark, H. (2009). Structural mapping of spliceosomes by electron microscopy. *Curr. Opin. Struct. Biol.* **19**, 96-102.
- 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. (1994). Dynamic RNA-RNA interactions in the spliceosome. *Annu. Rev. Genet.* **28**, 1-26.
- Manceau, V., Kielkopf, C.L., Sobel, A., and Maucuer, A. (2008). Different

requirements of the kinase and UHM domains of KIS for its nuclear localization and binding to splicing factors. *J. Mol. Biol.* **381**, 748-762.

Martin-Tumasch, S., Reiter, N.J., Brow, D.A., and Butcher, S.E. (2010). Structure and functional implications of a complex containing a segment of U6 RNA bound by a domain of Prp24. *RNA* **16**, 792-804.

McConnell, T.S., and Steitz, J.A. (2001). Proximity of the invariant loop of U5 snRNA to the second intron residue during pre-mRNA splicing. *EMBO J.* **20**, 3577-3586.

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.

Muto, Y., Krummel, D.P., Oubridge, C., Hernandez, H., Robinson, C.V., Neuhaus, D., and Nagai, K. (2004). The structure and biochemical properties of the human spliceosomal protein U1C. *J. Mol. Biol.* **341**, 185-198.

Nelissen, R.L., Will, C.L., van Venrooij, W.J., and Luhrmann, 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., and Nagai, K. (2010). Structural studies of the spliceosome: blind men and an elephant. *Curr. Opin. Struct. Biol.* **20**, 82-89.

Nielsen, H., and Johansen, S.D. (2009). Group I introns: Moving in new directions. *RNA Biol.* **6**, 375-383.

Noble, S.M., and Guthrie, C. (1996). Identification of novel genes required for yeast pre-mRNA splicing by means of cold-sensitive mutations. *Genetics* **143**, 67-80.

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.

Pandit, S., Lynn, B., and Rymond, B.C. (2006). Inhibition of a spliceosome turnover pathway suppresses splicing defects. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 13700-13705.

Pelham, R.J., and Chang, F. (2002). Actin dynamics in the contractile ring during cytokinesis in fission yeast. *Nature* **419**, 82-86.

- Pena, V., Jovin, S.M., Fabrizio, P., Orlowski, J., Bujnicki, J.M., Lührmann, R., and Wahl, M.C. (2009). Common design principles in the spliceosomal RNA helicase Brr2 and in the Hel308 DNA helicase. *Mol. Cell* **35**, 454-466.
- Perriman, R., and Ares Jr, M. (2010). Invariant U2 snRNA Nucleotides Form a Stem Loop to Recognize the Intron Early in Splicing. *Molecular Cell* **38**, 416-427.
- Pomeranz Krummel, D.A., Oubridge, C., Leung, A.K.W., Li, J., and Nagai, K. (2009). Crystal structure of human spliceosomal U1 snRNP at 5.5 Å resolution. *Nature* **458**, 475-480.
- Pyle, A.M. (2010). The tertiary structure of group II introns: implications for biological function and evolution. *Crit. Rev. Biochem. Mol. Biol.* **45**, 215-232.
- Query, C.C., Bentley, R.C., and Keene, J.D. (1989). A common RNA recognition motif identified within a defined U1 RNA binding domain of the 70K U1 snRNP protein. *Cell* **57**, 89-101.
- 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.
- Reidt, U., Wahl, M.C., Fasshauer, D., Horowitz, D.S., Luhrmann, R., and Ficner, R. (2003). Crystal structure of a complex between human spliceosomal cyclophilin H and a U4/U6 snRNP-60K peptide. *J. Mol. Biol.* **331**, 45-56.
- Rio, D.C. (1993). Splicing of pre-mRNA: mechanism, regulation and role in development. *Curr. Opin. Genet. Dev.* **3**, 574-584.
- Ritchie, D.B., Schellenberg, M.J., and MacMillan, A.M. (2009). Spliceosome structure: Piece by piece. *Biochim. Biophys. Acta.* **1789**, 624-633.
- Sander, B., Golas, M.M., Makarov, E.M., Brahm, H., Kastner, B., Lührmann, R., and Stark, H. (2006). Organization of core spliceosomal components U5 snRNA loop I and U4/U6 di-snRNP within U4/U6.U5 tri-snRNP as revealed by electron cryomicroscopy. *Mol. Cell* **24**, 267-278.
- Sashital, D.G., Cornilescu, G., and Butcher, S.E. (2004). U2-U6 RNA folding reveals a group II intron-like domain and a four-helix junction. *Nat. Struct. Mol. Biol.* **11**, 1237-1242.
- Schellenberg, M.J., Edwards, R.A., Ritchie, D.B., Kent, O.A., Golas, M.M., Stark, H., Luhrmann, R., Glover, J.N., and MacMillan, A.M. (2006). Crystal structure of a core spliceosomal protein interface. *Proc. Natl. Acad. Sci. USA* **103**, 1266-1271.

Schellenberg, M.J., Dul, E.L., and MacMillan, A.M. (2010). Structural model of the p14/SF3b155 branch duplex complex. *RNA* **17**, 155-165.

Scherly, D., Boelens, W., Dathan, N.A., van Venrooij, W.J., and Mattaj, I.W. (1990). Major determinants of the specificity of interaction between small nuclear ribonucleoproteins U1A and U2B'' and their cognate RNAs. *Nature* **345**, 502-506.

Schmidt, U., Podar, M., Stahl, U., and Perlman, P.S. (1996). Mutations of the two-nucleotide bulge of D5 of a group II intron block splicing in vitro and in vivo: phenotypes and suppressor mutations. *RNA* **2**, 1161-1172.

Schuwirth, B.S., Borovinskaya, M.A., Hau, C.W., Zhang, W., Vila-Sanjurjo, A., Holton, J.M., and Cate, J.H. (2005). Structures of the bacterial ribosome at 3.5Å resolution. *Science* **310**, 827-834.

Schwer, B. (2008). A conformational rearrangement in the spliceosome sets the stage for Prp22-dependent mRNA release. *Mol. Cell* **30**, 743-754.

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

Selmer, M., Dunham, C.M., Murphy, F.V.4th, Weixlbaumer, A., Petry, S., Kelley, A.C., Weir, J.R., and Ramakrishnan, V. (2006). Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* **313**, 1935-1942.

Sengoku, T., Nureki, O., Nakamura, A., Kobayashi, S., and Yokoyama, S. (2006). Structural basis for RNA unwinding by the DEAD-Box protein Drosophila Vasa. *Cell* **125**, 287-300.

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.

Stahley, M.R., and Strobel, S.A. (2005). Structural evidence for a two-metal-ion mechanism of group I intron splicing. *Science* **309**, 1587-1590.

Stahley, M.R., Adams, P.L., Wang, J., and Strobel, S.A. (2007). Structural metals in the group I intron: A ribozyme with a multiple metal ion core. *J. Mol. Biol.* **372**, 89-102.

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 Prp28p. *Mol. Cell* **3**, 55-64.

Staley, J.P., and Woolford, J.L. (2009). Assembly of ribosomes and spliceosomes: complex ribonucleoprotein machines. *Curr. Opin. Cell Biol.* **21**, 109-118.

Stark, H., Dube, P., Luhrmann, R., and Kastner, B. (2001). Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle. *Nature* **409**, 539-542.

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., Fedorova, O., Rajashankar, K., Wang, J., and Pyle, A.M. (2009). Tertiary architecture of the *Oceanobacillus iheyensis* group II intron. *RNA* **16**, 57-69.

Toor, N., Keating, K.S., and Pyle, A.M. (2009). Structural insights into RNA splicing. *Curr. Opin. Struct. Biol.* **19**, 260-266.

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.

Valadkhan, S. (2010). Role of the snRNAs in spliceosomal active site. *RNA Biol.* **7**, 345-353.

Valadkhan, S. (2005). snRNAs as the catalysts of pre-mRNA splicing. *Curr. Opin. Chem. Biol.* **9**, 603-608.

Valadkhan, S., and Jaladat, Y. (2010). The spliceosomal proteome: At the heart of the largest cellular ribonucleoprotein machine. *Proteomics* **10**, 4128-4141.

Valadkhan, S., and Manley, J.L. (2001). Splicing-related catalysis by protein-free snRNAs. *Nature* **413**, 701-707.

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.

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.

Valverde, R., Edwards, L., and Regan, L. (2008). Structure and function of KH domains. *FEBS J.* **275**, 2712-2726.

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

Weber, G., Trowitzsch, S., Kastner, B., Lührmann, R., and Wahl, M.C. (2010). Functional organization of the Sm core in the crystal structure of human U1 snRNP. *EMBO J.* **29**, 4172-4184.

Will, C.L., and Lührmann, R. (2010). Spliceosome Structure and Function. *Cold Spring Harb. Perspect. Biol.*

Will, C.L., Rumpler, S., Klein Gunnewiek, J., van Venrooij, W.J., and Lührmann, R. (1996). In vitro reconstitution of mammalian U1 snRNPs active in splicing: the U1-C protein enhances the formation of early (E) spliceosomal complexes. *Nucleic Acids Res.* **24**, 4614-23.

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.

Wolf, E., Kastner, B., Deckert, J., Merz, C., Stark, H., and Lührmann, R. (2009). Exon, intron and splice site locations in the spliceosomal B complex. *EMBO J.* **28**, 2283-2292.

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, D., Nouraini, S., Field, D., Tang, S.J., and Friesen, J.D. (1996). An RNA-dependent ATPase associated with U2/U6 snRNAs in pre-mRNA splicing. *Nature* **381**, 709-713.

Xu, Y.Z., and 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.

Yeh, T.C., Liu, H.L., Chung, C.S., Wu, N.Y., Liu, Y.C., and Cheng, S.C. (2010). Splicing factor Cwc22 is required for the function of Prp2 and for the spliceosome

to escape from a futile pathway. *Mol. Cell. Biol.* **31**, 43-53.

Zhang, L., Xu, T., Maeder, C., Bud, L.O., Shanks, J., Nix, J., Guthrie, C., Pleiss, J.A., and Zhao, R. (2009). Structural evidence for consecutive Hel308-like modules in the spliceosomal ATPase Brr2. *Nat. Struct. Mol. Biol.* **16**, 731-739.

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

Chapter 1

Prp8: the pivotal spliceosomal protein at the catalytic center

1-1. Introduction

Prp8 is a component of U5 snRNP and U4/U6•U5 tri-snRNP (Lossky et al., 1987; Gottschalk et al., 1999; Stevens et al., 2001). It functions in both U2- and U12-dependent splicing (Luo et al., 1999). PRP8 was first identified in a screen for temperature sensitive (ts) mutations in *S. cerevisiae* (Hartwell, 1967; Hartwell et al., 1970). When shifted from the permissive temperature (23 °C) to the restrictive temperature (36 °C), these mutants undergo an accumulation of RNA. Ten different complementation groups, *ma2-11* were identified, which were later renamed PRP to indicate their involvement with pre-mRNA processing (Vijayraghavan et al., 1989). Many additional studies have highlighted the importance of PRP8 gene in splicing. For example, the temperature-sensitive mutant *prp8-1* yeast extract was unable to form the 40S spliceosome and splice the actin pre-mRNA *in vitro*. Splicing can be restored with the addition of other *prp* mutants or the wild type extract (Lustig et al., 1986; Lin et al., 1987).

Prp8 is conserved in both its large size and sequence as well as in evolution, with a >60% sequence identity between yeast and human genes. In different organisms, Prp8 varies from ~230 kD to ~280 kD, which makes it the largest spliceosome component. Prp8 has been shown to cross-link to all three substrate sites within mRNA: the 5'SS, the BPS and the 3'SS (Wyatt et al., 1992; MacMillan et al., 1994; Teigelkamp et al., 1995; Umen and Guthrie a, 1995; Reyes et al., 1996). It was also shown to associate with the U5 (Dix et al., 1998) and the U6 snRNAs (Vidal et al., 1999). In addition, numerous PRP8 mutants

were indicated to suppress the mutations of the 5'SS, 3'SS, BPS and PPT, as well as the mutations of U4 and U6 snRNAs (Grainger and Beggs, 2005; and references therein). To date, Prp8 is the only spliceosomal protein known to interact with all the above RNA sequences. Since all of them are considered to be in or proximal to the catalytic center of the spliceosome, Prp8 has been proposed as the best candidate as a cofactor in RNA-mediated catalysis (Grainger and Beggs, 2005). In addition, Prp8 is also associated with different protein components of several snRNPs. The real role of Prp8 in the catalytic active site is not clear. It may serve as a scaffold for other components or it may play a more direct role in catalysis of one or both steps of the splicing event (Abelson, 2008).

The functional importance of the PRP8 gene has been shown by gene knockout and knockdown studies. Yeast PRP8 knockout is lethal. In *C. elegans*, injection of siRNA corresponding to PRP8 results in the arrest of embryogenesis at the late-gastrulation stage (Takahashi et al., 2001). PRP8 mutations have also been linked to human disease. Several mutations within the C-terminus of Prp8 have been identified in retinitis pigmentosa (RP), a genetic disorder characterized by progressive degeneration of the peripheral retina and loss of visual fields (McKie et al., 2001; Martinez-Gimeno et al., 2010; Towns et al., 2010).

This chapter summarizes recent biochemical and structural studies on Prp8. In order to distinguish the yeast and human Prp8 proteins, yPrp8 and hPrp8 will be used respectively. If a statement is equally relevant to both homologues, Prp8 will be used.

1-2. Prp8 domains

The large size of Prp8 makes it difficult to study the role of this protein by high resolution structural analysis. Domain analysis is complicated by the fact that the primary amino acid sequence of Prp8 lacks any obvious conserved protein motif. In the past decade, computational analyses have suggested putative functions for several regions in Prp8. An early study designated five regions (a to e) important for the suppression of cold-sensitive (cs) mutations in yeast U4 snRNA (Kuhn and Brow, 2000; Figure 1-1 C). Later, three major domains (N-terminal, central, C-terminal) were identified by the evolutionary comparison of nucleolar proteins (Staub et al, 2004; Figure 1-1 D). Based on resistance to transposon insertion, a four-domain model (I to IV) was also proposed by Beggs and Newman and coworkers (Boon et al., 2006; Figure 1-1 B).

A recent study, however, provided a totally new domain organization model for Prp8 (Dlakic and Mushegian, 2011; Figure 1-1 A). The analysis in the following sections will be based on this new study, and the arrangement of the old model will also be discussed.

1-2.1. N-terminal region

The N-terminus of Prp8 contains three domains. Proline-rich tracts or P regions are found within all fungal Prp8 proteins and are absent in most other organisms. The non-essential spliceosomal protein Lin1p/Snu40 associates with

the P region and this interaction is believed to play roles in U5 snRNP biogenesis (Bialkowska and Kurlandzka, 2002; Stevens et al., 2001). A putative bipartite nuclear localization signal (NLS) is next to the downstream of P region. The NLS of Prp8 was shown to be sufficient for the localization of Prp8 and it was also shown to mediate the transportation of the U5 snRNP precursor from the cytoplasm into nucleus (Boon et al., 2007).

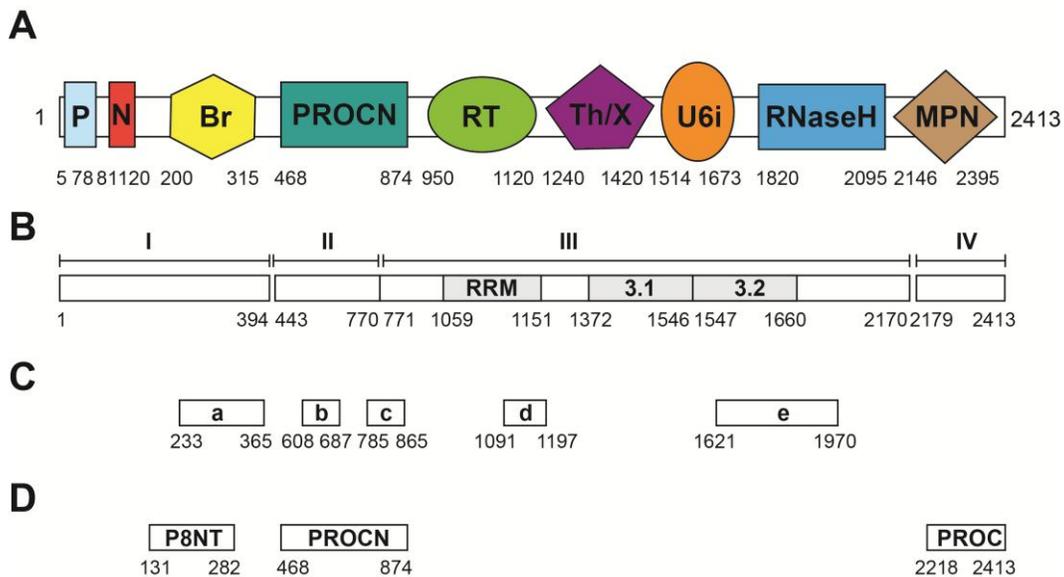


Figure 1-1. Diagrams of conserved domains in yeast Prp8. A) Model from Dlakic et al., 2011. P: proline-rich region; N: nuclear localization signal; Br: bromodomain; PROCN: PRO8 central domain; RT: reverse transcriptase-like palm-and-fingers domain; Th/X, conserved domain in Prp8 and a subset of fungal RT-like proteins, located at the same position as “maturase-specific” X/thumb domain; U6i, U6-interacting domain. B) Model from Boon et al., 2006. Domains I-IV indicate the positions at which Prp8 can be split in two and still function *in trans in vivo*. C) Model from Kuhn and Brow, 2000. Domains a-e indicate the locations of the five regions involved in suppression of the cold sensitive mutation U4-cs1. D) Model from Staub et al., 2004. P8NT: PRO8 N-terminal domain; PROC: PRO8 C-terminal domain.

Next in the N-terminal region is a newly-identified bromodomain or Br domain (Dlakic et al., 2011). This domain falls in the same region as the previously proposed PRO8 N-terminal domain (Staub et al., 2004) and it shares low sequence identity with known bromodomains (<14%). However, several residues important for substrate recognition and structure stabilization in bromodomains are conserved in Prp8. Known bromodomains adopt a four α -helical bundle and bind to acetylated lysine residues such as those on the N-terminus of histones (reviewed in: Zeng and Zhou, 2002). The charged residues in the ZA loop, which have been shown important to determine the substrate specificity, are highly conserved among Prp8 proteins in different organisms but not in chromatin-binding bromodomains. The N-terminus of Prp8 has been proposed to form intramolecular interactions with its own central domain; it also interacts directly with U5 snRNP components Brr2, Snu114, Prp40 and the U1 snRNP component Prp39 (Grainger et al., 2009). The bromodomain has been suggested to be a good candidate to mediate all these protein-protein interactions. This idea, however, should be regarded with care, since the sequence identity between Prp8 and other bromodomain proteins is very low.

1-2.2. Central region

The ~400 amino acid Prp8 central domain or PROCN was identified as a compact unit by sequence analysis; it also tolerates transposon insertions (Staub et

al., 2004; Boon et al., 2006). The physiological function of this region is still unknown.

The ~100 amino acids next to the PROCN in the middle of Prp8 were first proposed to form an RRM domain (Grainger and Beggs, 2005), adopting a canonical $\beta\alpha\beta\beta\alpha\beta$ ferredoxin-like fold. Most recently, this region along with flanking N-terminal and C-terminal sequences was predicted to be a reverse transcriptase or RT-like domain, with a structure similar to the *Tribolium* telomerase subunit TERT (Dlakic et al., 2011; Gillis et al., 2008). The corresponding region on TERT, containing the palm domain and parts of the fingers domain, is involved in substrate binding and the formation of the nucleotidyltransferase catalytic center (Mitchell et al., 2010). Interestingly, although TERT does contain an RNA-binding domain (TRBD), it is not located within the aligned sequence between Prp8 and TERT.

The active site of TERT is formed by three conserved aspartic acid residues coordinating the catalytic magnesium ion (Gillis et al., 2008). Two of these are from the consensus YUDD (U is a hydrophobic residue) of motif C and the other is from motif A of the palm region. In Prp8 only one Asp residue is conserved and is located in motif C. Based on the tertiary structure prediction, the other two Asp positions are replaced by Thr and Arg. Another conserved region, motif D of TERT palm domain which was shown to protonate the pyrophosphate (Castro et al., 2009), is also absent within Prp8. The RT-like domain of Prp8 has been shown to cross-link to the 5'SS and BPS (Turner et al., 2006). Considering the

extensive base-pairing of U6/5'SS and U2/BPS, it is tempting to assume the RT-like domain would recognize the dsRNA region in a similar manner as to TERT in the TERT/DNA RNA complex. The absence of several conserved residues within the RT-like domain of Prp8 possibly prevents metal coordination and renders it incapable of synthesizing nucleic acid. However the conserved Asp on motif C still leaves open the possibility that this modified active site is capable of chelating a metal ion and performing simpler reactions such as nucleotide or phosphate transfer, or hydrolysis of a phosphoester bond, either on its own or in complex with the RNA components of the spliceosome. Thus until the structure is determined, no conclusion can be made as to whether this region is an RT-like domain and whether a metal is chelated within the active site of the region corresponding to that of the TERT.

A large region of Prp8 between aa 1372-1660 was originally termed domain 3 since mutations in this region suppress 3'SS mutations (Umen and Guthrie, 1996). Domain 3 was further divided into two subdomains, domain 3.1 and domain 3.2, with domain 3.1 cross-linked to U5 snRNA and domain 3.2 to U6 snRNA. The majority of the suppression mutations reside in domain 3.2 (Grainger and Beggs, 2005; Turner et al., 2006). In the most recent model of Prp8 domain structure, domain 3.1 was suggested to be homologous to the region that is found in RT-like proteins encoded by nuclear genomes of fungi and by the green nonsulfur bacteria (Dlakic et al., 2011). The physiological function of the corresponding domain in group II and viral RT is poorly understood, and no high resolution structural

analysis is available. This domain was also proposed to be functionally similar to one in the group II reverse transcriptase (X/maturase domain) or eukaryotic retroviral reverse transcriptase (Th domain) (Blocker et al., 2005) and it has therefore been renamed Th/X domain. In the case of the retroviruses, a similar domain organization (RT/Th/Connection/RNaseH) to that proposed for Prp8 can be seen, indicating an evolutionary connection between Prp8 and eukaryotic viral reverse transcriptase (Kohlstaedt et al., 1992). Part of the Th domain in HIV-1 RT was shown to bind the DNA substrate, consistent with the cross-linking data indicating that this region of Prp8 makes contacts with U5 snRNA (Ding et al., 1998).

To the N-terminus of the Th/X domain is the U6 interaction domain (U6i). It has no sequence similarity to any known structures. A much higher predicted content of β -strands makes it unlikely to be an evolutionary homolog of the connection domain in viral RT. A Lys residue within this region (K1535 in yeast) was found to undergo acetylation, implying a potential intramolecular interaction between the Prp8 bromodomain and U6i (Choudhary et al., 2009).

1-2.3. C-terminal region

The C-terminus of Prp8 contains an RNaseH domain and an MPN (Mpr-1, Pad-1, N-terminal) domain. It is the best studied region in Prp8, with crystal structures available for both domains.

The RNase H domain, originally named domain IV, contains clusters of genetic features, including suppressors of pre-mRNA, U4-cs1, and U6 mutations, as well as synthetic lethal interactions with *prp28* mutations (reviewed in Grainger and Beggs, 2005). Konarska and coworkers have mapped a cross-link between the 5'SS GU di-nucleotide and hPrp8 during spliceosome assembly. The stretch of five amino acids QACLK in hPrp8 responsible for this interaction is not conserved in yeast, with the corresponding sequence being SAAMS. hPrp8 chimeras containing the SAAMS sequence can still cross-link to the 5'SS, suggesting this region might not be involved directly in the interaction but is rather nearby (Konarska et al., unpublished data; Reyes et al., 1999). All of the 5'SS suppressors within the domain IV are near this region (Siatecka et al., 1999). Umen and Guthrie have also shown that yeast alleles identified in screens for factors affecting PPT recognition are located in domain IV (Umen and Guthrie, 1995a). One of these alleles, *prp8-101* (E1960K) causes reduced cross-linking efficiency to both PPT and 3'SS (Umen and Guthrie, 1995a and 1995b). The crystal structure of domain IV provides a detailed view of this region and will be discussed in section 1-4.

The MPN domain of Prp8 resides in the C-terminus of Prp8. MPN domains are usually found in the N-terminus of proteins with various functions: proteasome regulatory subunits, eukaryotic initiation factor 3 (eIF3) subunits, the signalosome, and regulators of transcription factors (reviewed in Grainger and Beggs, 2005). A subset of MPN domains have been implicated in catalyzing the

hydrolytic removal of ubiquitin from target proteins in the presence of Zn^{2+} coordinated by the JAMM motif (reviewed in Bellare et al., 2006). In Prp8, however, a key His residue in this motif is not conserved and metal coordination does not occur (discussed below in the structure section). Instead, the MPN domain is proposed to mediate protein-protein interactions. Indeed, the MPN domain of Prp8 interacts with two factors involved in remodelling of the U5/U4 U6 tri-snRNP: Brr2 and Snu114. It has been shown that the MPN domain inhibits the Brr2 U4/U6-dependent ATPase activity (Maeder et al., 2009). The MPN domain was also shown to bind to ubiquitin and ubiquitinated Prp3 protein with an affinity comparable to other known ubiquitin-binding domains (UBDs) (Kuhn et al., 1999; Bellare et al., 2006; Song et al., 2010). Sontheimer and coworkers have also observed that when the I44A ubiquitin is introduced into yeast extract, U4/U6 U5 tri-snRNP levels are greatly reduced; this is proposed to be due to accelerated U4/U6 snRNA unwinding. In addition to binding ubiquitin, Prp8 itself also undergoes ubiquitination (Bellare et al., 2008). Ubiquitination may thus modify the ability of Prp8 to regulate the U4/U6 unwinding activity of Brr2. Although the modification site on Prp8 is currently unknown, it is tempting to assume that ubiquitinated Prp8 is recognized intramolecularly by interaction with its own MPN domain and that this association precludes the Brr2-Prp8 interaction.

As mentioned above, Prp8 mutations are observed in the patients with hereditary retinitis pigmentosa (RP) and all the observed mutations are within the MPN domain. Notably, the Prp8-Brr2 interaction is significantly weakened by

some of these mutations, suggesting the link between the dysregulation of splicing and human diseases (Maeder et al., 2008).

1-3. Prp8 and its interactions

1-3.1. Prp8-RNA interactions

In cross-linking studies, Prp8 is the only spliceosomal protein shown to directly contact all reactive sites within the intron: 5'SS, 3'SS, BPS and PPT (Wyatt et al., 1992; MacMillan et al., 1994; Teigelkamp et al., 1995; Umen and Guthrie a, 1995; Reyes et al., 1996; Figure 1-2 A). The interaction between Prp8 and different sequence elements of the pre-mRNA is dynamic and dependent on the stage of spliceosome assembly. The cross-linking between Prp8 and the 5'SS occurs in B complex, before the first step of splicing (Wyatt et al., 1992; Reyes et al., 1999). Prp8 cross-links to the BPS within complexes B and C (MacMillan et al., 1994) and to the PPT and 3'SS within complex C, likely after the first and before the second step (Teigelkamp et al., 1995; Chiara et al., 1997).

Prp8 was shown to interact with the 5'SS and at least eight nucleotides downstream before the first step of splicing occurs (Teigelkamp et al., 1995). This interaction occurs when Prp8 is part of the U4/U6 U5 tri-snRNP and is ATP-dependent (Maroney et al., 2000). Using protease treatment, the position of the Prp8 cross-link to the 5'SS was mapped to the five residues QACLK (aa 1894-1898 in hPrp8; Reyes et al., 1999). At the 3' end of intron, the cross-linking

between Prp8 and pre-mRNA spans a region from the BPS to 3'SS and includes bases in the 3' exon. This binding is not specific to the 3'SS sequence (McPheeters

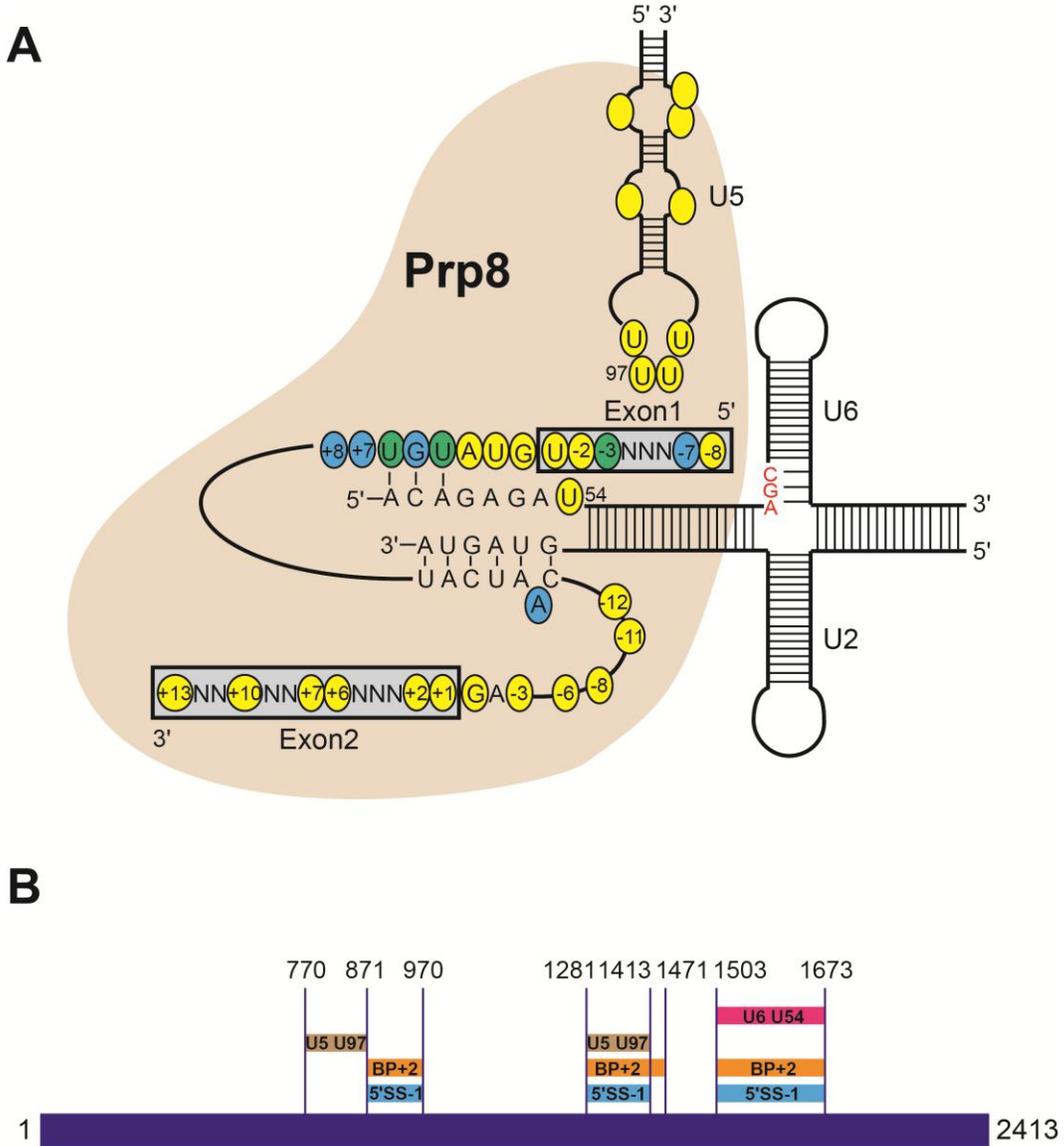


Figure 1-2. Prp8-RNA interactions. A) A summary of the RNA residues that are known to crosslink to Prp8 in the context of the spliceosome. Cross-linked residues are indicated by filled circles. Short-range cross-links using uridine and 4-thio-uridine are shown in yellow. Long-range cross-links using the reagent benzophenone are shown in blue. Residues where both short- and long-range cross-links are detected are shown in green (Adapted from Grainger and Beggs, 2005; Turner et al., 2006). B) The Prp8 regions that cross-link to crucial snRNA and pre-mRNA residues. (Adapted from Turner et al., 2006). All numbering relates to *S. cerevisiae*.

and Muhlenkamp, 2003). It has been proposed that the association of Prp8 with the 3'SS results in conformational changes which locks the free 5' exon into the spliceosome (Umen and Guthrie, 1995a). One of the DEAD/X helicases, Prp16, binds to the 3'SS and facilitates the Prp8 binding in an ATP-dependent manner (McPheeters and Muhlenkamp, 2003; reviewed in Grainger and Beggs, 2005).

More recent cross-linking experiments carried out in yeast and followed by proteolytic mapping suggest different regions of the protein interact with the 5'SS than previously identified (Turner et al., 2006). By randomly inserting the tobacco etch virus (TEV) protease site into yPrp8, the authors were able to map cross-links between yPrp8 and spliceosomal RNA. The 5'SS-interacting regions of yPrp8 were identified as amino acids 871-970, 1281-1413 and 1503-1673. In comparison, the five amino acids (aa 1966-1970 in yPrp8) identified previously to cross-link to the 5'SS (Reyes et al., 1999) were not observed in this study. The discrepancy may be due to the cross-linking being carried out at different stages of spliceosome assembly. Interestingly, the same three regions were also found to form interactions with the BPS, although the intensities of labeling differ between the 5'SS and the BPS. This is consistent with the idea that juxtaposition of the 5'SS and BPS is required for the first step of splicing to occur. Therefore, Prp8 has been proposed to play “lock and load” roles in both steps of splicing, with the conformational change within Prp8 resulting in the transition from the first step to the second step (reviewed in: Grainger and Beggs, 2005).

In the recent work of the Newman lab, the authors were also able to map the Prp8 regions responsible for interacting with the snRNAs (Figure 1-2 B). Prp8 has been shown to make extensive contacts with U5 snRNA, including the invariant loop 1 nucleotides γU^{96} - U^{99} (Dix et al., 1998). Both the loop 1 and Prp8 can be simultaneously cross-linked to the 5'SS and 3'SS in human and yeast, and this was thought to reflect the “lock and load” role of Prp8 for the second step (reviewed in: Grainger and Beggs, 2005). The interaction between Prp8 and U5 snRNA has been shown to be crucial for the normal function of Prp8. Three U5 loop 1 mutants were shown to reduce Prp8 stability and may affect interactions of Prp8 with pre-mRNAs. Depletion of U5 snRNA *in vivo* also significantly reduced Prp8 stability (Kershaw et al., 2009). Prp8 was also shown to contact position 54 of U6 snRNA, which is immediately downstream of the conserved ACAGAG motif that interacts with the 5'SS. In the TEV insertion assay, the U5 snRNA interacting regions were mapped to amino acids 770-871 and 1281-1413, while the U6 snRNA interacting domain is 1503-1673 (Turner et al., 2006). Notably, aa 1281-1413 interacts with the U5 snRNA, the 5'SS and BPS, while aa 1503-1673 interacts with the U6 snRNA, the 5'SS and BPS. These correspond to the Th/X domain (domain 3.1) and U6i domain (domain 3.2), respectively. Prp8 was also shown to interact with U2 snRNA. This interaction is restricted to the stem I of U2 snRNA and does not occur when the corresponding sequence forms helix II in the U2/U6 snRNA paired structure (Xu et al., 1998).

1-3.2. Prp8 protein interactions

Prp8-containing complexes have been identified by both yeast two-hybrid (YTH) screens and affinity purifications of Prp8. The most important two partners of Prp8 are Brr2 and Snu114. Both of these are U5 snRNP components and their direct interactions with Prp8 are indicated by both YTH and pull-down assays (Achsel et al., 1998; Dix et al., 1998). Both the N-terminus (aa 1-263) and the C-terminus (aa 2100-2413) of Prp8 were shown to pull out Brr2 in YTH, and amino acids 420-464 pulled out Snu114. The U1 snRNP protein Prp39 was shown to interact with the N-terminus (aa 1-263) of Prp8 (van Nues and Beggs, 2001) and ySnp1, the yeast homolog of U170K, interacts with the central region (aa 1166-1193) of Prp8 (Awasthi et al., 2001). Snp1 was also shown to associate with yExo84, a secretory protein affecting splicing, and Exo84 itself interacts with amino acids 750-930 of Prp8 (Fromont-Racine et al., 1997). An interaction between the N-terminus of yPrp8 (1-349) and the U1 snRNP protein Prp40 has been demonstrated (Abovich and Rosbash, 1997). As the N-terminus of Prp8 is associated with U1 snRNP proteins as well as with Brr2 and Snu114, it is possible that this region of Prp8 may couple the unwinding of the U1/5'SS interaction with U4/U6 unwinding by Brr2 (reviewed in: Grainger and Beggs, 2006). A yeast U1/U5 snRNP complex has been identified with unexpected results (Gottschalk et al., 2001). The purified U1 snRNP contains yPrp39, yPrp40 and other U1 snRNP components; the purified U5 snRNP, in contrast, lacks the majority of known Prp8-associated proteins, including Brr2. Instead, a distinctive protein, Aar2, was

found in the U1/U5 associated snRNP but is absent in the U4/U6 U5 tri-snRNP. Aar2 was also shown to be essential for growth; its depletion leads to a splicing defect. This U1/U5 complex may represent a step in the U5 snRNP assembly pathway.

In addition to being a U5 snRNP component, Prp8 has also been shown to be present in several other transcription-related complexes. Associations of Prp8 with a number of transcription factors have been reported. For example, the mammalian PRP4 kinase (PRP4K) was shown to co-purify with BRG1, a component of the conserved SWI/SNF remodelling complex, as well as the U5 snRNP proteins hPrp8, hBrr2 and hPrp6 (Dellaire et al., 2002). Thus PRP4K may couple transcriptional regulation to pre-mRNA splicing by modulating chromatin structures. Another example is the transcriptional co-activator hSKIP (yPrp45), also known as nuclear coactivator-62 kDa (NCoA62), which binds the vitamin D receptor (VDR) and acts as a remodelling protein when recruited to vitamin D-responsive elements in target genes (Baudino et al., 1998; Zhang et al., 2003). hSKIP was found to co-purify with hPrp8, PSF (Polypyrimidine tract-binding protein-associated splicing factor), hPrp28, hBrr2 and hSnu114, indicating the probable indirect association between Prp8 and VDR-activated transcription (Zhang et al., 2003). Interesting, hSKIP was shown to co-purify with the catalytically active spliceosome and the U5/Prp19 complex, but not with the U4/U6 U5 tri-snRNP, suggesting a possible role of SKIP and the Prp19 complex in remodelling the U5 snRNP to form the active spliceosome (Marakov et al.,

2002; see Figure I-2). There has also been growing evidence for the coupling of pre-mRNA splicing with the activity of RNA polymerase II (RNAP II). Immunoprecipitation experiments with antibodies against the C-terminal domain (CTD) of RNAP II showed its association with the U5 snRNP and U4/U6-U5 tri-snRNP and mRNA (Chabot et al., 1995; Vincent et al., 1996). Later this complex was shown to contain RNAP II, PSF, p54^{nrb}, hPrp8, and all five snRNAs (Kameoka et al., 2004). Notably, PSF and p54^{nrb} interact with each other and the stem 1b of U5 snRNA (Peng et al., 2002). The same region on U5 snRNA was shown to interact with yPrp8 (Dix et al., 1998).

1-4. Crystal structures of Prp8 domains

1-4.1. Structure of the MPN domain

The first crystal structure of a Prp8 domain was obtained in 2007. Two groups reported high-resolution structures of the MPN domain from budding yeast and *C. elegans* (Pena et al., 2007; Zhang et al., 2007; Figure 1-3 A).

The most typical example of the MPN superfamily is Rpn11, a proteasome component. The MPN of Rpn11 domain contains a conserved JAMM motif that contains the sequence EX_nHXHX₇SXXD. The conserved Glu, 2xHis, Ser, and Asp residues are responsible for coordinating a Zn²⁺ ion (Figure 1-3 C). Mutation of the conserved His and Asp residues has been shown to be lethal in yeast due to a defect in the deubiquitination activity carried out by Rpn11 (Tran et al., 2003).

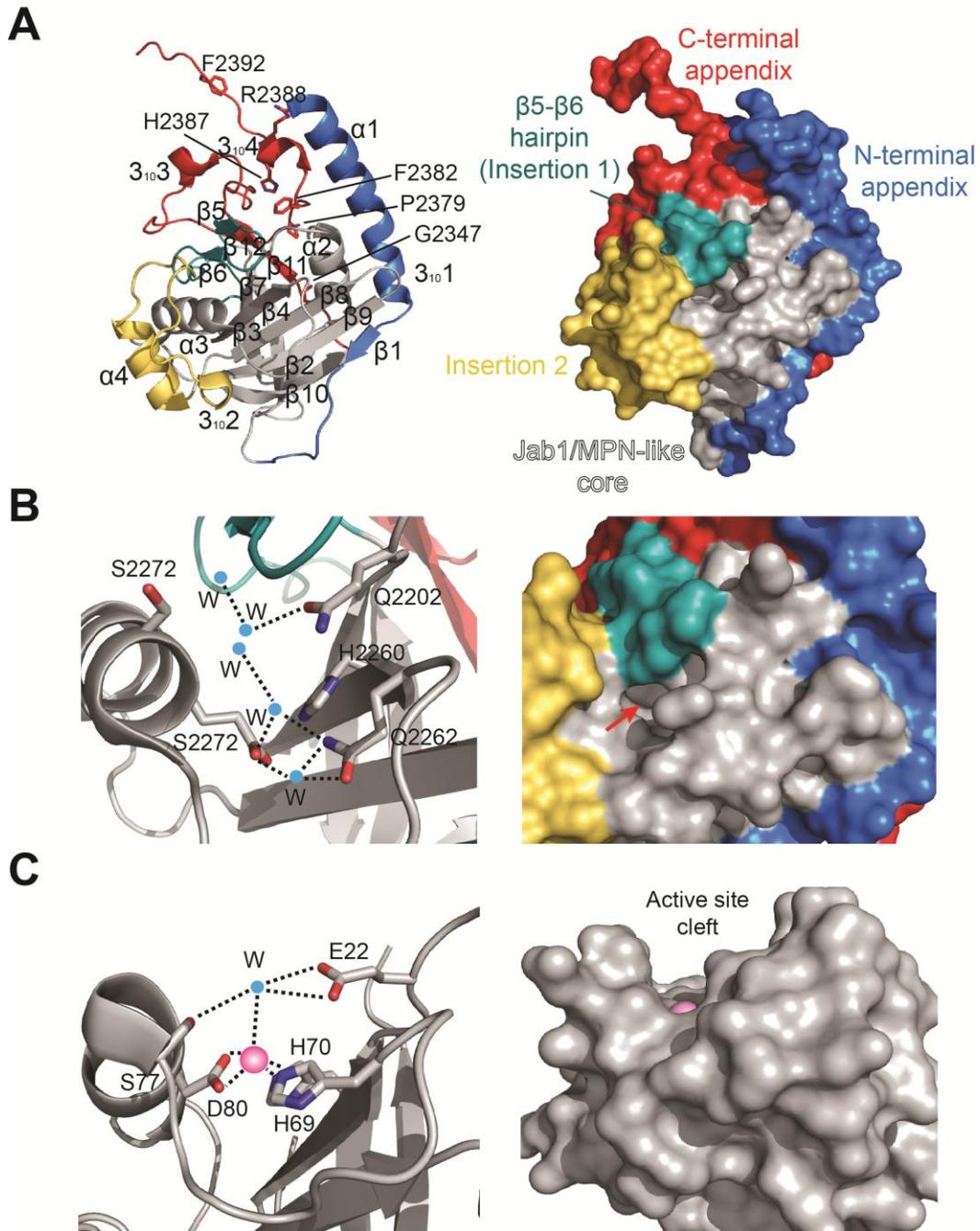


Figure 1-3. Crystal structures of Prp8 MPN domain. A) Ribbon plot (*left*) and surface view (*right*) of yPrp8 MPN domain. The RP-related residues are also shown. B) Ribbon plot (*left*) and surface view (*right*) of JAMM-motif in yPrp8 MPN domain. A putative Zn^{2+} binding site by superimposition of AF2198 is placed in the active site and indicated by arrow. C) Ribbon plot (*left*) and surface view (*right*) of JAMM-motif in AF2198. The Zn^{2+} is shown in purple (Adapted from Pena et al., 2008).

