

U6 snRNA Secondary Structure in Free U6 snRNPs

Elizabeth A. Dunn

B. Sc., University Of Northern British Columbia

Thesis Submitted In Partial Fulfillment Of

The Requirements For The Degree Of

Master Of Science

In

Mathematical, Computer, and Physical Sciences

(Chemistry)

The University of Northern British Columbia

July 2009

© Elizabeth A. Dunn, 2009



Library and Archives
Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file *Votre référence*
ISBN: 978-0-494-60837-1
Our file *Notre référence*
ISBN: 978-0-494-60837-1

NOTICE:

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

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

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

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

AVIS:

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

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

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

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


Canada

LAC Condensed Abstract – Elizabeth A. Dunn

A critical step in spliceosome assembly is the formation of the U4/U6 di-snRNP particle in which U4 and U6 snRNAs are engaged in an extensive intermolecular base pairing interaction. Little is known about how this particle is formed or its role in pre-mRNA splicing, and current models of U6 snRNA secondary structure in free U6 snRNP do not offer insight into these questions. We have generated a new model of U6 snRNA secondary structure in free U6 snRNPs that is consistent with existing structural and genetic data and we show support for this model through genetic analyses and oligonucleotide accessibility experiments. Our model predicts that catalytically important elements are unavailable for interaction with the pre-mRNA transcript until a large structural re-arrangement exposes these elements. We propose that U4 snRNA is responsible for promoting this re-arrangement through the establishment of an intermolecular interaction between U4 and U6.

TABLE OF CONTENTS

Abstract	ii
Table of Contents	iii
List of Tables	v
List of Figures	vi
Acknowledgement	viii
Chapter One – Introduction	1
1.1 Nuclear Pre-mRNA Splicing	1
1.2 Spliceosome Assembly	3
1.3 U6 snRNA	5
1.4 U6 snRNA in Free U6 snRNP	9
1.5 Motivation for a New Model of U6 snRNA Secondary Structure	12
Chapter Two – Introduction to the Dunn-Rader Model of U6 snRNA Structure in Free U6 snRNP	13
2.1 The Dunn-Rader Model of U6 snRNA Secondary Structure	13
2.2 Three Dimensional Modeling of U6 snRNA	15
2.2.1 Methods	15
2.2.2 U6 snRNA in Three Dimensions	15
2.3 Support for the Dunn-Rader Model from the Literature	18
2.3.1 Chemical Modification of Accessible U6 snRNA Bases	19
2.3.2 Compensatory Base Mutation	24
2.3.3 The U6 snRNA 3' Intramolecular Stem/loop (3' ISL)	26
2.3.4 Genetic Suppression Analysis – A62G and A79G	28
2.3.5 Exonuclease Degradation of U6 snRNA	31
Chapter Three – Structure Probing by Oligonucleotide Accessibility	33
3.1 Materials and Methods	33
3.2 Results and Discussion	36
Chapter Four – Genetic Support for the Lower Central Stem	46
4.1 Materials and Methods	46
4.2 Results and Discussion	50

Chapter Five – Co-variation of U6 snRNA Secondary Structure and Prp24 Homolog Architecture	58
5.1 Materials and Methods	59
5.2 Phylogenetic Conservation of U6 snRNA Secondary Structure	60
5.3 Evolutionary Relationship Based on U6 snRNA Sequence and Secondary Structure	69
5.4 Subtle Changes in U6 snRNA Sequence Correlate with Dramatic Changes in Prp24 Homolog Architecture	72
Chapter Six – The Dunn-Rader Model of U6 snRNA Secondary Structure Reveals the Primary Function of U4 snRNA in Pre-mRNA Splicing	77
Chapter Seven – Future Directions and Concluding Remarks	82
7.1 U6 snRNP Structure Determination	82
7.2 Genetic Studies	83
7.3 Structure Probing	84
7.4 U6 snRNA Structural Re-arrangements in Pre-mRNA Splicing	85
7.5 Concluding Remarks	86
Works Cited	ix

LIST OF TABLES

Table 1: DNA Oligonucleotide Sequence, GC Content, and Melting Temperature.	36
Table 2: Putative Prp24 Homologs.	60