In comparison, Prp8's MPN domain has a consensus QX_nHXQX₇SXXD sequence with the conserved Glu and His residues of Rpn11 replaced by Gln. In addition, it has been shown that the mutation of all four non-Ala residues in the JAMM motif of yPrp8 to Ala did not produce any phenotype at a permissive temperature (Bellare et al., 2006). Accordingly, this domain has lost ability to chelate Zn²⁺, a feature that is confirmed by the X-ray structures.

The MPN core is present in the yPrp8 MPN domain, with additional ~30 amino acid N-terminal and ~40 amino acid C-terminal extensions tightly wrapped around the core (Pena et al., 2007; Figure 1-3 A). About two-thirds of the interface between the extensional regions and the core are hydrophobic, suggesting the packing of the structure is quite stable. Not surprisingly, in this structure no metal ion was found to coordinate the JAMM motif residues. Instead, an array of ordered water molecules is seen in this portion of the structure (Figure 1-3 B). A striking feature of the yPrp8 MPN structure is the existence of two insertions compared to other known MPN structures. One of them, the β5-β6 insertion, runs exactly along the hydrophobic groove in the core domain, which in other MPN structures constitutes the substrate binding site (Ttan et al., 2003; Ambroggio et al., 2004). Therefore, the amino acid replacements in the JAMM motif of yPrp8 apparently impair Zn²⁺ binding and an inserted sequence completely masks the active site. The inaccessibility of the active site also rules out the possibility of the association between Prp8 and ubiquitin via the core domain.

In the report of the yPrp8 MPN domain structure, the authors also tested the effect of the RP-linked mutations. All these mutations lie in the C-terminal 35 residues of hPrp8 (Grainger and Beggs, 2005), and in the structure this corresponds to the C-terminal extension. The same region was also shown to mediate the interactions between hPrp8, hBrr2 and hSnu114. When the hPrp8 fragments containing these mutations were tested by yeast two-hybrid analysis, two of the RP-linked mutations (R2388K and R2388G) reduced the interaction of Prp8 with both Snu114 and Brr2; the F2392L mutation only reduced Prp8's interaction with Snu114; the F2382L mutation only affected the interaction with Brr2; the remaining three mutations (P2379T, H2387P and H2387R) had no effect (Pena et al., 2007). This result provides evidence for the link between impaired splicing and the development of RP. These mutations may either abolish the interactions between Prp8 and other spliceosomal components directly, or they could affect the structure of the module.

The structural analysis of the Prp8 MPN domain provides a novel model for a member of the MPN-like superfamily. Instead of acting as an enzyme, this domain is thought to mediate protein–protein interactions within the spliceosome. One potential role of Prp8 is that it serves as a platform for spliceosome assembly and the MPN domain is a major region of Prp8 responsible for this role.

1-4.2. Structure of the RNase H domain

Shortly after the structure of the MPN domain was solved, three independent groups, including our lab, reported the structures of the RNase H domain of hPrp8 and yPrp8 (Pena et al., 2008; Ritchie et al., 2008; Yang et al., 2008). These structures have furthered our understanding of the role Prp8 plays in splicing and provided some striking structures concerning the active centre of RNA splicing.

Both the human and yeast Prp8 RNaseH domain adopt very similar structures (Figure 1-4 A). In humans, the domain IV core is bipartite consisting of an N-terminal sub-domain (aa 1769- 1887), an RNase H fold, and a tightly packed C-terminal cluster of five helices (aa 1900-1990; Ritchie et al., 2008). The RNase H fold of Prp8 exhibits the same topology as other known RNase H structures, with a five-stranded mixed β -sheet buttressed by two α -helices surrounding it (Figure 1-4 C). This structural homology was not predicted by sequence alignment due to a 17 amino acid insertion (aa 1787-1803). In the crystal structure from our lab, two monomers are found in the asymmetric unit. In one of them these amino acids are well structured and form a two-stranded anti-parallel β -sheet ($\beta 2/\beta 3$) extruding from the RNase H core, while in the other monomer this region was disordered (Ritchie et al., 2008; the two monomers will be discussed in detail in Chapters 2 and 3). Superimposition of the hPrp8 and yPrp8 structures showed that the RNase H domain is rather invariant, while the β -hairpin region

and the C-terminal helical cluster adopt different orientations with respect to the RNase H fold. Therefore these elements are essentially mobile.

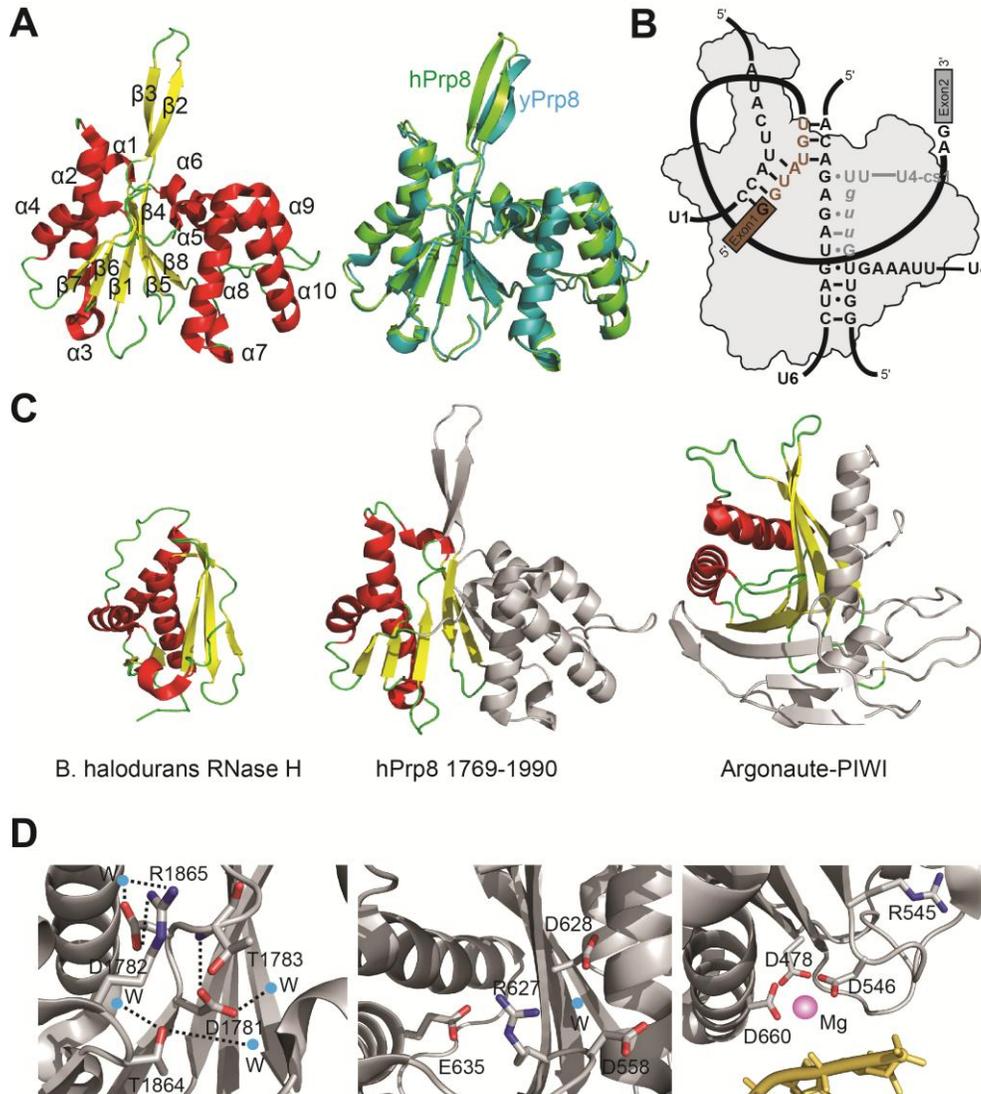


Figure 1-4. Crystal structures of Prp8 RNase H domain. A) Ribbon plot of hPrp8 RNase H domain (*left*) and alignment of hPrp8 (green) with yPrp8 (cyan) RNase H domain (*right*). B) Proposed RNase H-RNA interaction before catalytic activation (Adapted from Pena et al., 2008). C) Structure of the Prp8 domain IV core (*middle*) alongside *B. halodurans* RNase H (*left*) and the PIWI domain of *P. furiosus* Argonaute (*right*), with the RNase H fold indicated in red and yellow. D) The potential metal binding site in the core domain of hPrp8 (*left*) includes four hydrogen-bonding water molecules and is blocked by the side chain of Arg1865 hydrogen bonded to Asp1782. A similar organization is observed in the analogous *P. furiosus* PIWI domain site (*middle*). When the PIWI domain binds its DNA/RNA substrate, the blocking Arg is removed and a Mg^{2+} (pink) is observed in the active site (*right*) (Adapted from Ritchie et al., 2008; Pena et al., 2008; Song et al., 2004; Wang et al., 2008).

All but one of the yPrp8 domain IV alleles characterized so far correspond to amino acids on one face of the RNase H domain structure. A cluster of characterized alleles map to the β -hairpin. These include suppressors of mutations in the pre-mRNA and suppressors of hyperstabilization of U4/U6 (Ritchie et al., 2008). As mentioned above, the human Prp8 protein has been shown to cross-link to the 5'SS in B complex and this interaction was mapped to residues 1894-1898 (Reyes et al., 1999). In the crystal structure these amino acids are located in the $\alpha 5$ helix, which forms the base of the channel separating the RNase H domain and the C-terminal helical cluster. The surface around this area is positively charged, making it favourable to interact with the negatively charged phosphate backbone of RNA. As noted above, this five amino acid motif is not conserved between human and yeast (1894-QACLK-1898 and 1966-SAAMS-1979, respectively). But in both proteins, the different sequences adopt an almost identical extended 3_{10} -helical structure at the centre of the channel. Several residues have also been shown to be important for yPrp8/PPT interaction during suppression screens (Umen and Guthrie, 1996). These include F1834, R1922, V1946 and E1960. All of these residues are located on opposite sides of the channel in the RNase H fold below the β -hairpin region. Thus the binding surface for the PPT probably lies at the front entrance of the channel. Based on these observations, a model has been proposed for the binding of the pre-mRNA with Prp8 by comparison to the RNase H/substrate complex (Pena et al., 2008; Figure 1-4 B).

Our lab carried out gel-shift assays to check the binding between the RNase H domain and RNA. Three types of different substrates were tested: a single-stranded RNA representing the 5' or 3' splice sites; short duplex RNAs with a bulged RNA mimicking the pre-mRNA/U2 snRNA duplex; and several more complex RNAs based on proposed structures in the mature spliceosome. No sequence specificity was observed in this assay. But the RNase H core domain showed higher affinity for duplex RNAs over single-stranded ones, and a much greater affinity for the four-helix-junction RNA mimicking the interactions between the U2/U6 snRNAs and the 5'SS (Ritchie et al., 2008). The association between the four-helix-junction RNA and the RNase H domain protected the protein from trypsin cleavage at Arg1832, which is oriented toward the RNA binding surface of RNaseH, while leaving the cleavage at Trp1821, which is located on the opposite surface unaffected. These results thus confirm the association of the RNA with the RNase H fold. A cross-linking experiment was also performed to map the region on the RNase H fold responsible for its association with the four-helix-junction RNA. This region was mapped to amino acids 1869-1914 on hPrp8. This corresponds to the domain from $\alpha 4$ to $\alpha 6$ and includes the 5'SS interaction motif on the $\alpha 5$ helix (Ritchie et al., 2008).

On discovery of the RNase H fold in Prp8, an immediate question raised is whether the catalytic centre of RNase H is also conserved in Prp8. Canonical RNase H enzymes have three conserved carboxylates coordinating two catalytic metal ions $\sim 4\text{\AA}$ apart in their active sites (Nowotny et al., 2005). Only one of

these canonical metal binding sites is present in the Prp8 RNase H domain. Two Asp residues (D1781 and D1782 in hPrp8, D1853 and D1854 in yPrp8) and two Thr residues (T1783 and T1864 in hPrp8, T1855 and T1936 in yPrp8) form a site equivalent to that found in RNase H, although no metal was observed in either the human or yeast Prp8 structures. In both structures the side chain of an Arg residue (R1865 in hPrp8, R1937 in yPrp8) forms a hydrogen bond to one of the Asp residues (D1782 in hPrp8, D1854 in yPrp8) effectively blocking the potential metal binding site. All these conserved residues have been shown to be important for Prp8 functions by mutation assays in yeast. For example, both the D1853A single mutant and D1853AD1854A double mutant were lethal; D1853N shows both cold-sensitive (cs) and temperature-sensitive (ts) phenotypes; R1937A was lethal; D1853N, D1854A/N, T1855A and R1937K were all both cs and ts (Pena et al., 2008).

No Mg^{2+} ion was observed in the putative active site of human or yeast Prp8 structures, although the hPrp8 crystal was grown in 200 mM Mg^{2+} and the yPrp8 crystals were soaked in 20 mM Mg^{2+} . The occupation of the active site by Arg was proposed to block the binding of Mg^{2+} . In the crystal structure of *P. furiosus* Argonaute protein, the PIWI domain core belongs to the RNase H family (Song et al., 2004). A conserved DDE motif is found in the active site, and an Arg residue is also positioned at the center of the active site. As in other RNase H-like enzymes, Mg^{2+} is required for the RNA-induced silencing complex (RISC) activity. However, when the *T. thermophilus* Argonaute protein binds to its

DNA/RNA duplex substrate, a conformational change occurs and this Arg residue now moves out of the active site. A Mg^{2+} ion is now coordinated by the DDD motif in the active site (Wang et al., 2008; Figure 1-4 D). Another similar case is the protein transposon 5 (Tn5). In the crystal structure of Tn5 complexed with the DNA substrate, a conserved DDRE motif is found in the active site, with the DDE motif coordinating a Mn^{2+} ion and an analogous Arg residue contacting a phosphate group of the DNA substrate (Davies et al., 2000). Mutations of this Arg into Ala, Lys or Glu significantly reduced the complex formation (Naumann and Reznikoff, 2002). A similar organization of the active site has also been reported in EcoRV, with a Lys residue stabilizing the transition state (Horton et al., 1998).

The observation of an unexpected RNase H fold in Prp8 motivated our further studies of this domain. We have made a number of advances in understanding the role of this domain. Our results concerning key residues of the RNase H metal binding site will be the focus of the Chapter 2. A conformational change within the β -hairpin region coupling the two steps of splicing will be discussed in Chapter 3.

1-5. References

Abelson, J. (2008). Is the spliceosome a ribonucleoprotein enzyme? *Nat. Struct. Mol. Biol.* **15**, 1235-1237.

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

- 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 homologue of ribosomal elongation factor EF-2, and a novel WD-40 protein. *Mol. Cell. Biol.* **18**, 6756-6766.
- Ambroggio, X.I., Rees, D.C., and Deshaies, R.J. (2004). JAMM: a metalloprotease-like zinc site in the proteasome and signalosome. *PLoS Biol.* **2**, E2.
- Awasthi, S., Palmer, R., Castro, M., Mobarak, C.D., and Ruby, S.W. (2001). New roles for the Snp1 and Exo84 proteins in yeast pre-mRNA splicing. *J. Biol. Chem.* **276**, 31004-31015.
- Baudino, T.A., Kraichely, D.M., Jefcoat, S.C., Jr., Winchester, S.K., Partridge, N.C., and MacDonald, P.N. (1998). Isolation and characterization of a novel coactivator protein, NCoA-62, involved in vitamin D-mediated transcription. *J. Biol. Chem.* **273**, 16434-16441.
- Bellare, P., Kutach, A.K., Rines, A.K., Guthrie, C., and Sontheimer, E.J. (2006). Ubiquitin binding by a variant Jab1/MPN domain in the essential pre-mRNA splicing factor Prp8p. *RNA* **12**, 292-302.
- Bellare, P., Small, E.C., Huang, X., Wohlschlegel, J.A., Staley, J.P., and Sontheimer, E.J. (2008). A role for ubiquitin in the spliceosome assembly pathway. *Nat. Struct. Mol. Biol.* **15**, 444-451.
- Bialkowska, A., and Kurlandzka, A. (2002). Proteins interacting with Lin1p, a putative link between chromosome segregation, mRNA splicing and DNA replication in *Saccharomyces cerevisiae*. *Yeast* **19**, 1323-1333.
- Blocker, F.J., Mohr, G., Conlan, L.H., Qi, L., Belfort, M., and Lambowitz, A.M. (2005). Domain structure and three-dimensional model of a group II intron-encoded reverse transcriptase. *RNA* **11**, 14-28.
- Boon, K.L., Norman, C.M., Grainger, R.J., Newman, A.J., and Beggs, J.D. (2006). Prp8p dissection reveals domain structure and protein interaction sites. *RNA* **12**, 198-205.
- Boon, K.-L., Grainger, R.J., Ehsani, P., Barrass, J.D., Auchynnikava, T., Inglehearn, C.F., and Beggs, J.D. (2007). prp8 mutations that cause human retinitis pigmentosa lead to a U5 snRNP maturation defect in yeast. *Nat. Struct. Mol. Biol.* **14**, 1077-1083.

- Castro, C., Smidansky, E.D., Arnold, J.J., Maksimchuk, K.R., Moustafa, I., Uchida, A., Götte, M., Konigsberg, W., and Cameron, C.E. (2009). Nucleic acid polymerases use a general acid for nucleotidyl transfer. *Nat. Struct. Mol. Biol.* **16**, 212-218.
- Chabot, B., Bisotto, S., and Vincent, M. (1995). The nuclear matrix phosphoprotein p255 associates with splicing complexes as part of the [U4/U6.U5] tri-snRNP particle. *Nucleic Acids Res.* **23**, 3206-3213.
- Chiara, M.D., Palandjian, L., Feld Kramer, R., and Reed, R. (1997). Evidence that U5 snRNP recognizes the 3' splice site for catalytic step II in mammals. *EMBO J.* **16**, 4746-4759.
- Choudhary, C., Kumar, C., Gnad, F., Nielsen, M.L., Rehman, M., Walther, T.C., Olsen, J.V., and Mann, M. (2009). Lysine Acetylation Targets Protein Complexes and Co-Regulates Major Cellular Functions. *Science* **325**, 834-840.
- Davies, D.R., Goryshin, I.Y., Reznikoff, W.S., and Rayment, I. (2000). Three-dimensional structure of the Tn5 synaptic complex transposition intermediate. *Science* **289**, 77-85.
- Dellaire, G., Makarov, E.M., Cowger, J.M., Longman, D., Sutherland, H.G.E., Luhrmann, R., Torchia, J., and Bickmore, W.A. (2002). Mammalian PRP4 Kinase Copurifies and Interacts with Components of Both the U5 snRNP and the N-CoR Deacetylase Complexes. *Mol. Cell. Biol.* **22**, 5141-5156.
- Ding, J., Das, K., Hsiou, Y., Sarafianos, S.G., Clark, A.D., Jr., Jacobo-Molina, A., Tantillo, C., Hughes, S.H., and Arnold, E. (1998). Structure and functional implications of the polymerase active site region in a complex of HIV-1 RT with a double-stranded DNA template-primer and an antibody Fab fragment at 2.8 Å resolution. *J. Mol. Biol.* **284**, 1095-1111.
- Dix, I., Russell, C.S., O'Keefe, R.T., Newman, A.J., and Beggs, J.D. (1998). Protein-RNA interactions in the U5 snRNP of *Saccharomyces cerevisiae*. *RNA* **4**, 1239-1250.
- Dlakic, M., and Mushegian, A. (2011). Prp8, the pivotal protein of the spliceosomal catalytic center, evolved from a retroelement-encoded reverse transcriptase. *RNA* **17**, 799-808.
- Fromont-Racine, M., Rain, J.C., and Legrain, P. (1997). Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. *Nat. Genet.* **16**, 277-282.

- Gillis, A.J., Schuller, A.P., and Skordalakes, E. (2008). Structure of the *Tribolium castaneum* telomerase catalytic subunit TERT. *Nature* **455**, 633-637.
- Gottschalk, A., Kastner, B., Luhrmann, R., and Fabrizio, P. (2001). The yeast U5 snRNP coisolated with the U1 snRNP has an unexpected protein composition and includes the splicing factor Aar2p. *RNA* **7**, 1554-1565.
- Gottschalk, A., Neubauer, G., Banroques, J., Mann, M., Luhrmann, R., and Fabrizio, P. (1999). Identification by mass spectrometry and functional analysis of novel proteins of the yeast [U4/U6.U5] tri-snRNP. *EMBO J.* **18**, 4535-4548.
- Grainger, R.J., Barrass, J.D., Jacquier, A., Rain, J.C., and Beggs, J.D. (2009). Physical and genetic interactions of yeast Cwc21p, an ortholog of human SRm300/SRRM2, suggest a role at the catalytic center of the spliceosome. *RNA* **15**, 2161-2173.
- Grainger, R.J., and Beggs, J.D. (2005). Prp8 protein: at the heart of the spliceosome. *RNA* **11**, 533-557.
- Hartwell, L.H. (1967). Macromolecule synthesis in temperature-sensitive mutants of yeast. *J. Bacteriol.* **93**, 1662-1670.
- Hartwell, L.H., McLaughlin, C.S., and Warner, J.R. (1970). Identification of ten genes that control ribosome formation in yeast. *Mol. Gen. Genet.* **109**, 42-56.
- Horton, N.C., Newberry, K.J., and Perona, J.J. (1998). Metal ion-mediated substrate-assisted catalysis in type II restriction endonucleases. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 13489-13494.
- Kameoka, S., Duque, P., and Konarska, M.M. (2004). p54(nrb) associates with the 5' splice site within large transcription/splicing complexes. *EMBO J.* **23**, 1782-1791.
- Kershaw, C.J., Barrass, J.D., Beggs, J.D., and O'Keefe, R.T. (2009). Mutations in the U5 snRNA result in altered splicing of subsets of pre-mRNAs and reduced stability of Prp8. *RNA* **15**, 1292-1304.
- Kohlstaedt, L.A., Wang, J., Friedman, J.M., Rice, P.A., and Steitz, T.A. (1992). Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science* **256**, 1783-1790.
- Kuhn, A.N., and Brow, D.A. (2000). Suppressors of a cold-sensitive mutation in yeast U4 RNA define five domains in the splicing factor Prp8 that influence spliceosome activation. *Genetics* **155**, 1667-1682.

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

Lin, R.J., Lustig, A.J., and Abelson, J. (1987). Splicing of yeast nuclear pre-mRNA in vitro requires a functional 40S spliceosome and several extrinsic factors. *Genes Dev.* **1**, 7-18.

Lossky, M., Anderson, G.J., Jackson, S.P., and Beggs, J. (1987). Identification of a yeast snRNP protein and detection of snRNP-snRNP interactions. *Cell* **51**, 1019-1026.

Luo, H.R., Moreau, G.A., Levin, N., and Moore, M.J. (1999). The human Prp8 protein is a component of both U2- and U12-dependent spliceosomes. *RNA* **5**, 893-908.

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

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

Maeder, C., Kutach, A.K., and Guthrie, C. (2008). ATP-dependent unwinding of U4/U6 snRNAs by the Brr2 helicase requires the C terminus of Prp8. *Nat. Struct. Mol. Biol.* **16**, 42-48.

Makarov, E.M., Makarova, O.V., Urlaub, H., Gentzel, M., Will, C.L., Wilm, M., and Luhrmann, R. (2002). Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. *Science* **298**, 2205-2208.

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

Martinez-Gimeno, M., Gamundi, M.J., Hernan, I., Maseras, M., Milla, E., Ayuso, C., Garcia-Sandoval, B., Beneyto, M., Vilela, C., Baiget, M., Antinolo, G., and Carballo, M. (2003). Mutations in the pre-mRNA splicing-factor genes PRPF3, PRPF8, and PRPF31 in Spanish families with autosomal dominant retinitis pigmentosa. *Invest. Ophthalmol. Vis. Sci.* **44**, 2171-2177.

McKie, A.B., McHale, J.C., Keen, T.J., Tarttelin, E.E., Goliath, R., van Lith-Verhoeven, J.J., Greenberg, J., Ramesar, R.S., Hoyng, C.B., Cremers, F.P., Mackey, D.A., Bhattacharya, S.S., Bird, A.C., Markham, A.F., and Inglehearn, C.F. (2001). Mutations in the pre-mRNA splicing factor gene PRPC8 in

autosomal dominant retinitis pigmentosa (RP13). *Hum. Mol. Genet.* **10**, 1555-1562.

McPheeters, D.S., and Muhlenkamp, P. (2003). Spatial Organization of Protein-RNA Interactions in the Branch Site-3' Splice Site Region during pre-mRNA Splicing in Yeast. *Mol. Cell. Biol.* **23**, 4174-4186.

Mitchell, M., Gillis, A., Futahashi, M., Fujiwara, H., and Skordalakes, E. (2010). Structural basis for telomerase catalytic subunit TERT binding to RNA template and telomeric DNA. *Nat. Struct. Mol. Biol.* **17**, 513-518.

Naumann, T.A., and Reznikoff, W.S. (2002). Tn5 transposase active site mutants. *J. Biol. Chem.* **277**, 17623-17629.

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., Lührmann, 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.

Peng, R., Dye, B.T., Perez, I., Barnard, D.C., Thompson, A.B., and Patton, J.G. (2002). PSF and p54nrb bind a conserved stem in U5 snRNA. *RNA* **8**, 1334-1347.

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.

Reyes, J.L., Kois, P., Konforti, B.B., and Konarska, M.M. (1996). The canonical GU dinucleotide at the 5' splice site is recognized by p220 of the U5 snRNP within the spliceosome. *RNA* **2**, 213-225.

Ritchie, D.B., Schellenberg, M.J., Gesner, E.M., Raithatha, S.A., Stuart, D.T., and MacMillan, A.M. (2008). Structural elucidation of a PRP8 core domain from the heart of the spliceosome. *Nat. Struct. Mol. Biol.* **15**, 1199-1205.

- Siatecka, M., Reyes, J.L., and Konarska, M.M. (1999). Functional interactions of Prp8 with both splice sites at the spliceosomal catalytic center. *Genes Dev.* **13**, 1983-1993.
- Song, E.J., Werner, S.L., Neubauer, J., Stegmeier, F., Aspden, J., Rio, D., Harper, J.W., Elledge, S.J., Kirschner, M.W., and Rape, M. (2010). The Prp19 complex and the Usp4Sart3 deubiquitinating enzyme control reversible ubiquitination at the spliceosome. *Genes Dev.* **24**, 1434-1447.
- 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.
- Staub, E., Fiziev, P., Rosenthal, A., and Hinemann, B. (2004). Insights into the evolution of the nucleolus by an analysis of its protein domain repertoire. *BioEssays*. **26**, 567-581.
- Stevens, S.W., Barta, I., Ge, H.Y., Moore, R.E., Young, M.K., Lee, T.D., and Abelson, J. (2001). Biochemical and genetic analyses of the U5, U6, and U4/U6 x U5 small nuclear ribonucleoproteins from *Saccharomyces cerevisiae*. *RNA* **7**, 1543-1553.
- Takahashi, A., Muramatsu, H., Takagi, S., Fujisawa, H., Miyake, Y., and Muramatsu, T. (2001). A splicing factor, Prp8: preferential localization in the testis and ovary in adult mice. *J. Biochem.* **129**, 599-606.
- Teigelkamp, S., Whittaker, E., and Beggs, J.D. (1995). Interaction of the yeast splicing factor PRP8 with substrate RNA during both steps of splicing. *Nucleic Acids Res.* **23**, 320-326.
- Towns, K.V., Kipioti, A., Long, V., McKibbin, M., Maubaret, C., Vaclavik, V., Ehsani, P., Springell, K., Kamal, M., Ramesar, R.S., Mackey, D.A., Moore, A.T., Mukhopadhyay, R., Webster, A.R., Black, G.C.M., O'Sullivan, J., Bhattacharya, S.S., Pierce, E.A., Beggs, J.D., and Inglehearn, C.F. (2010). Prognosis for splicing factor PRPF8 retinitis pigmentosa, novel mutations and correlation between human and yeast phenotypes. *Hum. Mutat.* **31**, E1361-E1376.
- Tran, H.J., Allen, M.D., Lowe, J., and Bycroft, M. (2003). Structure of the Jab1/MPN domain and its implications for proteasome function. *Biochemistry* **42**, 11460-11465.
- Turner, I.A., Norman, C.M., Churcher, M.J., and Newman, A.J. (2006). Dissection of Prp8 protein defines multiple interactions with crucial RNA sequences in the catalytic core of the spliceosome. *RNA* **12**, 375-386.

Umen, J.G., and Guthrie, C. (1995). A novel role for a U5 snRNP protein in 3' splice site selection. *Genes Dev.* **9**, 855-868.

Umen, J.G., and Guthrie, C. (1995). Prp16p, Slu7p, and Prp8p interact with the 3' splice site in two distinct stages during the second catalytic step of pre-mRNA splicing. *RNA* **1**, 584-597.

Umen, J.G., and Guthrie, C. (1996). Mutagenesis of the yeast gene PRP8 reveals domains governing the specificity and fidelity of 3' splice site selection. *Genetics* **143**, 723-739.

van Nues, R.W., and Beggs, J.D. (2001). Functional contacts with a range of splicing proteins suggest a central role for Brr2p in the dynamic control of the order of events in spliceosomes of *Saccharomyces cerevisiae*. *Genetics* **157**, 1451-1467.

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

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

Vincent, M., Lauriault, P., Dubois, M.F., Lavoie, S., Bensaude, O., and Chabot, B. (1996). The nuclear matrix protein p255 is a highly phosphorylated form of RNA polymerase II largest subunit which associates with spliceosomes. *Nucleic Acids Res.* **24**, 4649-4652.

Wachtel, C., and Manley, J.L. (2009). Splicing of mRNA precursors: the role of RNAs and proteins in catalysis. *Mol. BioSyst.* **5**, 311.

Wang, Y., Juranek, S., Li, H., Sheng, G., Tuschl, T., and Patel, D.J. (2008). Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. *Nature* **456**, 921-926.

Wyatt, J.R., Sontheimer, E.J., and Steitz, J.A. (1992). Site-specific cross-linking of mammalian U5 snRNP to the 5' splice site before the first step of pre-mRNA splicing. *Genes Dev.* **6**, 2542-2553.

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

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

Zeng, L., and Zhou, M.M. (2002). Bromodomain: an acetyl-lysine binding domain. *FEBS Lett.* **513**, 124-128.

Zhang, C., Dowd, D.R., Staal, A., Gu, C., Lian, J.B., van Wijnen, A.J., Stein, G.S., and MacDonald, P.N. (2003). Nuclear coactivator-62 kDa/Ski-interacting protein is a nuclear matrix-associated coactivator that may couple vitamin D receptor-mediated transcription and RNA splicing. *J. Biol. Chem.* **278**, 35325-35336.

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

Chapter 2

Structural and functional studies of the Prp8 active site

The work presented in this chapter represents collaboration between authors. Tao Wu, Matt Schellenberg, Dustin Ritchie, and Karim Atta prepared proteins for crystallization, grew the crystals, prepared mutant yeast strains, and performed splicing assays. Matt Schellenberg, Dustin Ritchie, and Tao Wu performed the X-ray data collection. Matt Schellenberg solved the crystal structures.

2-1. Introduction

2-1.1. Two monomers of Prp8 RNase H domain

In chapter 1 the X-ray structure of hPrp8 RNase H domain from our lab was discussed. One intriguing feature of this structure is the existence of two monomers in the asymmetric unit. We originally considered this to be insignificant but later reconsidered this judgement. This chapter describes our efforts to further elucidate the role of Prp8 in splicing.

The Prp8 RNase H domain has a 17 amino acid insertion within the RNase H fold. In monomer a of the crystallographic asymmetric unit these amino acids are well structured and form a two-stranded anti-parallel β -sheet ($\beta 2/\beta 3$) extruding from the RNase H core, while in monomer b this region was disordered as a displaced loop (Ritchie et al., 2008). It has been shown that the largest proportion of suppressor alleles associated with the Prp8 RNase H domain map to the β -hairpin region (reviewed in: Grainger and Beggs, 2005) suggesting that this region is crucial for the function of Prp8 within the spliceosome.

Although we did not observe metal ions in our original structure, the question still remained whether the Prp8 RNase H domain is a metalloprotein. One of the two canonical RNase H metal-binding sites is present with two Asp residues spatially conserved with respect to Mg^{2+} coordinating Asp residues of RNase H. The mutation of these putative ligands in Prp8 to abolish the possibility of metal coordination results in growth defects or lethality in yeast (Pena et al.,

2008). All of this evidence points to metal binding within the potential active site. In this chapter this possibility will be further investigated.

2-1.2. One metal vs. two metal mechanisms

Two-metal ion mechanisms have been long proposed for a large group of protein-metallonucleases. However, the detailed mechanism of metallonuclease reactions, typically supported by Mg^{2+} , has a long and controversial history. Two-metal ion mechanisms have enjoyed much favour, based largely on the X-ray crystal structures of these enzymes with more than one metal ion in the active site. Nonetheless, this mechanism has been under challenge with new evidence reported (reviewed in: Dupureur, 2010).

To catalyze phosphodiester bond cleavage, four main active-site components are required: a nucleophile to which the phosphoryl group can be transferred; a general base to activate and position the nucleophile; one or more Lewis acids to stabilize the negative charge that develops in the transition state; and last, a general acid that can protonate the leaving group oxygen (reviewed in: Horton et al., 2004). The best-studied examples are the endonuclease family. These enzymes are rich in acidic residues, which are ideal ligands for multiple Mg^{2+} ions. But members from this group that are closely related structurally have been crystallized with differing numbers of metal ions. For example, a single metal ion was observed in EcoRI structures (pdb code 1QPS) in both pre-transition

and post-reaction ternary complexes of EcoRI-DNA-Mn²⁺ (Figure 2-1 A). In addition, a ternary complex of BglII revealed only one Mn²⁺ binding site per subunit (Lukacs et al., 2000). In contrast, BamHI, a close structural relative of EcoRI, has been crystallized with two metal ions (reviewed in Pingoud et al., 2009; Figure 2-1 B). A special case in the endonuclease family is EcoRV, which binds a single Mg²⁺ in the enzyme-substrate complex and binds two Mg²⁺ ions in the enzyme-product complex (Kostrewa and Winkler, 1995). Another endonuclease, I-PpoI features a catalytic site with a Mg²⁺ coordinated by a conserved Asn and three water molecules (Galburt et al., 1999).

A similar ambiguity remains when the structures of other nuclease families are examined. The RNase H domain of reverse transcriptase crystallized with two metal ions, while the RNase H from *E. coli* contains only one metal ion (Nowotny et al., 2005; Katayanagi et al., 1993; Figure 2-1 C). In a recent study, Berger and coworkers have proposed a mechanism for type II and IA topoisomerase where the transition state is stabilized by a single bound metal and conserved arginine (Schmidt et al., 2010; Figure 2-1 D). In this structure, a second metal ion is found, but it is only involved in anchoring of the (-1) phosphate. There are some differences in conserved active site groups among the above enzymes that might be used to support the idea of one or two metal binding sites, but clearly there are limits to the interpretation of structures when it comes to mechanism (reviewed in Dupureur, 2010).

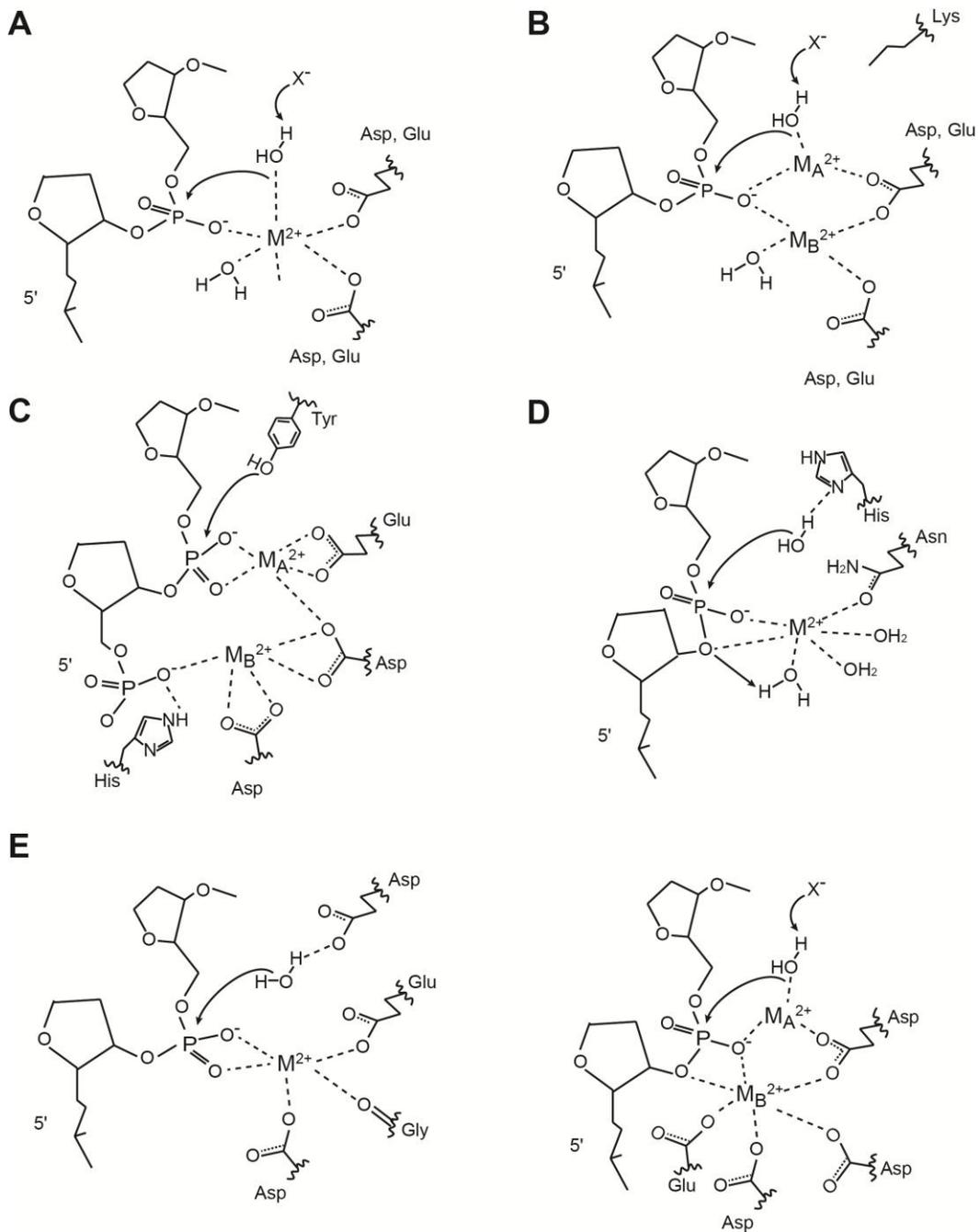


Figure 2-1. Examples of general metal ion driven mechanisms in protein metallonucleases. A) One-metal ion mechanism in EcoRI active site (adapted from Dupureur, 2008). (B) Two-metal ion mechanism in BamHI active site (adapted from Dupureur, 2008). (C) One-metal ion mechanism in I-PpoI active site (Adapted from Galburt et al., 1999). (D) One-metal ion mechanism in type II and IA topoisomerase active site (adapted from Schmidt et al., 2010). (E) (*left*) One-metal ion mechanism in *E. coli* RNase H active site (adapted from Katayanagi et al., 2003). (*right*) Two-metal ion mechanism in *B. halodurans* RNase H active site (adapted from Nowotny et al., 2005).

The most informative experimental data on the one- vs. two-metal mechanism issue comes from the global kinetic analysis of the PvuII endonuclease, an enzyme with two metal ion binding sites, which can cut DNA with one metal ion. The rate is enhanced by two orders of magnitude when a second metal ion is present (Xie et al., 2008). In this case it seems that the second metal makes the reaction more efficient and it is not absolutely required for the catalysis. Several other computational and biochemical studies have also revealed the unequal importance of the two metals and proposed the second metal ion in the active site to be auxiliary (Mones et al., 2007a; Mones et al., 2007b).

Whether each metallonuclease uses one- or two-metal mechanism remains elusive. But one thing is assured: nucleic acids can be hydrolyzed with one metal ion, as shown in aforementioned cases as well as numerous other examples.

2-1.3. Metals in the spliceosome

High resolution structural analysis of the group I and group II introns has supported a general two-metal ion mechanism proposed for catalytic RNA based on analysis of phosphoryl transfer by protein enzymes (Steitz and Steitz, 1993; Adams et al., 2004; Lipchock et al., 2008; Toor et al., 2008). In both structures, the two metals are well positioned to promote catalysis of the exon ligation reaction, with M_1 activating the nucleophile and M_2 stabilizing the leaving group, and both metals coordinating the scissile phosphate (Figure I-7 B). This model

suggests how the second step of the splicing reaction is carried out in group I and group II introns.

It has long been believed that the spliceosome is in essence a ribozyme based on the mechanistic similarity between self-splicing group II introns and the processing of nuclear pre-mRNAs (Sharp, 1991). 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). The mechanistic similarity regarding the function of the U2 and U6 snRNAs has led to the suggestion that pre-mRNA splicing is also an essentially RNA catalyzed process. Steitz and Steitz have suggested a two metal ion mechanism for the spliceosome (Steitz and Steitz, 1993; Figure 2-2). In this model, an RNA site serves as a scaffold to bind two catalytic metal ions. The mechanism postulates that the catalysis is facilitated by two divalent metal ions, with one of them activating the attacking water or sugar hydroxyl and the other coordinating and stabilizing the oxyanion leaving group. Both ions act as Lewis acids and stabilize the expected pentacovalent transition state.

There is evidence for coordination of one metal ion to the internal stem loop of U6 snRNA (U6 ISL; Yean et al., 2000). U6 snRNA with a sulphur substituted non-bridging phosphoryl oxygen of nucleotide U80 reconstitutes a fully assembled yet catalytically inactive spliceosome, while adding a thiophilic ion such as Mn^{2+} rescues the first transesterification reaction. Intriguingly, this region was proposed to bind metal ions in a manner similar to that of DV in group II

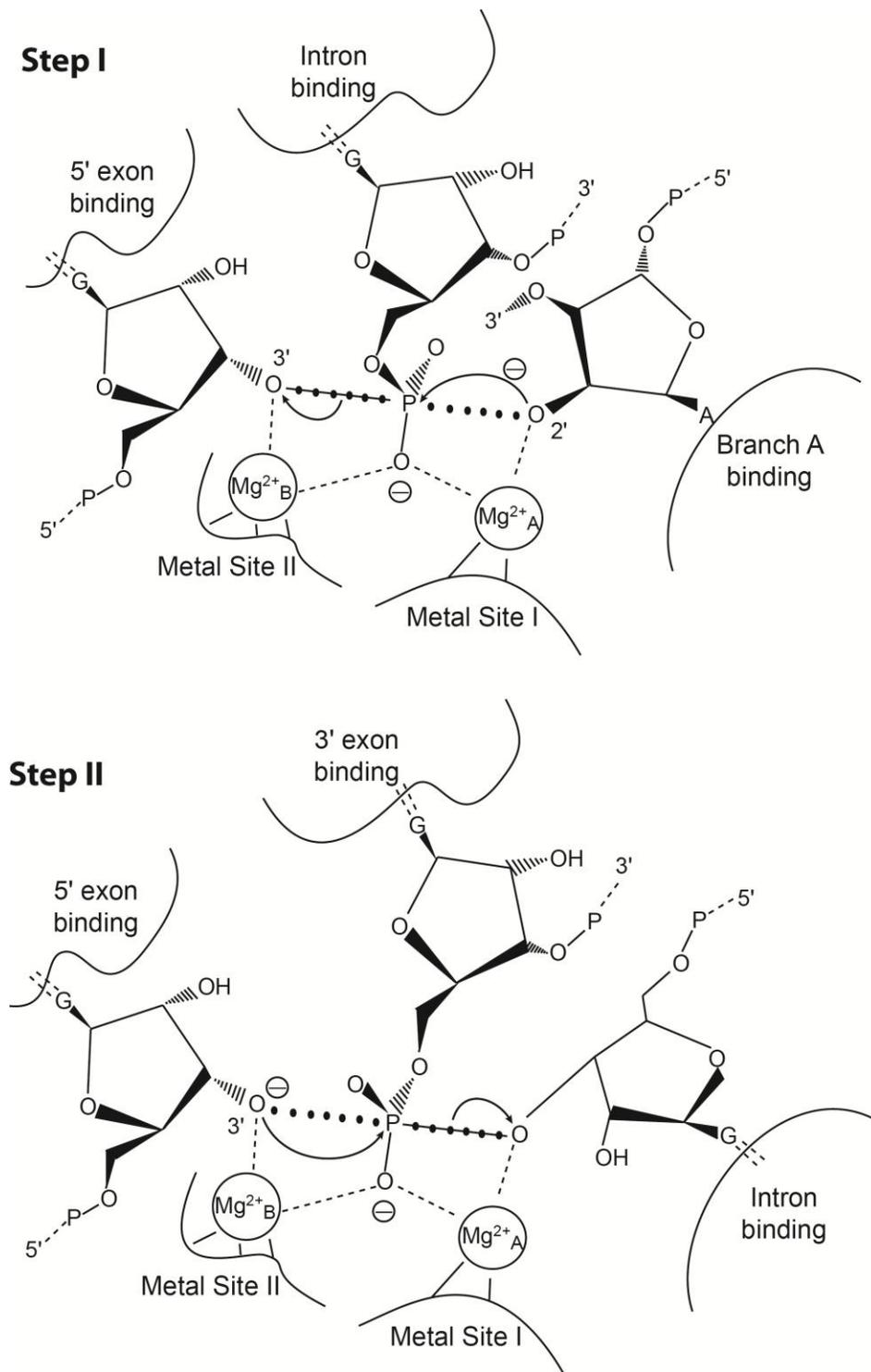


Figure 2-2. Proposed two metal mechanism for splicing. In both steps, divalent metal activates a nucleophile, and stabilizes the transition state/intermediate and leaving group. The identities of the substrate and metal-binding sites in the spliceosome are unknown and the active sites for the two steps may be distinct (Adapted from Steiz and Steiz, 1993).

intron (Toor et al., 2008). Metal coordination has also been reported involving the scissile phosphodiester linkages of the substrate (Sontheimer et al., 1997; Gordon et al., 2000). In this case, 3'-sulfur substitution at the 5'SS and 3'SS of a nuclear pre-mRNA causes a switch in metal specificity from Mg^{2+} to Mn^{2+} in the first and the second step of splicing respectively, which suggests that the spliceosome uses a catalytic metal ion to stabilize the 3'-oxyanion leaving group during both steps of splicing. The metal binding pattern is thus similar in the spliceosome to those observed in protein structures.

Considering the evidence that the splicing reaction in the protein-free system is significantly slower than that in spliceosome-catalyzed systems (See chapter 1; Valadkhan and Manley, 2001; Valadkhan et al., 2007), it is highly possible that the spliceosome active site is composed of both RNA and protein components. Indeed, crosslinking studies have revealed the intimate association of spliceosomal proteins with the splice sites and branch nucleotide throughout spliceosome assembly and perhaps through both steps of splicing (MacMillan et al., 1994; Umen and Guthrie, 1995; Query et al., 1996). In particular, Prp8 is the only spliceosomal protein shown to directly contact all reactive sites within the intron: 5'SS, 3'SS, BPS and PPT (Figure 1-2 A). Thus, the existence of the RNase H domain and conserved residues makes it the best candidate to carry out the role of catalysis within the active site of spliceosome.

2-2. Results

2-2.1. Mg^{2+} binds to the Prp8 active site within monomer b

No Mg^{2+} ion was observed in the putative active site in our original wild-type hPrp8 RNase H domain structure as well as the yPrp8 RNase H domain structures, despite growing the hPrp8 crystal in 200 mM Mg^{2+} and soaking the yPrp8 crystals in 20 mM Mg^{2+} (Pena et al., 2008; Ritchie et al., 2008; Yang et al., 2008; Figure 2-3 B). The positioning of R1865 in the active site was proposed to block the binding of Mg^{2+} (Ritchie et al., 2008). Similar mechanisms have been observed in the RNase H fold of the Argonaute PIWI domain and also in the Tn5 transposase (Song et al., 2004; Davies et al., 2000). In the Argonaute PIWI domain, conformational changes occur upon binding to the substrate, and the Arg residue is displaced to unmask the Mg^{2+} binding site (Wang et al., 2008). It is proposed that similar masking and unmasking events occur in Prp8 and the side chain of R1865 is replaced by RNA binding during spliceosome assembly and activation.

In order to test the hypothesis that R1865 blocks the metal binding, we expressed and crystalized the R1865A hPrp8 RNaseH mutant. The X-ray structure was solved at 1.15Å resolution. The R1865A mutant was essentially identical in structure to the wild-type protein with two monomers in the asymmetric unit. The β -hairpin region of monomer b, which is largely disordered in the crystals obtained with the wild-type protein, is now visible (Figure 2-3 C). In this structure the translation of the β -hairpin region between the two monomers becomes much

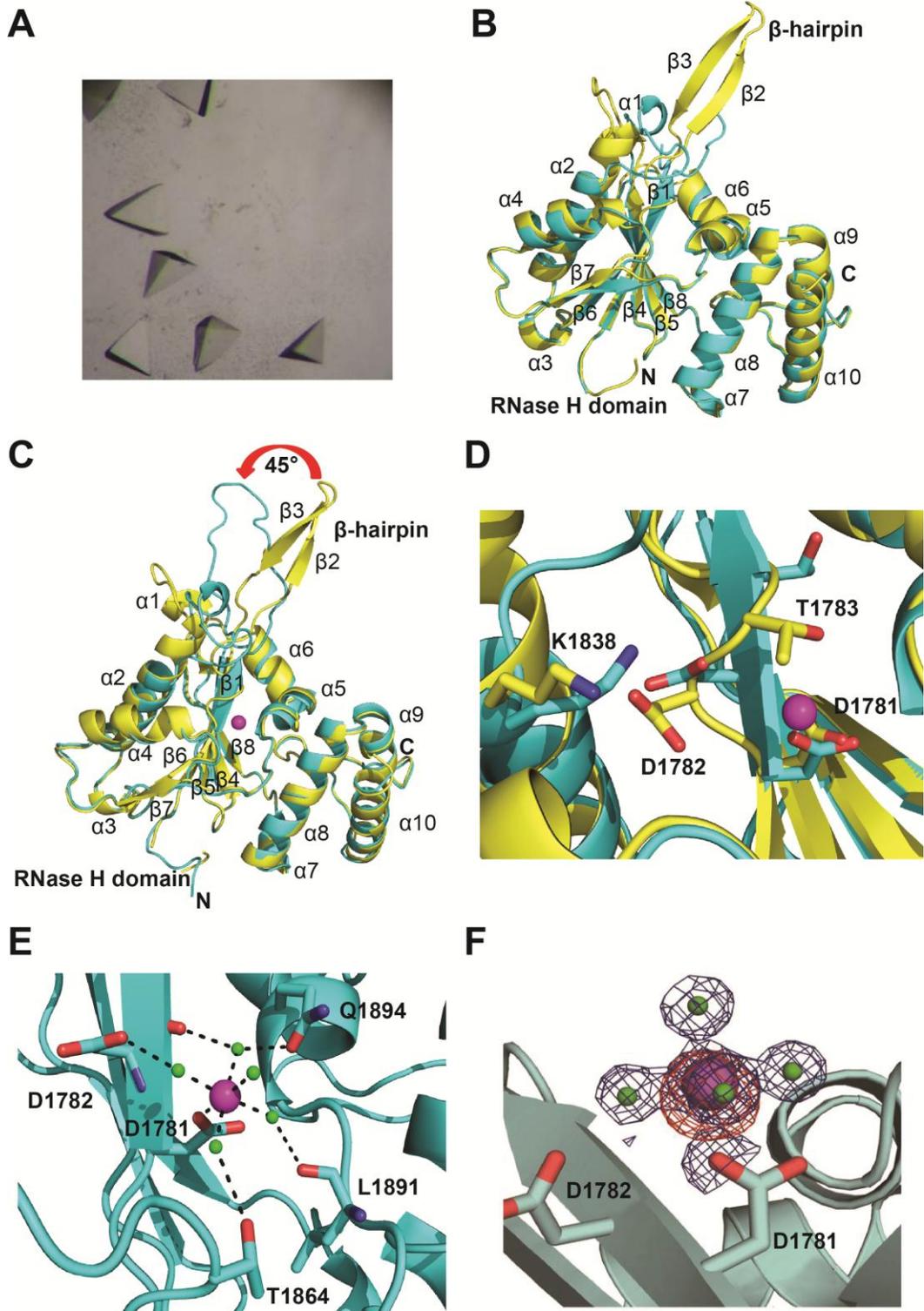


Figure 2-3. (Legend on next page)

clearer: 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 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 hairpin. As a consequence, D1782 moves closer to D1781 and T1783 is displaced $\sim 4 \text{ \AA}$ upwards (Figure 2-3 D). A Mg^{2+} ion is observed in the active site at high occupancy only in monomer b. Coordination of this Mg^{2+} ion involves inner sphere contact with the side chain of D1781 and outer sphere coordination via five ordered water molecules to the carboxylate of D1782, carbonyls of D1782 and L1891, the amide carbonyl of Q1894, and the hydroxyl of T1864 (Figure 2-3 E, F). This assignment was confirmed by soaking the R1865A crystals in CoCl_2 and observing strong anomalous diffraction from Co^{2+} bound in the active site. The new crystallization conditions have 300mM MgCl_2 , compared to 100-200 mM MgCl_2 in our old conditions for the wild-type proteins crystals. Therefore we soaked R1865A crystals in decreasing concentrations of Mg^{2+} before data collection. We observed the Mg^{2+} ion in concentrations of MgCl_2 as low as 50 mM, but not in 20 mM.

Figure 2-3. Conformational plasticity in the Prp8 RNase H domain. A) Crystals of wild-type Prp8 RNase H domain and Prp8 RNase H domain R1865A proteins in the new crystallization conditions. B) Overlay of two monomers from the asymmetric unit of the wild-type Prp8 RNase H domain crystal structure. In monomer b (*cyan*), the β -hairpin is disrupted. C) Overlay of two monomers from the asymmetric unit of the Prp8 RNase H domain R1865A crystal structure. In monomer b (*cyan*), the β -hairpin is more ordered and translated $\sim 45^\circ$, the $\alpha 2$ helix shortened and $\beta 1$ strand extended. The Mg^{2+} is shown in purple. D) Displacement of T1783 upon conformational rearrangement facilitates binding of Mg^{2+} (*purple*) at the canonical site. Shown is overlay of monomer a (*yellow*) and monomer b (*cyan*). E) Detail of coordination of Mg^{2+} ion (*purple*) and its inner-sphere waters (*green*) in monomer b of Prp8 Domain IV. F) Detail of 1.15 \AA 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.

This apparent K_d is higher than the physiological concentration under which pre-mRNA splicing occurs (~2 mM), but additional Mg^{2+} ligands (e.g. phosphodiester RNA backbone) in the spliceosomal context likely contribute to enhance Mg^{2+} ion occupancy at this site.

Interestingly, under the new crystallization conditions, which contain higher concentrations of Mg^{2+} and 10-15% of PEG 4000 instead of NaCl as precipitant, we were able to obtain crystals from wild-type protein with partial occupancy of this site by Mg^{2+} , Mn^{2+} and Co^{2+} observed in monomer b. We repeated the same Mg^{2+} titration for crystals before data collection. No Mg^{2+} ion was bound at 50 mM $MgCl_2$, and even in as high as 1 M $MgCl_2$, less than a full occupancy of the Mg^{2+} site was observed. The weaker Mg^{2+} association may be explained by the presence of the R1865 positive charge. It thus appears that it is the larger conformational change observed between monomer a and b that is required for binding of divalent metal to the Prp8 RNase H site, although rearrangement of R1865 may contribute to enhanced metal-binding within the context of the spliceosome.

2-2.2. Mutations of metal coordinating residues cause a splicing defect

If the Asp residues (D1781 and D1782) are important for Mg^{2+} coordination, one would expect that mutations of these residues would impair splicing as well as the growth of cells. Previous reports have shown several mutations of these

residues to be lethal or severely deleterious (Pena et al., 2008; Yang et al., 2008). For example, mutation of the outer sphere ligand, D1854 to alanine (D1782A in human), results in severe growth defects at both 16 °C and 37 °C, while the mutation of the inner sphere D1853 to alanine (D1781A in human) is lethal.

In order to further investigate the possible role of metal-binding in Prp8 function, we first repeated the above mutations (Figure 2-4 A). In our strain background, the D1854A mutant has no obvious growth defect and D1853A is lethal. We then performed more mutations on the inner sphere ligand D1853. As expected, mutations of D1853 to Met and Lys are lethal. Mutation to Cys causes lethality at 37 °C and slower growth (~half of the rate of wild-type) at 30 °C. Mutation to Ser causes severely slower growth at 37 °C and slightly slower growth at 30 °C, while mutation to Asn, which would still preserve significant metal binding potential, is only mildly impaired at 37 °C. The conservative D1853 to Glu mutation does not cause any apparent growth defect. Interestingly, mutation of D1853 to His, which has an imidazole ring and interacts with an Mg²⁺ at the active center of different enzymes (reviewed in: Harding, 2004), does not result in an obvious growth defect.

We next examined *in vitro* splicing of an ACT1 pre-mRNA comparing extracts prepared from wild-type and PRP8 D1853 mutant yeast strains. Of all the four mutants we tested, D1853E and D1853H exhibited similar splicing efficiency

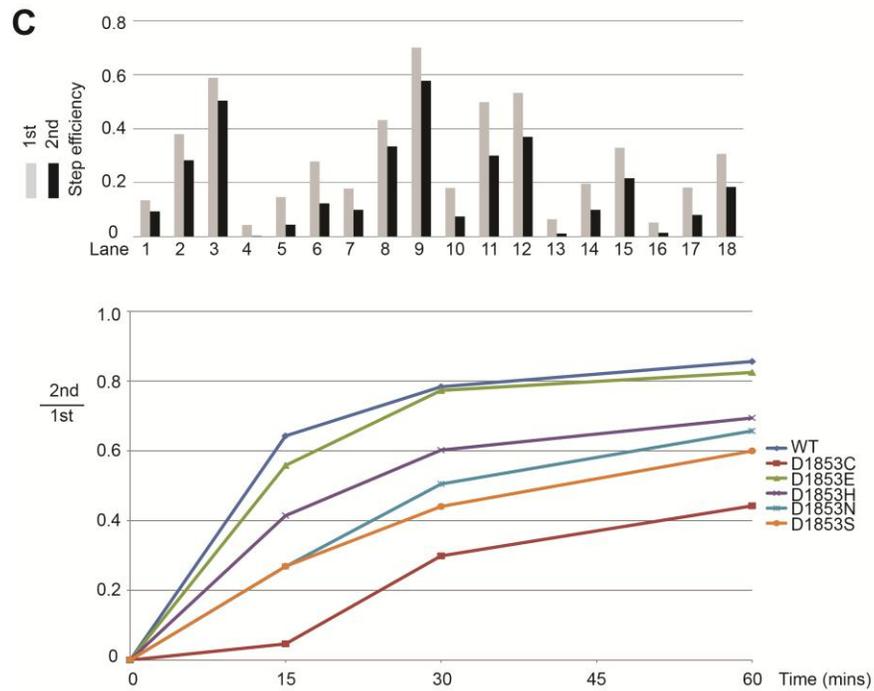
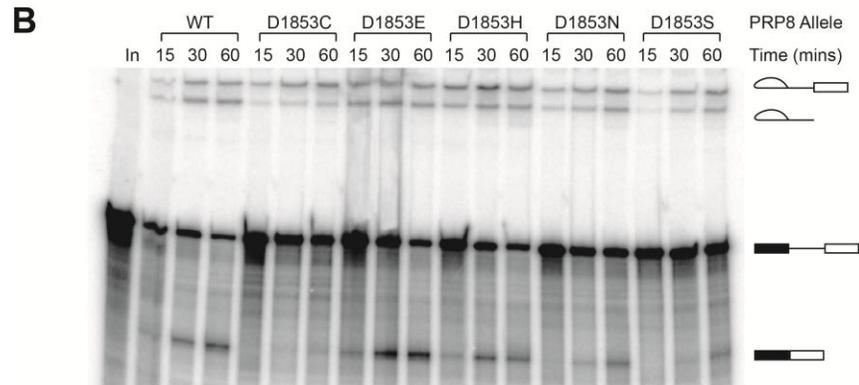
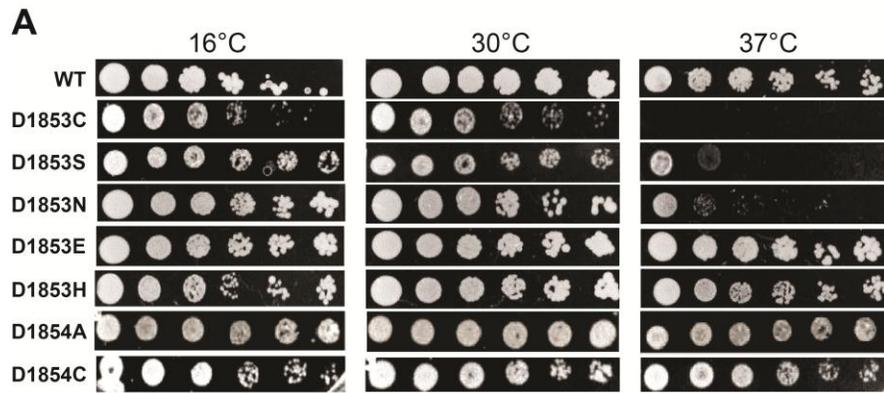


Figure 2-4. (Legend on next page)

in both steps compared to the wild-type. D1853N and D1853S were mildly impaired for both steps of splicing. D1853C was mildly impaired in the first step and exhibited a strong defect in the second step (Figure 2-4 B, C).

2-2.3. The impairment of D1853C function is not due to any change in structure

The splicing defect of D1853C brings forward the possibility of metal switch experiment within the active site. A Cys residue has not been reported in the literature as coordinating Mg^{2+} ; while in the crystal structure of phosphoenolpyruvate carboxykinase (PEPCK), a cysteine residue was shown to coordinate the active site Mn^{2+} (reviewed in: Harding, 2004; Holyoak et al., 2006). Cys residues have also been reported to coordinate Zn^{2+} and Co^{2+} in numerous crystal structures (reviewed in: Harding, 2004). In addition, it has also been shown that cysteine residue substitutions at each of the DDE motif residues in Tn10 transposase result in a change in the divalent metal ion requirements for catalysis from Mg^{2+} to Mn^{2+} (Allingham et al., 1999). Therefore, we checked the

Figure 2-4. PRP8 D1853 mutations impair the second step of splicing. A) Spot test of D1853 mutants. D1853A and D1853L mutants are not shown since they are both lethal. B) Pre-mRNA splicing with D1853 mutants. 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 exon-lariat intermediate. C) (*upper*) Quantification of the first and second step efficiency of D1853 mutants. (*bottom*) Quantification of the second step efficiency over the first step efficiency. The number indicates the percentage of RNA substrate undergoing the second step after finishing the first step.

splicing of D1853C in the presence of different divalent metal ions (Figure 2-5 C). However, no metal rescue was observed when Mn^{2+} , Cd^{2+} , Zn^{2+} or Co^{2+} were added into the D1853C reactions.

We also solved the structure of the corresponding mutation in the human protein, D1781C, and a D1781C/R1865A double mutant. In both cases, the Prp8 RNase H domain and the architecture of the metal binding site is preserved. Consistent with the *in vitro* splicing result, Mg^{2+} does not bind to this site. And in contrast to the wild-type protein, we were unable to observe bound Mn^{2+} following crystal soaks (Figure 2-5 A, B). When the crystal was soaked in Zn^{2+} , we were able to detect metal in the active site. However, the coordination of the Zn^{2+} involves an amide nitrogen (data not shown). Although this unusual structure has been reported in small molecules (Mohamadou and Gérard, 2001), it is highly unlikely the Zn^{2+} in D1781C is relevant to the function of the active site.

2-2.4. The impairment of D1853C function is not due to defects in complex formation, non-snRNP protein functions or the first step

To investigate whether the impaired second step of splicing in D1853C is due to the deficiency of spliceosome assembly, we checked the spliceosomal complex formation *in vitro*. In yeast, the spliceosome assembly involves the formation of six complexes: CC, A, B1, B2, C1, C2, corresponding to E, A, B, B*, C and post-spliceosomal complexes respectively in mammalian cells (Cheng and

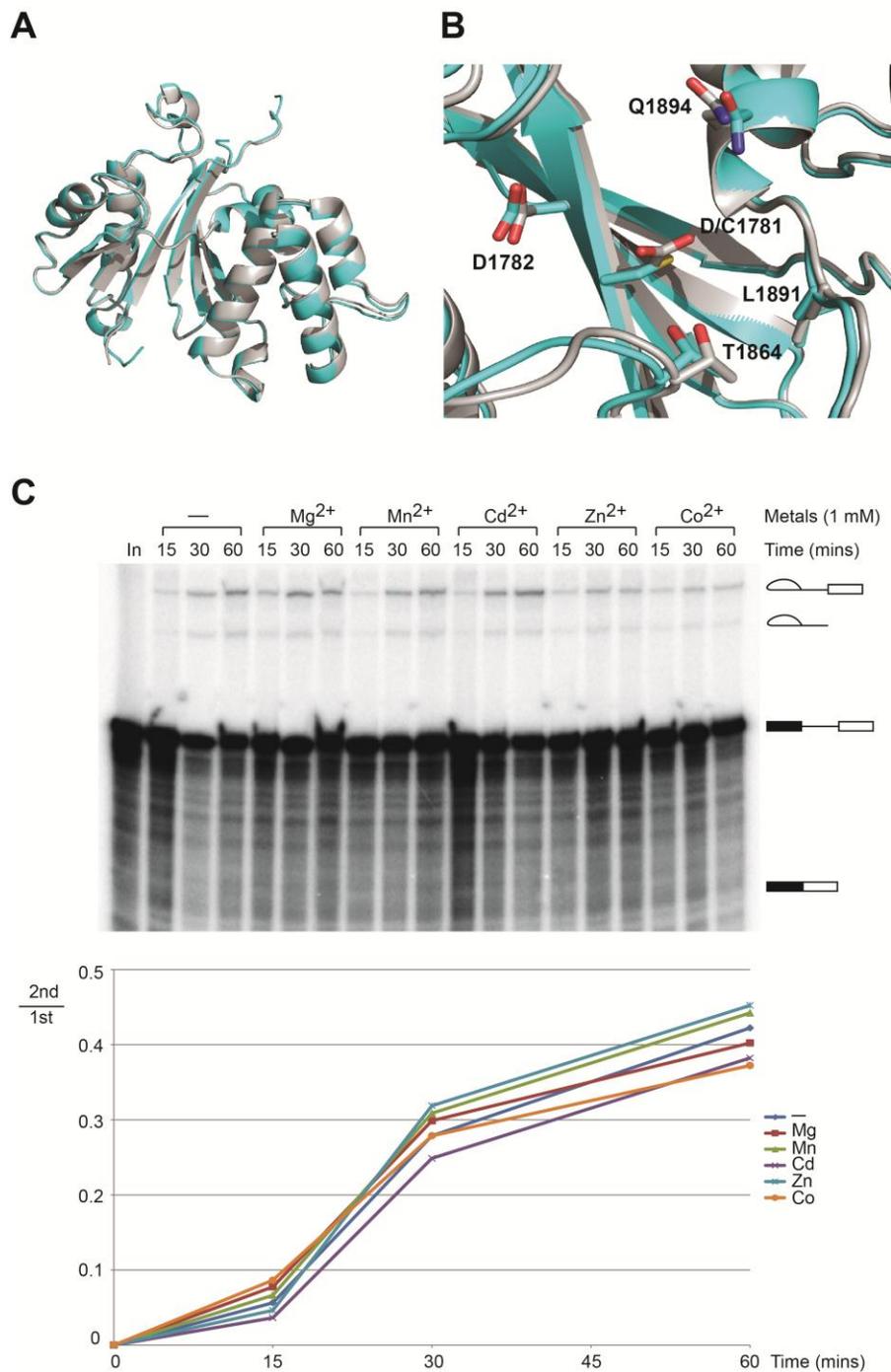


Figure 2-5. PRP8 D1853C does not bind metal. A) Overlay of monomer b from the wild-type (*grey*) and D1853C (*cyan*) RNase H domain crystal structures. In both of them, the β -hairpin is disrupted. B) Overlay of the active site of monomer b from the wild-type (*grey*) and D1853C (*cyan*) RNase H domain crystal structures. All the conserved residues are in the same orientation. C) (*upper*) *In vitro* splicing of D1853C extract in the addition of 1 mM divalent metal ions. In, Input; -, without additional metals. (*bottom*) Quantification of the second step efficiency over the first step efficiency. No metal rescue of the second step is observed.

Abelson, 1987; Ruby and Abelson, 1991; Figure I-2). We compared the complex formation on the ACT1 pre-mRNA within the wild-type and D1853C yeast extracts. D1853C extract formed spliceosomal complexes at a comparable rate to the wild-type extract, indicating the spliceosome assembly is not compromised in the mutant (Figure 2-6 A).

Another explanation for the second step deficiency in D1853C is that the non-snRNP proteins which are required for the exon ligation step are compromised in the mutant. To test this hypothesis, we designed a chase experiment (Figure 2-6 B). The wild-type or D1853C extract were first incubated with labeled ACT1 mRNA. The reaction was then chased with the wild-type extract in which the snRNP proteins were saturated with excessive cold mRNA. If non-snRNP proteins were defective in the D1853C mutant, we should see a rescue of the second step efficiency. However, there was no difference between the non-chased and chased reactions. Thus, defects in non-snRNP proteins are not responsible for the impairment in D1853C.

Recently, Staley and colleagues have reported a proofreading mechanism in splicing (Koodathingal et al, 2010). Before 5'SS cleavage, Prp16 antagonizes splicing by competing with 5'SS cleavage and thereby permits rejection of suboptimal pre-mRNA through a kinetic proofreading mechanism. After 5'SS cleavage, Prp16 monitors the first step product and promotes splicing by allowing exon ligation of the optimal substrate. If the D1853C mutation results in the formation of suboptimal structures at the 5'SS or the lariat intermediate, this may

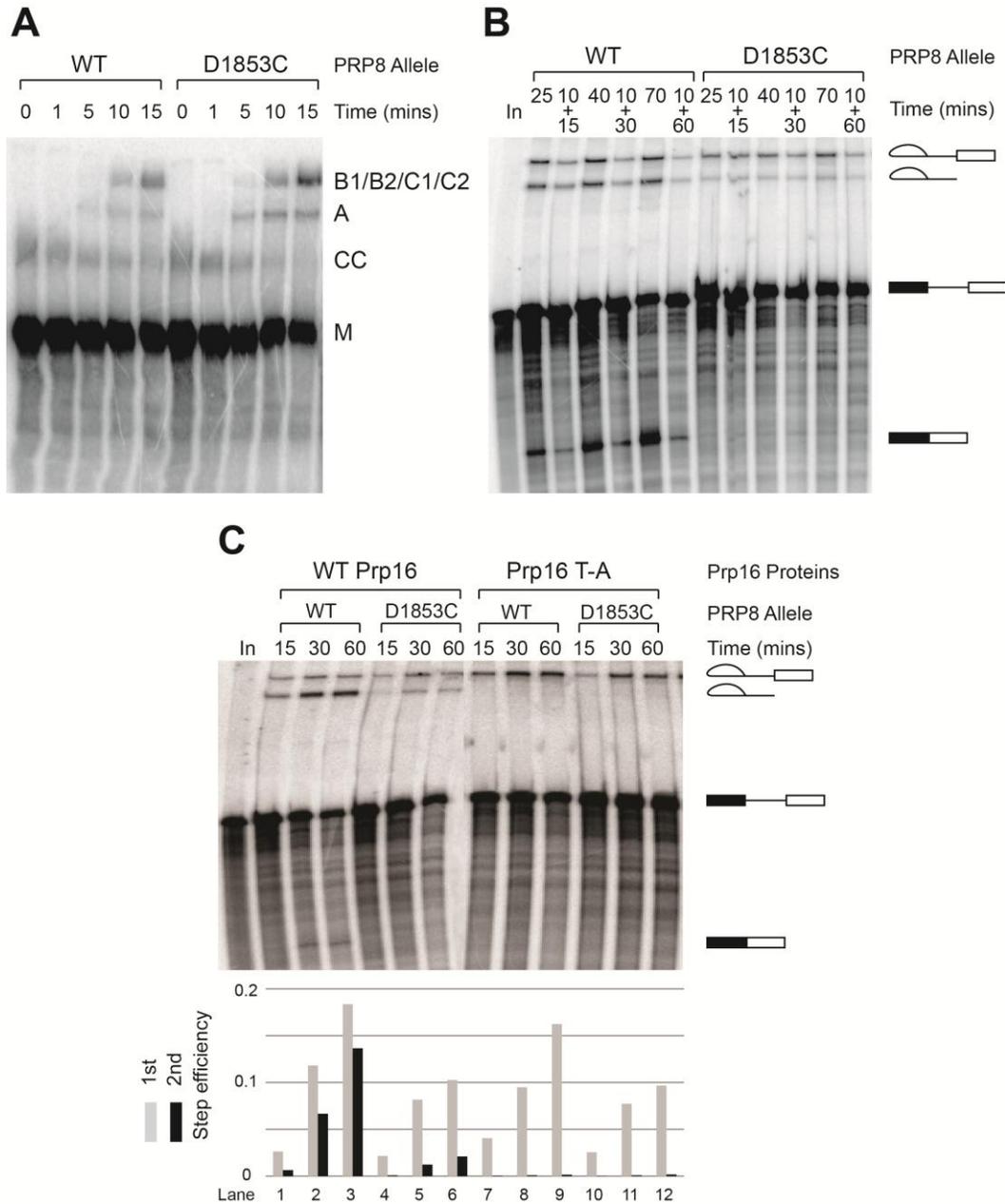


Figure 2-6. The second step defect of D1853C is not due to defects in complex formation, non-snRNP protein functions or the first step. A) D1853C mutant splicing extracts form spliceosomal complexes on the ACT1 pre-mRNA at a comparable rate to splicing extract from wild-type yeast. M, pre-mRNA; CC, commitment complex; A, A complex; and B, B complex. B) Chase experiment with wild-type and D1853C extracts. The upper number indicates the incubation time before adding the competitor and the lower number indicates the incubation time after that. The addition of the splicing system saturated with cold mRNA substrate did not rescue the second step defect of D1853C. C) (*upper*) Splicing in the presence of wild-type Prp16 and Prp16 T-A mutant. (*bottom*) Quantification of the first step and second step efficiency.

explain why the first step is mildly impaired while the second step is severely affected. In order to test this hypothesis, we repeated splicing in the addition of wild-type Prp16 or Prp16 T380A mutant (Figure 2-6 C). This T-to-A mutation within motif I of Prp16 has been shown to abolish the ATP binding activity (Hotz and Schwer, 1998). If the 5'SS formation or the first step product is suboptimal in D1853C and degraded by the Prp16 activated pathway, an increase in the first step splicing efficiency should be observed when the Prp16 T-A mutant is added. We did not observe any change in this experiment, indicating the first step splicing is normal in D1853C.

The spliceosome undergoes extensive conformational changes during the transition from the first step to the second step (Umen and Guthrie, 1995; Chua and Reed, 1999). It has been hypothesized that the conformational changes may be rate-limiting rather than the actual chemical step of exon ligation and that in a metal-rescue study the metal-specificity in the second step was masked by this (Gordon et al., 2000). In order to more closely examine the exon ligation step, isolated from between step conformational change, we carried out a bimolecular exon-ligation assay (Anderson and Moore, 1997), which divides the full-length ACT1 pre-mRNA substrate into two fragments. The labelled 5' RNA substrate (373 nt) is derived from the ACT1 pre-mRNA and contains an exon, consensus 5'SS, branch site, and the polypyrimidine tract. The synthetic 3' RNA substrate (25 nt) used in this report contains the last 5 nt of the intron (UUUAG) and is

unlabelled. While the second step was observed in the wild-type extract, the D1853C extract was unable to carry out the exon ligation (Figure 2-7).

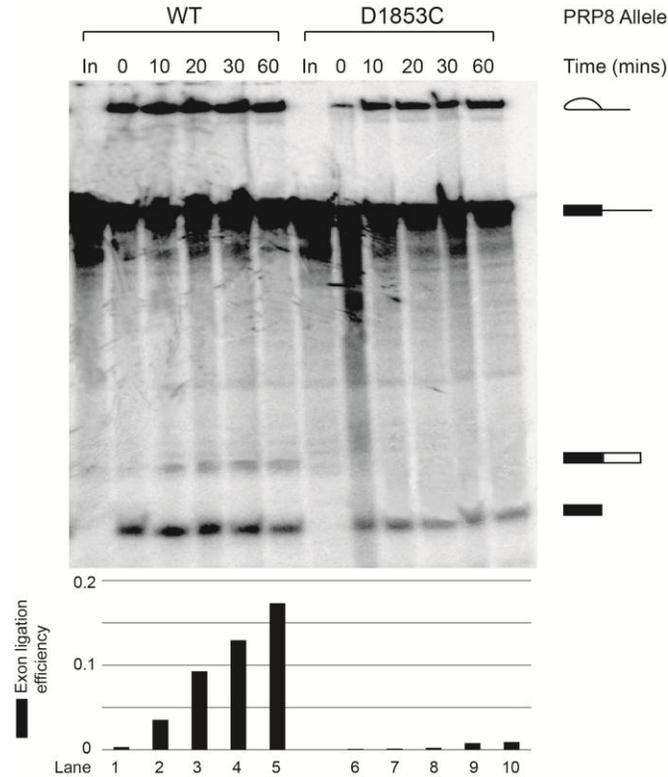


Figure 2-7. Bimolecular exon-ligation assay with WT and D1853C. (*upper*) Exon ligation assay with wild-type and D1853C extracts. Labelled 5' portion of the ACT1 pre-mRNA was pre-incubated with the extract for 30 min, followed by the addition of the unlabelled 3' portion. On the right side of the gel, from bottom to top are indicated the 5' exon, mRNA, 5' portion of ACT1 pre-mRNA and 5' exon-lariat intermediate. (*bottom*) Quantification of the exon ligation efficiency.

The 5' RNA substrate used in the bimolecular exon-ligation assay is also a good tool to assess the kinetics of the first step splicing. We incubated the wild-type and D1853C extract with this substrate for up to 2 hrs. Although the final efficiency in D1853C was only ~70% of that in wild-type, both of them displayed similar enzyme kinetics. The half time or $t_{1/2}$ in this first-order reaction is almost the same in both the wild-type and the mutant (~30 minutes) and after 50 minutes

both of them reached the plateau (Figure 2-8). Therefore the observed mild impairment in D1853C was not due to the loss in the enzyme activity; rather it was caused by other factors. This is probably true because although the protein concentration is roughly the same for all our yeast extracts (~20 mg/mL), the snRNA concentration in D1853C is significantly lower (about half of that in wild-type; data not shown).

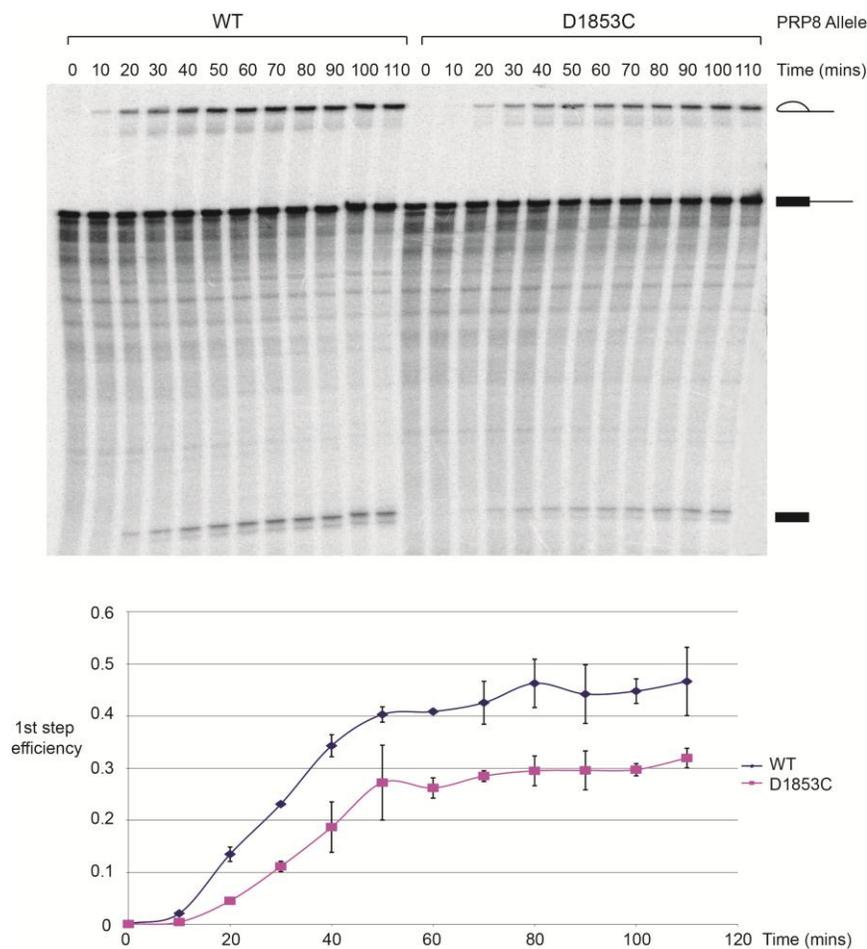


Figure 2-8. Kinetics of the first step of splicing. (*upper*) Wild-type and D1853C extracts were incubated with the first-step substrate for up to 110 minutes. (*bottom*) The quantification of the first-step efficiency. In both extracts the reaction plateaus after 50 minutes. The efficiency of D1853C is lower than the wild-type, but no significant difference was observed for the kinetics of the first step of splicing. Data from two individual assays were used to plot the curve.

2-2.5. The exon ligation step in yeast is ATP-independent but concentration-dependent

The bimolecular exon-ligation assay has only been carried out in human cell extract previously. We thus decided to do more studies with this system in yeast extracts. First we asked whether Prp16 and Prp22 are involved in the catalysis of the second step of splicing. Moore and co-workers have shown that ATP is required for the exon-ligation in human splicing extracts and they proposed that ongoing ATP hydrolysis by Prp16 or Prp22 is required to maintain the spliceosome in a conformational state capable of binding the AG at the 3'SS (Anderson and Moore, 2000). We tried the two-step splicing assay with inactive Prp16 and Prp22 mutants. As expected, when the Prp16 mutant was added before the first step occurred, no exon-ligation was observed due to the absence of the transition from the first step to the second step. Lower exon-ligation efficiency was obtained when Prp22 mutant was added before the first step due to the inability of the spliceosome to release the mature mRNA. Surprisingly, when the mutants were added to chase the wild-type helicases after the first-step occurred, no difference was observed for the second step (Figure 2-9). This means Prp16 and Prp22 are not involved in the processing or the maintenance of the second step.

Next, we tested whether ATP was required for the exon-ligation in yeast. We depleted free ATP by gel filtration over Zeba™ desalt spin columns, which resulted in $\geq 95\%$ retention of salts and other small molecules smaller than 1 kD.

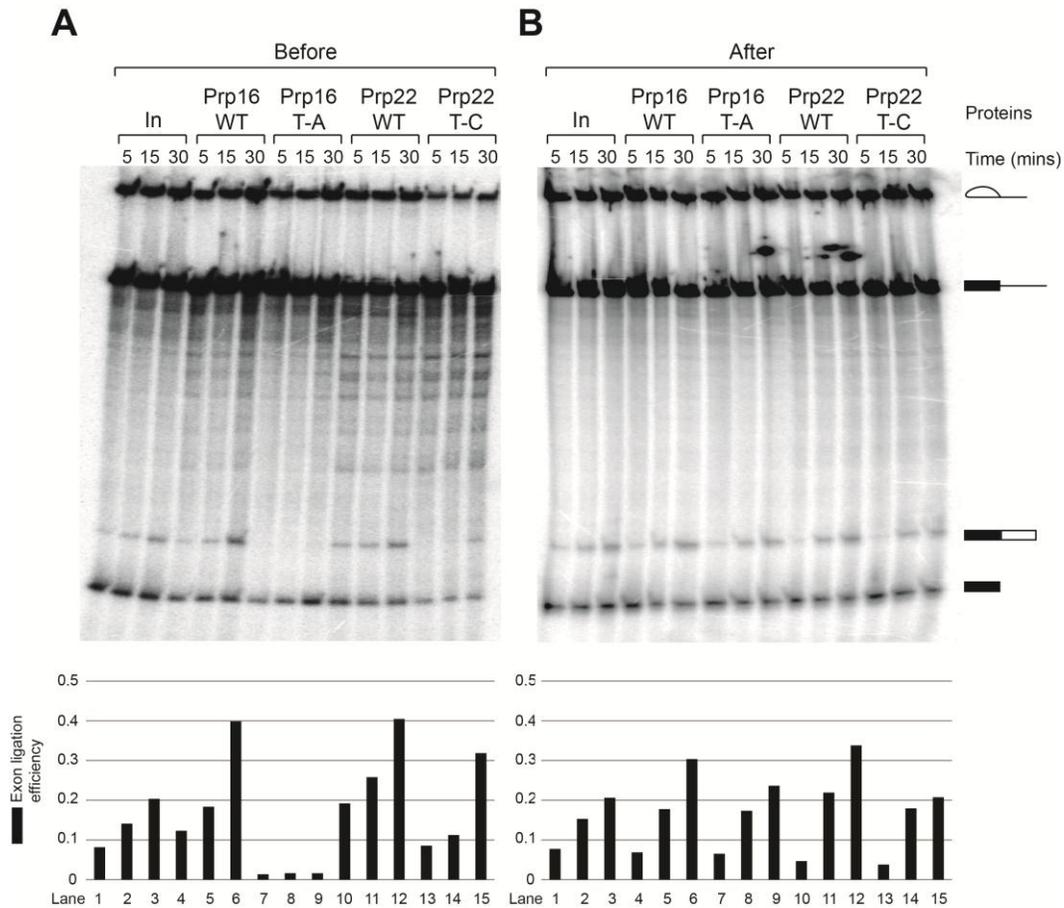


Figure 2-9. The exon ligation step does not depend on Prp16 or Prp22. A) (*upper*) Exon ligation assay with wild-type extract. The wild-type Prp16 protein (lanes 4-6), the Prp16 T-A mutant (lanes 7-9), the wild-type Prp22 protein (lanes 10-12) and the Prp22 T-C mutant (lanes 13-15) were added before the first-step substrate. (*bottom*) Quantification of the exon ligation efficiency. B) The same assay as in A) except that the proteins were added along with the second-step substrate after the extract was incubated with the first-step substrate for 30 minutes.

Although the ATP-depletion resulted in ~50% lower efficiency, it did not completely inhibit the exon-ligation activity (Figure 2-10 A). This is consistent with previous report that ATP was not required during the second cleavage-and-ligation reaction in yeast (Horowitz and Abelson, 1993).