LIST OF FIGURES

Figure 1. Nuclear pre-mRNA splicing is catalyzed by the spliceosome and is achieved through two transesterification reactions.	2
Figure 2. U6 snRNA conformational re-arrangements throughout the splicing cycle.	7
Figure 3. The Dunn-Rader model of U6 snRNA structure in the free U6 snRNP.	14
Figure 4. U6 snRNA can fold into a closed and open conformation without disrupting the base pairing interactions proposed in the Dunn-Rader model.	17
Figure 5. 2D TOCSY NMR spectrum of <i>in vitro</i> transcribed U6 snRNA.	18
Figure 6. Chemical modification of naked U6 snRNA and U6 snRNA in U6 snRNPs.	22
Figure 7. The Karaduman et al. (2006) chemical modification data for <i>in vitro</i> transcribed U6 snRNA and U6 snRNPs mapped to the proposed closed and open Dunn-Rader models.	23
Figure 8. A62G and A79G suppressor mutations mapped onto U6 snRNA secondary structures in the free U6 snRNP, di-snRNP, and active spliceosome.	29
Figure 9. U6 snRNA secondary structure probing by oligonucleotide accessibility.	37
Figure 10. Construction of U6 snRNA central stem double mutants.	52
Figure 11. Growth phenotypes and U4/U6 assembly phenotypes of the central stem double mutants, a wild type control, and a control for cold sensitivity, U6-A62G.	55
Figure 12. U6 snRNA and U6atac sequence alignment.	62
Figure 13. <i>S. cerevisiae</i> and <i>H. sapiens</i> stem/loops A and B can fold into similar three-dimensional structures.	69
Figure 14. Subtle changes in U6 snRNA sequence are accompanied by substantial changes in Prp24 homolog architecture.	71

Figure 15. U6 snRNA is tightly regulated through riboswitch-like conformational changes during spliceosome assembly and activation.

80

ACKNOWLEDGEMENT

First and foremost I would like to thank my supervisor Dr. Stephen Rader for introducing me to the world of research and providing mentorship and support throughout the duration of my graduate work. I would also like to thank my graduate thesis committee members Dr. Andrea Gorrell, Dr. Daniel Erasmus, and Dr. Saif Zahir for helpful discussions and project direction. Special thanks to Dr. Phillip Johnson for collaborating with us for the NMR experiments and Dr. Fred Damberger for helping us with the interpretation of the NMR data. I would like to thank all past and present members of the Rader Lab for helpful discussions and guidance.

There are many others outside of academics that I would like to extend a special thanks to: William Dunn for his support, understanding and encouragement over the last seven years, and my parents, Rick and Kim Chester who have provided encouragement and support from a very young age. I would also like to thank my sisters, Adrienne and Jen, and brother, Chris, for their understanding and entertainment over the years.

This work was funded through an NSERC Discovery Grant awarded to Dr. Stephen Rader and UNBC administered research project awards awarded to Elizabeth Dunn.

Chapter One – Introduction

A critical step in eukaryotic gene expression is the processing of precursor messenger RNA (pre-mRNA) transcripts to generate mature mRNAs that can subsequently be translated into the appropriate polypeptide chain. One such processing event, known as nuclear pre-mRNA splicing, involves the removal of non-coding RNA and the subsequent joining of the remaining regions of coding RNA. This complex process is catalyzed by the spliceosome, a large ensemble of over 100 proteins and five small nuclear RNAs (snRNA). Although splicing has been studied extensively over the past three decades, very little is known about the function of many of these essential splicing factors. U6 snRNA in particular has been studied considerably and although experimental data suggest that U6 plays a central role throughout the splicing reactions, the mechanism by which it may do so has yet to be elucidated. The major goal of this thesis was to develop a model of the structure of U6 snRNA in its inactive free small nuclear ribonucleoprotein (snRNP) particle as a step toward understanding how U6 becomes activated for splicing.

1.1 Nuclear Pre-mRNA Splicing

Splicing of pre-mRNA transcripts was first proposed following the finding that a mature mRNA transcript produced in cells infected by Adenovirus 2 was generated through the joining of several RNA fragments (Berget et al. 1977). Since this discovery, a general mechanism for the excision of intervening sequences (introns) and joining of coding sequences (exons) has been developed (Fig. 1). In the first of two transesterification reactions, the 2' hydroxyl of a bulged adenosine, located at the branch