We also tried a titration of the 3' RNA. Higher exon-ligation efficiency was observed with increasing RNA concentration (Figure 2-10 B). Thus, the only

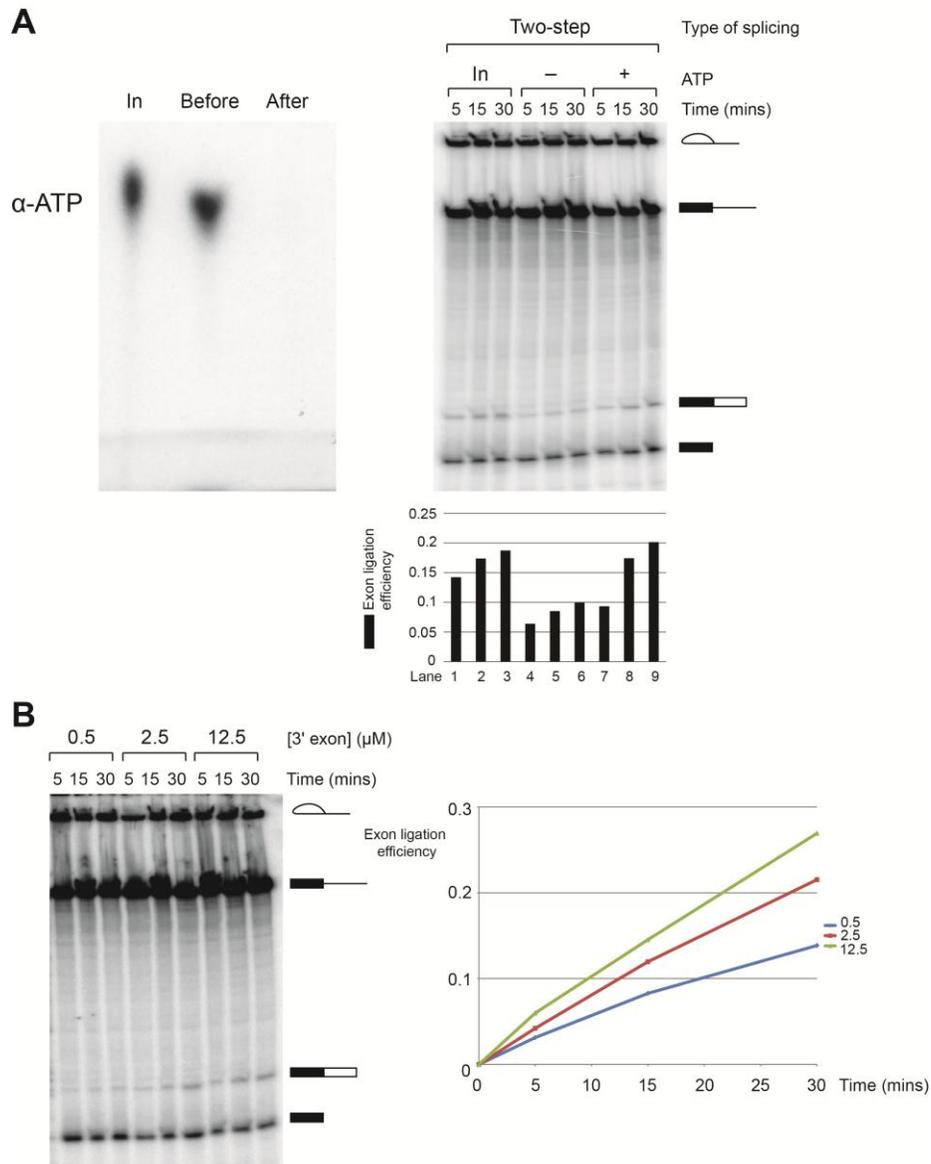


Figure 2-10. The exon ligation step is not ATP-dependent but concentration-dependent. A) (*left*) TLC before and after ATP depletion. No free α -ATP was detected after passing through the desalt column. (*right*) Exon ligation after ATP depletion. The second step was mildly reduced when ATP was depleted (lanes 4-6) and returned to normal in the addition of ATP (lanes 7-9). B) Exon ligation with titration of the second-step substrate. With a series of five-fold titration (0.5 mM, 2.5 mM and 12.5 mM), it was indicated the exon ligation efficiency depends on the binding rate of the substrate.

factor determining efficiency of the second step in yeast observed by us is the concentration or the binding rate of the substrate. Once the first step occurs and

the ATP-dependent conformational change is finished, the spliceosome is now committed for the second step and ready to carry out the exon-ligation reaction. We concluded that the yeast splicing is more straightforward than that in human and the protein factors required for each step are less complicated.

2-3. Discussion

In this study, using X-ray crystallography we have observed two distinct conformations of the Prp8 RNase H domain. One of them - monomer b - permits coordination of Mg^{2+} at the conserved RNase H metal-binding site. The results of mutagenesis of the inner-sphere D1853 indicated that monomer b is involved in the second step or exon ligation of splicing.

We tested a series of mutations of the conserved residues in the RNase H motif and only observed phenotypes with D1853 mutants. At almost the same time when we started our research, Staley and coworkers observed exactly the same result with the D1853C mutant (Sebastian and Staley, unpublished data). Since D1853 is the only residue to make direct inner sphere contact with Mg^{2+} , this result is consistent with a metal binding activity of Prp8 in the spliceosome. An intriguing fact is when Asp1781 is mutated to His, no phenotype was observed. We crystalized the human D1781H mutant. After soaking in Mg^{2+} or Mn^{2+} no metal ion was found in the active site (data not shown). The N3 on the imidazole ring interacts with the side chain of Q1894, which might block entry of the metal

ions. When Q1894 is mutated to Ser, however, we were still unable to detect any metal in the active site. Histidines have been shown to coordinate all the divalent metal ions in known protein structures, although it prefers soft metals like Mn^{2+} (reviewed in: Harding, 2004). Thus we cannot rule out the possibility that D1853H mutant is still able to coordinate metals within the context of the spliceosome.

Of all the D1853 mutants, D1853C and D1853S mutations result in the most severe impairment. This is consistent with our expectation since the softer cysteine ligand strongly disfavours binding a hard metal ion and this mutation should abrogate Mg^{2+} coordination. The mild defect in the first step splicing with D1853C can be explained by the slow growth of the cells at 30 °C. In fact, when the snRNA concentration was checked by RT-PCR, a notable difference was obtained between wild-type and D1853C extract (data not shown). However, considering the protein structure, complex formation, the first step product, the non-snRNP protein functions and the transition from first step to second step are all normal in the D1853C mutant, it is unlikely that the severe defect in the second step is due to the decrease in snRNA concentration.

Our attempted rescue of the second step in the presence of the thiophilic divalent manganese ion was unsuccessful. However, it has been shown that recovery of enzymatic activity where Mg^{2+} protein ligands (aspartate or glutamate) have been mutated to cysteine is not always possible (Gao et al., 2004). In addition, U6 snRNA has been shown to contribute to RNA splicing through

metal-ion coordination, as Mn^{2+} rescues the first transesterification step of splicing in phosphorothioate-modified U6 snRNA where the substitution disrupts the critical U80 Mg^{2+} binding site (Yean et al., 2000). Given that Mn^{2+} is more thiophilic than Mg^{2+} , this could be an indication that during the second step of splicing, position 1853 in Prp8 is involved in formation of the metal binding site to which the residues on snRNAs may also contribute and a combination of mutations may be required for the metal switch assay. Since no direct metal rescue has been observed by us, more evidence needs to be obtained to solve this problem in the future.

Taken together, the studies described here provide functional evidence for involvement of the D1853 residue in the Prp8 RNaseH-like active site during the catalytic stage of splicing. In the absence of a crystal structure of an active spliceosome, the approach delineated here constitutes a powerful method for investigating the configuration of the spliceosomal active site.

2-4. Materials and Methods

2-4.1. Protein expression, purification, and crystallization

A cDNA representing an N-terminal extension of hPrp8 1831-1990 encoding hPrp8 amino acids 1769-1990 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 Prp8 fragment. All mutants

were generated using the overlapping PCR method. The resulting MBP•Prp8 fusion proteins were expressed in *E. coli* and purified by sequential amylose resin, size exclusion chromatography, TEV cleavage and anion exchange.

Crystals of hPrp8 aa1769-1990 were grown at 23 °C using the hanging drop vapor diffusion technique. Crystals of native protein were grown by mixing one μL of 10 mg/mL protein solution (10 mM Tris, pH 8.0, 0.1 mM EDTA, 5 mM DTT, 0.02% NaN_3) with one μL of precipitant (2.5 M NaCl, 100 mM Tris, pH 7.0, 100 mM MgCl_2). Crystals were transferred to precipitant containing 20% glycerol and frozen in liquid nitrogen for data collection (Ritchie et al., 2008).

For the new crystallization conditions, crystals were grown using the same procedure but with a different precipitant solution (10-14% PEG 4000, 100 mM Tris pH 8.0, and 300 mM MgCl_2), and frozen in the same precipitant with 16% glycerol for data collection. For soaks with other metals, the crystals were soaked in the same precipitant, with the MgCl_2 replaced with 100 mM of the indicated metal chloride salt (except for ZnCl_2 , which is 10 mM) for one hour prior to freezing in the same solution with 16% added glycerol.

2-4.2. 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).

2-4.3. 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).

2-4.4. Construction of RNA substrates

The ACT1 pre-mRNA was made by *in vitro* transcription with [α -³²P]-ATP. The 25 μ L reaction was incubated at 37 °C for 4 h, with the final concentrations of the various components as follows: 40 mM Tris pH 8.0, 6 mM MgCl₂, 10 mM NaCl, 2 mM spermidine, 0.5 mM CTP, 0.5 mM GTP, 0.5 mM UTP, 60 μ M ATP, 1.32 μ M [α -³²P]-ATP (PerkinElmer), 5 ng DNA template, 10 mM DTT, and 1 μ L

T7 RNA polymerase. The reaction was then separated on a 7%, 19:1 Acrylamide: Bis-acrylamide, 8 M Urea PAGE.

2-4.5. Creation of mutant *S. cerevisiae* strains and splicing extracts

Mutant PRP8 genes were created using gap repair of plasmid pJU186 (Umen and Guthrie, 1995), which contains a HIS selectable marker and PRP8 gene. Mutant plasmids were transformed into strain JDY8.06 (*ura3-52, leu2-3,-112, ade2, his3-A1, trpl-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). After transformation with the pJU186 plasmid, 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 (Boeke et al, 1987). 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 to contain desired mutation by sequencing. Splicing extract was prepared from wildtype and mutant yeasts as described (Umen and Guthrie, 1995).

For the spot test analysis, yeast strains were inoculated in YPD media (1% yeast extract, 2% peptone, 2% dextrose) and were grown over-night at 30 °C with

shaking. The following day, we checked the concentration of each culture and performed serial dilutions in order to reach 1×10^6 , 2×10^5 , 4×10^4 , 8×10^3 , 1.6×10^3 and 3.2×10^2 cell/ml. From each one of these dilutions we spotted 10 μ L drop on YPD plates. Plates were photographed after 2 days at 30 °C, 3 days at 37 °C and 4 days at 16 °C.

2-4.6. Splicing assays

Splicing reactions (10 μ L) were performed as described (Lin et al., 1985), with the final concentrations of the various components as follows: 2 mM ATP, 2.5 mM MgCl₂, 3% (w/v) PEG 8000, 60 mM potassium phosphate pH 7.0, 20 mM KCl, 8 mM HEPES, 8% (v/v) glycerol, 80 μ M EDTA, 0.2 mM DTT, and 1 nM actin pre-mRNA. Yeast proteins were present at about 10-12 mg/ml. The reaction was stopped by the addition of 10 μ L of stop solution (1 mg/mL proteinase K, 50 mM EDTA, 1% SDS) and the mixture was incubated at 65 °C for 15 min. Each sample was extracted once with phenol/chloroform/isoamyl alcohol and once with chloroform, and was ethanol precipitated. The pellet was resuspended in loading dye 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).

The quantification was performed using ImageQuant (Molecular Dynamics). The volume of each band was adjusted by subtracting the background noise and

normalized by the number of adenosines in its sequence. The first- and second-step efficiencies were then calculated as shown in Equation 1 and 2.

$$\text{1st step efficiency} = \frac{[\text{ariat intermediate}] + [\text{mRNA}]}{[\text{ariat intermediate}] + [\text{mRNA}] + [\text{pre-mRNA}]} \quad (\text{Eq. 1})$$

$$\text{2nd step efficiency} = \frac{[\text{mRNA}]}{[\text{ariat intermediate}] + [\text{mRNA}] + [\text{pre-mRNA}]} \quad (\text{Eq. 2})$$

2-4.7. Native gel analysis

Aliquots of standard splicing reactions containing ³²P-labeled mRNA were taken at the indicated times and adjusted to 0.7 mg/mL heparin and 12% glycerol with a trace of bromophenol blue. To separate individual spliceosomal complexes, samples were analyzed on non-denaturing 4% polyacrylamide gels (80:1 acrylamide/bis-acrylamide) at 160 V for 3 hours in TGM buffer (50 mM Tris base, 50 mM glycine, 2 mM MgCl₂).

2-4.8. Exon-ligation assay and chase experiment

For the chase experiment, labelled ACT1 pre-mRNA was incubated in 10 μL splicing reactions with different extracts. Excessive unlabelled ACT1 pre-mRNA (1 μM) was incubated in 10 μL splicing reactions with wild-type simultaneously. After 10 min, 1 μL of the unlabelled reaction was added into 10 μL labelled reaction and then continued for 15, 30 and 60 min.

For the exon ligation assay, labelled 5' portion of the ACT1 pre-mRNA was pre-incubated for 30 min in 10 μ L splicing reactions. Exon ligation was initiated by addition of the 3' portion of the ACT1 RNA substrate to splicing reactions and incubation was then continued for 10, 20, 30 and 60 min. For the Prp16 and Prp22 chase experiment, 1 μ g purified mutant was added into the 10 μ L reaction either with the 5' RNA or 3' RNA substrates.

The ATP depletion assay was performed by buffer exchange using 1 mL bed volume Zeba™ desalt spin columns (Thermo scientific). Columns were pre-equilibrated three times with splicing buffer without ATP. Splicing reactions (100 μ L per column) that had been pre-incubated with 5' substrate for 30 min were applied to washed columns and centrifuged for 1 min at 1,500 \times g. To determine the extent to which reactions had been depleted of free nucleotides, splicing reactions were supplemented with trace amounts of [α -³²P] ATP prior to loading on the column. A fluorescent PEI-cellulose chromatography plate was spotted with 1 μ L of the spiked reaction before and after ATP depletion, as well as with [α -³²P] ATP as the control.

2-5. References

Adams, P.L., Stahley, M.R., Gill, M.L., Kosek, A.B., Wang, J., and Strobel, S.A. (2004). Crystal structure of a group I intron splicing intermediate. *RNA* **10**, 1867-1887.

Allingham, J.S., Pribil, P.A., and Haniford, D.B. (1999). All three residues of the Tn 10 transposase DDE catalytic triad function in divalent metal ion binding. *J.*

Mol. Biol. **289**, 1195-1206.

Anderson, K., and Moore, M.J. (1997). Bimolecular exon ligation by the human spliceosome. *Science* **276**, 1712-1716.

Anderson, K., and Moore, M.J. (2000). Bimolecular exon ligation by the human spliceosome bypasses early 3' splice site AG recognition and requires NTP hydrolysis. *RNA* **6**, 16-25.

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

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

Cheng, S.C. (1994). Formation of the yeast splicing complex A1 and association of the splicing factor PRP19 with the pre-mRNA are independent of the 3' region of the intron. *Nucleic Acids Res.* **22**, 1548-1554.

Cheng, S.C., and Abelson, J. (1987). Spliceosome assembly in yeast. *Genes Dev.* **1**, 1014-1027.

Chua, K., and Reed, R. (1999). Human step II splicing factor hSlu7 functions in restructuring the spliceosome between the catalytic steps of splicing. *Genes Dev.* **13**, 841-850.

Davies, D.R., Goryshin, I.Y., Reznikoff, W.S., and Rayment, I. (2000). Three-dimensional structure of the Tn5 synaptic complex transposition intermediate. *Science* **289**, 77-85.

Dupureur, C.M. (2008). Roles of metal ions in nucleases. *Curr. Opin. Chem. Biol.* **12**, 250-255.

Dupureur, C.M. (2010). One is enough: insights into the two-metal ion nuclease mechanism from global analysis and computational studies. *Metallomics* **2**, 609.

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

Galburt, E.A., Chevalier, B., Tang, W., Jurica, M.S., Flick, K.E., Monnat, R.J., Jr., and Stoddard, B.L. (1999). A novel endonuclease mechanism directly visualized for I-PpoI. *Nat. Struct. Biol.* **6**, 1096-1099.

Gao, K., Wong, S., and 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.

Gordon, P.M., Sontheimer, E.J., and Piccirilli, J.A. (2000). Metal ion catalysis during the exon-ligation step of nuclear pre-mRNA splicing: extending the parallels between the spliceosome and group II introns. *RNA* **6**, 199-205.

Harding, M.M. (2004). The architecture of metal coordination groups in proteins. *Acta. Crystallogr. D. Biol. Crystallogr.* **60**, 849-859.

Holyoak, T., Sullivan, S.M., and Nowak, T. (2006). Structural insights into the mechanism of PEPCK catalysis. *Biochemistry* **45**, 8254-8263.

Horowitz, D.S., and Abelson, J. (1993). Stages in the second reaction of pre-mRNA splicing: the final step is ATP independent. *Genes Dev.* **7**, 320-329.

Horton, J.R., Blumenthal, R.M., and Cheng, X. (2004). Restriction endonucleases: structure of the conserved catalytic core and the role of metal ions in DNA cleavage. In: *Restriction Endonucleases* (A. Pingoud, ed.). Berlin: Springer-Verlag, **14**:361-392.

Hotz, H.R., and Schwer, B. (1998). Mutational analysis of the yeast DEAH-box splicing factor Prp16. *Genetics* **149**, 807-815.

Katayanagi, K., Okumura, M., and Morikawa, K. (1993). Crystal structure of Escherichia coli RNase HI in complex with Mg²⁺ at 2.8 Å resolution: proof for a single Mg(2+)-binding site. *Proteins* **17**, 337-346.

Koodathingal, P., Novak, T., Piccirilli, J.A., and 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.

Kostrewa, D., and Winkler, F.K. (1995). Mg²⁺ binding to the active site of EcoRV endonuclease: a crystallographic study of complexes with substrate and product DNA at 2 Å resolution. *Biochemistry* **34**, 683-696.

Lin, R.J., Newman, A.J., Cheng, S.C., and Abelson, J. (1985). Yeast mRNA splicing in vitro. *J. Biol. Chem.* **260**, 14780-14792.

Lipchock, S.V., and Strobel, S.A. (2008). A relaxed active site after exon ligation by the group I intron. *Proc. Natl. Acad. Sci. U.S.A.* **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.

- Lukacs, C.M., Kucera, R., Schildkraut, I., and Aggarwal, A.K. (2000). Understanding the immutability of restriction enzymes: crystal structure of BglII and its DNA substrate at 1.5 Å resolution. *Nat. Struct. Biol.* **7**, 134-140.
- 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.
- Mohamadou, A., and Gérard, C. (2001). Synthesis and characterisation of zinc(ii) complexes of tripodal N7 ligands involving pyridine and amine or amide nitrogen donors. Crystal structure of a zinc(ii) complex. *J. Chem. Soc. Dalton Trans.* **2001**, 3320-3328.
- Mones, L., Kulhanek, P., Florian, J., Simon, I., and Fuxreiter, M. (2007). Probing the two-metal ion mechanism in the restriction endonuclease BamHI. *Biochemistry* **46**, 14514-14523.
- Mones, L., Simon, I., and Fuxreiter, M. (2007). Metal-binding sites at the active site of restriction endonuclease BamHI can conform to a one-ion mechanism. *Biol. Chem.* **388**, 73-78.
- 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.
- 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.
- Pingoud, V., Wende, W., Friedhoff, P., Reuter, M., Alves, J., Jeltsch, A., Mones, L., Fuxreiter, M., and Pingoud, A. (2009). On the Divalent Metal Ion Dependence of DNA Cleavage by Restriction Endonucleases of the EcoRI Family. *J. Mol. Biol.* **393**, 140-160.
- Potterton, E., Briggs, P., Turkenburg, M., and Dodson, E. (2003). A graphical user interface to the CCP4 program suite. *Acta Crystallogr. D. Biol. Crystallogr.* **59**, 1131-1137.
- Query, C.C., Strobel, S.A., and Sharp, P.A. (1996). Three recognition events at the branch-site adenine. *EMBO J.* **15**, 1392-1402.

Ritchie, D.B., Schellenberg, M.J., Gesner, E.M., Raithatha, S.A., Stuart, D.T., and MacMillan, A.M. (2008). Structural elucidation of a PRP8 core domain from the heart of the spliceosome. *Nat. Struct. Mol. Biol.* **15**, 1199-1205.

Ruby, S.W., and Abelson, J. (1991). Pre-mRNA splicing in yeast. *Trends Genet.* **7**, 79-85.

Schmidt, B.H., Burgin, A.B., Deweese, J.E., Osheroff, N., and Berger, J.M. (2010). A novel and unified two-metal mechanism for DNA cleavage by type II and IA topoisomerases. *Nature* **465**, 641-644.

Scott, W.G., Murray, J.B., Arnold, J.R., Stoddard, B.L., and Klug, A. (1996). Capturing the structure of a catalytic RNA intermediate: the hammerhead ribozyme. *Science* **274**, 2065-2069.

Sharp, P.A. (1991). "Five easy pieces". *Science* **254**, 663.

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.

Sontheimer, E.J., Sun, S., and Piccirilli, J.A. (1997). Metal ion catalysis during splicing of premessenger RNA. *Nature* **388**, 801-805.

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. (2003). 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.

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.

Wang, Y., Juranek, S., Li, H., Sheng, G., Tuschl, T., and Patel, D.J. (2008).

Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. *Nature* **456**, 921-926.

Xie, F., Qureshi, S.H., Papadakos, G.A., and Dupureur, C.M. (2008). One- and two-metal ion catalysis: global single-turnover kinetic analysis of the PvuII endonuclease mechanism. *Biochemistry* **47**, 12540-12550.

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

Chapter 3

The conformational change of Prp8 is coupled to the transition between the two steps of splicing

The work presented in this chapter represents collaboration between authors. Tao Wu, Matt Schellenberg, and Karim Atta prepared proteins for crystallization, grew the crystals, prepared mutant yeast strains, and performed splicing assays. Matt Schellenberg performed the X-ray data collection and solved the crystal structures.

3-1. Introduction

3-1.1. Transition between the first and the second step of splicing

Splicing catalysis proceeds through two consecutive transesterification reactions involving three sites of the intron. During the first step, the branch adenosine attacks the 5' splice site (5'SS), producing a lariat intermediate and cleaved 5' exon. During the second reaction, the 5' exon attacks the 3' splice site (3'SS), yielding spliced mRNA and lariat intron products (reviewed in: Konarska and Query, 2005). Considering that the substrates for the two chemical reactions are different, some rearrangements of either the substrates or the spliceosome at the catalytic center are required. If the spliceosome uses a single active site for both catalytic steps, the lariat intermediate formed in the first step would need to be displaced to allow positioning of the 3'SS for the second step (Steiz and Steiz, 1993). The stereochemistry of splicing provides evidence for two active sites in the spliceosome, with the first step similar to that of the group II introns and the second step similar to that of both the group I and II introns (Moore and Sharp, 1993). In this case, the spliceosome would exist in two distinct conformations during catalysis and bind the substrates differently for the two steps.

Based on the suppression of mutations of the RNA substrates by PRP8 alleles, Query and Konarska proposed a model of rearrangements within the spliceosome at the time of transition between the first and second step of splicing (Query and Konarska, 2004; also see Figure 3-1). Specifically they suggest that there are two different active site conformations in competition with each other

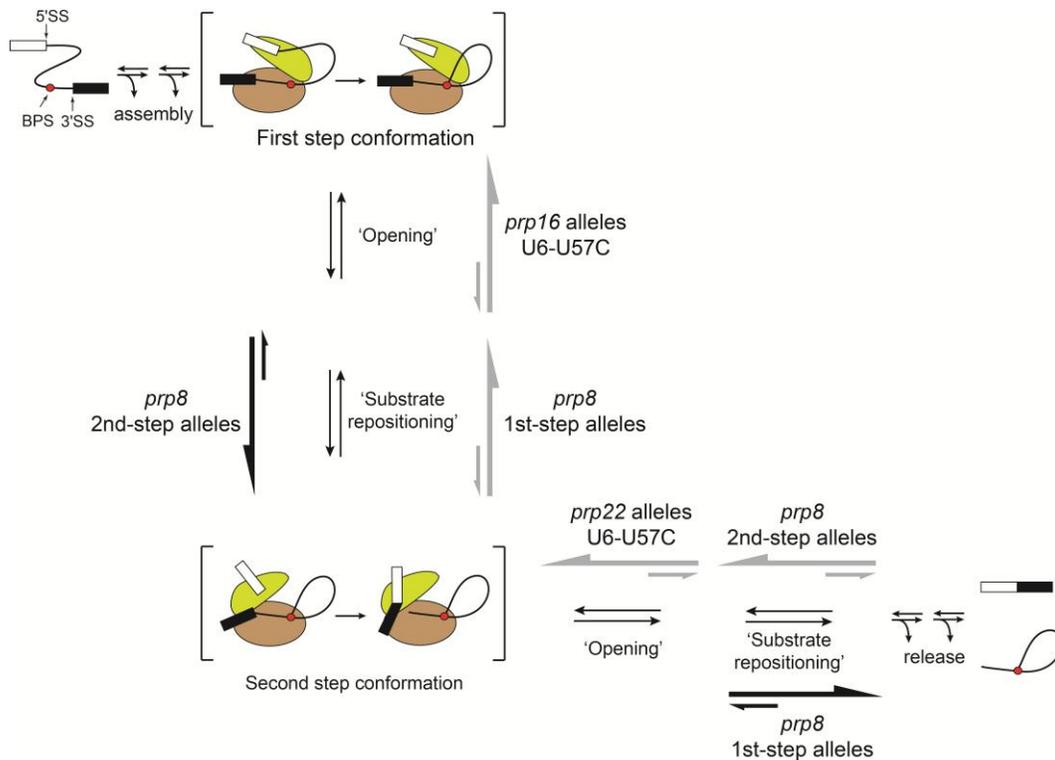


Figure 3-1. Schematic representation of splicing pathway and assignment of mutant alleles of prp8, prp16, prp22 and U6 snRNA that modulate spliceosomal transitions. The first and second catalytic steps require different conformations of the spliceosome shown by green and brown ovals. The equilibrium between these conformations is modulated by interactions of Prp16, Prp22, Prp8 and U6 snRNA. As indicated by the grey and black arrows, certain alleles of these factors improve the first step and inhibit the second, and other alleles act oppositely (Adapted from Konarska et al., 2005; Liu et al., 2007).

and that the equilibrium between these conformations is modulated by a number of factors. Conformations that favor the first step inhibit the second step and vice versa. Recently, experiments from the Cheng lab showed that changes in spliceosome active site conformation between the two steps of splicing are thermodynamically close in free energy and may not be complex (Tseng and Cheng, 2008). By adjusting cation concentrations of the buffer, they were able to reverse both chemical steps of splicing within isolated spliceosomes which had undergone both of the chemical steps of splicing but were stalled before mRNA

release. The splicing equilibrium was also sensitive to divalent metal ions (Mg^{2+} and Mn^{2+}), with subtle changes in the ion concentration altering the direction of the forward and reverse reactions. Therefore the transition between different steps of splicing does not seem to be mechanically complex, but rather depends on only a few key factors.

Among these factors, U6 snRNA, Prp16 and Prp8 are all thought to be involved in direct interactions with the catalytic center, although the specific events in the transition between the two steps are still poorly understood. Konarska and co-workers have demonstrated the importance of repositioning of the reaction substrates during this transition. The 5'SS is positioned for the first step by pairing with the invariant U6 snRNA-ACAGAG site. The authors demonstrated that this pairing interaction must be disrupted to allow transition to the second step. 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 confirmed 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). In addition, mutations of U57 of U6 snRNA in helix 1a have been shown to have opposing effects on splicing. U57A inhibits the first step and improves the second step, whereas U57C improves the first and inhibits the second step of BS mutants (McPheeters, 1996; Figure 3-1).

Prp16 is a DEAH-box ATPase and is thought to facilitate the transition between the first and second steps of splicing. Prp16 has been shown to unwind RNA duplexes in vitro (Wang et al., 1998). It has also been shown that the Prp16-induced conformational change protects the 3'SS from RNase H cleavage (Schwer and Guthrie, 1992), but it still remains unclear whether this helicase activity is directly related to its function in the transition between different steps. Mutant Prp16 alleles are defective for the second step of splicing but are able to suppress the A-to-C mutation at the branch site (BSC), suggesting these alleles stabilize the first step conformation (Burgess and Guthrie, 1993; Figure 3-1). Another DEAH-box ATPase, Prp22, is required for the spliceosome disassembly after the splicing reaction. Therefore, PRP22 alleles with reduced ATPase activity favor the second-step conformation and inhibit mRNA release (Schwer and Gross, 1998; Wagner et al., 1998; Figure 3-1).

3-1.2. PRP8 alleles affecting the first and second step of splicing

In addition to the factors introduced above, Prp8 has been placed in a central position with respect to spliceosome activation. A large number of mutations in PRP8 were identified that suppress the U4-cs1 cold-sensitive mutant as well as the mutations in the DExD/H helicases Prp28 and Brr2, which are required for the unwinding of the U1 snRNA/5'SS and U4/U6 RNA respectively (Kuhn et al., 1999; de la Cruz et al., 1999; Kuhn and Brow, 2000; Kuhn et al., 2002; Figure 3-2

A; Table 3-1). This suggested that Prp8 plays a role in spliceosome activation. PRP8 alleles also suppress a large number of intron mutations at the branch site and both 5' and 3' splice sites, indicating that individual amino acids or regions of Prp8 may function at the catalytic core of the spliceosome (Umen and Guthrie, 1996; Siatecka et al., 1999; Kuhn and Brow, 2000; Query and Konarska, 2004; reviewed in Grainger and Beggs and references therein; Figure 3-2 A; Table 3-1).

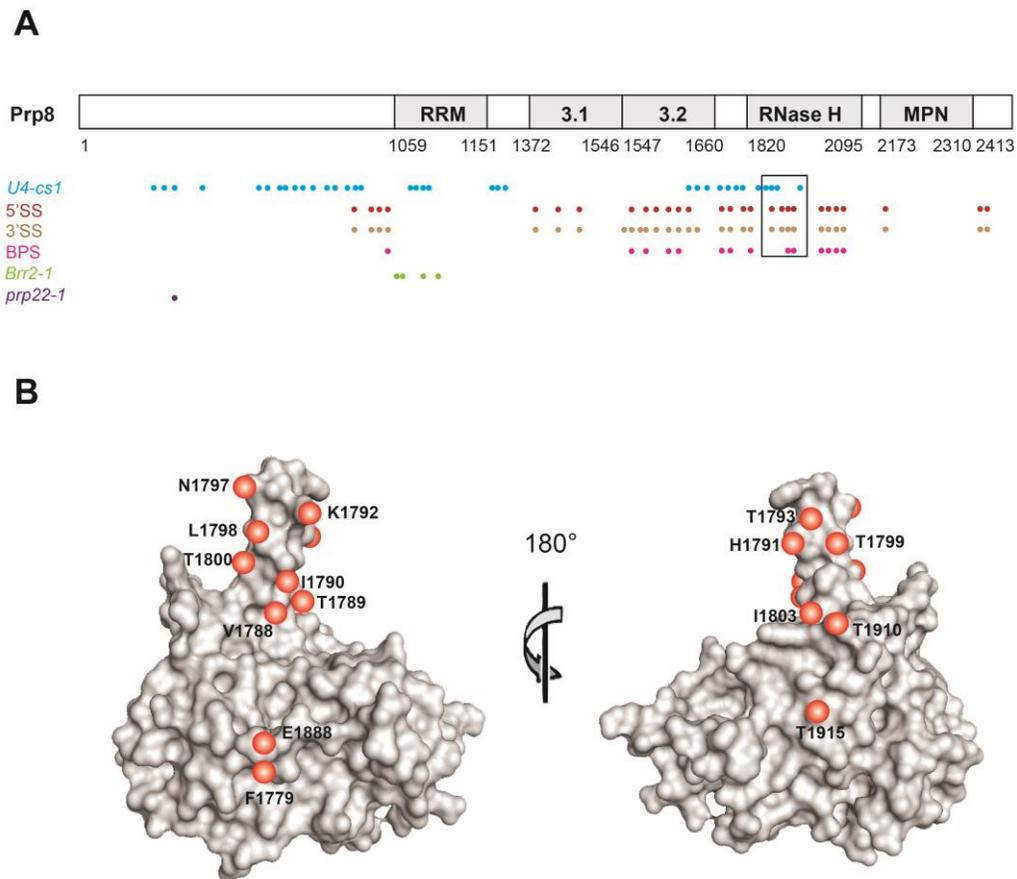


Figure 3-2. Diagram of Prp8 mapping the known yeast alleles. A) The numbers along the length of Prp8 represent the yeast amino acid number. The mutations suppressing U4-cs1 (blue), 5'SS (red), 3'SS (brown), BPS (pink), Br2-1 (green) and Prp22-1 (purple) are mapped with color circles. The region corresponding to the β -hairpin of the RNase H domain is shown in box (Adapted from Grainger and Beggs, 2005). The proposed RRM, domain 3.1 and 3.2, RNase H domain and MPN domain are shown for reference. B) Mapping of yeast mutant alleles in the context of the structure of human Prp8 RNase H domain. Shown are the positions of amino acids corresponding to mutant alleles (Adapted from Ritchie et al., 2008).

Table 3-1. Summary of the Prp8 alleles mapping to the RNase H region

Yeast alleles (Human number)	Supression of mutations in					Allele types		Preferred monomer structure	
	5'SS	3'SS	BPS	PPT	U4- cs1	First step	Second step	a	b
F1834S/L (Y1762)				×					
F1851L (F1779)					×				
V1860D (V1788)					×	×√		√	
V1860N (V1788)					×				
T1861P (T1789)	×	×			×		√		√
V1862D (I1790)					×				
V1862Y (I1790)					×		√		√
H1863E (H1791)		×	×				×√		
K1864E (K1792)	×	×							
T1865K (T1793)						×√			
N1869D (N1797)	×	×	×				×√		√
V1870N (L1798)	×	×	×				×√		√
A1871E (T1799)						×√			
T1872E (T1800)						×√		√	
I1875T (I1803)					×				
E1960K (E1888)				×		×√			
T1982A (T1910)	×	×	×				×		
V1987A (T1915)	×	×	×				×		

Amino acid numbering is for the *S. cerevisiae* protein (human amino acid number shown in brackets). “×”= displays this phenotype in previous reports, “√” = confirmed or observed in this thesis. (Adapted from Grainger and Beggs, 2005; Liu et al., 2007; Yang et al., 2008)

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 steps respectively (Query and Konarska, 2004; Liu et al., 2007). Unlike Prp16 and Prp22, Prp8 mutations can shift the equilibrium in different directions. 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 (Figure 3-1). In general, these alleles do not affect splicing of a wild-type pre-mRNA substrate. But they have opposing effects on three other pre-mRNA substrates. The first is a pre-mRNA with the branch adenosine of the intron mutated to a cytosine (BSC). This mutation has been shown to cause a severe defect in both the first and second steps of splicing. It has been shown that “first-step” alleles of PRP8 will partially alleviate the defect in the first step of splicing, while the second-step alleles will make the splicing of BSC even worse (Burgess and Guthrie, 1993). Another pre-mRNA has the branch adenosine mutated to a guanosine (BSG). This mutation has a defect in the second step of splicing. The third substrate has an adenosine to uridine mutation at the 3'SS (UuG) and is also impaired for the second step. PRP8 second-step alleles attenuate the second-step defect and the first-step alleles aggravate it (reviewed in Liu et al., 2007). All these three substrates were used in the following studies described in this chapter.

The long list of PRP8 alleles able to suppress mutations at multiple positions in a pre-mRNA strongly supports the proposal of two active site

conformations and puts Prp8 in a central place of the spliceosome. The fact that a large proportion of these mutations are clustered in the β -hairpin in the RNase H domain, which adopts two different conformations in the two monomers in our crystal structures, is especially intriguing to us (Figure 3-2 B; Table 3-1). In the studies outlined below, we used structural and molecular methods to show that 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.

3-2. Results

3-2.1. Structures of first-step alleles

In order to obtain the structural evidence for previously identified and potential PRP8 alleles, we tried to create and crystallize several previously identified mutants. Six mutants were capable of forming crystals. X-ray diffraction data of them was obtained, and their structures were solved.

The PRP8 mutant V1860D was first identified to suppress the cold sensitivity caused by U4-cs1, and was later shown to be a first-step allele (Kuhn and Brow, 2000; Yang et al., 2008). We solved the 2.0 Å crystal structure of the human V1788D mutant corresponding to yeast V1860D. 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 that of D1788 is exposed to the solvent and would not make any extra intramolecular hydrogen bonds (Figure 3-3 A). This argues that the mutant allele

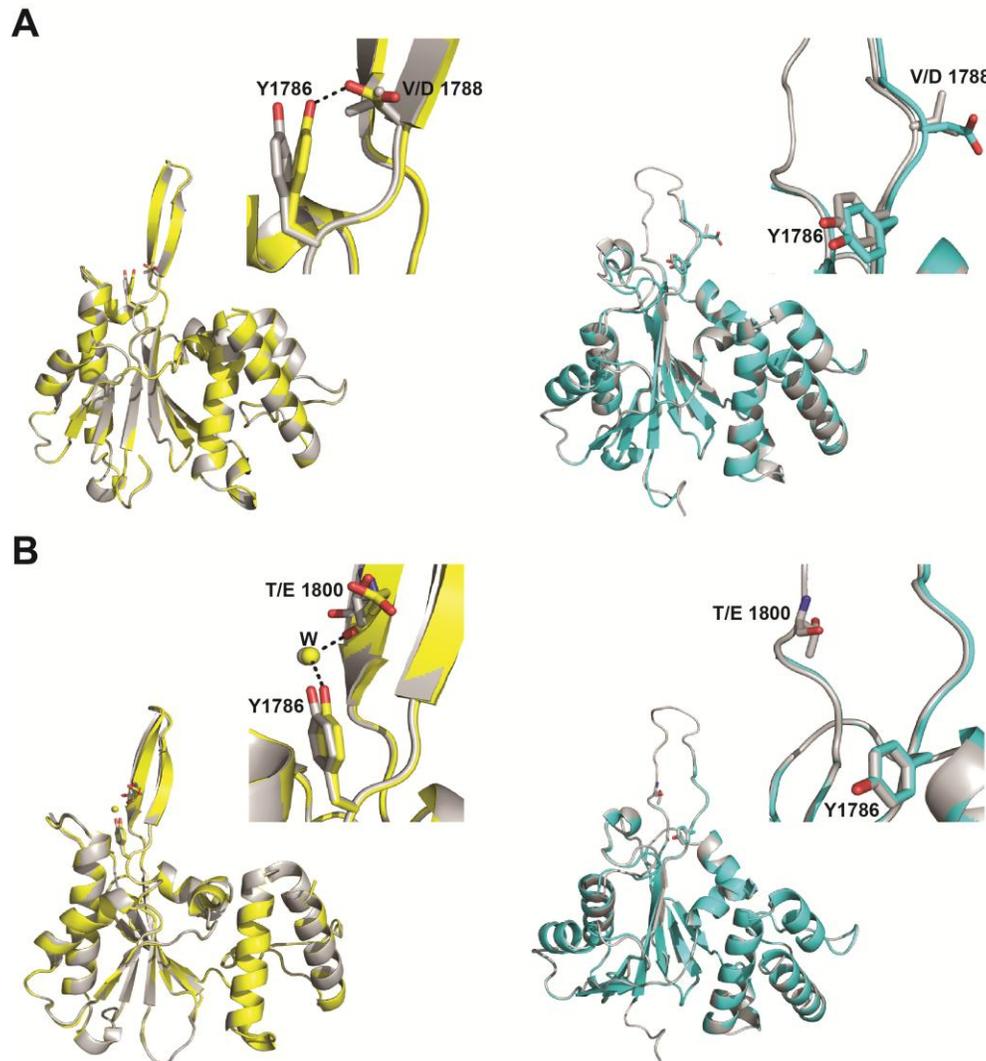


Figure 3-3. Structure of the PRP8 first-step alleles. A) Details of the 2 Å X-ray structure of V1788D hPrp8 RNase H domain corresponding to the yeast first-step allele V1860D. Shown are wild-type (grey) and mutant for monomers a (yellow) and b (cyan). The non-metal binding β -hairpin conformation of monomer a is stabilized by a hydrogen-bond formed between D1788 and Y1786. B) Details of the 1.5 Å X-ray structure of T1800E hPrp8 RNase H domain corresponding to the yeast first-step allele T1872E. The non-metal binding β -hairpin conformation of monomer a is stabilized by the extra water-mediated hydrogen bond of E1800 to Y1786.

favours the β -hairpin/non-metal binding conformation. The D1788 and Y1786 interaction effectively anchors the β -strand in position, stabilizing it relative to the more extended and translated loop structure.

We then solved the 1.5 Å crystal structure of the human T1800E mutant corresponding to yeast T1872E. This mutant has been shown to be a first-step allele (Yang et al., 2008). In this case, monomer a is stabilized by an additional hydrogen-bonding interaction between the side chains of E1800 and Y1786, mediated through a water molecule. The metal binding conformation of monomer b accommodates this mutation with no changes in structure, in which E1800 and Y1786 are no longer close to each other and no hydrogen bond can be formed (Figure 3-3 B). Therefore the V1860D and T1872E mutants use the same mechanism to favor the non-metal binding conformation over the other.

3-2.2. Structures of second-step alleles

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 *prp8-151* allele contains an 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 4.4 Å from one conformation of a flexible glutamate (E1795), suggesting there might be the possibility of charge-charge repulsion in this conformation for the mutant. In the N1797D structure, E1795 is more ordered and only favors the conformation which is further from N1797D (Figure 3-4 A). The combination of these two effects would be expected to stabilize the monomer b conformation relative to monomer a.

The PRP8 mutant V1862Y, which corresponds to I1790Y in human protein, was first identified to suppress the cold sensitivity caused by U4-cs1 (Kuhn and Brow, 2000). Examination of the wild-type structure indicates this mutation would form an additional hydrogen bond between the side chains of N1797 and Y1790 (human numbers) in monomer b. We solved the 1.75 Å structure of the I1790Y crystal and the prediction turns out to be true. In monomer b these two residues are 2.6 Å apart and the hydrogen bond between them makes the disordered hairpin region much more structured than in the wild-type protein structure (Figure 3-4 B). In monomer a, the side chains of N1797 and Y1790 face opposite directions in the β-hairpin region. No extra hydrogen bond was observed and the I1790Y mutation was accommodated. No difference was observed between the wild-type protein and the mutant. Therefore, the I-to-Y mutation

would be expected to stabilize the monomer b conformation relative to monomer a and is a potential second-step allele. Our later assays confirmed this hypothesis.

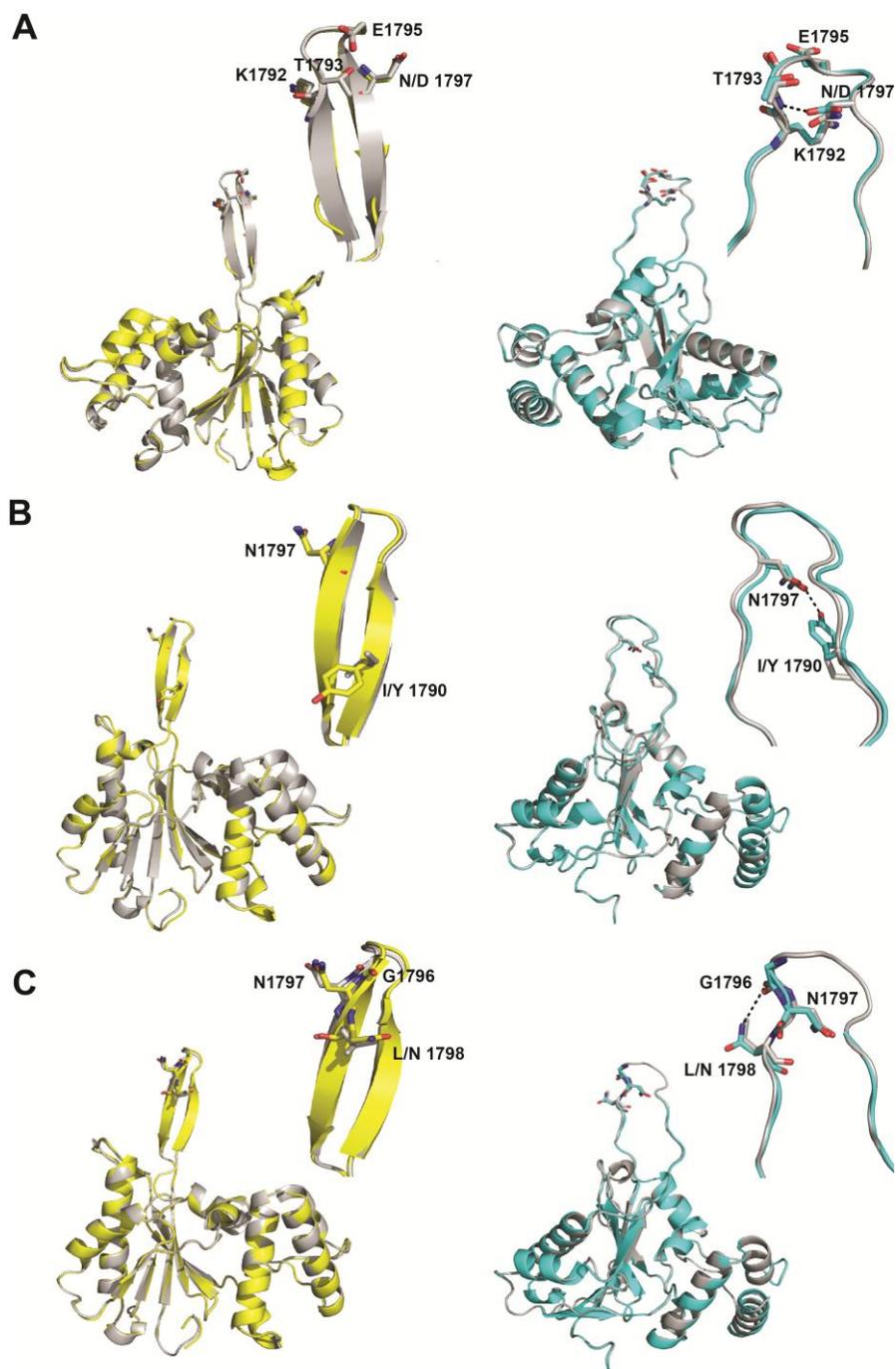


Figure 3-4. (Legend on next page)

The *prp8-162* allele contains a V1870N mutation, and displays a strong second-step phenotype (Query and Konarska, 2004; Liu et al., 2007). An analysis of the structure of the R1865A protein whose loop is visible in monomer b shows that the C δ of the corresponding amino acid, L1798, is 3.1 Å away from the peptide carbonyl of a nearby G1796 (Figure 3-4 C). This suggests that mutation of this amino acid to Asn (which has a similar overall geometry) would create an additional hydrogen bond in monomer b. In monomer a, the side chain 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 side chain, 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 3-4. Structure of the PRP8 second-step alleles. A) 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. (*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. B) Details of the 1.55 Å X-ray structure of the human I1790Y mutant corresponding to the yeast V1862Y second-step allele. (*left*) The non-metal binding β -hairpin conformation of monomer a accommodates the mutation with no changes in structure. (*right*) Y1790 forms an additional hydrogen bond to N1797 in monomer b. C) 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.

The *prp8-201* allele contains a T1861P mutation, which corresponds to T1789 in the human Prp8 structure. This mutant has been shown to suppress the cold-sensitive U4-cs1 and mutations at 5'SS and 3'SS (Kuhn et al., 1999; Collins and Guthrie, 1999). In order to examine this potential second-step allele (Guthrie et al., unpublished data), 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. Magnesium ion binding is unchanged by this mutation, with partial Mg²⁺ ion bound to D1781 in T1789P and full occupancy in T1789P/R1865A. 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 3-5 A). We performed a crystallographic mixing experiment where equal concentrations of R1865A and T1789P/R1865A double mutant protein were mixed in the 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 crystal lattice and improves resolution. 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 3-5 B). This demonstrates the pronounced preference of this proline for the non-ordered

hairpin conformation, implying that monomer b conformation corresponds to the second-step conformation of the spliceosome.

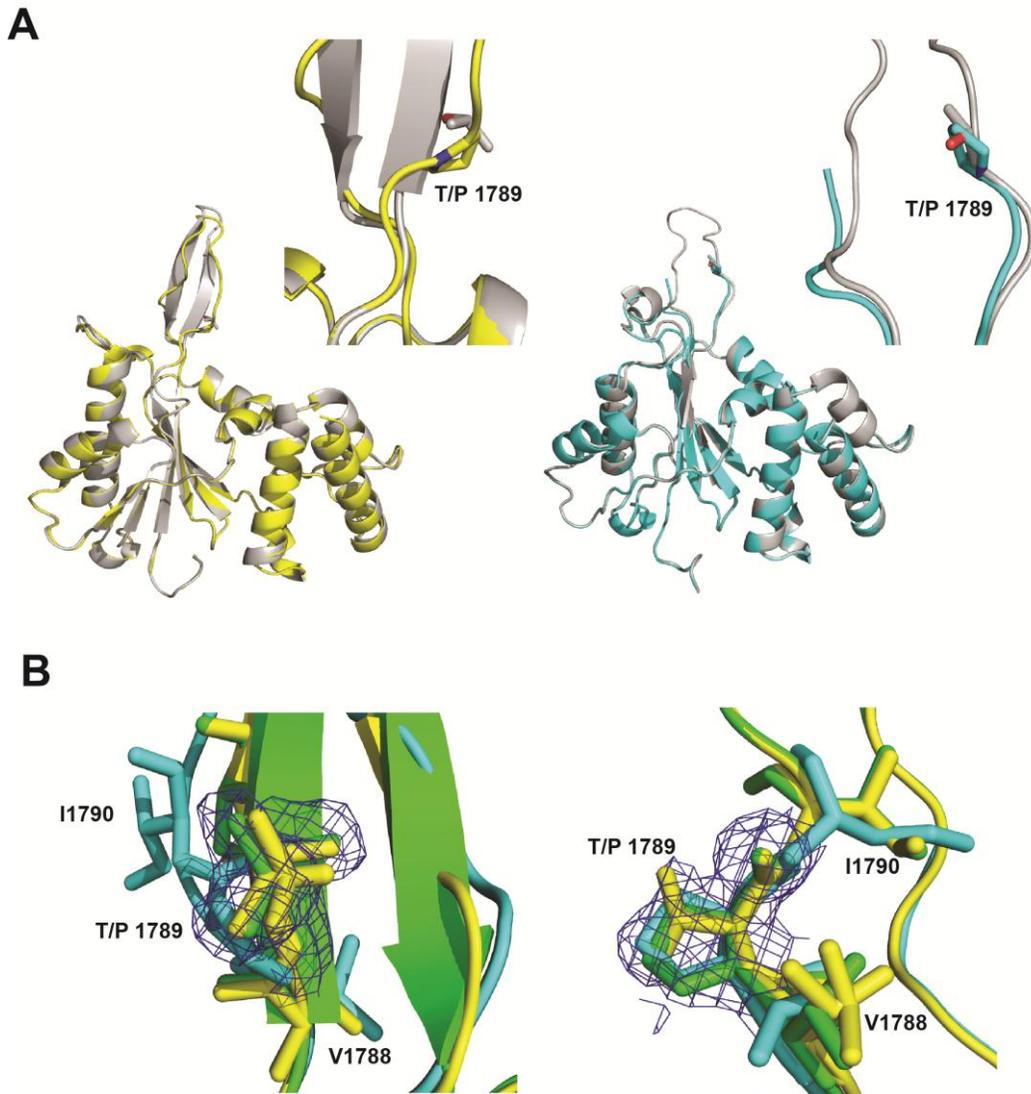


Figure 3-5. Structure of the second-step allele T1789P. A) Details of the 1.65 Å X-ray structure of the human T1789P mutant corresponding to the yeast T1861P second-step allele. (*left*) The non-metal binding β -hairpin conformation of monomer a is destabilized by the introduction of proline. (*right*) The T1789P mutation is accommodated in the disrupted β -hairpin of monomer b. B) 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. 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.

3-2.3. *In vivo* splicing of the first- and second-step alleles

Next we performed growth tests for all the first- and second-step alleles. As previously reported (Yang et al., 2008; Query and Konarska, 2004; Kuhn et al., 2000), T1872E, V1870N and V1862Y grew at a similar rate as the wild-type at both 16 °C and 37 °C. E1960K, V1860D and T1781P have all been reported to be ts (Umen and Guthrie, 1995; Kuhn et al., 2000). In our strain background, V1860D and T1861P were severely sick at 37 °C, while E1960K was both cs and ts (Figure 3-6).

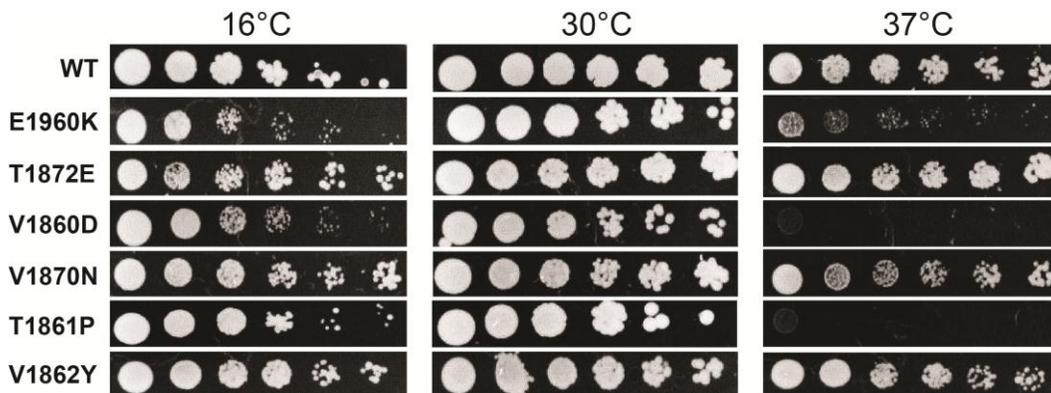


Figure 3-6. Growth phenotypes of strains containing PRP8 first- and second-step alleles. The first-step allele E1960K showed both cs and ts phenotype. The first-step allele V1860D and the second-step allele T1861P were both ts. Other alleles showed similar growth rate as the wild-type.

Two of the alleles used in the structural studies (T1861P and V1862Y) have not been verified to affect first- or second-step splicing before. Therefore we analyzed *in vivo* splicing of all these alleles with the ACT1-CUP1 reporter system. The same method has been shown to allow for quantitative monitoring of splicing efficiency by conferring copper resistance to yeast. The splicing can also be

analyzed in detail by primer extension (Lesser and Guthrie, 1993). To perform the copper-resistant growth test, an ACT1-CUP1 reporter containing the A-to-C mutation at the branch site (BSC) was co-transformed into the yJU75 strain along with the plasmid containing different PRP8 mutations. As discussed above, the A-to-C mutation causes a severe defect in both the first and second steps of splicing. For the wild-type PRP8, a partial growth inhibition was observed at 0.15 mM Cu²⁺ and a complete inhibition was gained at 0.25 mM Cu²⁺. As expected, all the first-step alleles aggravated the growth defect of the BSC mutation, with no growth observed at 0.15 mM Cu²⁺. The second-step alleles alleviated the growth defect to varying degrees. Although all of them conferred complete resistance at 0.15 mM Cu²⁺, the V1870N also conferred complete resistance even at 0.25 mM Cu²⁺; the T1861P conferred partial resistance at this concentration; and no resistance was obtained with V1862Y mutation at 0.25 mM Cu²⁺ (Figure 3-7 A). Thus the two new second-step alleles (T1861P and V1862Y) were confirmed and the strength of the three second-step alleles seems to be in the order of V1870N>T1861P>V1862Y.

We also tried primer extension with the total RNA isolated from the above strains, using a primer complementary to the 3'exon sequence in the reporter mRNA to reveal levels of pre-mRNA, mRNA, and lariat intermediate. In this experiment we used ACT1-CUP1 reporters containing the wild-type intron sequence, the A-to-G mutation at the branch site (BSG) as well as the BSC mutation. As discussed above, the BSG mutation is defective in the second step of

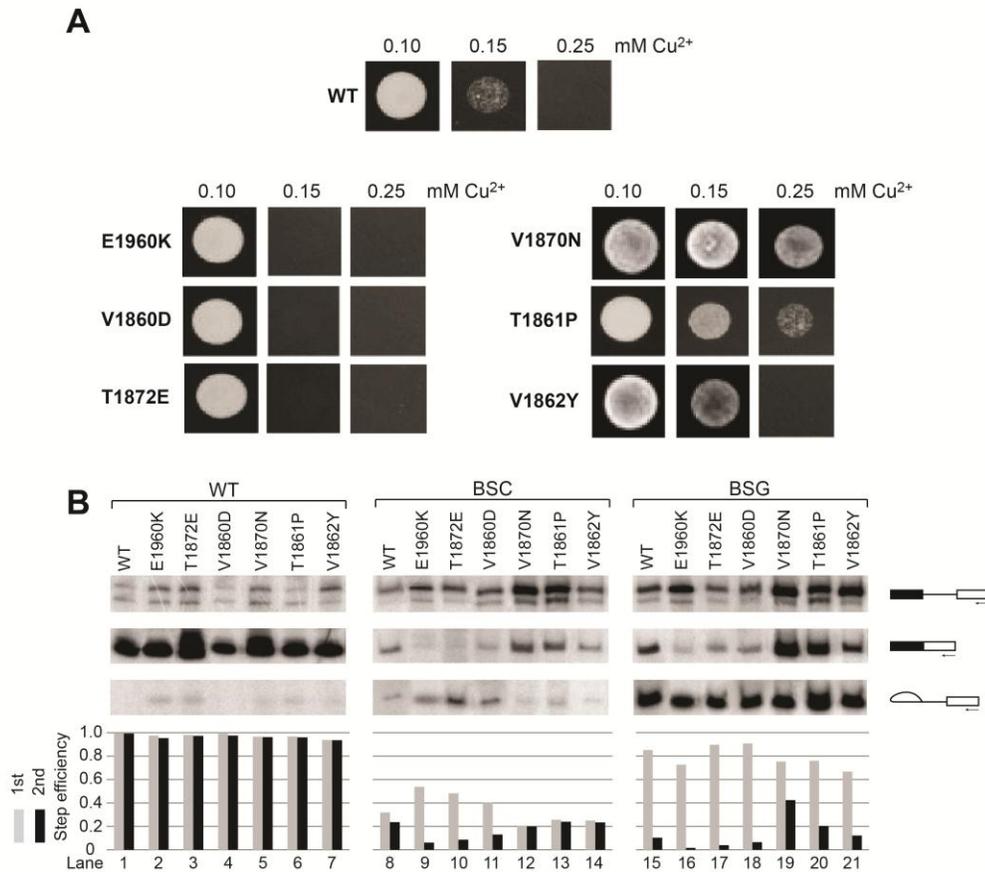


Figure 3-7. *In vivo* splicing with first- and second-step PRP8 alleles. A) Copper growth phenotypes of strains carrying BSC ACT1-CUP1 reporter plasmid and PRP8 wild-type (*upper*), first- (*bottom left*) or second-step (*bottom right*) alleles. Representative copper concentrations are shown. B) (*upper*) Primer extension analysis of RNA recovered from cells with wild-type, BSC and BSG ACT1-CUP1 plasmid and PRP8 alleles. (*bottom*) Quantification of the first and second step efficiency.

splicing. It should be noted that the values obtained from the primer extension represent steady-state levels of RNA *in vivo*, and some products may be under-represented because of discard pathways. However, they provide useful information of the relative differences between the tested alleles. All the PRP8 alleles tested completely spliced the wild-type reporter. When the BSC reporter was used, the first-step alleles showed reduced abundance of mRNA and

increased abundance of lariat intermediates. The strength of the three first-step alleles could also be ordered as E1960K>T1872E>V1860D. As for the BSG reporter, an increased level of mRNA was observed for the second-step alleles, while a decreased level of mRNA was obtained with the first-step alleles. The order of the strength of the second-step alleles was confirmed as V1870N>T1861P>V1862Y (Figure 3-7 B). These results are consistent with the copper growth test.

3-2.4. *In vitro* splicing of the first- and second-step alleles

We next examined *in vitro* splicing with extracts prepared from all the first- and second-step alleles. First we checked the first-step alleles with the ACT1 pre-mRNA containing the BSC mutation. As expected, E1960K - the strong first-step allele - improved the level of the lariat-exon significantly compared to the wild-type. But surprisingly, both T1872E and V1860D showed lower first step efficiency, although the second-step was almost completely abolished as expected. However, when we tried to chase the mutant with the wild-type extract at a 1:1 ratio, increased first-step rate (about 20% increase for V1860D and two-fold increase for the stronger first-step allele T1872E) was observed (Figure 3-8). Therefore the T1872E and V1860D mutants seem to be impaired for some step(s) before the first transesterification reaction is complete; while in the presence of the wild-type extract these first-step alleles can carry out the first-step reaction

with higher efficiency compared to the wild-type alone. At this point we do not know which step is compromised. It is also likely a factor other than Prp8 is present at reduced levels in the mutants, because otherwise the WT extract would not recover activity.

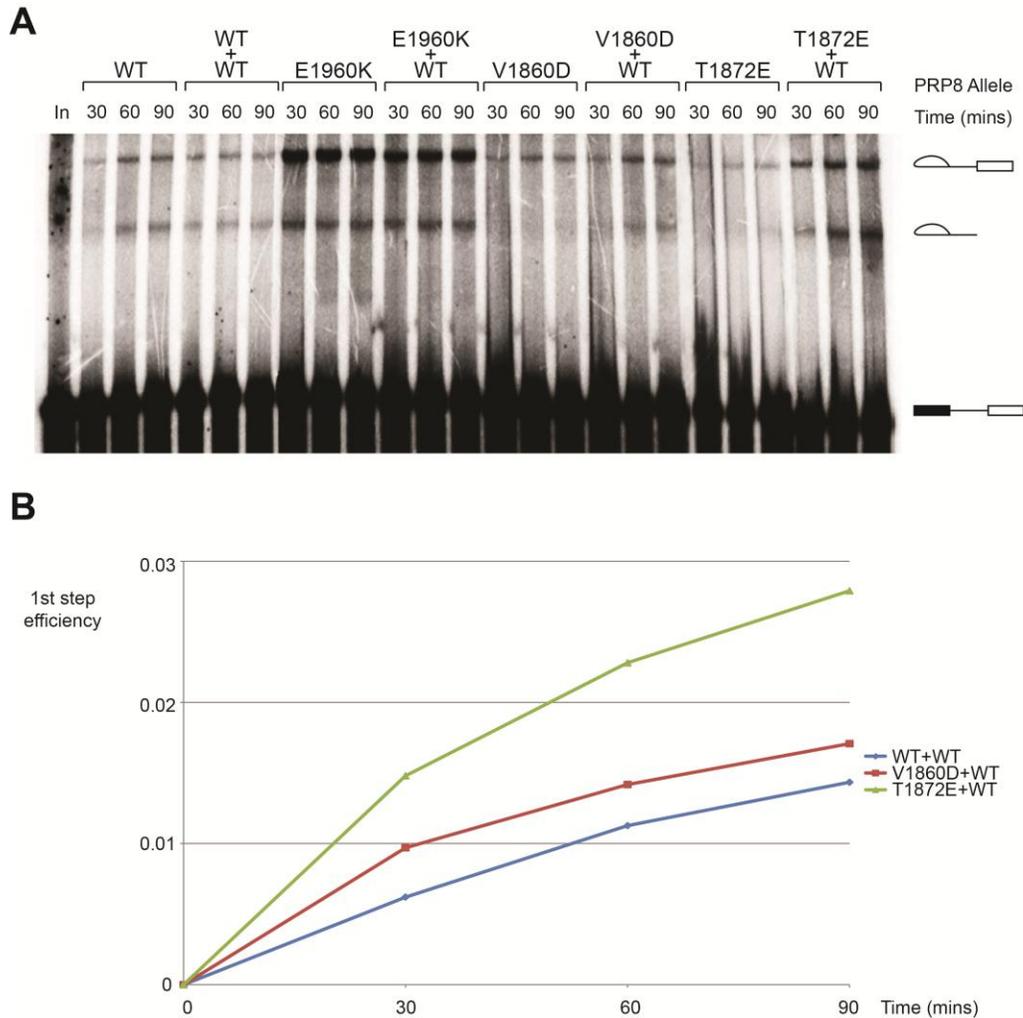


Figure 3-8. PRP8 first-step alleles improve the first step splicing of the BSC substrate. A) The yeast nuclear extract containing wild-type PRP8 showed defects for both the first and second step splicing of the BSC substrate *in vitro*. The first-step allele E1960K suppressed the defect of the first step. The V1860D and T1872E alleles alone showed lower first step efficiency and no second step splicing. When the first-step alleles were chased with wild-type, however, the first step efficiency was increased compared to the wild-type alone. B) Quantification of the first step efficiency of wild-type+wild-type, wild-type+1860D and wild-type+T1872E. The first-step alleles chased by wild-type increased the first step efficiency.

We also examined the *in vitro* splicing with the second-step alleles using the ACT1 pre-mRNA containing the A-to-U mutation at the 3'SS (UuG). This mutation is not impaired for the first-step but is significantly impaired for the second-step. In agreement with the *in vivo* results, all of the three second-step alleles increased the amount of the mRNA formed. The order of the increase is also consistent with the copper growth test and the primer extension (Figure 3-9).

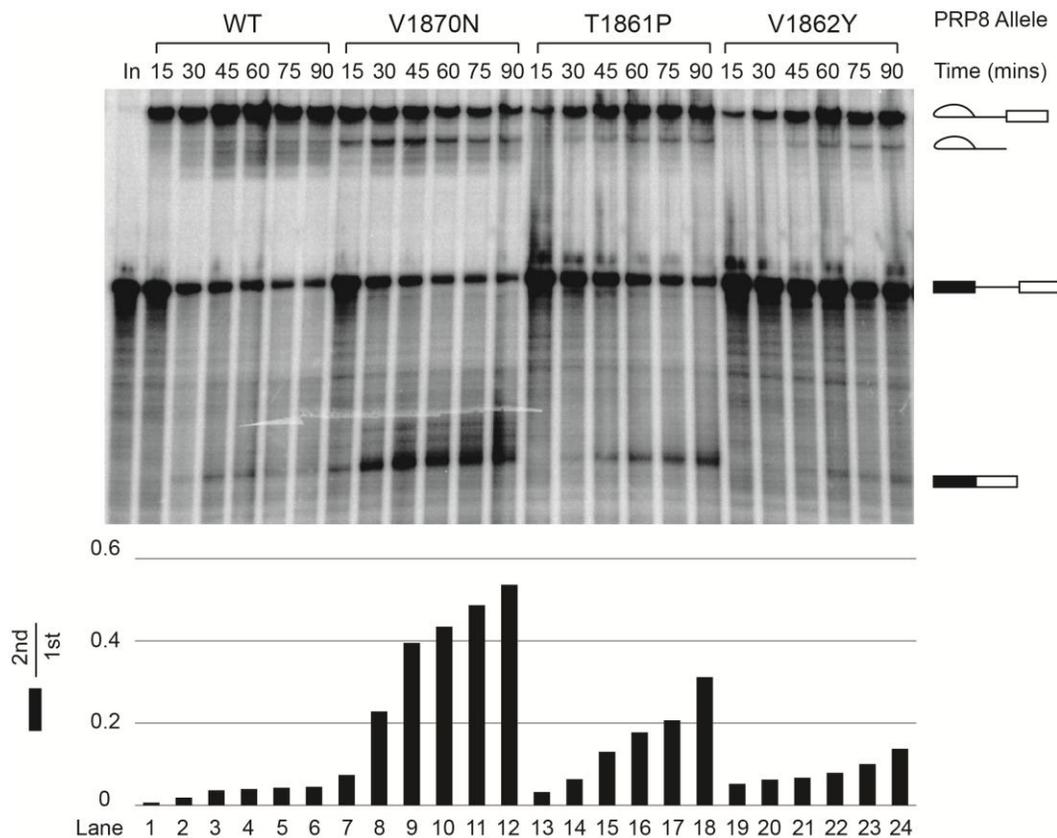


Figure 3-9. PRP8 second-step alleles improve the second step splicing of the UuG substrate. (*upper*) The yeast nuclear extract containing wild-type PRP8 (lane 1-6) showed a defect for the second step splicing of the UuG substrate *in vitro*, while second-step alleles PRP8 V1870N (lane 7-12), Prp8 T1871P (lane 13-18) and PRP8 V1862Y (lane 19-24) increased the second step efficiency. (*bottom*) Quantification of the second step efficiency over the first step efficiency. The number indicates the percentage of RNA substrate undergoing the second step after finishing the first step.

3-2.5. D1853C is not a first-step allele but can be rescued by second-step alleles

At this point the defect of the second step of splicing by the D1853C PRP8 mutation (see chapter 2) brings forward the question as to whether the impairment is caused by this mutation functioning as a first-step allele. In order to determine if this is the case, we compared the *in vitro* splicing of the BSC substrate with wild-type, D1853C and E1960K alleles of PRP8. As expected, E1960K improves the efficiency of the first step of splicing of this substrate. However, D1853C impairs the first step of splicing of this substrate, even in the presence of the wild-type extract (Figure 3-10 A). This indicates that the severe second-step defect caused by D1853C is not caused by stabilizing the first-step conformation of the spliceosome.

It has been shown that a double-mutant containing both first- and second-step mutations in a single Prp8 molecule cancelled suppression by the individual mutations and yielded a phenotype similar to that observed with wild-type (Liu et al., 2007). This can be explained by our two-monomer model that both conformations are now stabilized and the equilibrium between the two steps is thus less affected. We therefore assumed that although the defect of the second step in D1853C is not due to altered stability of the two monomers and the equilibrium of the splicing, a combination of D1853C with a second-step allele mutation may still stabilize the monomer b conformation and thereby rescue the second-step defect in D1853C.

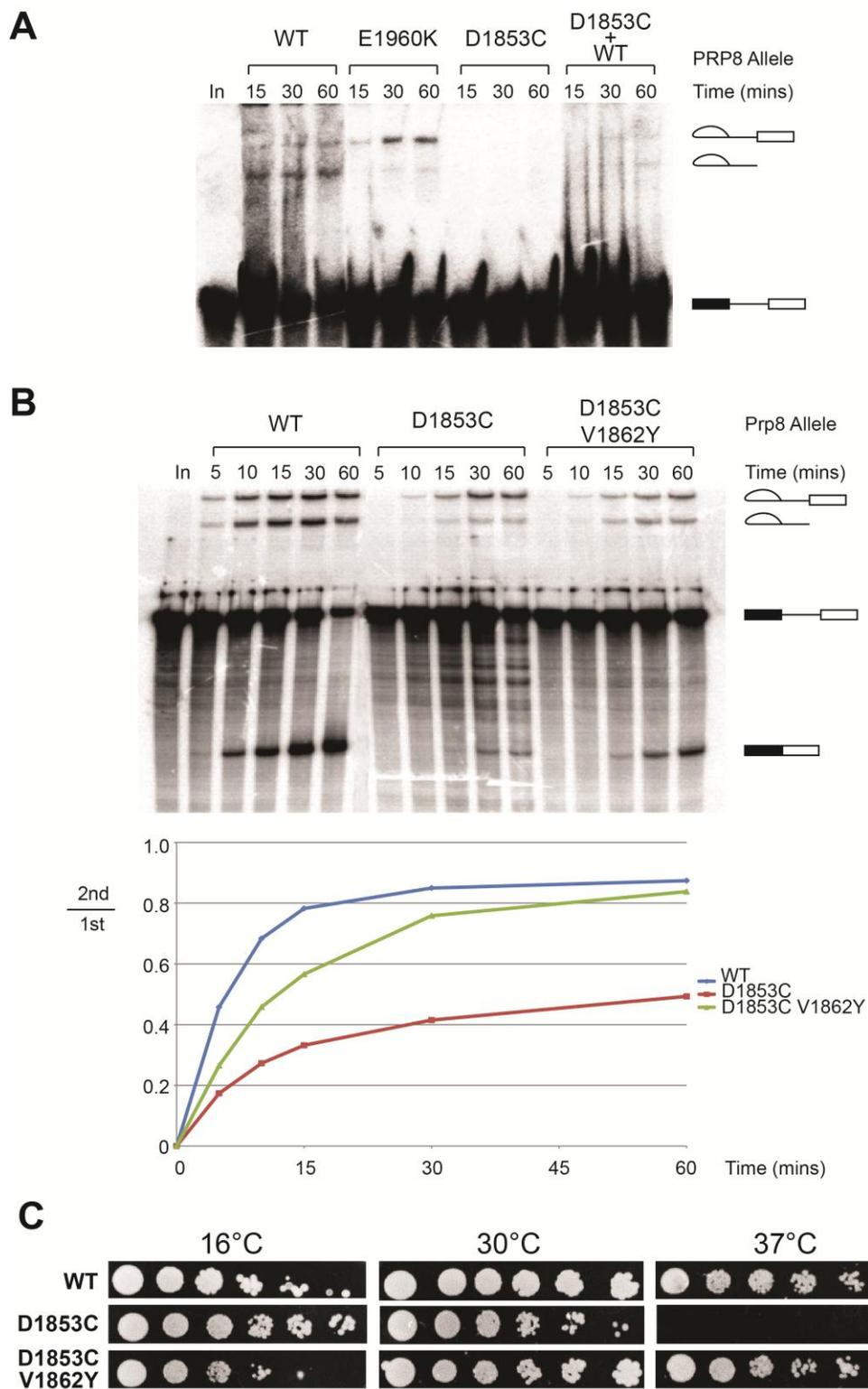


Figure 3-10. (Legend on next page)

D1853C impairs the second step of splicing like BSG or UuG, therefore a second step allele should suppress the D-to-C mutation, just as it suppresses other second-step defects. A combination of the D1853C with a second-step mutation should stabilize monomer b and therefore move the equilibrium towards the second step. This double-mutant should rescue the second-step deficiency and increase the amount of the spliced mRNA, even though not to the same level as the wild-type. This hypothesis turned out to be true, as shown with D1853C/V1862Y mutant. The second-step efficiency is substantially raised *in vitro*, although still at a lower level compared to the wild-type (Figure 3-10 B). In addition, this double mutation also rescued the lethality of D1853C at 37 °C and it showed a growth phenotype similar to the wild-type (Figure 3-10 C).

3-3. Discussion

In the studies described here, we have observed crystallographically two distinct conformations of the Prp8 RNase H domain. We have shown that several

Figure 3-10. D1853C is not a first-step allele but can be rescued by second-step alleles. A) The yeast nuclear extract were tested for the splicing of the BSC substrate *in vitro*. The first-step allele E1960K increases the first-step but decreases the second-step efficiency. In contrast, D1853C further impairs the first step of splicing for this substrate, indicating it is not a first-step allele. D1853C chased by the wild-type extract did not show increased first-step efficiency compared to the wild-type alone. B) (*upper*) *In vitro* splicing with the wild-type pre-mRNA substrate. D1853C is severely impaired for the second step, while the double mutation D1853C V1862Y rescues the second step. (*lower*) Quantification of the second step efficiency over the first step efficiency. The number indicates the percentage of RNA substrate undergoing the second step after finishing the first step. C) Spot test of D1853C and D1853C/V1862Y. The growth of D1853C is mildly impaired at 30 °C and is lethal at 37 °C, while D1853C/V1862Y displays similar growth rate to the wild-type at all temperatures.

reported first- or second-step alleles either form extra hydrogen bonds to stabilize or disrupt one conformation of the Prp8 RNase H domain. Therefore we propose that this domain undergoes a similar conformational change between the two steps of splicing. Two alleles among those discussed here, T1861P and V1862Y, have not been demonstrated as first- or second-step alleles before. Based on their crystal structures, we proposed that these two mutations both lead to the preference for the monomer b conformation and were therefore potentially second-step alleles. This hypothesis was confirmed by *in vivo* and *in vitro* splicing assays. Four other known alleles have also been confirmed by structural and biochemical information. In addition, the structure of T1789P/R1865A + R1865A mixture strongly argues that the T-to-P mutation moves the equilibrium towards monomer b corresponding to the second-step conformation of the spliceosome.

Combined with the data in chapter 2, the structures of proteins corresponding to first- and second-step alleles also argue that the non-metal binding conformation corresponds to a first-step conformation and divalent metal coordination by Prp8 is coupled to the second step of splicing. Our results provide more evidence supporting the model of two active sites within the spliceosome. They also explain why the splicing equilibrium was sensitive to the concentrations of salt and divalent metal ions (Tseng and Cheng, 2008), since the protein conformation can easily be affected thus shifting the equilibrium.

Altogether four first-step and three second-step alleles in the β -hairpin region have been reported before (Table 3-1). In this study we were able to

provide structural information for four of them. The other three, the second-step allele H1863E and the first-step alleles T1865K and A1871E (Yang et al., 2008) have also been confirmed by our *in vitro* splicing (data not shown). Although we were able to grow crystals for all three human proteins, no conclusion can be reached because the regions containing these mutations were disordered in one monomer or the other. Based on analysis of the wild-type Prp8 structure, the T1799E mutation in human Prp8 could be proposed to form an additional hydrogen bond with K1801 and stabilize monomer a. No clue can be found as to the conformational preference in H1863E and T1865K. The structural basis for these alleles remains to be elucidated.

The transition between the first- and second-step states in the spliceosome is more complex than that described here as evidenced by alleles located outside the β -hairpin region involved with the conformational change. For instance, in addition to the β -hairpin region, there are some other hot spots on the RNase H domain whose mutations suppress the mutants in the pre-mRNA sequence. For example, the T1982A and V1987A all suppress the 5'SS, BPS and 3'SS mutations and result in the second-step alleles. Another mutation, F1851L, suppresses the U4-cs1 mutation (Umen and Guthrie, 1995; Siatecka et al., 1999; Table 3-1). All these three mutations are within or around the five-stranded parallel/antiparallel β -sheet region in the RNase H core. One of them, T1782A, is located close to the β -hairpin region (Figure 3-2 B). However, none of these mutations are predicted to affect the structure of the β -hairpin region by simple inspection of our original

structure. Therefore there seems to be additional factors affecting the conformational change between the splicing steps. We tried to solve the structure of human E1888K protein, which is a strong first-step allele in yeast. Unfortunately there is no change in its structure compared to the wild-type (data not shown). It should also be mentioned that two first-step alleles (R1793K and L1557F) and four second-step alleles (*prp8-B141* (I857T, Y923C, E942G), P986T, *prp8-C133* (I1444V, T1565A, V1621A) and W1575R) lying outside the RNase H domain have been reported (Collins and Guthrie, 1999; Query and Konarska, 2004). The *prp8-C133* and W1575R alleles are located in the proposed domain 3 or the Th/X and U6i domains, while the *prp8-B141* and P986T are not within any known or predicted domain (Grainger and Beggs, 2005; Dlakic and Mushegian, 2011; Figure 1-1). Therefore additional domains seem to be involved in the transition between the splicing steps. It will be appealing to discover additional mechanisms by which Prp8 regulates splicing.

It has been observed that many PRP8 mutations suppress the U4-cs1 cold-sensitive phenotype, but the relationship between splice-site suppression and U4-cs1 suppression by PRP8 is still not well understood. Notably, U4-cs1 suppression appears to occur prior to 5'SS cleavage (Kuhn et al. 1999). It is also intriguing that the U4 cs alleles are not correlated with the first- and second-step alleles. Within the RNase H region, the first-step allele V1860D and the second-step allele T1861P both suppress the U4-cs1 mutation, while all the other first- and second-step alleles have no effect. Furthermore, PRP8 alleles suppressing the

Brr2, Prp28 and Prp22 mutations are also not correlated specifically with the first- and second-step alleles. Intriguingly, none of these alleles fall into the RNase H region. It thus remains to be shown how Prp8 can accomplish other functions and it will be exciting to see whether the same or perhaps another conformational change is involved in U4/U6 unwinding and Brr2 activation.

To summarize, our data provides for the first time the structural evidence supporting the two-conformation model for the transition of the two splicing steps. The two monomers in the RNase H region may only present a small part of the whole story. 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 the metal ion also implicates it as a key regulatory mechanism in catalysis of the second step of splicing. Previous research has placed Prp8 at the heart of the spliceosome and our experiments provide new insights into its crucial role in splicing.

3-4. Materials and Methods

3-4.1. Protein expression, purification, and crystallization

All mutants were generated using the overlapping PCR method. Cloning, expression, and purification of all human mutant proteins was performed as described in section 2-4.1. Crystals were grown using the procedure described in

section 2-4.1 under the new crystallization conditions (10-14% PEG 4000, 100 mM Tris pH 8.0, and 300 mM MgCl₂).

Data collection and processing was performed as described in section 2-4.2. Model building and refinement was carried out as described in section 2-4.3.

3-4.2. Construction of RNA substrates

The ACT1 pre-mRNA was made by *in vitro* transcription with [α -³²P]-ATP. The template was the PCR product off pBS-ACT1 plasmid. The templates for BSC and UuG pre-mRNA substrates were generated using the overlapping PCR method as described in section 2-4.4.

3-4.3. Creation of mutant *S. cerevisiae* strains and splicing extracts

Mutant *S. cerevisiae* strains and splicing extracts were prepared with the procedure described in 2-4.5.

For the creation of the copper-resistant strains, the ACT1-CUP1 plasmid containing either the wild-type sequence, BSC or BSG mutations was co-transformed along with pJU186 containing wild-type or mutant PRP8 into yJU75 (*MATa, ade2 cup1D::ura3 his3 leu2 lys2 prp8D::LYS2 trp1, pJU169 (PRP8 URA3 CEN ARS)*); (Umen and Guthrie, 1996; Courtesy of Dr. Jonathan Staley). 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.

3-4.4. Temperature and copper growth tests

Spot tests were carried out as described in section 2-4.5. For the copper growth assay, cultures containing the ACT1-CUP1 and PRP8 mutant plasmids were grown overnight in -Leu-His medium and diluted to $A_{600}=0.2$ and equal volumes were spotted onto -Leu-His plates containing 0–2.0 mM CuSO_4 (Lesser and Guthrie, 1993). Plates were photographed after 3 days at 30 °C.

3-4.5. Primer extension

The yeast RNA was extracted using hot phenol (modified from Kohrer and Domdey, 1991). Cultures containing the ACT1-CUP1 and PRP8 mutant plasmids were grown overnight in -Leu-His medium and diluted to 5 mL $A_{600}=0.2$. The diluted cells were grown in -Leu-His medium for additional 6 hours to $A_{600}=1.0$. The cells were then spun down and resuspended in 400 μL of AE buffer with 10% SDS (50 mM NaOAc, 10mM EDTA, pH=5.0). 400 μL hot phenol/AE solution was immediately added into the solution and incubated at 65 °C for 30 min. After spinning down the pellet, the upper phase was transferred into clean tubes and the

hot phenol/AE treatment was repeated once. The upper phase was then extracted three to four times with phenol/chloroform/isoamyl alcohol and chloroform, followed by ethanol precipitation. The pellet was resuspended in 10 μL ddH₂O.

Primer extensions were carried out using the YAC6 primer complementary to exon 2 of ACT1 (Query and Konarska, 2004). The primer was end-labelled with [γ -³²P]-ATP.

Primer extensions were performed using the RevertAidTM H minus first strand cDNA synthesis kit (Fermentas). First 12 μL of the mixture (1 μg total RNA, 2 pmol labelled primer) was heated to 70 $^{\circ}\text{C}$ for 10 min and slowly cooled down to 40 $^{\circ}\text{C}$. The reaction was then chilled on ice and supplemented to 20 μL (4 μL reaction buffer, 1 u RNase inhibitor, 1 mM dNTP mix, 10 u reverse transcriptase). The mixture was incubated for 5 min at 37 $^{\circ}\text{C}$ and then 55 min at 42 $^{\circ}\text{C}$. The reaction was terminated and the RNA was degraded with 0.5 M NaOH at 70 $^{\circ}\text{C}$. Extension products were extracted with phenol/chloroform/isoamyl alcohol and chloroform, followed by ethanol precipitation. The pellet was resuspended, separated in 7% polyacrylamide, 8 M urea gels and visualized by autoradiography.

3-4.6. Splicing assays

Splicing reactions were performed as described in section 2-4.6.

3-5. References

- Burgess, S.M., and Guthrie, C. (1993). A mechanism to enhance mRNA splicing fidelity: the RNA-dependent ATPase Prp16 governs usage of a discard pathway for aberrant lariat intermediates. *Cell* **73**, 1377-1391.
- 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.
- de la Cruz, J., Kressler, D., and Linder, P. (1999). Unwinding RNA in *Saccharomyces cerevisiae*: DEAD-box proteins and related families. *Trends Biochem. Sci.* **24**, 192-198.
- Grainger, R.J., and Beggs, J.D. (2005). Prp8 protein: at the heart of the spliceosome. *RNA* **11**, 533-557.
- Kohrer, K., and Domdey, H. (1991). Preparation of high molecular weight RNA. *Methods Enzymol.* **194**, 398-415.
- Konarska, M.M., and Query, C.C. (2005). Insights into the mechanisms of splicing: more lessons from the ribosome. *Genes Dev.* **19**, 2255-2260.
- Konarska, M.M., Vilardell, J., and Query, C.C. (2006). Repositioning of the Reaction Intermediate within the Catalytic Center of the Spliceosome. *Mol. Cell* **21**, 543-553.
- Kuhn, A.N., and Brow, D.A. (2000). Suppressors of a cold-sensitive mutation in yeast U4 RNA define five domains in the splicing factor Prp8 that influence spliceosome activation. *Genetics* **155**, 1667-1682.
- Kuhn, A.N., Li, Z., and Brow, D.A. (1999). Splicing factor Prp8 governs U4/U6 RNA unwinding during activation of the spliceosome. *Mol. Cell* **3**, 65-75.
- Kuhn, A.N., Reichl, E.M., and Brow, D.A. (2002). Distinct domains of splicing factor Prp8 mediate different aspects of spliceosome activation. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 9145-9149.
- Lesser, C.F., and Guthrie, C. (1993). Mutational analysis of pre-mRNA splicing in *Saccharomyces cerevisiae* using a sensitive new reporter gene, CUP1. *Genetics* **133**, 851-863.
- 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., and Sharp, P.A. (1993). Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. *Nature* **365**, 364-368.

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.

Schwer, B., and Gross, C.H. (1998). Prp22, a DExH-box RNA helicase, plays two distinct roles in yeast pre-mRNA splicing. *EMBO J.* **17**, 2086-2094.

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

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

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

Strauss, E.J. and Guthrie, C. (1991). A cold-sensitive mRNA splicing mutant is a member of the RNA helicase gene family. *Genes & Dev.* **5**, 629-641.

Tseng, C.K., and Cheng, S.C. (2008). Both Catalytic Steps of Nuclear Pre-mRNA Splicing Are Reversible. *Science* **320**, 1782-1784.

Tseng, C.K., Liu, H.L., and Cheng, S.C. (2010). DEAH-box ATPase Prp16 has dual roles in remodeling of the spliceosome in catalytic steps. *RNA* **17**, 145-154.

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.

Umen, J.G., and Guthrie, C. (1996). Mutagenesis of the yeast gene PRP8 reveals domains governing the specificity and fidelity of 3' splice site selection. *Genetics* **143**, 723-739.

Wagner, J.D., Jankowsky, E., Company, M., Pyle, A.M., and Abelson, J.N. (1998). The DEAH-box protein PRP22 is an ATPase that mediates ATP-dependent mRNA release from the spliceosome and unwinds RNA duplexes. *EMBO J.* **17**, 2926-2937.

Wang, Y., Wagner, J.D., and Guthrie, C. (1998). The DEAH-box splicing factor Prp16 unwinds RNA duplexes in vitro. *Curr. Biol.* **8**, 441-451.

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

Chapter 4

Conclusions and future directions

The work presented in this chapter represents collaboration between authors. Tao Wu prepared proteins for crystallization, grew the crystals, and performed the limited proteolysis assay. Matt Schellenberg performed the X-ray data collection and solved the crystal structures.

4-1. Conclusions

In this thesis we show that the RNase H domain of Prp8 binds a magnesium ion which is important for the second step of splicing as the mutation of the key D1853 residue which abrogates metal binding severely inhibits the second step of splicing. We also show that yeast Prp8 first- and second-step alleles favor the conformation of one of two distinct monomers observed in the crystal structure of Prp8 domain IV. These data provide structural and functional evidence that the RNase H domain of Prp8 undergoes a conformational change between the first and second step of splicing, and that this unmask a metal-binding site which is required for the catalysis of the second step. This model of splicing regulation is summarized in Figure 4-1.