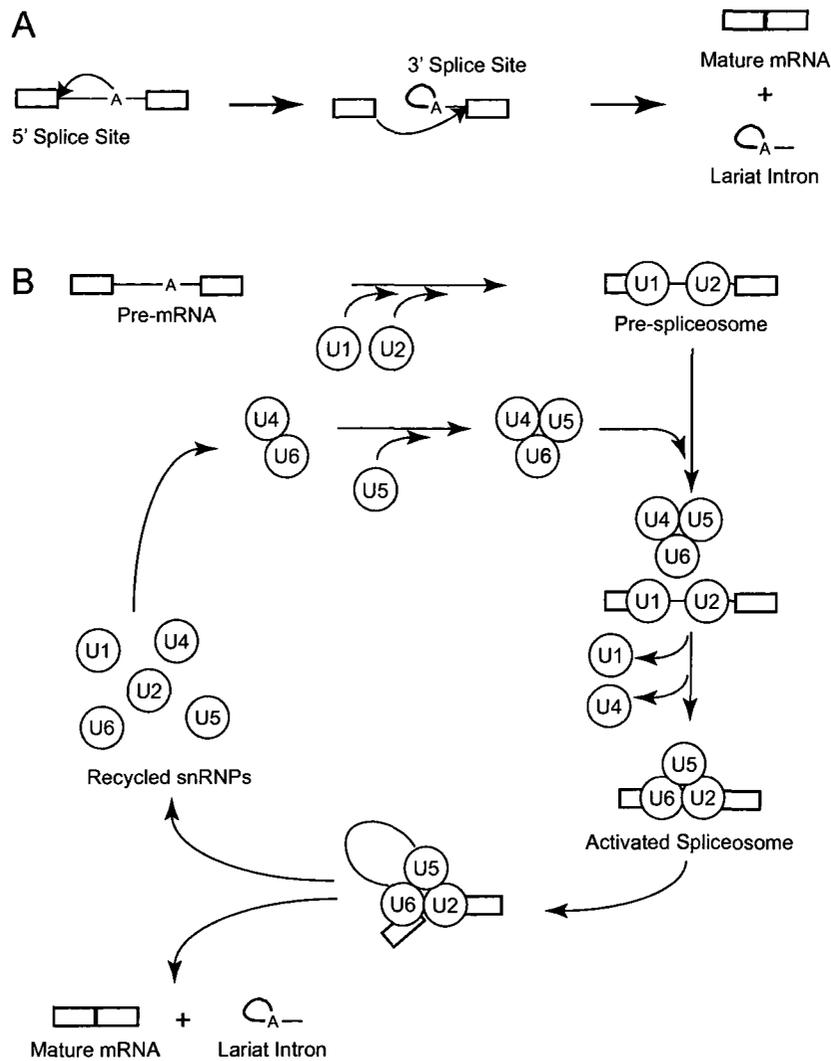


Figure 1. Nuclear pre-mRNA splicing is catalyzed by the spliceosome and is achieved through two transesterification reactions. The splicing reaction (A) and stepwise spliceosome assembly (B) are shown. Exons are represented by open boxes and the intron is shown with a line containing the branch point adenosine.

site in the pre-mRNA intron, reacts with the 5' splice site (Padgett et al. 1984, Konarska et al. 1985). This results in the liberation of the 5' exon concomitantly with the formation of a 2'-5' phosphodiester linkage, which joins the 5' end of the intron to the branch residue (Padgett et al. 1984, Konarska et al. 1985). In the second reaction, the 3' hydroxyl

of the 5' exon reacts with the 3' splice site, cleaving away the lariat intron and joining the 5' and 3' exons through a 3'-5' phosphodiester linkage (Padgett et al. 1984, Konarska et al. 1985). The inversion of stereochemistry at the chiral phosphate in each step suggests that both of these transesterifications occur through an in line S_N2 nucleophilic reaction (Maschhoff & Padgett 1993, Moore & Sharp 1993).

1.2 Spliceosome Assembly

Despite the simplicity of the splicing reactions, recognition of the 5' and 3' splice sites and the branch point adenosine to achieve high fidelity splicing is a complex process. While the splice sites and branch site in each pre-mRNA transcript are defined by consensus sequences, these sequences are short and in some cases quite degenerate (Irimia et al. 2007). The length of intervening sequences from one transcript to another also varies considerably, and the presence of alternative splice sites in more complex eukaryotic organisms complicates the process further (Bon et al. 2003, Irimia et al. 2007). Consequently the splicing reactions have been proposed to be facilitated by a massive, and highly dynamic, macromolecular complex known as the spliceosome.

The spliceosome is thought to be responsible for a number of events including splice site and branch point recognition and positioning of the pre-mRNA transcript in a suitable orientation for the splicing reactions to take place (Brody & Abelson 1985, Grabowski et al. 1985). The spliceosome is composed of more than one hundred proteins and five snRNAs, U1, U2, U4, U5, and U6 (Stevens et al. 2002, Reviewed in Jurica & Moore 2003). Each snRNA associates with a specific set of proteins to generate an snRNP particle, and these particles undergo extensive structural re-arrangements

throughout the splicing cycle (Fortner et al. 1994, Staley & Guthrie 1999, Hilliker et al. 2007, Perriman & Ares 2007). Although there have been reports of a pre-assembled “holospliceosome” containing all five splicing snRNPs (Stevens et al. 2002), the assembly of the spliceosome has traditionally been thought to occur *de novo* on each new pre-mRNA transcript through the ordered association of the snRNP particles (Fig .1, Bindereif & Green 1987, Cheng & Abelson 1987, Konarska & Sharp 1987).