Our experiments convincingly establish Prp8 as a metalloprotein which promotes the exon ligation reaction; thus it supports the hypothesis that spliceosomes are ribonucleoprotein enzymes. As discussed in the introduction, although a protein-free system was able to carry out the second step of splicing, this reaction is significantly slower than that in spliceosome-catalyzed systems (Valadkhan and Manley, 2001; Valadkhan et al., 2007). This can be explained by our results: protein factors not only help to stably position the snRNA elements but also participate in the formation of the active site. At this point the RNA components of the active site are still unknown, but several studies have provided candidates. U6 snRNA of the spliceosome is structurally similar to the domain V (DV) of the group II intron and its bulge region on the 3'ISL has been shown to

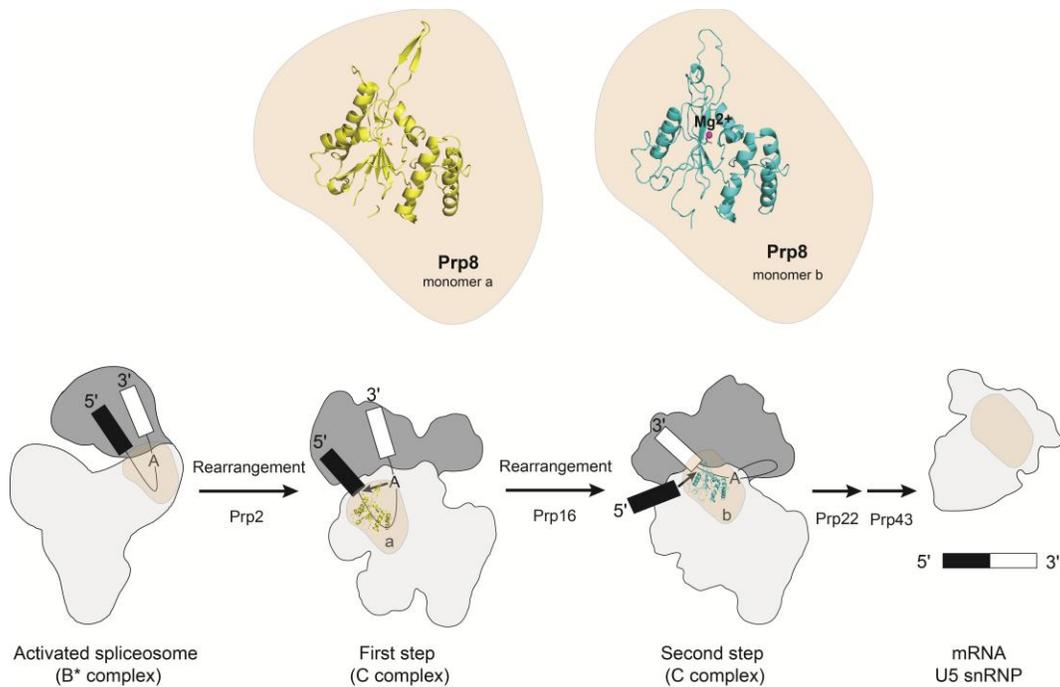


Figure 4-1. Model of the role of Prp8 RNase H domain in splicing. The conformational change within the β -hairpin of Prp8 converts monomer a to monomer b. This results in the binding of Mg^{2+} (purple) to the active site and the transition from the first step into the second step of splicing. The shape of the complexes and the position of the Prp8 and pre-mRNA within them are all based on the EM structure studies (Boehringer et al., 2004; Golas et al., 2010). The U5-snRNP region is distinguished from other regions of the complex B* and C by light grey.

contain a metal binding site critical for splicing activity (Yean et al., 2000). Mutagenesis studies have also suggested multiple roles of the U6 AGC trinucleotide for splicing and one of these functions is the potential binding of a catalytic metal (Hilliker and Staley, 2004). As noted in chapter 1 and above, a protein-free system containing only U2/U6 snRNA and two short RNA substrates was shown to be able to carry out the second step of splicing (Valadkhan et al., 2009). All these data strongly argue that U6 snRNA contributes to the metal coordination and further mutagenesis assays should be focused on the two hot spots noted here. It has also long been debated whether the spliceosome has a

single active site for both steps of splicing or two separate sites for different steps. Our result supports the latter model. The mutation of D1853 does not affect the first step; hence the RNase H domain of Prp8 does not seem to be directly involved in the first step catalysis. As more evidence becomes available, it will be intriguing to investigate the components of different active sites.

As this thesis was prepared, a paper was published concerning the structure of Cwc2, a component of the yeast nineteen complex (NTC) (Lu et al., 2011). The NTC contains Prp19 and other related proteins and helps to stabilize association of U5 and U6 with the B* complex after the dissociation of U1 and U4 snRNP (Chan et al., 2003). The Cwc2 structure has a Torus domain, an RNA recognition motif (RRM) and a zinc-finger (ZnF) and they are tightly integrated in a compact folding unit. It also has a protruding element connecting the ZnF and RRM domains, which was shown to be involved in the RNA binding. It is interesting that this protruding loop can be traced in only one of the two molecules from the asymmetric unit (Lu et al., 2011). This is reminiscent of the two monomers of the Prp8 RNase H domain. There is additional evidence that the NTC complex regulates the second step of splicing since the depletion of NTC complex caused mild first step defect but severe second-step defect (Ajuh et al., 2000). Plus, Cwc2 has been shown to pull-down Prp8 (Gahura et al., 2009). It is intriguing to speculate whether a conformational change in Cwc2 is related to that in Prp8. Therefore the transition from the first step to the second step might involve a network of proteins undergoing conformational changes. With more structures

available for individual components we will have a clearer view of the larger splicing machinery.

4-2. Future directions

4-2.1. Designing new mutant alleles for the RNase H domain

The conformational change of Prp8 shown in chapter 4 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 also be useful for stalling spliceosomes at certain stages of the splicing reaction for other biochemical experiments.

As an initial trial, we tried to design a double cysteine mutant which might lock the spliceosome at a certain stage through disulfide formation. The same technique has been very successfully used in the stabilization of intramolecular interactions as well as characterization of protein protein, protein DNA and protein RNA complexes (Matsumura et al., 1989; Mansfeld et al., 1997; Huang et al., 1998; Corn and Berger, 2007; Schellenberg et al., 2011). Two residues in the β -hairpin region, I1790 and T1800, are in close proximity in monomer a. In monomer b they are distant from each other and they face opposite sides of the loop region. Therefore, we created the I1790C/T1800C double cysteine mutant. Based on the mutagenesis structure prediction, the two cysteines should be able to

form a disulfide bridge under non-reducing conditions, which would stabilize the β -hairpin region and the first-step conformation.

We were able to grow crystals in reducing conditions, which we could then soak in oxidising or only weakly reducing conditions to see disulfide bond formation. The disulfide was completely oxidised in 1 mM BME, but was reduced in 1 mM TCEP (Figure 4-2). We then made the yeast double mutant and prepared cell extract for *in vitro* splicing assay. As an attempt to form a stable disulfide bond which would trap the Prp8 RNase H domain in the first-step conformation, we attempted a treatment with different oxidizing reagents up to the point that the splicing was completely blocked. We also tried the mild oxidation by dialyzing the extract into buffer saturated with oxygen by constantly bubbling air through it. The same method has been used to successfully create disulfide bond cross-linked ribosomes (Horan and Noller, 2007). Unfortunately, we were not able to observe any shift towards the first step in both cases (data not shown). The overnight dialysis completely quenched the splicing reaction. Because of the susceptibility of the double cysteine mutant to oxidation in the isolated crystallized domain, we would expect a similar susceptibility in Prp8 in intact spliceosomes. It seems that this double cysteine mutant is no more sensitive to oxidation than wild-type spliceosomes, and it is a likely possibility that the Prp8 protein is buried deeply in the core and not easily accessible for oxidation. Although our attempt to design new mutants was not successful in producing the desired tool, it suggests a helpful strategy for future studies.

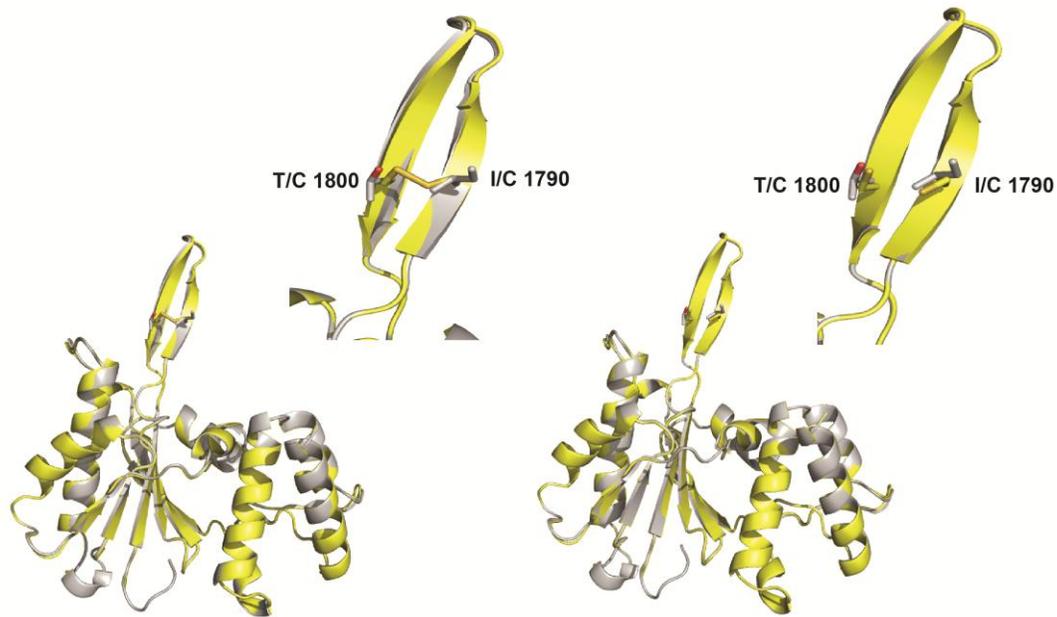


Figure 4-2. Structure of the Prp8 I1790C/T1800C mutant. (*left*) Details of the 1.9 Å X-ray structure of the double cysteine mutant soaked in non-reducing buffer. Shown are the monomers a of the wild-type (grey) and the mutant (yellow). The non-metal binding β -hairpin conformation of monomer a is stabilized by a disulfide bridge between C1790 and C1800. (*right*) Details of the 2 Å X-ray structure of the double cysteine mutant in reducing buffer containing 1 mM TCEP. The disulfide bridge is now totally reduced.

4-2.2. Mapping of protein–RNA crosslinking sites within the RNase H region

As discussed in chapter 1, Prp8 cross-links to the 5'SS before the first step of splicing and the region responsible for this interaction is not conserved between human and yeast (1894-QACLK-1898 and 1966-SAAMS-1979, respectively; Reyes et al., 1999). Four residues (F1834, R1922, V1946 and E1960) in the RNase H region have also been shown to be important for yPrp8/PPT interaction during suppression screens (Umen and Guthrie, 1996). Based on these data and the structure of the RNase H domain, a model of has been proposed regarding the interactions between the RNase H domain and the 5'SS as well as the U1, U5 and

U6 snRNAs (Wyatt et al., 1992; Reyes et al., 1996; Reyes et al., 1999; Pena et al., 2008; Figure I-4). However, the crosslinking data noted above was obtained from the precatalytic B complex and the key Prp8 residues involved in this interaction are still unknown.

The UV-induced protein-RNA crosslinking called zero-length cross-link utilizes the natural reactivity of the nucleic acid bases after irradiation at a wavelength of 254 nm. It relies on the UV-induced formation of covalent bonds between RNA bases and amino acid residues in close spatial proximity (Karmar et al., 2011). This technique could be used to produce more detailed information about Prp8-RNA contacts. It will be intriguing to investigate whether the first- and second-step alleles or U4 cs alleles change the crosslinking between Prp8 and the substrate.

For a future Prp8 crosslinking experiment, a good RNA substrate candidate will be the four-helix-junction RNA used in a domain IV gel-shift assay from our lab (Ritchie et al., 2008). This RNA mimics the interaction between the U2 and U6 snRNAs and between the 5'SS and U6 snRNA at the catalytic site of the spliceosome and displayed the highest affinity for the RNase H domain during the mobility shift assay (Sashital et al., 2004; Ritchie et al., 2008), making it the best candidate for crosslinking and mass spectrometry studies. It would also be worthwhile to design other RNA substrates around the BPS and the 3'SS for crosslinking studies since we have shown the RNase H region active site is important for the catalysis of the second step. These studies will provide detailed

information about the interactions between the RNase H domain and its RNA substrate. Also, a comparison between the data obtained for the 5'SS and 3'SS may verify our model of the conformational change.

4-2.3. Investigating the role of other Prp8 domains in different steps of splicing

The RNase H region and the well-studied MPN domain encompass less than one third of mutations known to suppress the mutations in 5'SS, 3'SS, BPS, U4 and U6 snRNAs as well as the mutations in Brr2, Prp22 and Prp28 proteins. There are several other hot regions on Prp8. For example, a cluster of eighteen mutations suppressing the U4-cs1 mutations are located in the N-terminus region ranging from aa 236-861. Three of these are within a proposed bromo (Br) domain and the others are within the Prp8 central (PROCN) domain. In a proposed reverse transcriptase (RT-like) domain, five alleles suppress the pre-mRNA, U4-cs1 and brr2-1 mutations respectively. Two second-step alleles have also been confirmed in this region. The proposed Th/X and U6i regions (or the proposed domain 3) contain more than a dozen alleles suppressing the pre-mRNA mutations and two confirmed second-step alleles as well as three alleles suppressing the U4-cs1. There are also several other alleles in the sequences with no proposed roles (Grainger and Beggs, 2005 and references therein; Dlakic et al., 2011; Figure 1-1 and 3-2). Therefore the physiological functions of Prp8 involve a much more

complicated network between different regions than discussed in this thesis. High-resolution structural analysis of other domains will help to elucidate a complete understanding of Prp8-RNA and Prp8-protein interactions.

As an initial trial to solve the structures of other Prp8 domains, we examined the proposed RT domain. This region also contains the previously proposed RRM motif (Grainger and Beggs, 2005). The corresponding RT-like region on the telomerase subunit TERT is involved in substrate binding and the formation of the nucleotidyltransferase catalytic center (Mitchell et al., 2010). The RT-like domain of Prp8 has been shown to cross-link to the 5'SS and BPS (Turner et al., 2006). Considering the extensive base-pairing of U6/5'SS and U2/BPS, it is tempting to assume the RT-like domain would recognize the dsRNA region in a similar manner as to TERT in the TERT/DNA RNA complex. An MBP fused yPrp8 (aa 965-1189) was expressed in *E.coli* and purified with an apparent size corresponding to a dimer in buffer containing detergent. However after the MBP tag was cleaved, the protein aggregated, indicating this construct itself is only partially stable. Limited proteolysis of the protein aggregate yielded a stable fragment of ~17kD (Figure 4-3). We are currently working to identify and express this region.

The proposed domain 3 is also of special interest in light of the structural studies. It was originally divided into two regions, the more C-terminal domain 3.2 (1547–1660 in yeast) being most highly conserved (72% identity between yeast and human) and containing all but two of the pre-mRNA suppressor alleles

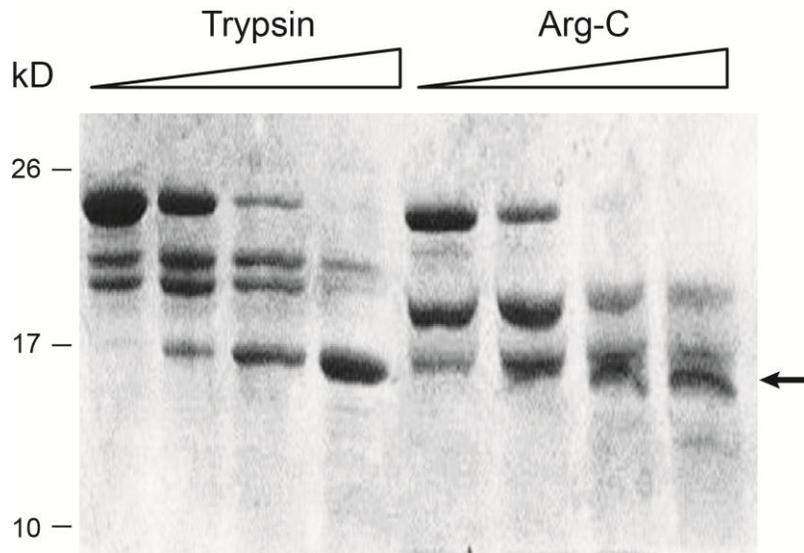


Figure 4-3. Limited proteolysis of yPrp8 965-1189. The protein was incubated with the proteases for 1 hr at room temperature. A stable ~17 kD fragment is shown by arrow.

found in domain 3 (Grainger and Beggs, 2005). In the most recent models it was split into a Th/X domain and a U6 interaction (U6i) domain. There are no clues in the primary sequence as to the possible three-dimensional structure of domain 3. We tried to express more than twenty constructs of domain 3 in *E.coli* and none of them was soluble (data not shown). Genetic studies have also implied that domain 3.2 may be involved in an intramolecular interaction with a more N-terminal region of yPrp8 as a slow growth phenotype caused by a H659P mutation is suppressed by the domain 3.2 mutation L1634F (Kuhn and Brow, 2000). However, our co-expression of constructs from these two regions did not produce soluble protein either (data not shown). Based on the secondary structure prediction, domain 3 contains a large proportion of loops, which suggests it may

need RNA or other proteins for proper folding and stable expression, similar to the N-terminus of U1-70K in the U2 snRNP crystal structure . Future work should be focused to solve this issue. Hopefully, analyses of the other proposed domains will further clarify the role of Prp8; we can anticipate exciting new developments in the near future.

4-3. Materials and Methods

4-3.1. Protein expression, purification, and crystallization

The I1790C/T1800C mutation was generated using the overlapping PCR method. Cloning, expression, and purification of the proteins were done as described in section 2-4.1. Crystals were grown using the procedure described in section 2-4.1 under the new crystallization conditions (10-14% PEG 4000, 100 mM Tris pH 8.0, and 300 mM MgCl₂) with the inclusion of 1 to 4 mM TCEP (tris(2-carboxyethyl)phosphine). The purification of the yPrp8 aa 965-1189 was performed in buffer containing 0.1% Tween-20.

X-ray diffraction data collection and processing was done as described in section 2-4.2. Model building and refinement was done as described in section 2-4.3.

4-3.2. Limited proteolysis

Purified 20 µg of yPrp8 aa 965-1189 was mixed with increasing amounts of trypsin or endoprotease Arg-C (Sigma-Aldrich) and incubated for 1 h at RT. The reactions were stopped by the addition of loading buffer and analyzed on a 16% (200:1) SDS-PAGE gel.

4-4. References

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

Boehringer, D., Makarov, E.M., Sander, B., Makarova, O.V., Kastner, B., Lührmann, R., and Stark, H. (2004). Three-dimensional structure of a pre-catalytic human spliceosomal complex B. *Nat. Struct. Mol. Biol.* **11**, 463-468.

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.

Corn, J.E., and Berger, J.M. (2007). FASTDXL: A Generalized Screen to Trap Disulfide-Stabilized Complexes for Use in Structural Studies. *Structure* **15**, 773-780.

Gahura, O.e., Ahrhámová, K.i., Skružný, M., Valentová, A., Munzarová, V., Folk, P., and Půta, F.e. (2009). Prp45 affects Prp22 partition in spliceosomal complexes and splicing efficiency of non-consensus substrates. *J. of Cell. Biochem.* **106**, 139-151.

Golas, M.M., Sander, B., Bessonov, S., Grote, M., Wolf, E., Kastner, B., Stark, H., and Lührmann, R. (2010). 3D Cryo-EM Structure of an Active Step I Spliceosome and Localization of Its Catalytic Core. *Mol. Cell* **40**, 927-938.

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

Hilliker, A.K., and Staley, J.P. (2004). Multiple functions for the invariant AGC

triad of U6 snRNA. *RNA* **10**, 921-928.

Horan, L.H., and Noller, H.F. (2007). Intersubunit movement is required for ribosomal translocation. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 4881-4885.

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

Kramer, K., Hummel, P., Hsiao, H.-H., Luo, X., Wahl, M., and Urlaub, H. (2011). Mass-spectrometric analysis of proteins cross-linked to 4-thio-uracil- and 5-bromo-uracil-substituted RNA. *Int. J. Mass Spectro.* **304**, 184-194.

Kuhn, A.N., and Brow, D.A. (2000). Suppressors of a cold-sensitive mutation in yeast U4 RNA define five domains in the splicing factor Prp8 that influence spliceosome activation. *Genetics* **155**, 1667-1682.

Lu, P., Lu, G., Yan, C., Wang, L., Li, W., and Yin, P. (2011). Structure of the mRNA splicing complex component Cwc2: insights into RNA-recognition. *Biochem. J.* [Epub ahead of print].

Mansfeld, J., Vriend, G., Dijkstra, B.W., Veltman, O.R., Van den Burg, B., Venema, G., Ulbrich-Hofmann, R., and Eijsink, V.G. (1997). Extreme stabilization of a thermolysin-like protease by an engineered disulfide bond. *J. Biol. Chem.* **272**, 11152-11156.

Matsumura, M., Signor, G., and Matthews, B.W. (1989). Substantial increase of protein stability by multiple disulphide bonds. *Nature* **342**, 291-293.

Mitchell, M., Gillis, A., Futahashi, M., Fujiwara, H., and Skordalakes, E. (2010). Structural basis for telomerase catalytic subunit TERT binding to RNA template and telomeric DNA. *Nat. Struct. Mol. Biol.* **17**, 513-518.

Pomeranz Krummel, D.A., Oubridge, C., Leung, A.K.W., Li, J., and Nagai, K. (2009). Crystal structure of human spliceosomal U1 snRNP at 5.5 Å resolution. *Nature* **458**, 475-480.

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.

Ritchie, D.B., Schellenberg, M.J., Gesner, E.M., Raithatha, S.A., Stuart, D.T., and MacMillan, A.M. (2008). Structural elucidation of a PRP8 core domain from the heart of the spliceosome. *Nat. Struct. Mol. Biol.* **15**, 1199-1205.

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

Schellenberg, M.J., Dul, E.L., and MacMillan, A.M. (2010). Structural model of the p14/SF3b155 branch duplex complex. *RNA* **17**, 155-165.

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. (1996). Mutagenesis of the yeast gene PRP8 reveals domains governing the specificity and fidelity of 3' splice site selection. *Genetics* **143**, 723-739.

Valadkhan, S., and Manley, J.L. (2001). Splicing-related catalysis by protein-free snRNAs. *Nature* **413**, 701-707.

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.

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.

Wyatt, J.R., Sontheimer, E.J., and Steitz, J.A. (1992). Site-specific cross-linking of mammalian U5 snRNP to the 5' splice site before the first step of pre-mRNA splicing. *Genes Dev.* **6**, 2542-2553.

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

Appendix I ⁽¹⁾

Characterization of *in vivo* dimerization of ADAR proteins

¹Adapted from Chilibeck et al., (2006) *J. Biol. Chem.* **24**, 10630-10635

The work presented in this chapter represents collaboration between authors. Kaari Chilibeck prepared the constructs and performed the fluorescence microscopy and FRET assay with the wild-type ADAR proteins. Tao Wu and Chao Liang prepared the constructs and performed the fluorescence microscopy FRET assay with the ADAR mutants.

I-1. Introduction

Newly transcribed RNA molecules undergo a number of modifications other than splicing which alter their makeup and affect the final protein products. These processes, termed RNA editing, include base insertions, deletions or substitutions (Zinshteyn and Nishikura, 2009). The most common RNA editing in eukaryotes is the conversion of adenosine (A) to inosine (I), carried out by the ADAR (adenosine deaminase that acts on RNA) enzyme family (Figure I-1).

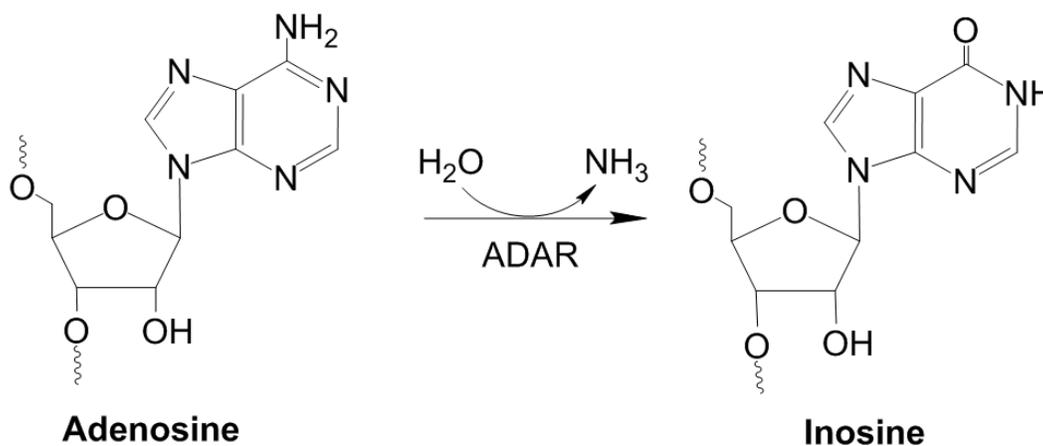


Figure I-1. Chemistry of deamination reaction. A hydrolytic deamination reaction converts adenosine to inosine.

The A-to-I editing may change the genetic information of the RNA molecule, as inosine is read as guanosine (G) by the translational machinery. A number of examples have been reported. These include the tissue-specific editing of the 2C subtype of serotonin receptor ($5\text{-HT}_{2\text{C}}\text{R}$), resulting in an amino acid change that causes more than tenfold reduction in the interaction between

receptors and their G coupled proteins (Burns et al., 1997). Two A-to-I modification sites within the neural-specific AMPA class of glutamate-gated (GluR) ion channels lead to the Q/R and R/G mutations and affect the recovery rate of the channels from desensitization (Sommer et al., 1991; Lomel et al., 1997). In Hepatitis δ virus, the conversion of UAG stop codon within its antigenomic RNA sequence to a UIG tryptophan codon is required for the transition from replication into packaging (Polson et al., 1996). The A-to-I editing may also affect alternative splicing by creating new 5' and 3' splice sites or abolishing 3'SS and the branch site. The best known example of editing is the self-regulation of ADAR2, which edits its own mRNA and thus modulates its own expression (Reuter et al., 1999). Recent studies have also indicated that A-to-I editing occurs in the miRNA sequence, thus altering its target specificity (Blow et al., 2006).

Despite the functional importance, the well-studied protein-coding editing only accounts for a very small portion of inosines detected in mammalian RNAs. It is estimated one inosine is present for every 17,000 nucleotides in brain mRNAs (Paul and Bass, 1996). Bioinformatic studies have mapped more than 12,000 editing sites, most of which occur in noncoding regions, typically in Alu repeats (Levanon et al., 2004; Maas et al., 2003). The editing of noncoding sequences may affect the stability and turnover of the RNA substrates. Extensive A-to-I editing may result in degradation of dsRNA, and Tudor staphylococcal nuclease (Tudor-SN), a component of RNA-induced silencing complex (RISC), has been shown to cleave dsRNA containing multiple I U base pairs (Scadden and

Smith, 2001; Scadden, 2005). It has also been suggested A-to-I editing within the 3' UTR results in the nuclear retention of mRNA (Prasanth et al., 2005; Chen et al., 2008). To date, the real roles of editing noncoding sequences remain largely unknown.

The ADAR activity was first identified in *Xenopus laevis* eggs and embryos with dsRNA unwinding activity (Bass and Weintraub, 1987). Three ADAR proteins, ADAR1-3 are found in humans, each with two or three dsRNA-binding domains (dsRBD) and a C-terminal catalytic deaminase domain (Figure I-2). ADAR1 also has one (short isoform) or two (long isoform) N-terminal Z-DNA binding domains, which prefer Z-form over B-form dsDNA or RNA and bind to the target with twofold symmetry. The physiological roles of these domains remain unknown (reviewed in: Seo et al., 2010; Schwartz et al., 1997). ADAR3 has an N-terminal arginine-rich single-stranded RNA binding (R domain). The ADAR3 deaminase activity has not been reported and recent phylogenetic studies proposed that ADAR3 may have evolved from ADAR2 (Jin et al., 2009).

The physiological importance of ADAR genes has been demonstrated. Deletion of ADAR1 is embryonic lethal in mice, and fibroblasts derived from ADAR1^{-/-} embryos are subject to stress-induced apoptosis (Wang et al., 2003). ADAR2 knockout mice showed a normal embryonic growth but died during infancy (Higuchi et al., 2000). In addition, ADAR deficiencies have been shown to play roles in several human diseases, including dyschromatosis symmetrica

hereditaria, amyotrophic lateral sclerosis (ALS) and gliomas (reviewed in: Nishikura, 2010).

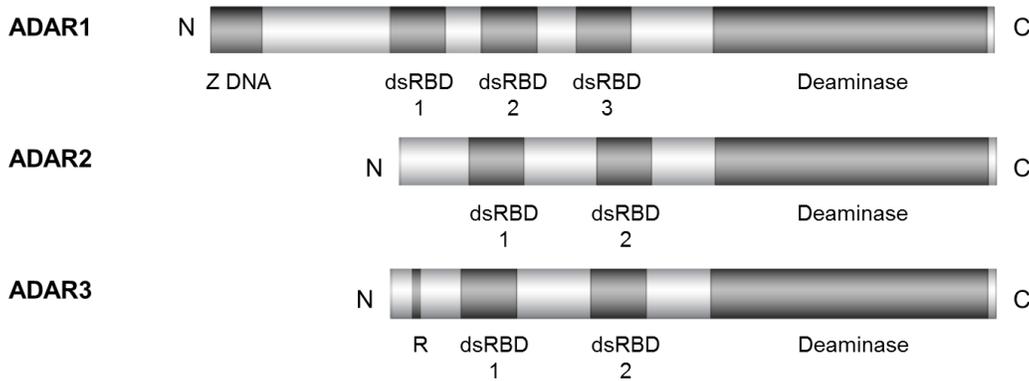


Figure I-2. Domain structure and localization of human ADAR proteins. Each ADAR contains two or three dsRBDs. The short isoform of ADAR1 (p110) has an N-terminal Z-DNA binding domain. ADAR3 has an N-terminal arginine-rich single-stranded RNA binding (R domain).

The structures of several ADAR2 domains have been reported. The C-terminal deaminase domain has a zinc ion in the active site and an inositol hexakisphosphate (IP₆) molecule buried within the enzyme core. IP₆ contributes to the protein fold and is required for the deaminase activity (Macbeth et al., 2005). The structure of the two dsRBDs of ADAR2 has also been resolved by NMR. The dsRBD1 was found to recognize a conserved pentaloop and dsRBD2 binds two bulged bases adjacent to the editing site (Stefl et al., 2006).

In vitro assays have presented evidence as to how ADAR proteins recognize the RNA substrate and perform the activity. Our lab reported the RNA-dependent dimerization of human ADAR2 on the RNA substrate, which is required for the

efficient editing activity (Jaikaran et al., 2002). In a later paper, the homodimer was also detected with both recombinant ADAR1 and ADAR2, but not ADAR3. Nonetheless, the dimerization seemed to be RNase-resistant in this case (Cho et al., 2003). The dimerization was also observed with *Drosophila* ADAR. However, when two Ala residues within dsRBD1 are mutated to Glu, the RNA binding was blocked and no dimerization was observed (Gallo et al., 2003). Taken together, the formation of homodimers of ADAR1 and ADAR2 has been confirmed by all studies, but they are in disagreement in regards to whether RNA is required for this dimerization.

In order to provide insights into the mechanisms of A-to-I editing, we decided to study *in vivo* ADAR dimerization by fluorescence resonance energy transfer (FRET). This method takes advantage of the energy transfer between cyan fluorescence protein (CFP) and yellow fluorescence protein (YFP) and has been proven to be effective in determining the intermolecular associations *in vivo* (Bastiaens et al., 1996).

I-2. Results

I-2.1. Cellular localization of ADAR1 and ADAR2

ADAR1 and ADAR2 were cloned into the mammalian expression vectors ECFP and EYFP to produce ADAR fusion proteins. N-terminal CFP-ADAR1 and CFP-ADAR2 were then individually transfected into HeLa cells. As expected,

ADAR proteins are predominantly localized in nucleolus. In addition, we found that extra-nucleolar structures formed in ADAR1 transfected cells (Figure I-3).

These extra-nucleolar structures will be discussed in detail in Appendix II.

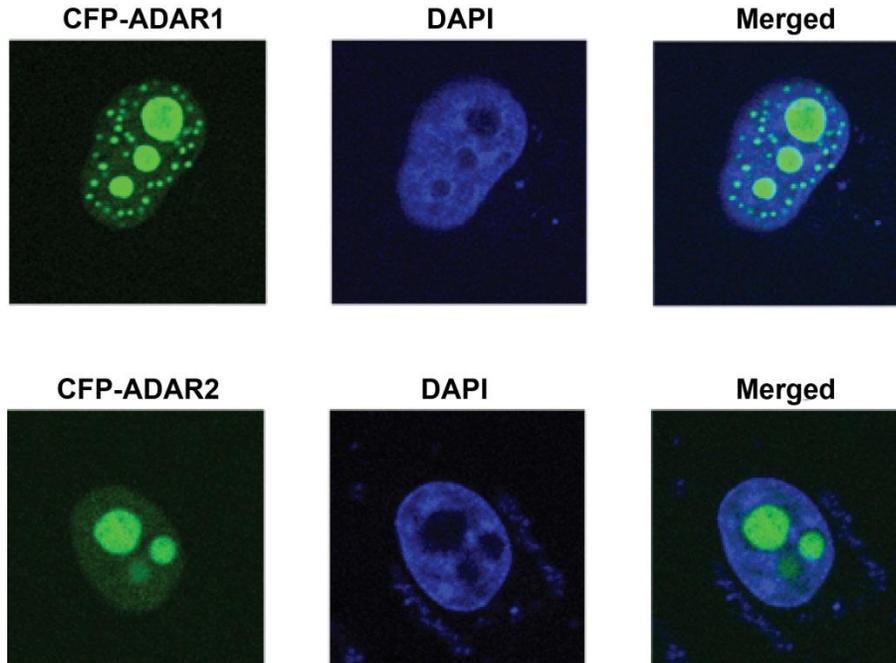


Figure I-3. Localization of human ADAR proteins. Confocal fluorescence imaging of HeLa cells transfected with CFP-ADAR1 or CFP-ADAR2 (*left*) showing DAPI imaging of the nucleus (*middle*). A merged image (*right*) indicates both ADAR1 and ADAR2 are localized to the nucleoli.

I-2.2. Homodimerization of ADAR1 and ADAR2

N-terminal CFP- and YFP-ADAR1 or ADAR2 were co-transfected into HeLa cells. For FRET assays, the CFP signal was measured at 475 nm after excitation at 458 nm. The YFP signal was measured at 600 nm, after excitation at 514 nm. YFP was then selectively photobleached at 514 nm; the increase in CFP

fluorescence corresponds to the energy transferred from CFP to YFP. The FRET efficiency can then be calculated by the difference in CFP signal intensity before and after the photobleaching.

When a plasmid expressing tandem CFP-YFP fused protein was transfected into HeLa cells, the FRET efficiency was ~35%. This corresponds to the largest FRET possibly observed. In contrast, when CFP and YFP were co-transfected, no FRET was observed (data not shown). When CFP-ADAR1 and YFP-ADAR1 expressing plasmids were co-transfected, an 18% FRET efficiency was measured. A 19% efficiency was also obtained from CFP-ADAR2 and YFP-ADAR2 co-transfection (Figure I-4; Table I-1). Thus, as assayed by FRET, both ADAR1 and ADAR2 form homodimers in the nucleolus.

Table I-1. FRET efficiency of ADARs in HeLa cells

Constructs	-RNase (%)	+RNase (%)
CFP-ADAR1 YFP-ADAR1	18 ±4	18 ±4
CFP-ADAR2 YFP-ADAR2	19 ±3	18 ±3
CFP-ADAR2 YFP-ADAR1	12 ±3	15 ±2

Since FRET signal varies with r^6 (Förster, 1948), we constructed C-terminal CFP- and YFP-ADAR1 and ADAR2 and repeated the FRET assay, trying to elucidate the disposition of ADAR monomers with respect to each other (i.e.

head-to-head vs. head-to-tail dimerization). However, no significant difference was observed between N-terminal and C-terminal constructs. Thus the FRET results are not useful to reveal the orientation of the ADAR monomers with respect to each other.

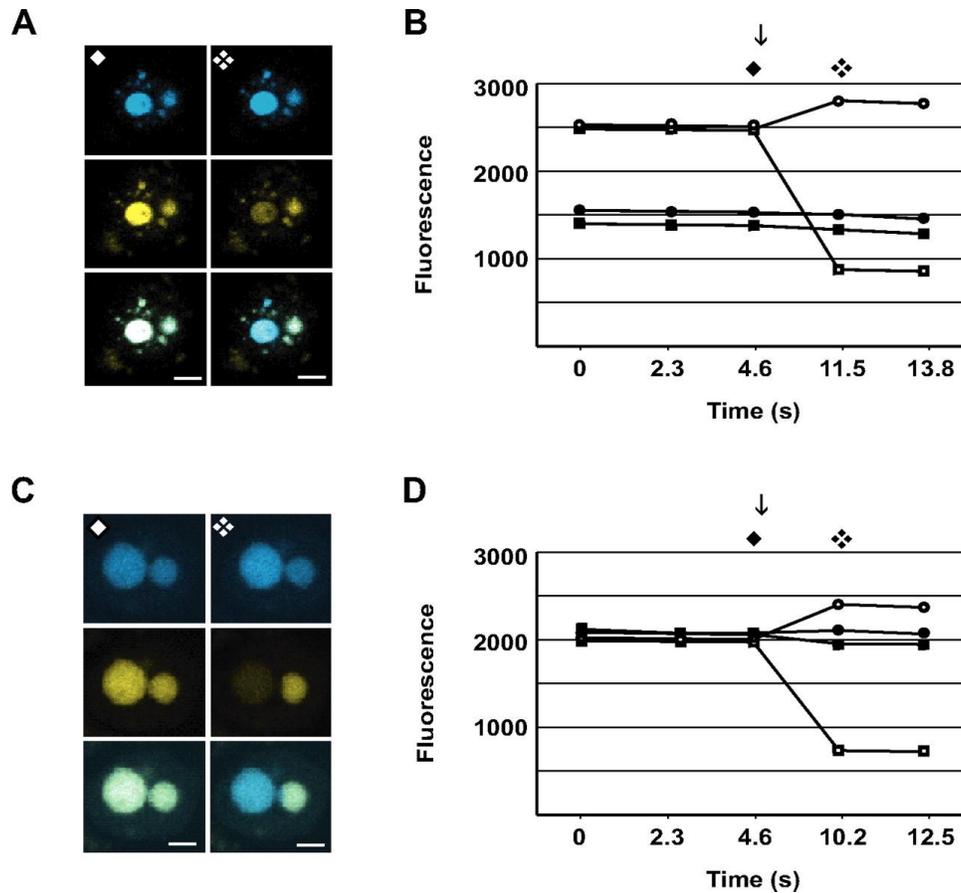


Figure I-4. Homodimerization of ADAR1 and ADAR2. A) FRET studies of HeLa cells transfected with CFP-ADAR1 and YFP-ADAR1. Shown are emissions filtered at 462-484nm (top), 580-612nm (middle) and merged images (bottom). Images were taken before (left) and after (right) photobleaching YFP within one nucleolus. Bar, 5 μ m. B) Quantification of emissions before and after photobleaching (indicated by arrow). CFP fluorescence is measured with the photobleached nucleolus (open circles) and a non-bleached nucleolus (closed circles). YFP fluorescence is measured in the same photobleached nucleolus (open squares) and non-bleached nucleolus (closed circles). C) FRET studies as in panel A but cotransfected with CFP-ADAR2 and YFP-ADAR2. D) Quantification of panel C. The time points of images in panel A and C are indicated in panel B and D (solid diamond, broken diamond)

I-2.3. Heterodimerization of ADAR1 and ADAR2

We performed further experiments to investigate the possibility of heterodimerization between ADAR1 and ADAR2. N-terminal CFP-ADAR2 was co-transfected with N-terminal YFP-ADAR1 into HeLa cells, and the FRET efficiency observed between them was 12%, which is consistent with the formation of ADAR heterodimers in the nucleolus (Figure I-5).

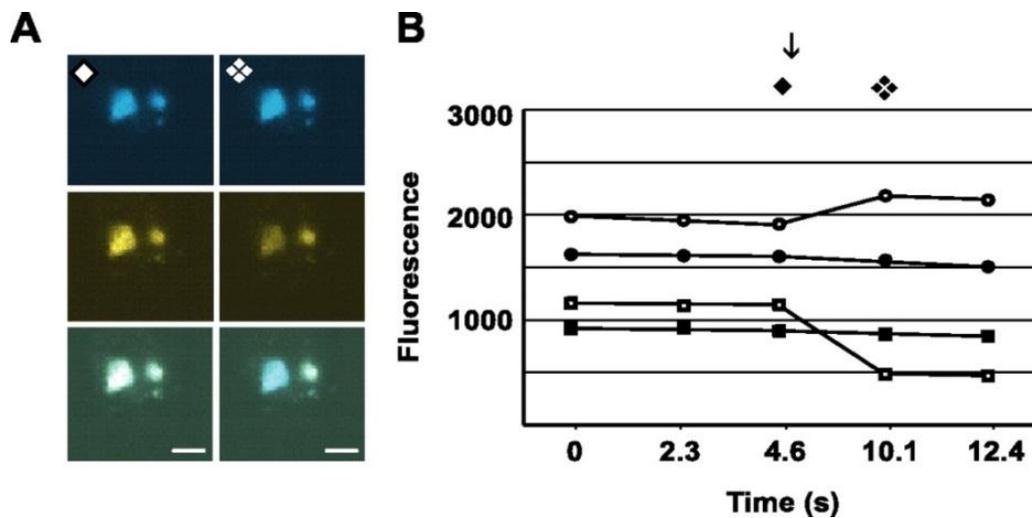


Figure I-5. Heterodimerization of ADAR1 and ADAR2. A) FRET studies of HeLa cells transfected with CFP-ADAR2 and YFP-ADAR1. Shown are emissions filtered at 462-484nm (*top*), 580-612nm (*middle*) and merged images (*bottom*). Images were taken before (*left*) and after (*right*) photobleaching YFP within one nucleolus. Bar, 5 μ m. B) Quantification of emissions before and after photobleaching (indicated by *arrow*). CFP fluorescence is measured within the photobleached nucleolus (*open circles*) and a non-bleached nucleolus (*closed circles*). YFP fluorescence is measured in the same photobleached nucleolus (*open squares*) and non-bleached nucleolus (*closed circles*). The time points of images in panel A are indicated in panel B (*solid diamond*, *broken diamond*)

We also tried the co-transfection of plasmids expressing C-terminal CFP-ADAR2 and N-terminal YFP-ADAR1. As in the previous assay, there was no

significant difference between oppositely oriented constructs. Again, it seems impossible to determine the disposition of ADAR monomers in this way.

I-2.4. RNA-independent ADAR dimerization

As mentioned above, whether ADAR dimerization requires RNA is the subject of debate. One reason studies of ADAR dimerization are complicated is that the endogenous RNA substrates for ADARs are unknown. Thus we wished to examine whether abolishing RNA binding affected the formation of homo- or heterodimers.

We repeated co-transfection experiments with plasmids expressing N-terminal CFP- and YFP-ADAR1 or ADAR2 as described above. Transfected cells were then permeabilized and treated with a mixture of RNases. Consistent with earlier studies (Sansam et al., 2003), the RNase treatment resulted in the translocation of both ADAR proteins from the nucleolus into the nucleus (Figure *I-6 A and C*). In the case of co-transfection of CFP-ADAR2 and YFP-ADAR1, both proteins were observed throughout the nucleus but they still remained co-localized (Figure *I-6 E*).

We next performed the FRET experiments on RNase-treated cells within extra-nucleolar regions of the cell. In all cases FRET efficiencies were similar in both untreated and treated cells (Figure *I-6*; Table *I-1*). These results indicate that

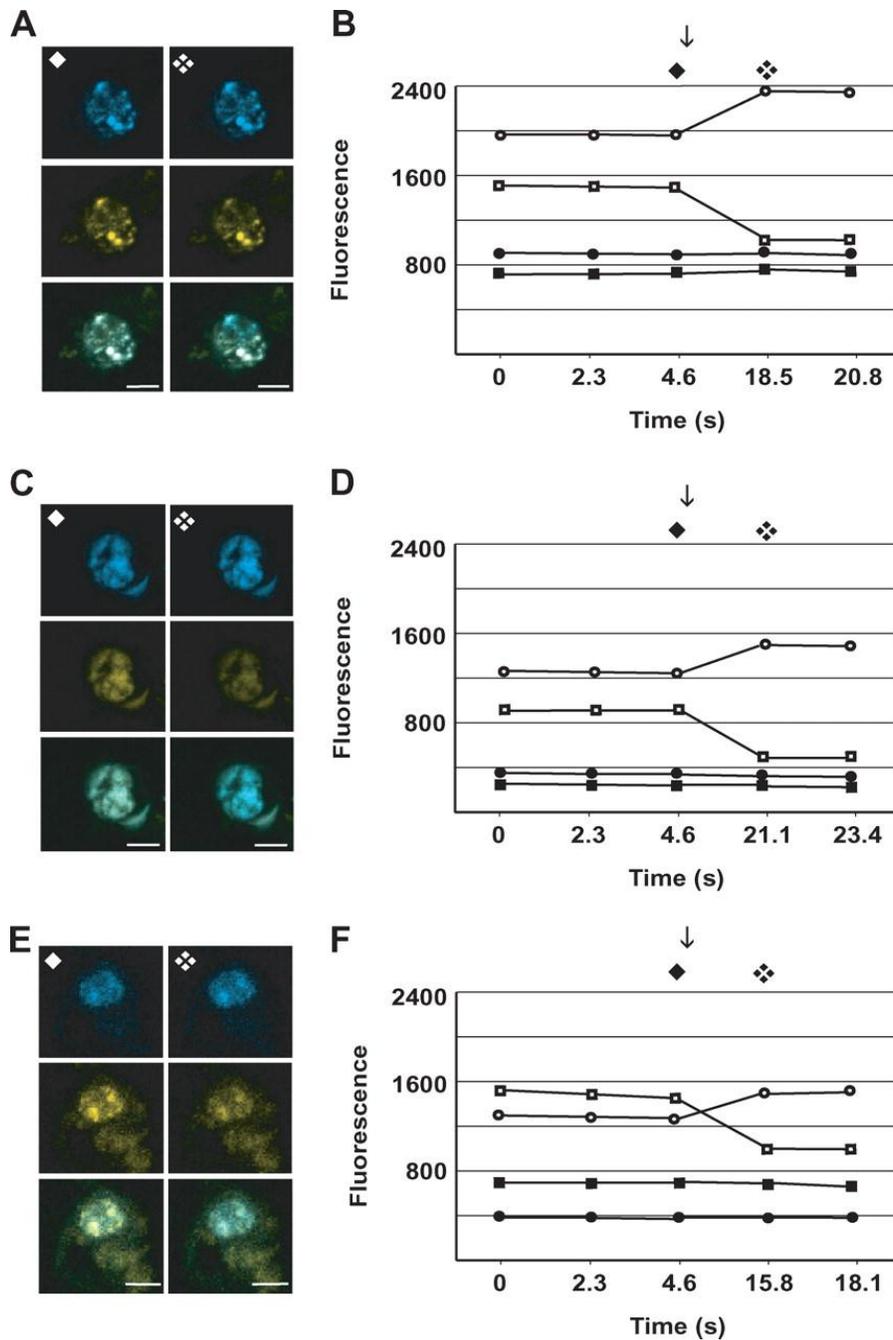


Figure I-6. RNA-independent homodimerization and heterodimerization of ADAR1 and ADAR2. A) FRET studies of HeLa cells transfected with CFP-ADAR1 and YFP-ADAR1 and then treated with RNase. Shown are emissions filtered at 462-484nm (*top*), 580-612nm (*middle*) and merged images (*bottom*). Images were taken before (*left*) and after (*right*) photobleaching YFP within one nucleolus. *Bar*, 5 μ m. B) Quantification of emissions before and after photobleaching (indicated by *arrow*). C) FRET studies as in panel A but cotransfected with CFP-ADAR2 and YFP-ADAR2. D) Quantification of panel C. E) FRET studies as in panel A but cotransfected with CFP-ADAR2 and YFP-ADAR1. F) Quantification of panel E.

neither homo- nor heterodimerization of ADAR1 and ADAR2 is dependent on RNA binding.

An earlier report has determined the minimal region for ADAR2 dimerization is the first dsRNA-binding domain (dsRBD1), and single point mutations within dsRBD1 abolish RNA-binding activity and ADAR2 dimerization (Gallo et al., 2003). Our lab also confined the ADAR2 dimerization region to aa 81-181, including dsRBD1 and part of the linker between dsRBD1 and dsRBD2 (Dul and Schellenberg, unpublished data). We performed FRET with several ADAR2 constructs to test this theory. As expected, N-terminal CFP- and YFP- ADAR2 deaminase constructs (aa 299-702) did not localize to the nucleolus (instead expressed throughout the cell) and did not undergo FRET (data not shown). In contrast, all the constructs containing the dsRBD1 showed a considerable FRET efficiency, although lower than full-length ADAR2 (Figure I-7; Table I-2). However, the point mutations within dsRBD1 (E114K/V and S141K/V) which are expected to abolish the RNA binding activity did not affect the FRET. Again, the dimerization of ADAR proteins does not appear to be RNA dependent.

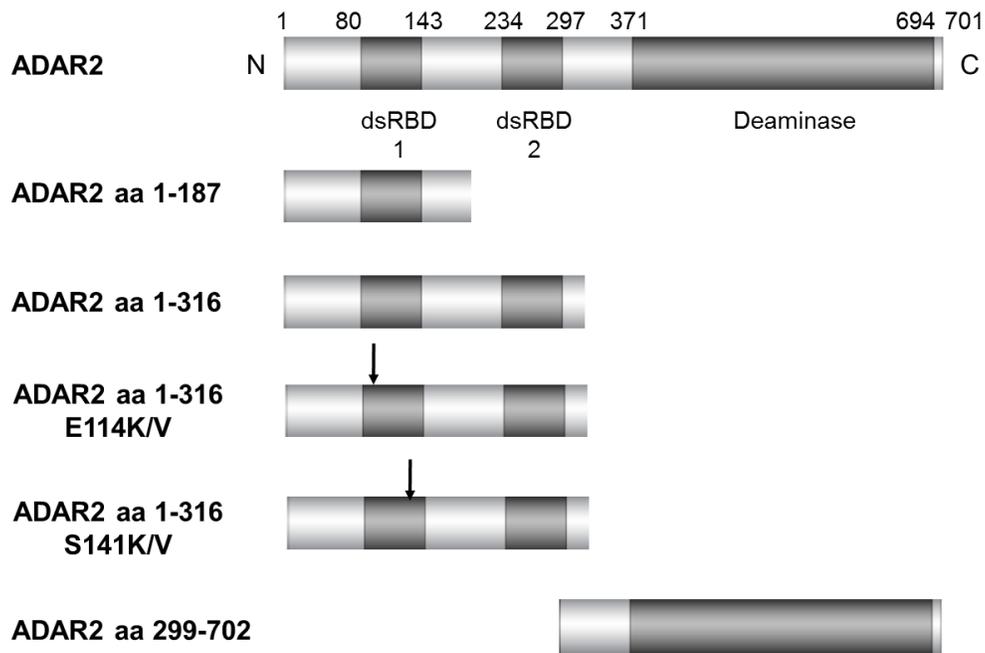
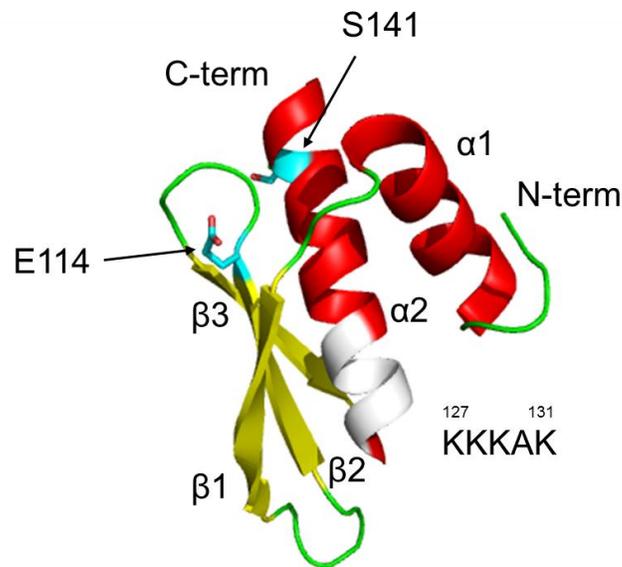
A**B**

Figure I-7. ADAR2 constructs and mutations for FRET. A) ADAR2 constructs used in this study. The point mutations in dsRBD1 are indicated with arrows. B) NMR structure of dsRBD1 with canonical α - β - β - α motif (Adapted from Stefl et al., 2006). A consensus KKKXXK motif is shown in white. The two mutated residues are indicated by arrows.

Table I-2. FRET efficiency of ADAR2 constructs

Constructs	FRET efficiency (%)
CFP-ADAR2 YFP-ADAR2	19 \pm 3
CFP-ADAR2 aa 1-187 YFP-ADAR2 aa 1-187	9 \pm 2
CFP-ADAR2 aa 1-316 YFP-ADAR2 aa 1-316	10 \pm 3
CFP-ADAR2 1-316 E114K YFP-ADAR2 1-316 E114K	9 \pm 3
CFP-ADAR2 1-316 E114V YFP-ADAR2 1-316 E114V	10 \pm 2
CFP-ADAR2 1-316 S141K YFP-ADAR2 1-316 S141K	8 \pm 2
CFP-ADAR2 1-316 S141V YFP-ADAR2 1-316 S141V	10 \pm 2

I-3. Discussion

In this report we examined the localization and dimerization of human ADAR1 and ADAR2 in HeLa cells by transient expression of fluorescent fusion proteins. Our results revealed the primary localization of both ADAR1 and ADAR2 to the nucleolus. After RNase treatment, this localization was completely abolished. Therefore, the nucleolar localization of ADARs is RNA dependent. The following FRET assays have also confirmed both ADAR1 and ADAR2 form homodimers. The formation of heterodimers between ADAR1 and ADAR2 contradicts the previous evidence (Cho et al., 2003). The observed lower FRET efficiency (12% vs. 19%) may be explained by the different distance between

CFP and YFP in heterodimers; or it may indicate this association is weaker than homodimerization. The high salt condition used in the latter study might have disrupted heterodimer formation.

The localization of ADAR1 and ADAR2 within the nucleolus has been proposed to be mediated through their binding to ribosomal RNA (rRNA) or to small nucleolar RNA (snoRNA). When rRNA synthesis is inhibited, ADAR2 is translocated into nucleoplasm (Sansam, 2003). Therefore, the ADAR proteins may be stored temporarily in the nucleolus but move into the nucleoplasm when substrate dsRNAs appear. This substrate-dependent translocation has been confirmed by the fact that when GluR-B mRNA is overexpressed, the transcript accumulated in the nucleoplasm and redirected ADARs to these non-nucleolar sites (Desterro et al., 2003). The shuttle of ADARs between subnuclear compartments may represent an important regulatory mechanism and remains to be further studied.

ADAR dimerization has been under debate for many years, and reports from different groups make this topic even more complicated. Our initial conclusion of RNA dependent dimerization was based on the observation of an RNA dependent ADAR2-ADAR2 crosslink as well as the sequential formation of monomer- and dimer-bound RNA complexes (Jaikaran et al., 2002). When compared to the results here, it is likely that the crosslink was the result of an RNA dependent rearrangement when the dimer binds to its substrate. After our result was published, a later paper came out suggesting the RNA independent dimerization

of ADAR proteins (Valente and Nishikura, 2007). In this study, the authors mutated the KKXXK motif within dsRBD domains. The RNA binding and deaminase activity were completely abolished with the mutants, while the dimer still remained intact during purification. To this date we have no definite explanation as to why ADAR proteins have displayed monomeric properties in these studies. It may be due to different purification conditions. Also, the AtoE mutations used in two previous studies may not be a good candidate since this mutation could potentially affect the hydrophobic core of the α -helix and disrupt the dsRBD structure (Gallo et al., 2003; Poulsen et al., 2006). In contrast, the Glu and Ser residues mutated in our study and the Lys residue on the KKXXK motif mutated in the other report are all on the exposed surface. Taken together, our *in vivo* FRET studies present strong evidence in support of the RNA independent manner of dimerization. It will be interesting to see more experimental evidence, both *in vivo* and *in vitro*, which may shed more light on this debated topic.

I-4. Materials and Methods

I-4.1. Mammalian Expression Constructs

Human ADAR1 p110 was amplified by PCR from pJEL/hADAR1/H6 using primers containing HindIII and BamHI restriction sites. Human ADAR2 was PCR amplified from a previously described template (Jaikaran et al., 2002) using primers containing EcoRI and SalI restriction sites. ADAR2 point mutations were

made by overlapping PCRs. ADAR1 and ADAR2 were then inserted into pEYFP-C1, pEYFP-N1, pECFP-C1, and pECFP-N1 (Clontech) using the appropriate restriction enzymes. Plasmids were transformed into chemically competent DH5 α Escherichia coli, amplified, and purified using a plasmid mini-prep kit (Sigma).

I-4.2. Cell Culture and Transfection

HeLa cells were cultured as monolayers in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum, 50 units/ml of penicillin, and 50 μ g/ml of streptomycin at 37 $^{\circ}$ C and 5% CO $_2$. Cells (2×10^5) were plated on 25-mm coverslips (Fisher Scientific) in six-well tissue culture dishes and were allowed to adhere for 24 h. The cells were transiently transfected with 1-5 μ g of plasmid DNA using Perfectin (Gene Therapy Systems) according to the manufacturer's protocol and analyzed 20-24 h post transfection. To examine the RNA dependence of localization and ADAR association, HeLa cells transfected with 1-5 μ g of plasmid DNA were permeabilized in Gal-Screen Buffer B (Applied Biosystems) for 10 min. They were then washed twice with phosphate-buffered saline (PBS) before being treated with an RNase mixture (RNase A, 200 μ g/ μ L; RNase T1, 100 units/ μ L; RNase V1, 0.5 units/ μ L) for 30 min.

I-4.3. Fluorescence Microscopy

Transiently transfected cells on coverslips were rinsed three times in phosphate-buffered saline and fixed for 15 min in freshly prepared 4% formaldehyde (Sigma) at room temperature. Coverslips were washed a further three times in phosphate-buffered saline before being mounted onto slides using Vectashield with DAPI (Vector Laboratories). Images were collected with a Zeiss laser scanning confocal microscope (LSM 510 NLO Mea) mounted on a Zeiss Axiovert 200 M inverted microscope with a $\times 40$ F-fluar lens (N.A. 1.3) equipped with four visible lasers with five laser lines and a spectral Meta detector. The 458- and 514-nm laser lines (emitted from a 25-milliwatt argon laser) were used to image CFP and YFP. Band pass filters of 462-484 and 580-612 nm were used in collecting emission from CFP and YFP, respectively.

I-4.4. Fluorescence Resonance Energy Transfer

FRET experiments were performed on fixed cells using the donor recovery after acceptor photobleach method (Bastiaens et al., 1996). First, images were obtained in the CFP and YFP channels and the intensities of the signals were calculated. YFP was then selectively photobleached at 514 nm in a defined region of the cell. A second set of images was then obtained using the same conditions as prior to photobleaching. FRET efficiency was calculated as shown in Equation 1

$$\text{FRET efficiency} = \frac{(D_{\text{post}} - B_{\text{post}}) - (D_{\text{pre}} - B_{\text{pre}})}{(D_{\text{post}} - B_{\text{post}})} \quad (\text{Eq. 1})$$

where D is the donor channel intensity, B is the background intensity, and “pre” and “post” indicate measurements before and after photobleaching. A non-bleached area of the same cell served as an internal control.

I-5. References

Bass, B.L., and Weintraub, H. (1987). A developmentally regulated activity that unwinds RNA duplexes. *Cell* **48**, 607-613.

Bastiaens, P.I., Majoul, I.V., Verveer, P.J., Soling, H.D., and Jovin, T.M. (1996). Imaging the intracellular trafficking and state of the AB5 quaternary structure of cholera toxin. *EMBO J.* **15**, 4246-4253.

Blow, M.J., Grocock, R.J., van Dongen, S., Enright, A.J., Dicks, E., Futreal, P.A., Wooster, R. and Stratton, M.R. (2006). RNA editing of human microRNAs. *Genome Biol.* **7**, R27.

Burns, C.M., Chu, H., Rueter, S.M., Hutchinson, L.K., Canton, H., Sanders-Bush, E., and Emeson, R.B. (1997). Regulation of serotonin-2C receptor G-protein coupling by RNA editing. *Nature* **387**, 303-308.

Chen, L.L., DeCerbo, J.N., and Carmichael, G.G. (2008). Alu element-mediated gene silencing. *EMBO J.* **27**, 1694-1705.

Cho, D.S., Yang, W., Lee, J.T., Shiekhatar, R., Murray, J.M., and Nishikura, K. (2003). Requirement of dimerization for RNA editing activity of adenosine deaminases acting on RNA. *J. Biol. Chem.* **278**, 17093-17102.

Desterro, J.M., Keegan, L.P., Lafarga, M., Berciano, M.T., O'Connell, M., and Carmo-Fonseca, M. (2003). Dynamic association of RNA-editing enzymes with the nucleolus. *J. Cell Sci.* **116**, 1805-1818.

Gallo, A., Keegan, L.P., Ring, G.M., and O'Connell, M.A. (2003). An ADAR that edits transcripts encoding ion channel subunits functions as a dimer. *EMBO J.* **22**,

3421-3430.

Higuchi, M., Maas, S., Single, F.N., Hartner, J., Rozov, A., Burnashev, N., Feldmeyer, D., Sprengel, R., and Seeburg, P.H. (2000). Point mutation in an AMPA receptor gene rescues lethality in mice deficient in the RNA-editing enzyme ADAR2. *Nature* **406**, 78-81.

Jaikaran, D.C., Collins, C.H., and MacMillan, A.M. (2002). Adenosine to inosine editing by ADAR2 requires formation of a ternary complex on the GluR-B R/G site. *J. Biol. Chem.* **277**, 37624-37629.

Jin, Y., Zhang, W., and Li, Q. (2009). Origins and evolution of ADAR-mediated RNA editing. *IUBMB Life* **61**, 572-578.

Levanon, E.Y., Eisenberg, E., Yelin, R., Nemzer, S., Hallegger, M., Shemesh, R., Fligelman, Z. Y., Shoshan, A., Pollock, S.R., Szybel, D., Olshansky, M., Rechavi, G., and Jantsch, M.F. (2004). Systematic identification of abundant A-to-I editing sites in the human transcriptome. *Nat. Biotechnol.* **22**, 1001-1005.

Lomeli, H., Mosbacher, J., Melcher, T., Hoyer, T., Geiger, J.R., Kuner, T., Monyer, H., Higuchi, M., Bach, A., and Seeburg, P.H. (1994). Control of kinetic properties of AMPA receptor channels by nuclear RNA editing. *Science* **266**, 1709-1713.

Maas, S. (2010). Gene regulation through RNA editing. *Discov. Med.* **10**, 379-386.

Maas, S., Rich, A., and Nishikura, K. (2003). A-to-I RNA editing: recent news and residual mysteries. *J. Biol. Chem.* **278**, 1391-1394.

Macbeth, M.R., Schubert, H.L., Vandemark, A.P., Lingam, A.T., Hill, C.P., and Bass, B.L. (2005). Inositol hexakisphosphate is bound in the ADAR2 core and required for RNA editing. *Science* **309**, 1534-1539.

Nishikura, K. (2010). Functions and regulation of RNA editing by ADAR deaminases. *Annu. Rev. Biochem.* **79**, 321-349.

Paul, M.S., and Bass, B.L. (1998). Inosine exists in mRNA at tissue-specific levels and is most abundant in brain mRNA. *EMBO J.* **17**, 1120-1127.

Polson, A.G., Bass, B.L., and Casey, J.L. (1996). RNA editing of hepatitis delta virus antigenome by dsRNA-adenosine deaminase. *Nature* **380**, 454-456.

Poulsen, H., Jorgensen, R., Heding, A., Nielsen, F.C., Bonven, B., and Egebjerg, J. (2006). Dimerization of ADAR2 is mediated by the double-stranded RNA binding

domain. *RNA* **12**, 1350-1360.

Prasanth, K.V., Prasanth, S.G., Xuan, Z., Hearn, S., Freier, S.M., Bennett, C.F., Zhang, M.Q., and Spector, D.L. (2005). Regulating gene expression through RNA nuclear retention. *Cell* **123**, 249-263.

Rueter, S.M., Dawson, T.R., and Emeson, R.B. (1999). Regulation of alternative splicing by RNA editing. *Nature* **399**, 75-80.

Sansam, C.L., Wells, K.S., and Emeson, R.B. (2003). Modulation of RNA editing by functional nucleolar sequestration of ADAR2. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 14018-14023.

Scadden, A.D. (2005). The RISC subunit Tudor-SN binds to hyper-edited double-stranded RNA and promotes its cleavage. *Nat. Struct. Mol. Biol.* **12**, 489-496.

Scadden, A.D., and Smith, C.W. (2001). RNAi is antagonized by A \rightarrow I hyper-editing. *EMBO Rep.* **2**, 1107-1111.

Seo, Y.J., Ahn, H.C., Lee, E.H., Bang, J., Kang, Y.M., Kim, H.E., Lee, Y.M., Kim, K., Choi, B.S., and Lee, J.H. (2010). Sequence discrimination of the Zalpha domain of human ADAR1 during B-Z transition of DNA duplexes. *FEBS Lett.* **584**, 4344-4350.

Sommer, B., Kohler, M., Sprengel, R., and Seeburg, P.H. (1991). RNA editing in brain controls a determinant of ion flow in glutamate-gated channels. *Cell* **67**, 11-19.

Stefl, R., Xu, M., Skrisovska, L., Emeson, R.B., and Allain, F.H. (2006). Structure and specific RNA binding of ADAR2 double-stranded RNA binding motifs. *Structure* **14**, 345-355.

Ule, J., Jensen, K.B., Ruggiu, M., Mele, A., Ule, A., and Darnell, R.B. (2003). CLIP identifies Nova-regulated RNA networks in the brain. *Science* **302**, 1212-1215.

Valente, L., and Nishikura, K. (2007). RNA binding-independent dimerization of adenosine deaminases acting on RNA and dominant negative effects of nonfunctional subunits on dimer functions. *J. Biol. Chem.* **282**, 16054-16061.

Wang, Q., Miyakoda, M., Yang, W., Khillan, J., Stachura, D.L., Weiss, M.J., and Nishikura, K. (2004). Stress-induced apoptosis associated with null mutation of ADAR1 RNA editing deaminase gene. *J. Biol. Chem.* **279**, 4952-4961.

Zinshteyn, B., and Nishikura, K. (2009). Adenosine-to-inosine RNA editing. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **1**, 202-209.

Appendix II

Characterization of ADAR1-containing extra-nucleolar bodies

The work presented in this chapter represents collaboration between authors. Tao Wu prepared the constructs and performed the fluorescence microscopy and FRAP assay. Tao Wu and Chao Liang prepared the constructs and performed the pull-down assay with the ADAR proteins. Andy Lo from Dr. Liang Li's lab performed the mass spectrometry assay.

II-1. Introduction

The vast majority of cellular genetic information in eukaryotes is stored in the nucleus, which has been shown to be both highly organized and dynamic. A prominent feature of the nucleus is the existence of discrete subnuclear regions with specialized functions termed nuclear bodies (reviewed in: Handwerger and Gall, 2006). To date, numerous nuclear bodies have been characterized and the most important ones are summarized below.

The nucleolus is a non-membrane bound structure composed of ~700 proteins and nucleic acids found within the nucleus (reviewed in: Lam et al., 2005). Ribosomal RNA (rRNA) is transcribed and assembled within the nucleolus. Another group of RNA molecules, small nucleolar RNAs (snoRNAs), are also found in the nucleolus and serve as guides for the chemical modification of rRNA (reviewed in: Scott and Ono, 2011). The localization of ADAR1 and ADAR2 in the nucleolus through competitive binding to rRNA or snoRNA was proposed to negatively regulate ADAR activity (Sansam et al., 2003). The ADAR proteins may be stored temporarily in the nucleolus but move into the nucleoplasm when substrate dsRNAs appear. The shuttle of ADARs between subnuclear compartments may represent an important regulatory mechanism.

Nuclear speckles, also called splicing speckles or speckles, are subnuclear structures that are enriched in pre-mRNA splicing factors, including snRNPs and other transcription and splicing-related proteins (reviewed in: Handwerger and

Gall, 2006). Each nucleus contains about 25-50 speckles and they appear as irregular, punctate structures and vary in size and shape (reviewed in: Mao et al., 2011). Speckles are very dynamic structures, with their protein and RNA components cycling quickly between speckles and other nuclear locations, including active transcription sites. The physical state of speckles is very sensitive to the phosphorylation of their components (reviewed in: Lamond and Spector, 2003).

Paraspeckles are relatively newly identified subnuclear bodies ~0.5-1.0 μm in diameter. The name derives from their proximity to nuclear speckles. There are 10-20 paraspeckles in each nucleus (reviewed in: Mao et al., 2011). The major components of paraspeckles are PSPs (paraspeckle proteins) and p54^{nrb}. Paraspeckles are dynamic structures that are altered in response to changes in cellular metabolic activity. They are transcription dependent and in the absence of RNA Pol II transcription they disappear (Fox et al., 2002). The role of paraspeckles is still unclear but they have been shown to be involved in nuclear retention of some A-to-I hyperedited mRNAs (reviewed in: Bond and Fox, 2009).

The promyelocytic leukaemia (PML) protein is a key organizer of PML nuclear bodies. Also called nuclear dots, they are dynamic structures 0.1-1.0 μm in diameter found in most cell-lines and many tissues. There are 1-30 PML bodies per nucleus on average, depending on the cell type and cell-cycle stage (Dellaire and Bazett-Jones, 2004). The PML protein recruits a number of other proteins. The only common feature of these recruited proteins is their sumoylation, a

protein modification process similar to ubiquitination (Bernardi and Pandolfi, 2007). The physiological function of PML bodies remain largely unknown, but they have been shown to regulate several diverse cellular functions, including DNA-damage responses, apoptosis, cellular senescence and angiogenesis (reviewed in Lallemand-Breitenbach and de Thé 2010).

The last major group of major nuclear bodies are Cajal bodies (CBs). CBs are spherical structures of 0.3-1.0 μm in diameter and they are less abundant than the previously described bodies with less than 10 in each nucleus. They occur both free in the nucleoplasm and physically associated with histone and snRNA gene loci (reviewed in Handwerger and Gall, 2006). The marker components of CBs are the proteins Coilin/p80 and SMN (survival of motor neuron). CBs have been implicated in important RNA-related metabolic processes such as snRNA and snoRNA biogenesis and modification, snRNP and snoRNP assembly and trafficking, histone mRNA processing and telomere maintenance (reviewed in Morries, 2008).

It has been reported that overexpression of exogenous ADAR1 leads to the appearance of discrete bodies in the nucleoplasm, which are reminiscent of nuclear speckles (Desterro et al., 2003). However, double-labeling experiments using antibodies against splicing factors show that these structures are distinct from the nuclear speckles where spliceosomal components accumulate; the authors concluded that these structures represented protein aggregates. Since we also observed the presence of these structures during our FRET studies and they

are exclusive to ADAR1 expression, we decided to further analyze them. We also used pull-down assays to investigate the ADAR-associated proteins and potential ADAR-containing complexes.

II-2. Results

II-2.1. ADAR1 specific extra-nucleolar bodies

In Appendix I we reported the observance of the extra-nucleolar structures when ADAR1 is overexpressed *in vivo* (Figure I-3). Interestingly, when ADAR2 was expressed alone, it was exclusively localized in the nucleolus (Figure I-3; Figure II-1 A). However, when fluorescently labelled ADAR1 and ADAR2 were both over-expressed, ADAR2 was also observed in those ADAR1 specific structures (Figure II-1 A). This is consistent with the heterodimerization between ADAR proteins.

To exclude the possibility that these structures are not tiny nascent nucleoli, I tried the co-expression of ADAR1 with nucleolin, a prominent component of nucleolus (Bugler et al., 1982). As expected, nucleolin signals are focused in three to four foci which do not stain with DAPI, consistent with the average amount of nucleoli within cells, while ADAR1 signals stand alone in the extra-nucleolar regions (Figure II-1 B). Thus these structures are not newly formed nucleoli.

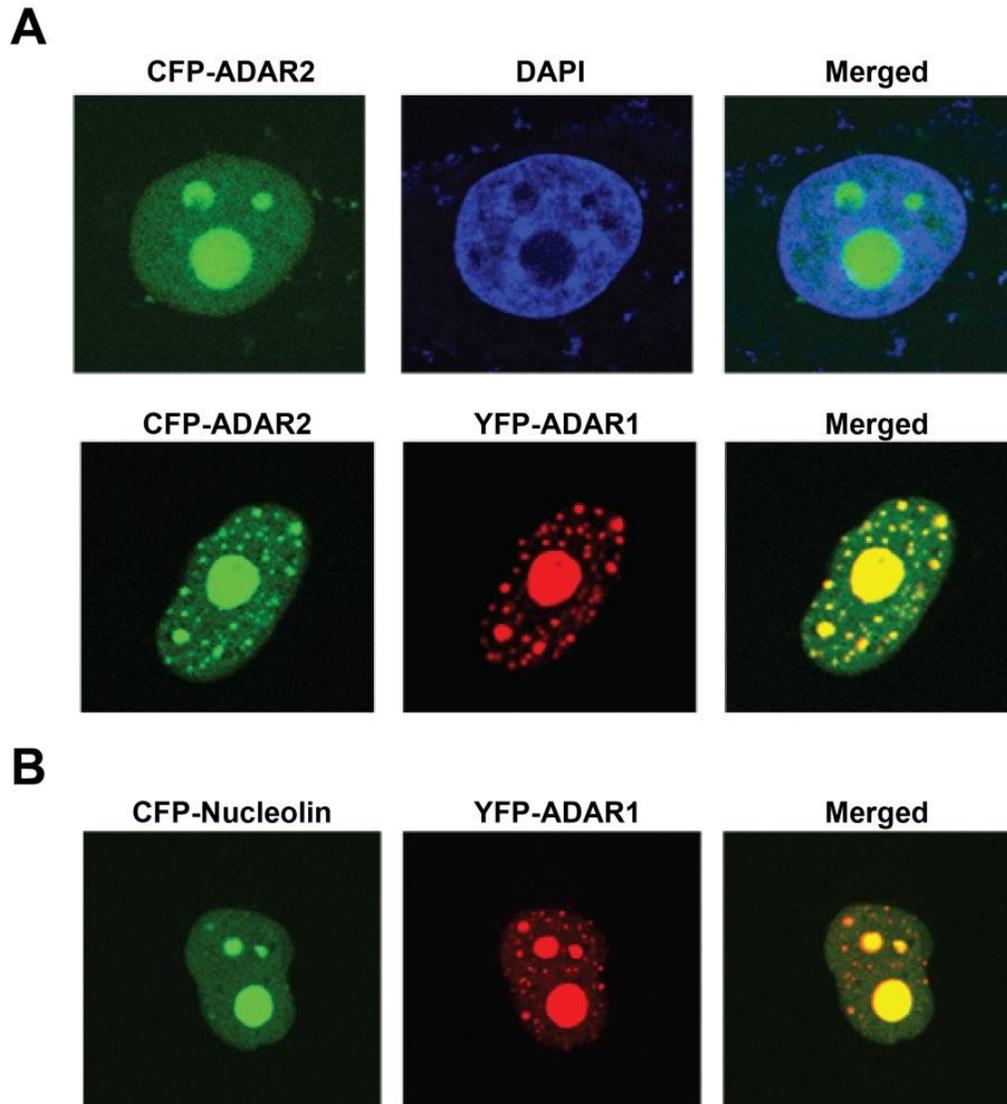


Figure II-1. ADAR1-specific extra-nucleolar bodies. A) ADAR2 is localized in nucleoplasm but is pulled into the extra-nucleolar bodies by ADAR1. B) ADAR1 co-localizes with nucleolin in nucleolus, but not in the extra-nucleolar bodies.

From now on these ADAR1 specific structures will be referred to as AdB (ADAR1 bodies) for convenience.

II-2.2. FRAP of AdB

Since the ADAR1 protein was over-expressed *in vivo* for fluorescence studies, the first question is whether the non-nucleolar structures are due to the protein aggregation. In order to answer this question, we performed a FRAP (fluorescence recovery after photobleaching) experiment within ADAR1 over-expressed HeLa cells.

FRAP has been proven to be useful in assessing protein solubility *in vivo* (Phair and Misteli, 2000; Desterro et al., 2003). The ability to analyze the mobility of individual proteins is very useful in studying the dynamics of nuclear proteins (Carrero et al., 2002; Houtsmuller, 2005). Here we expressed CFP-ADAR1 in HeLa cells. The nucleolar region or the extra-nucleolar structures were then photobleached and recovery of CFP signals in the photobleached region was recorded. Based on the quantification in the FRAP experiment, there is no significant difference in recovery between the nucleolus and the extra-nucleolar bodies, suggesting that there is exchange of ADAR1 proteins between the AdBs and that the ADAR1 in those bodies is highly mobile (Figure *II-2*). We concluded that the AdBs are unlikely to be protein aggregates. Instead, they probably represent nuclear organelles where endogenous ADAR1 proteins carry out physiological functions.

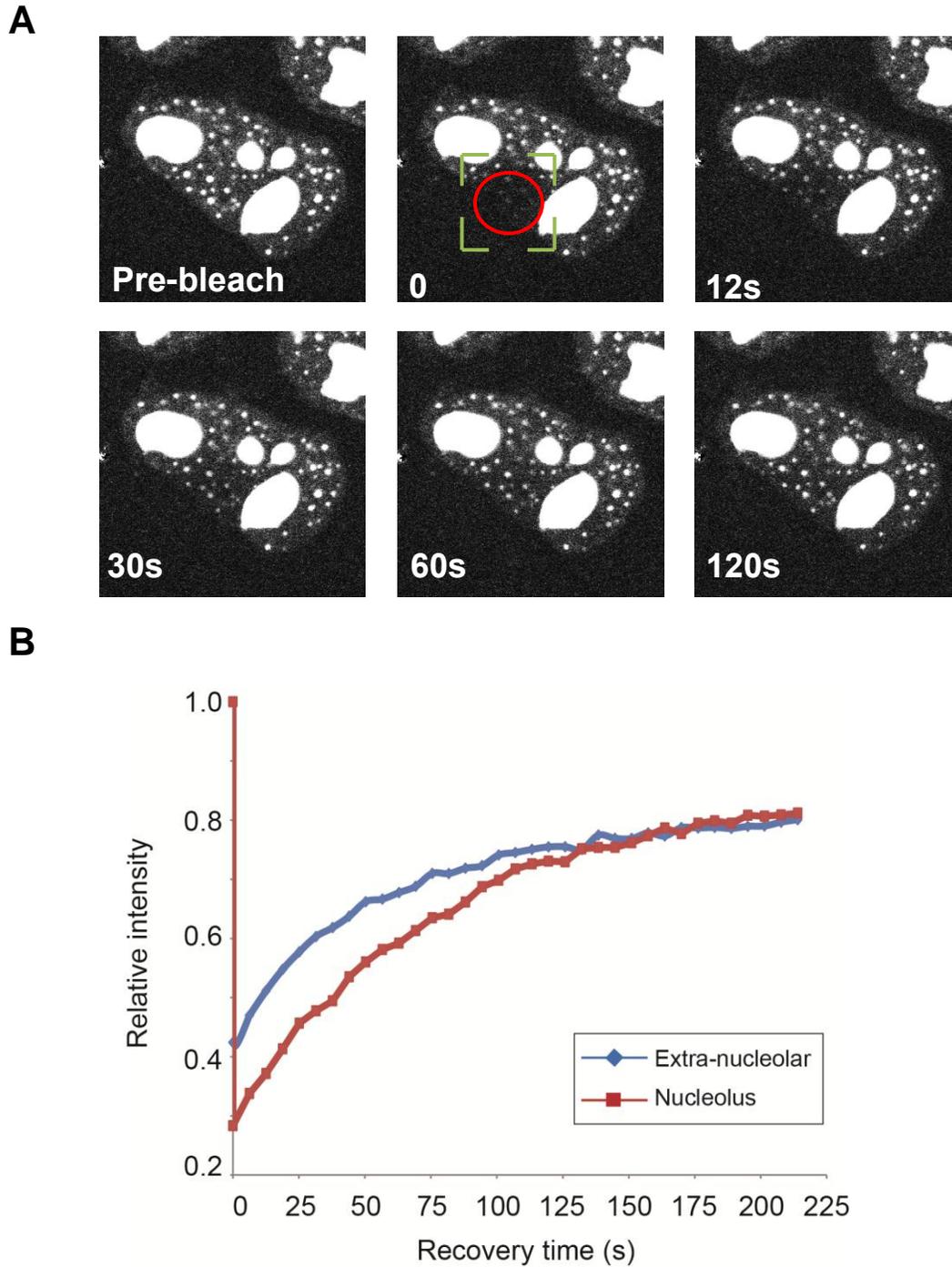


Figure II-2. FRAP of AdBs. A) Images of CFP-ADAR1 expressing cells before and during recovery after bleaching of an extra-nucleolar region. B) Kinetics of recovery after bleaching of an extra-nucleolar region (blue) and a nucleolus region (red).

II-2.3. AdBs are not known major nuclear bodies

In order to determine whether the AdBs correspond to any of the well-studied nuclear bodies described above, we co-expressed ADAR1 with major components of these bodies and determined their localization. In each assay, ADAR1 was expressed as a YFP-fusion and the nuclear body component was fused to CFP. Previous work has shown ADAR1 does not co-localize with splicing speckles (Desterro et al., 2003). Thus we checked whether ADAR1 is localized to other major nuclear bodies.

Three marker proteins were cloned and expressed for this purpose. p54^{nrb}, coilin/p80 and PML have been shown to be hallmark proteins of paraspeckles, Cajal bodies and PML bodies respectively (reviewed in: Handwerger and Gall, 2006). In the event, co-localization of ADAR1 was not observed with any of these three proteins (Figure *II-3 A*). FRET was also carried out in some areas where signals of two proteins seemed to overlap, but only a background level of FRET efficiency was observed (data not shown).

Several findings suggest a tight co-transcriptional coupling of splicing and A-to-I editing (Raitskin et al., 2001; Schoft et al., 2007). For example, ADAR2 edits its own mRNA, creating a new 3'SS. This process is co-transcriptional and the C-terminal domain (CTD) of the large subunit of RNA polymerase II (pol II) is essential for efficient co-transcriptional auto-editing (Laurencikiene et al., 2006). In order to examine whether the AdBs were RNA pol II dependent

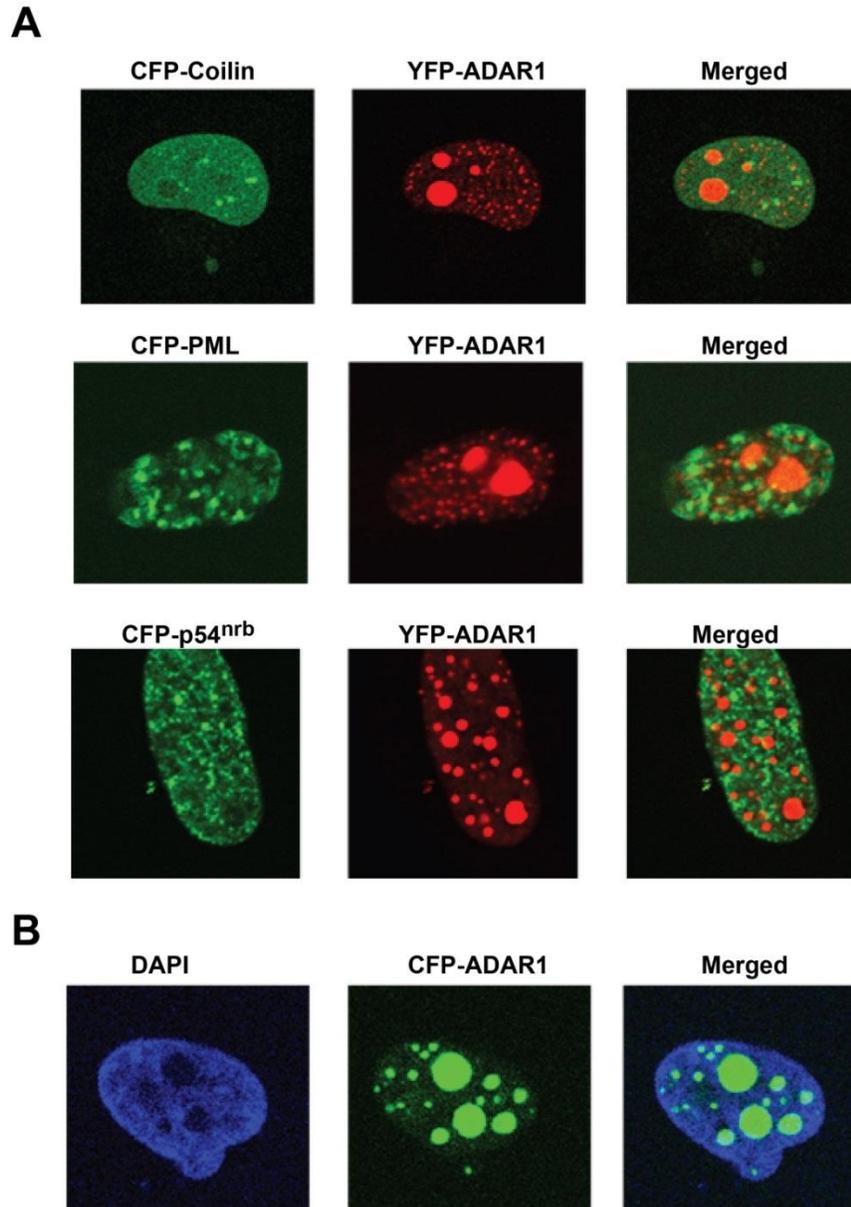


Figure II-3. AdBs do not correspond to any known structure. A) Co-expression of ADAR1 with components of Cajal bodies (*upper*), PML bodies (*middle*) and paraspeckles (*bottom*). B) AdBs do not disappear after α -amanitin treatment.

transcription sites, HeLa cells were treated with α -amanitin, an effective inhibitor of RNA pol II (Dundr et al., 1995; Gong et al., 2004). After being incubated in α -amanitin for five hours, the AdBs were still observed with no change in

morphology and amount (Figure II-3 B). Therefore these structures do not correspond to co-transcriptional editing sites.

II-2.4. Isolation of ADAR-containing complexes

In order to further investigate ADAR1-containing bodies, we carried out pull-down assays using both FLAG-tagged ADAR1 and ADAR2 proteins. The same technique has been used to search for partners of RNA-processing enzymes (Gregory et al., 2004; Colmenares et al., 2007). This was accomplished by developing HEK-293-derived stable cell lines expressing FLAG-tagged ADAR1 and ADAR2. FLAG-ADARs were isolated by immunoaffinity chromatography. The eluate contained a rich harvest of proteins. ADAR1 and ADAR2 were identified by SDS-PAGE gel and also confirmed by western blotting (Figure II-4 A). The ADAR1 and ADAR2 mediated editing activity within the pulldowns was also tested with a 78-nucleotide RNA containing the R/G editing site of the human GluR-B pre-mRNA (Jaikaran et al., 2002). Both pulldowns showed equivalent editing activity (~10% efficiency; Figure II-4 B).

The affinity-purified samples were subjected to trypsin digestion and LC-MALDI analysis. This identified numerous proteins and only those with more than one peptide identified are listed here (Table II-1). Not surprisingly, the majority of the proteins pulled down by ADARs are ribosomal proteins due to the interaction between ADARs and rRNA. Another major group identified is the

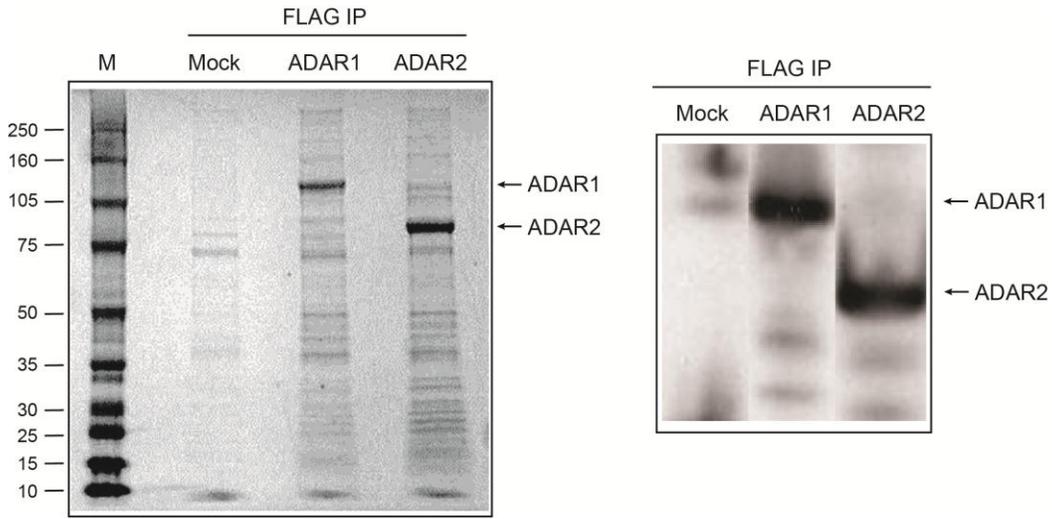
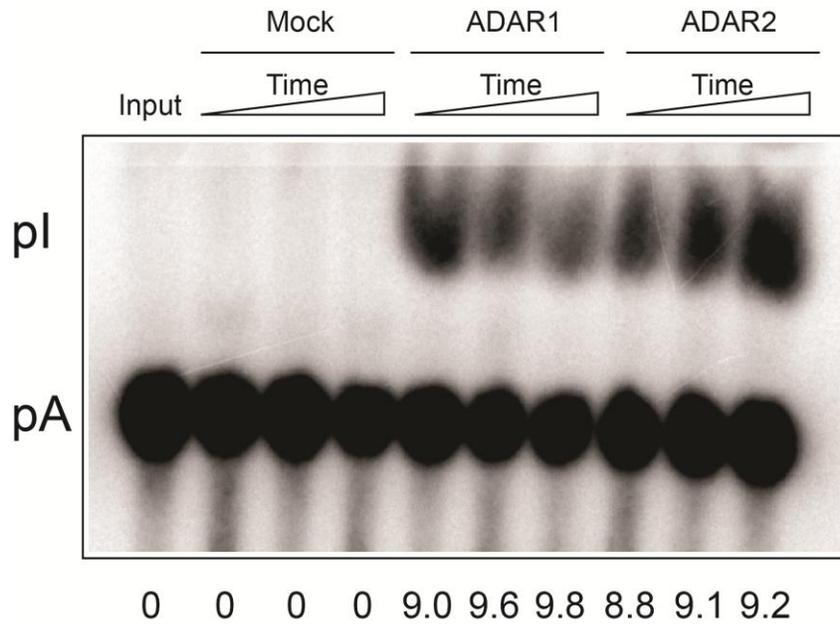
A**B**

Figure II-4. Purification of ADAR1- and ADAR2-containing complexes. A) Immunoaffinity elution from anti-FLAG beads resolved by SDS-PAGE. FLAG-ADARs were shown by silver staining (*left*) and western blotting with anti-FLAG antibodies (*right*). B) Editing reactions with FLAG-tagged ADAR1 and ADAR2 using a 78-nucleotide RNA containing the R/G site internally labeled. The editing efficiency is shown below.

Table II-1. Proteins pulled down by ADAR1 and ADAR2

Samples	Proteins identified		Number of peptides identified
ADAR1 & ADAR2	Heat shock proteins	Hsp70	9
		Hsp90	4
	Nuclear factor of activated T-cells (NFAT)	ILF3	7
		ILF2	3
	Ribosomal proteins	60S ribosomal protein P2	3
		60S ribosomal protein L7a	3
		60S ribosomal protein L18	3
		40S ribosomal protein S3	3
		60S ribosomal protein L7	2
		60S ribosomal protein L3	2
	Heterogeneous nuclear ribonucleoprotein (hnRNPs)	hnRNP C1/C2	5
		hnRNP U	3
		hnRNP F	2
		hnRNP A2/B1	2
		hnRNP R	2
	Other proteins	RNA helicase A	3
Nucleolin		3	
HuR		2	
SRP9		1	
ADAR1	Histones	Histone H 1.3	2
		Histone H2BF	1
		Histone H2AX	1
	hnRNPs	hnRNP Q	2
		hnRNP G	2
ADAR2	Histones	Histone H1.2	1
	Other proteins	Vimentin	3
		Nucleolar RNA helicase 2	1
		Nucleophosmin	1

heterogeneous nuclear ribonucleoproteins (hnRNPs) family. hnRNPs are complexes of RNA and proteins present in the cell nucleus, which bind to mRNA and the splicing apparatus. Among other functions, they are involved in preventing folding of pre-mRNA into secondary structures that may inhibit its

interactions with other proteins and transportation of mature mRNA out of the nucleus (reviewed in Krecic and Swanson, 1999). hnRNP-L has been shown to bind to the hepatitis δ virus RNA, which is also a substrate for A-to-I editing (Sikora et al., 2009). Therefore the interaction between ADARs and hnRNPs may be transient and RNA-dependent. We checked the co-expression of hnRNP-F with ADAR1 and they did not co-localize with each other (Figure II-5). We observed the interaction of several histone proteins with ADAR1 in the MS experiment, as well as the RNA helicase A (RHA). This is consistent with a previous report of inosine-rich RNA (I-RNA) pull-down assays (Wang et al., 2005). In that study, another protein, Vigilin, was also detected and it was proposed to form a complex with ADAR and RHA. In the presence of the edited RNA, the Vigilin complex recruits the DNA-PKcs enzyme, which phosphorylates a set of targets to participate in heterochromatin formation (reviewed in Fernandex et al., 2005). Interestingly, we did not detect Vigilin in our pull-down assay. Also the co-expression of ADAR1 with H2AX did not show co-localization (Figure II-5). Therefore, ADAR may have been an artifact in the Vigilin complex, or our tag may interfere with the interaction between ADAR and Vigilin.

Several other possible partners for ADARs identified in the pull-down assay were also tested for interaction in co-expression studies. Hu-antigen R (HuR) contains three RNA-binding domains and binds cis-acting AU-rich elements. It destabilizes mRNAs and regulates gene expression (Masuda et al., 2011). Another RNA-associated protein, signal recognition particle 9 (SRP9), forms a

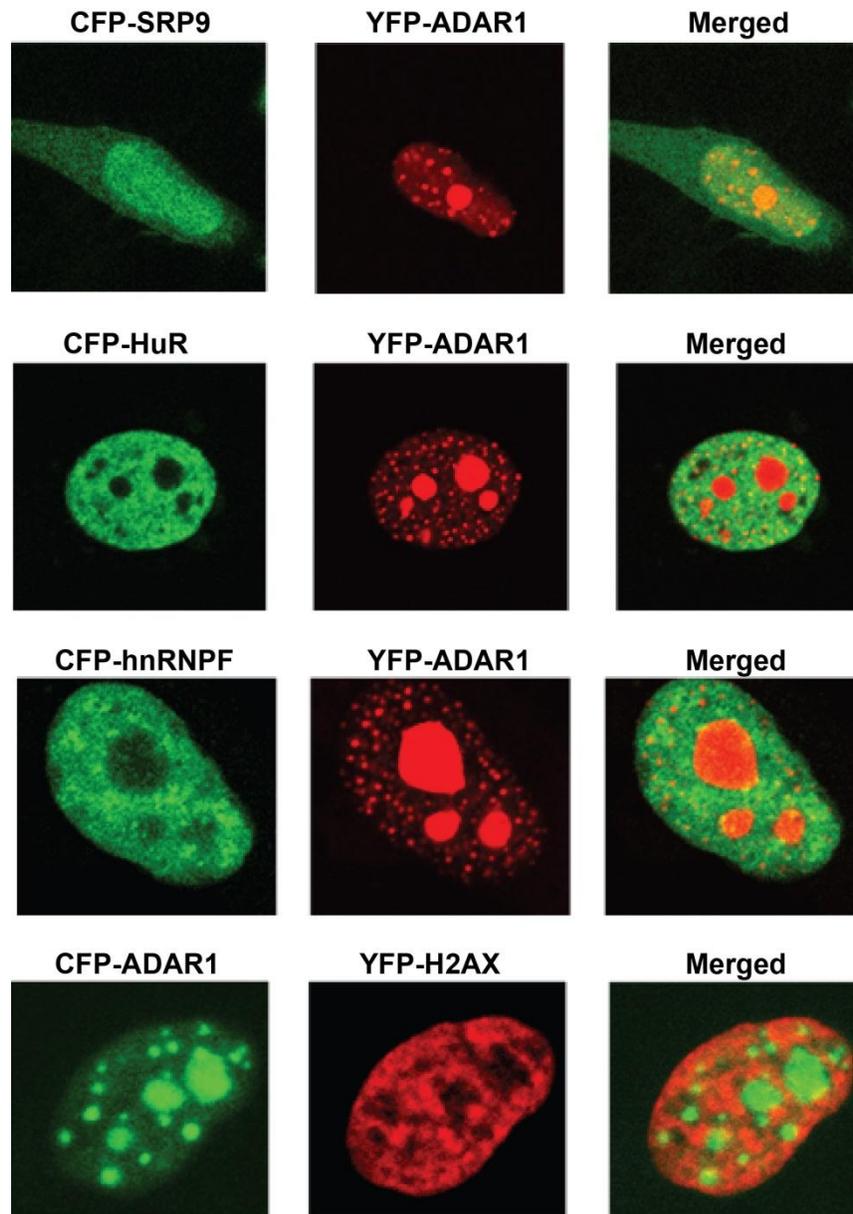


Figure II-5. Proteins pulled down by ADAR1 do not co-localize with ADAR1. Four proteins pulled down by ADAR1 were co-expressed with ADAR1 and the localization was shown.

heterodimer with SRP14 and binds Alu elements of RNA and regulates mRNA translation (Häsler and Strub, 2006). When co-expressed with ADAR1, neither of these showed co-localization, indicating their interaction with ADAR1 is indirect

and perhaps RNA-dependent (Figure II-5). In addition, we observed both subunits of the nuclear factor of activated T-cells (NFAT) – ILF2 and ILF3 in the pull-down assay. More than three peptides from each protein are detected, suggesting their association with ADARs. NFAT transcription factors play a central role in initiating T-cell activation through the induction of immediate-early T-cell specific genes including interleukin-2 (IL-2). The larger subunit ILF3, also known as NF90, contains two dsRBDs and has been shown to be capable of binding rRNAs (Langland et al., 1999). When expressed alone, ILF2 was found throughout the nucleus. When ADAR1 was co-expressed with ILF2, however, several co-localization foci appeared (Figure II-6). These foci include the

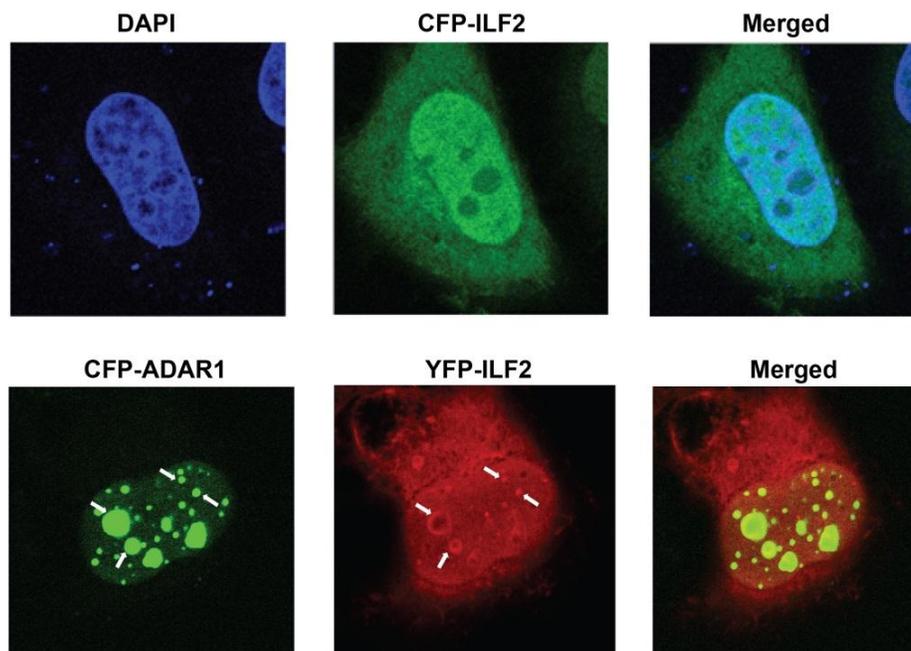


Figure II-6. Association of ADAR1 with ILF2. When expressed alone, the CFP-ILF2 signal is throughout the nucleus (*upper*). When ILF2 is co-expressed with ADAR1, ILF2 is translocated into nucleolus and extra-nucleolar regions (*bottom*; indicated by arrows).

periphery region of nucleolus, as well as some, if not all, of the ADBs. This confirmed the interaction between ADAR1 and NFAT.

We also carried out the glycerol gradient separation of ADAR1-complexes. This technique has been used in the separation of ribosomal subunits, as well as numerous other complexes (Bousquet-Antonelli et al., 2000). The majority of ADAR1 was found to be within a ~200 kD complex, consistent with the size of an ADAR1 dimer (Figure II-7). Therefore, it is likely that the ADAR1-specific structures contain only ADAR1 protein and all the other proteins identified in the pull-down assay only form indirect and RNA-dependent association with ADAR1.

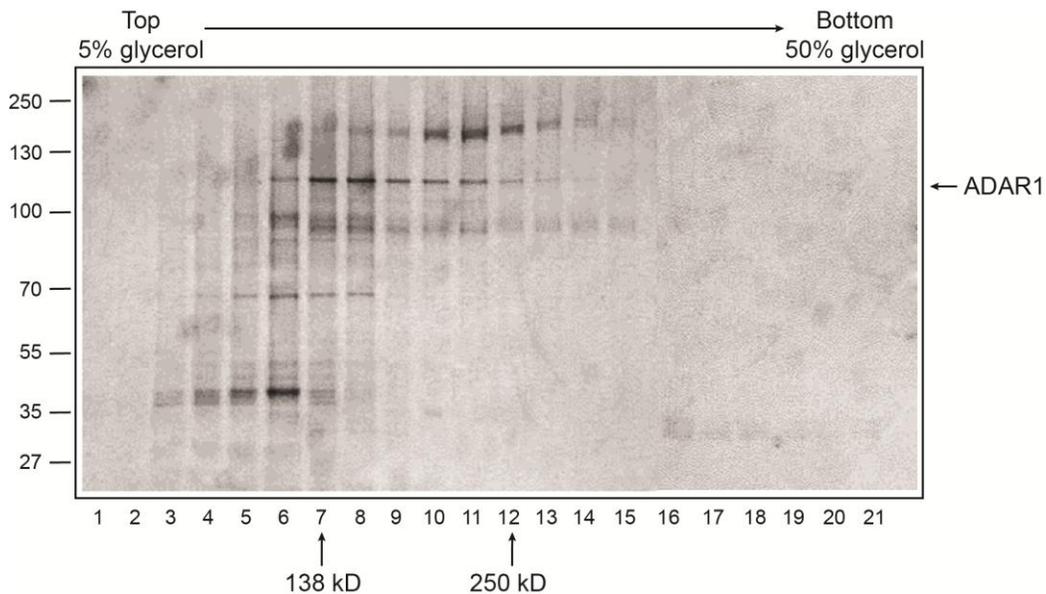


Figure II-7. Glycerol gradient separation of ADAR1-containing complexes. HEK 293 cell lysate was loaded onto a 5–50% glycerol gradient and fractionated by ultracentrifugation. Fractions collected were then revealed by SDS-PAGE and visualized by silver staining.

II-3. Discussion

In this study we tried to characterize the ADAR1 specific extra-nucleolar bodies. Four major nuclear bodies were checked, as well as the pol II transcription site. We did not investigate other nuclear bodies, including perinucleolar compartments, polycomb bodies, nuclear stress granules (nSG) and clastosomes (reviewed in Map et al., 2011). Based on our pull-down results, none of the components of other known nuclear bodies were detected. Therefore it is unlikely that the AdBs belong to any known nuclear organelle.

Although the florescence signal from the AdBs showed a similar recovery rate to nucleoli in our FRAP studies; we are still not able to rule out the possibility of these structures being protein aggregates. Although that the majority of them look tiny, a few large spherical bodies exist in most cells. This is reminiscent of well-studied nuclear aggresomes, formed by poly-glutamine (poly-Q) containing proteins. These aggresomes contain microtubules, proteasome and molecular chaperones. FRAP experiments in these bodies showed a constant exchange between aggregates and nucleoplasm, and these bodies were found to be associated with PML bodies (Fu et al., 2005). Considering that we did not observe co-localization between ADAR1 and PML bodies, and that none of the characteristic components of aggresomes were found in our pull-down assays, the ADAR1-specific bodies do not seem to correlate with aggresomes. Another reported example of nuclear protein aggregates is GFP-tagged histone deacetylase (HDAC4). When overexpressed in mouse cells, numerous spherical domains are

scattered throughout the nucleoplasm (Carrero et al., 2002). The authors did not perform the FRAP assays within these domains. Similar structures have also been observed in other reports, their formation depends on the interaction between HDAC4 and myocyte enhancer binding factor 2 (MEF2) proteins (Borghi et al., 2001). HDAC4 was shown to repress the MEF2 activity in nucleus (Miska et al., 1999). Therefore, even if the HDAC4 and AdBs are protein aggregates, they still may represent the functional site where endogenous enzymes carry out physiological roles.

The pull-down assays with ADARs yielded some interesting results. Although we documented numerous proteins, none is specific to ADAR1 except for some hnRNP and histone proteins. In addition, all the proteins identified can be found in the nucleolus (Andersen et al., 2002), and our co-expression experiments indicated none of the proteins on the list were co-localized with ADAR1, except for NFAT subunits. The glycerol gradient separation also suggested that the majority of ADAR1, if not all, exists as a homodimer *in vivo*. Hence the possibility is either that the AdBs only account for a very low percentage of ADAR1 protein or ADAR1 is the only protein component within those structures. The association between ADAR1 and NFAT requires further study.

The nature of the AdBs remains elusive. Shortly after we finished this study, a paper was published regarding the localization of DGCR8, the partner of Drosha (Shiohama et al., 2007). Surprisingly, DGCR8 showed a very similar localization

to ADAR1, and the pull-down assay with DGCR8 yielded almost identical results as that with the ADAR proteins. Upon incubation with α -amanitin, however, DGCR8 translocated from the nucleolus into the extra-nucleolar bodies, which differs from our α -amanitin treatment result. The authors assumed the extra-nucleolar foci represented the site where DGCR8 binds its pri-miRNA substrate immediately after splicing occurs. The complex is then transferred into the nucleolus, where the pri-miRNA is processed into pre-miRNA. It is then passed to DGCR8/ILF3 complex and ILF3 shuttles the pre-miRNA to Exportin 5 (XPO5), which moves it into the cytosol. This model requires more evidence. If it is true, ADAR1 and DGCR8 may share a common mechanism for processing RNA molecules within the nucleus, even though their substrates are different. Although we cannot elucidate the characteristic of the ADBs, our effort provides some initial evidence for future studies.

II-4. Materials and Methods

II-4.1. Mammalian Expression Constructs

All genes were amplified from human total RNA (Clontech). Genes used for fluorescence studies were inserted into pEYFP-C1, pEYFP-N1, pECFP-C1, and pECFP-N1 (Clontech) using the appropriate restriction enzymes. ADAR1 and ADAR2 genes used for pull-down assays were cloned into p3 \times FLAG-CMV-10 (Sigma).

II-4.2. Cell Culture and Transfection

Cells culture and transfections were carried out with the same procedures as described in Appendix I. For the α -amanitin treatment assays, HeLa cells were incubated with 40 $\mu\text{g}/\text{mL}$ α -amanitin for 5 hrs before microscopy.

II-4.3. Fluorescence Microscopy

Fluorescence microscopy was carried out with the same procedures as outlined in Appendix I.

II-4.4. Fluorescence Recovery After Photobleaching

Live cells were imaged at 37 °C maintained by a heating/cooling frame (LaCon, Germany) in conjunction with an objective heater (PeCon, Germany). Images were collected with a Zeiss laser scanning confocal microscope (LSM 510 NLO Mea) mounted on a Zeiss Axiovert 200 M inverted microscope with a $\times 40$ F-fluar lens (N.A. 1.3) equipped with four visible lasers with five laser lines and a spectral Meta detector. Each FRAP analysis started with five image scans, followed by a bleach pulse of 0.5 seconds on a region using the 458 nm laser line at 100% laser power. Single section images were collected at 6.3 second intervals. For imaging, the laser power was attenuated to 0.1-0.2% of the bleach intensity. The average fluorescence in the nucleus $T(t)$ and the average fluorescence in the

bleached region $I(t)$ were calculated for each background subtracted image at time t after bleaching. FRAP recovery curves were normalized according to Phair and Misteli (Phair and Misteli, 2000). The fluorescence signal measured in a region of interest normalized to the change in total fluorescence was determined as:

$$I_{rel} = \frac{T_0 I_t}{T_t I_0}$$

where T_0 is total cellular intensity during prebleach, T_t the total cellular intensity at timepoint t , I_0 the average intensity in the region of interest during pre-bleach, I_t the average intensity in the region of interest in timepoint t .

II-4.5. Affinity purification of FLAG-ADAR1 and FLAG-ADAR2

Full-length ADAR1 and ADAR2 genes were cloned into p3×FLAG-CMVTM-10 expression vector (Sigma-aldrich) and transfected into HEK-293 human embryonic kidney cells. Transfected cells were grown in the presence of 2.5 µg/mL G418, and individual colonies were isolated and analysed for FLAG-ADARs expression. To purify the ADAR1 and ADAR2 complex, nuclear extract generated from 100 15-cm plates (4×10^9 cells or ~150 mg of nuclear extract) was incubated with anti-FLAG M2 affinity resin (Sigma). After two washes with buffer A (20 mM Tris-HCl pH 7.9, 0.5 M KCl, 10% glycerol, 1 mM EDTA, 5 mM dithiothreitol, 0.5% NP40, 0.2 mM phenylmethylsulphonyl fluoride), the affinity column was eluted with buffer A containing FLAG peptide (400 µg/mL)

in accordance with the manufacturer's instructions (Sigma). The samples were then subjected to trypsin digestion and LC-MALDI analysis (in collaboration with Dr. Li's lab, Department of Chemistry, University of Alberta).

For the glycerol gradient separation, 200 μ L of HEK 293 cell lysate was loaded on a 5-50% (v/v) glycerol density gradient (10 mL) prepared in a buffer containing 20 mM Tris-HCl pH 7.9, 60 mM KCl, 1mM EDTA, 5 mM dithiothreitol and 0.2 mM phenylmethylsulphonyl fluoride and centrifuged at 160,000 g for 18 h at 4 $^{\circ}$ C. 20 fractions were then collected and acetone precipitated. The proteins within each fraction were resolved by 4%-12% Bis-Tris SDS-PAGE gel (Invitrogen).

II-4.6. Editing reactions

R/G editing substrates were synthesized by T7 RNA polymerase transcription using templates generated by PCR from a substrate cloned into pBS (Stratagene). The internal label was introduced by the ligation of upstream and ³²P-end-labeled, ApG-primed, downstream T7 transcription products in the presence of a bridging oligonucleotide using T4 DNA ligase (Jaikaran et al., 2003). Editing reactions contain 20 mM Hepes pH 8.0, 100 mM KCl, 0.5 mM DTT, 20% glycerol, 0.01% Nonidet P-40, RNasin (1 unit/ μ L, invitrogen), 50 nM FLAG-tagged ADAR1 or ADAR2 and 0.5 nM RNA. The substrate RNA was denatured at 80 $^{\circ}$ C for 2 min and allowed to renature at 30 $^{\circ}$ C for 10 min just prior

to addition to the editing reactions. At each time point, an aliquot was removed and quenched by its addition to stop solution (6% SDS, 250 mM Tris, pH 8.0, 25 mM EDTA) at 80 °C. Each sample was extracted once with phenol/chloroform/isoamyl alcohol and once with chloroform, and was precipitated with ethanol. The pellet was resuspended in 10 µL of double distilled water and digested with Nuclease P1 (0.5 units/reaction; Amersham Biosciences), and the products were resolved by thin layer chromatography (saturated (NH₄)₂SO₄:0.1 M NaOAc:isopropanol; 79:19:2, v/v) using cellulose-polyethyleneimine chromatography plates (J. T. Baker).

II-5. References

- Andersen, J.S., Lyon, C.E., Fox, A.H., Leung, A.K., Lam, Y.W., Steen, H., Mann, M., and Lamond, A.I. (2002). Directed proteomic analysis of the human nucleolus. *Curr. Biol.* **12**, 1-11.
- Bernardi, R., and Pandolfi, P.P. (2007). Structure, dynamics and functions of promyelocytic leukaemia nuclear bodies. *Nat. Rev. Mol. Cell Biol.* **8**, 1006-1016.
- Bond, C.S., and Fox, A.H. (2009). Paraspeckles: nuclear bodies built on long noncoding RNA. *J. Cell Biol.* **186**, 637-644.
- Borghi, S., Molinari, S., Razzini, G., Parise, F., Battini, R., and Ferrari, S. (2001). The nuclear localization domain of the MEF2 family of transcription factors shows member-specific features and mediates the nuclear import of histone deacetylase 4. *J. Cell Sci.* **114**, 4477-4483.
- Bousquet-Antonelli, C., Vanrobays, E., Gelugne, J.P., Caizergues-Ferrer, M., and Henry, Y. (2000). Rrp8p is a yeast nucleolar protein functionally linked to Gar1p and involved in pre-rRNA cleavage at site A2. *RNA* **6**, 826-843.
- Bugler, B., Caizergues-Ferrer, M., Bouche, G., Bourbon, H., and Amalric, F.

(1982). Detection and localization of a class of proteins immunologically related to a 100-kDa nucleolar protein. *Eur. J. Biochem.* **128**, 475-480.

Carrero, G., McDonald, D., Crawford, E., de Vries, G., and Hendzel, M.J. (2003). Using FRAP and mathematical modeling to determine the in vivo kinetics of nuclear proteins. *Methods* **29**, 14-28.

Colmenares, S.U., Buker, S.M., Buhler, M., Dlakić, M., and Moazed, D. (2007). Coupling of Double-Stranded RNA Synthesis and siRNA Generation in Fission Yeast RNAi. *Mol. Cell* **27**, 449-461.

Dellaire, G., and Bazett-Jones, D.P. (2004). PML nuclear bodies: dynamic sensors of DNA damage and cellular stress. *BioEssays* **26**, 963-977.

Desterro, J.M., Keegan, L.P., Lafarga, M., Berciano, M.T., O'Connell, M., and Carmo-Fonseca, M. (2003). Dynamic association of RNA-editing enzymes with the nucleolus. *J. Cell Sci.* **116**, 1805-1818.

Fernandez, H.R., Kavi, H.H., Xie, W., and Birchler, J.A. (2005). Heterochromatin: On the ADAR Radar? *Curr. Biol.* **15**, R132-R134.

Fox, A.H., Lam, Y.W., Leung, A.K., Lyon, C.E., Andersen, J., Mann, M., and Lamond, A.I. (2002). Paraspeckles: a novel nuclear domain. *Curr. Biol.* **12**, 13-25.

Fu, L., Gao, Y.S., Tousson, A., Shah, A., Chen, T.L., Vertel, B.M., and Sztul, E. (2005). Nuclear aggresomes form by fusion of PML-associated aggregates. *Mol. Biol. Cell* **16**, 4905-4917.

Gong, X.Q., Nedialkov, Y.A., and Burton, Z.F. (2004). Alpha-amanitin blocks translocation by human RNA polymerase II. *J. Biol. Chem.* **279**, 27422-27427.

Gregory, R.I., Yan, K.P., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N., and Shiekhattar, R. (2004). The Microprocessor complex mediates the genesis of microRNAs. *Nature* **432**, 235-240.

Handwerger, K.E., and Gall, J.G. (2006). Subnuclear organelles: new insights into form and function. *Trends Cell Biol.* **16**, 19-26.

Houtsmuller, A.B. (2005). Fluorescence recovery after photobleaching: application to nuclear proteins. *Adv. Biochem. Eng. Biotechnol.* **95**, 177-199.

Jaikaran, D.C., Collins, C.H., and MacMillan, A.M. (2002). Adenosine to inosine editing by ADAR2 requires formation of a ternary complex on the GluR-B R/G site. *J. Biol. Chem.* **277**, 37624-37629.

- Krecic, A.M., and Swanson, M.S. (1999). hnRNP complexes: composition, structure, and function. *Curr. Opin. Cell. Biol.* **11**, 363-371.
- Lallemand-Breitenbach, V., and de The, H. (2010). PML Nuclear Bodies. *Cold Spring Harb. Perspect. Biol.* **2**, a000661-a000661.
- Lam, Y.W., Trinkle-Mulcahy, L., and Lamond, A.I. (2005). The nucleolus. *J. Cell Sci.* **118**, 1335-1337.
- Lamond, A.I., and Spector, D.L. (2003). Nuclear speckles: a model for nuclear organelles. *Nat. Rev. Mol. Cell Biol.* **4**, 605-612.
- Langland, J.O., Kao, P.N., and Jacobs, B.L. (1999). Nuclear factor-90 of activated T-cells: A double-stranded RNA-binding protein and substrate for the double-stranded RNA-dependent protein kinase, PKR. *Biochemistry* **38**, 6361-6368.
- Lippincott-Schwartz, J., and Patterson, G.H. (2003). Development and use of fluorescent protein markers in living cells. *Science* **300**, 87-91.
- Mao, Y.S., Zhang, B., and Spector, D.L. (2011). Biogenesis and function of nuclear bodies. *Trends Genet.* [Epub ahead of print].
- Masuda, K., Abdelmohsen, K., Kim, M.M., Srikantan, S., Lee, E.K., Tominaga, K., Selimyan, R., Martindale, J.L., Yang, X., Lehrmann, E., Zhang, Y., Becker, K.G., Wang, J.-Y., Kim, H.H., and Gorospe, M. (2011). Global dissociation of HuR-mRNA complexes promotes cell survival after ionizing radiation. *EMBO J.* **30**, 1040-1053.
- Miska, E.A., Karlsson, C., Langley, E., Nielsen, S.J., Pines, J., and Kouzarides, T. (1999). HDAC4 deacetylase associates with and represses the MEF2 transcription factor. *EMBO J.* **18**, 5099-5107.
- Morris, G.E. (2008). The Cajal body. *Biochim. Biophys. Acta.* **1783**, 2108-2115.
- Phair, R.D., and Misteli, T. (2000). High mobility of proteins in the mammalian cell nucleus. *Nature* **404**, 604-609.
- Raitskin, O., Cho, D.S., Sperling, J., Nishikura, K., and Sperling, R. (2001). RNA editing activity is associated with splicing factors in hnRNP particles: The nuclear pre-mRNA processing machinery. *Proc. Natl. Acad. Sci. U.S.A* **98**, 6571-6576.
- Rudd, M.D., and Luse, D.S. (1996). Amanitin greatly reduces the rate of transcription by RNA polymerase II ternary complexes but fails to inhibit some transcript cleavage modes. *J. Biol. Chem.* **271**, 21549-21558.

- Sansam, C.L., Wells, K.S., and Emeson, R.B. (2003). Modulation of RNA editing by functional nucleolar sequestration of ADAR2. *Proc. Natl. Acad. Sci. U.S.A* **100**, 14018-14023.
- Schoft, V.K., Schopoff, S., and Jantsch, M.F. (2007). Regulation of glutamate receptor B pre-mRNA splicing by RNA editing. *Nucleic Acids Res.* **35**, 3723-3732.
- Scott, M.S., and Ono, M. (2011). From snoRNA to miRNA: dual function regulatory non-coding RNAs. *Biochimie*. [Epub ahead of print].
- Shiohama, A., Sasaki, T., Noda, S., Minoshima, S., and Shimizu, N. (2007). Nucleolar localization of DGCR8 and identification of eleven DGCR8-associated proteins. *Exp. Cell Res.* **313**, 4196-4207.
- Sikora, D., Greco-Stewart, V.S., Miron, P., and Pelchat, M. (2009). The hepatitis delta virus RNA genome interacts with eEF1A1, p54nrb, hnRNP-L, GAPDH and ASF/SF2. *Virology* **390**, 71-78.
- Wang, Q., Zhang, Z., Blackwell, K., and Carmichael, G.G. (2005). Vigilins Bind to Promiscuously A-to-I-Edited RNAs and Are Involved in the Formation of Heterochromatin. *Curr. Biol.* **15**, 384-391.
- White, J., and Stelzer, E. (1999). Photobleaching GFP reveals protein dynamics inside live cells. *Trends Cell Biol.* **9**, 61-65.

Appendix III

Sequences of oligonucleotides used in this thesis

III-1. Primer sequences 5' → 3'

III-1.1. Human Prp8

hPrp8_1769_fwd: gcgcgcgaattcgaaaattgtattttcaaggtgagctcttccaac

hPrp8_1990_rev: gcgcgcaagcttcatcagtcagccaagatcag

hPrp8_D1781C_fwd: attatctggttggtgacaccaacgtctac

hPrp8_D1781C_rev: gtagacgttggtgacacaaaccagataat

hPrp8_V1788D_fwd: accaacgtctacagagacactattcacaagacc

hPrp8_V1788D_rev: ggtcttggttaagtgtctctgtagacgttggt

hPrp8_T1789P_fwd: aacgtctacagagtgccattcacaagaccttt

hPrp8_T1789P_rev: aaaggtcttggttagggcactctgtagacgtt

hPrp8_I1790Y_fwd: gtctacagagtgacttaccacaagacctttgaa

hPrp8_I1790Y_rev: ttcaaaggtcttggttaagtcactctgtagac

hPrp8_H1791E_fwd: tacagagtgactattgaaaagacctttgaaggg

hPrp8_H1791E_rev: cccttcaaaggtctttcaatagtcactctgta

hPrp8_T1793K_fwd: gtgactattcacaagaaattgaagggaacttg

hPrp8_T1793K_rev: caagttccctcaaattcttgaatagtcac

hPrp8_N1797D_fwd: aagacctttgaagggacttgacaaccaagccc

hPrp8_N1797D_rev: gggcttggttgattgtccccttcaaaggtctt

hPrp8_L1798N_fwd: acctttgaagggaacaatacaaccaagcccatc

hPrp8_L1798N_rev: gatgggcttggtgtattgttccttcaaaggt

hPrp8_T1799E_fwd: ttgaagggaacttgagaccaagcccatcaac

hPrp8_T1799E_rev: gttgatgggcttggtcctcaagttcccttcaaa

hPrp8_T1800E_fwd: gaagggaacttgacagaaaagcccatcaacgga

hPrp8_T1800E_rev: tccgttgatgggcttttctgtcaagttcccttc

hPrp8_R1865A_fwd: cagatcattgtcaccgccaaggcatgctggac

hPrp8_R1865A_rev: gtccagcatgcccttggcggtgacaatgatctg

hPrp8_E1888K_fwd: gtcacaaaggatcgaagctccaactcccttc

hPrp8_E1888K_rev: gaaaggagttggagcttcgatcctttgatgac

hPrp8_Q1894S_fwd: ctccaactcccttcagcgcgtgtctcaagtg

hPrp8_Q1894S_rev: caccttgagacacgcgctgaaaggagttggag

hPrp8_I1790CT1800C_fwd:

gtctacagagtgactgccacaagaccttgaagggaacttgacatgcaagcccatcaacgga

hPrp8_I1790CT1800C_rev:

tccgttgatgggcttgcatgtcaagttcccttcaaaggtcttgggcaagtcactctgtagac

III-1.2. Yeast Prp8

yPrp8_965_fwd: gcgcgcaattcgaaaattgtattttcaaggaaagcctgtgatata

yPrp8_1189_rev: gcgcgctcgactcatcactctgccaagtattc

yPrp8_recombination_fwd: gcacgaggatgaaaaattaat

yPrp8_recombination_rev: tgacctggcagcttctaattc

yPrp8_D1853A_fwd: attaaactcttcgttgccgacactaatgtgtat

yPrp8_D1853A_rev: atacacattagtgtcggcaacgaagagttaat

yPrp8_D1853C_fwd: attaaactcttcgtttgcgacactaatgtgtat

yPrp8_D1853C_rev: atacacattagtgtcgaacgaagagttaat

yPrp8_D1853E_fwd: attaaactcttcgttgaagacactaatgtgtat

yPrp8_D1853E_rev: atacacattagtgtctcaacgaagagttaat

yPrp8_D1853H_fwd: attaaactcttcgttcatgacactaatgtgtat

yPrp8_D1853H_rev: atacacattagtgtcatgaacgaagagttaat

yPrp8_D1853M_fwd: attaaactcttcgttatggacactaatgtgtat

yPrp8_D1853M_rev: atacacattagtgtccataacgaagagttaat

yPrp8_D1853N_fwd: attaaactcttcgttaacgacactaatgtgtat

yPrp8_D1853N_rev: atacacattagtgtcgtaacgaagagttaat

yPrp8_D1853S_fwd: attaaactcttcgtttctgacactaatgtgtat

yPrp8_D1853S_rev: atacacattagtgtcagaaacgaagagttaat

yPrp8_D1854A_fwd: aaactcttcggtgacgccactaatgtgtataga

yPrp8_D1854A_rev: tctatacacattagtggcgtcaacgaagagttt

yPrp8_D1854C_fwd: aaactcttcggtgactgcactaatgtgtataga

yPrp8_D1854C_rev: tctatacacattagtgacgtcaacgaagagttt

yPrp8_V1860D_fwd: actaatgtgtatagagatactgtccacaagact

yPrp8_V1860D_rev: agtcttgggacagtatctctatacacattag

yPrp8_T1861P_fwd: aatgtgtatagagttcctgtccacaagactttt

yPrp8_T1861P_rev: aaaagtcttgggacaggaactctatacacatt

yPrp8_V1862Y_fwd: gtgtatagagttacttaccacaagacttttgaa

yPrp8_V1862Y_rev: ttcaaaagtcttggtaagtaactctatacac

yPrp8_H1863E_fwd: tatagagttactgtcgaaaagacttttgaaggg

yPrp8_H1863E_rev: cccttcaaaagtcttttcgacagtaactctata

yPrp8_T1865K_fwd: gttactgtccacaagaagtttgaagggaacgtt

yPrp8_T1865K_rev: aacgttccttcaaaccttcttggacagtaac

yPrp8_N1869D_fwd: aagacttttgaaggggatgttgctacaaaagca

yPrp8_N1869D_rev: tgctttttagcaacatccccctcaaaagtctt

yPrp8_V1870N_fwd: acttttgaagggaacaatgctacaaaagcaatt

yPrp8_V1870N_rev: aattgctttttagcattgttccttcaaaagt

yPrp8_A1871E_fwd: tttgaagggaacgttgaacaaaaagcaattaat

yPrp8_A1871E_rev: attaattgcttttgttcaacgttccttcaaa

yPrp8_T1872E_fwd: gaagggaacgttgctgaaaaagcaattaatggt

yPrp8_T1872E_rev: accattaattgcttttcagcaacgttccttc

yPrp8_E1960K_fwd: gcaatcaggccaacaaagctgcgactaccattt

yPrp8_E1960K_rev: aatggtagtcgcagctttgttgctgattgc

III-1.3. ACT1 transcription template

T7: taatacgaactcactatagg

PBS_rev: gcgcgcaagcttgggctgcaggtcgac

ACT1_fwd: taatacgaactcactataggg

ACT1_rev: gcgcgcaagcttgggctgcaggtcgac

ACT1_5'_exon_fwd: taatacgaactcactataggg

ACT1_5'_exon_rev: aacatataatatagc

ACT1_BSC_fwd: gtgtctcatgtactaccatcgattgcttcat

ACT1_BSC_rev: atgaagcaatcgatggtagtagtacatgagacac

ACT1_BSG_fwd: gtgtctcatgtactagcatcgattgcttcat

ACT1_BSG_rev: atgaagcaatcgatgctagtagtacatgagacac

ACT1_UuG_fwd: ctatattatgtttgaggttgctgctttg

ACT1_UuG_rev: caaagcagcaacctcaaacatataatag

III-1.4. Primer extension

YAC6: ggcactcatgacctc

III-1.5. ADAR proteins

ADAR1_FRET_fwd: gcgcgccaagcttatggccgagatcaaggag

ADAR1_FRET_rev: gcgcgcggatccctactatactgggcagagata

ADAR1_pulldown_fwd: gcgcgcatagatctatggccgagatcaaggag

ADAR1_pulldown_rev: gcgcgcggatccctactatactgggcagagata

ADAR2_FRET_fwd: gcgcgctcgaattcatggatatagaagatgaa

ADAR2_FRET_rev: gcgcgctcgactcatcagggcgtgagtgagaactg

ADAR2_pulldown_fwd: gcgcgcatagatctatggatatagaagatgaa

ADAR2_pulldown_rev: gcgcgcggtacctcatcagggcgtgagtgagaa

ADAR2_187_rev: gcgcgcaagctttcatcactccgcctgtcaggagt

ADAR2_316_rev: gcgcgcaattcctactactgaagaccctcact

ADAR2_E114K_fwd: gccattcaccttcacagacatgacaaa

ADAR2_E114K_rev: atgtctgtgaaggtgaatggccaggtt

ADAR2_E114V_fwd: gccattcaccaccacagacatgacaaa

ADAR2_E114V_rev: atgtctgtggtggtgaatggccaggtt

ADAR2_S141K_fwd: ctgaacgaacttctcaaggccttctc

ADAR2_S141K_rev: gccttgaggaagttcgttcagtttct

ADAR2_S141V_fwd: ctgaacgaacaccctcaaggccttctc

ADAR2_S141V_rev: gccttgagggtgttcgttcagtttct

III-1.6. Other proteins

Coil_fwd: gcgcgctctcgag atggcagctccgagacg

Coil_rev: gcgcgcggatccctacta ggcaggttctgtact

H2AX_fwd: gcgcgcaattcaatgtcgggccgcggcaag

H2AX_rev: gcgcgcggatccttattagtactcctgggagggc

hnRNPF_fwd: gcgcgcagatctatgatgctgggccctgag

hnRNPF_rev: gcgcgcaagcttctactagtcatagccacccat

HuR_fwd: gcgcgcagatctatgtctaattggttatgaa

HuR_rev: gcgcgcaattcttattatttgggacttgtt

ILF2_fwd: gcgcgcagatctatgaggggtgacagaggc

ILF2_rev: gcgcgcaattctcatcactcctgagtttccat

NCL_fwd: gcgcgagatctatggtgaagctcgcgaag

NCL_rev: gcgcggaattcctactattcaaactcgtctt

PML_fwd: gcgcgcaagcttcgatggagcctgcacccgcc

PML_rev: gcgcggaattctcatcagaggcctgcttgacg

p54nrb_fwd: gcgcggaattctatgcagagtaataaaact

p54nrb_rev: gcgcgcgatccttattagtatcggcgacgttt

SRP9_fwd: gcgcgagatctatgccgcagtaccagacc

SRP9_rev: gcgcggaattctcatcactcagtttccatggt

III-2. RNA sequences 5' → 3'

ACT1_3'_exon: uuuagagguugcugcuuugguuauu

R/G substrate:

ggguccauuaagguggguggaauaguauaacaauaugcucaauguuguuauaguauccaccuaccc
ugauguguc