In the traditional view of step-wise assembly of the spliceosome, the U1 and U2 snRNPs first recognize and associate with the pre-mRNA transcript through base pairing interactions with the 5' splice site and branch sequence respectively (Fig. 1, Black et al. 1985, Parker et al. 1987). Following formation of this pre-spliceosome complex, the U4, U5 and U6 snRNPs associate with the spliceosome as a preformed triple snRNP complex (Cheng & Abelson 1987, Konarska & Sharp 1987). U4/U6 base pairing is then disrupted and U6 replaces U1 at the 5' splice site (Wassarman & Steitz 1992, Lesser & Guthrie 1993, Staley & Guthrie 1999). At the same time, U4 and U1 become loosely associated with the spliceosome and may even be released while U2, U5 and U6 become tightly associated with the pre-mRNA transcript to form the catalytically active spliceosome (Cheng & Abelson 1987, Konarska & Sharp 1987). A number of conformational rearrangements of the active spliceosome then allow the splicing reactions to proceed, resulting in the production of a mature mRNA transcript and the excised lariat intron (Staley & Guthrie 1999, Hilliker et al. 2007).

1.3 U6 snRNA

It is not yet clear whether nuclear pre-mRNA splicing is catalyzed by the snRNA or the protein components of the spliceosome, however it is becoming apparent that both the snRNAs and the proteins are required to perform critical functions throughout the splicing cycle (Guthrie 1991, Collins & Guthrie 2000, Valadkhan 2007, Abelson 2008). Many of the proteins serve as a structural scaffold around which the spliceosome can correctly assemble while many others are involved in promoting structural re-arrangement of the snRNAs (Ohi et al. 2005, Raghunathan & Guthrie 1998b, Staley & Guthrie 1999). However it is largely the snRNAs that recognize and base pair to the splice sites and branch sequence of pre-mRNA transcripts (Black et al. 1985, Parker et al. 1987). The close proximity of U2 and U6 snRNAs to the catalytic center of the spliceosome, and their tight association with the pre-mRNA transcript in functionally activated spliceosomes, has led to the speculation that splicing may be an RNA catalyzed event (Cech 1986, Madhani & Guthrie 1992).

Of the five snRNAs, U6 is the most likely to play a direct role in pre-mRNA splicing catalysis. It is the most highly conserved in size, with essentially all size variation residing in the phylogenetically conserved 5' stem/loop, where stem length can vary by several base pairs (Brow & Guthrie 1988). The size conservation is accompanied by a highly conserved nucleotide sequence that shows approximately 80% sequence identity across the middle third of yeast and human U6 snRNAs, and approximately 60% identity across the entire U6 sequence of these same two species (Brow & Guthrie 1988, Guthrie & Patterson 1988). Steric constraints within the catalytically active spliceosome

may explain the strict size conservation while the high level of sequence conservation may reflect nucleotide specific requirements during spliceosome assembly and activation, and/or throughout the splicing reactions (Brow & Guthrie 1988).

While many mutations throughout U6 snRNA result in only weak conditional growth phenotypes or show no observable deviation from wild type growth at all, one stretch of nucleotides, known as the ACAGAGA box, is unexpectedly intolerant of mutation (Fabrizio & Abelson 1990, Madhani et al. 1990, Wolff et al. 1994). There are a few known examples of variation in this sequence element in which deviation at the second position to a U appears so far to be limited to the Trypanosomatidae family (Xu et al. 1994). Point mutations within this stretch of seven nucleotides result in lethality *in vivo* and severe inhibition of splicing *in vitro* with various levels of first and/or second step blocks (Fabrizio & Abelson 1990, Madhani et al. 1990). The ACAGAGA sequence has been shown to genetically interact with, as well as to cross-link to, the 5' splice site region of the pre-mRNA and has also been proposed to play an as of yet undefined role throughout the splicing reactions (Fig. 2, Fabrizio & Abelson 1990, Wassarman & Steitz 1992, Lesser & Guthrie 1993, Kandels-Lewis & Séraphin 1993, Wolff et al. 1994, Kim & Abelson 1996).

A second highly conserved sequence, the AGC triad, is located approximately six nucleotides downstream of the ACAGAGA box in mature U6 snRNAs. The AGC triad has been shown to interact with U2 snRNA through the formation of helix Ib, a structure that has been proposed to play a role in exon ligation (Fig. 2, Madhani & Guthrie 1992, Hilliker & Staley 2004). Interestingly, group II self-splicing introns also contain an AGC

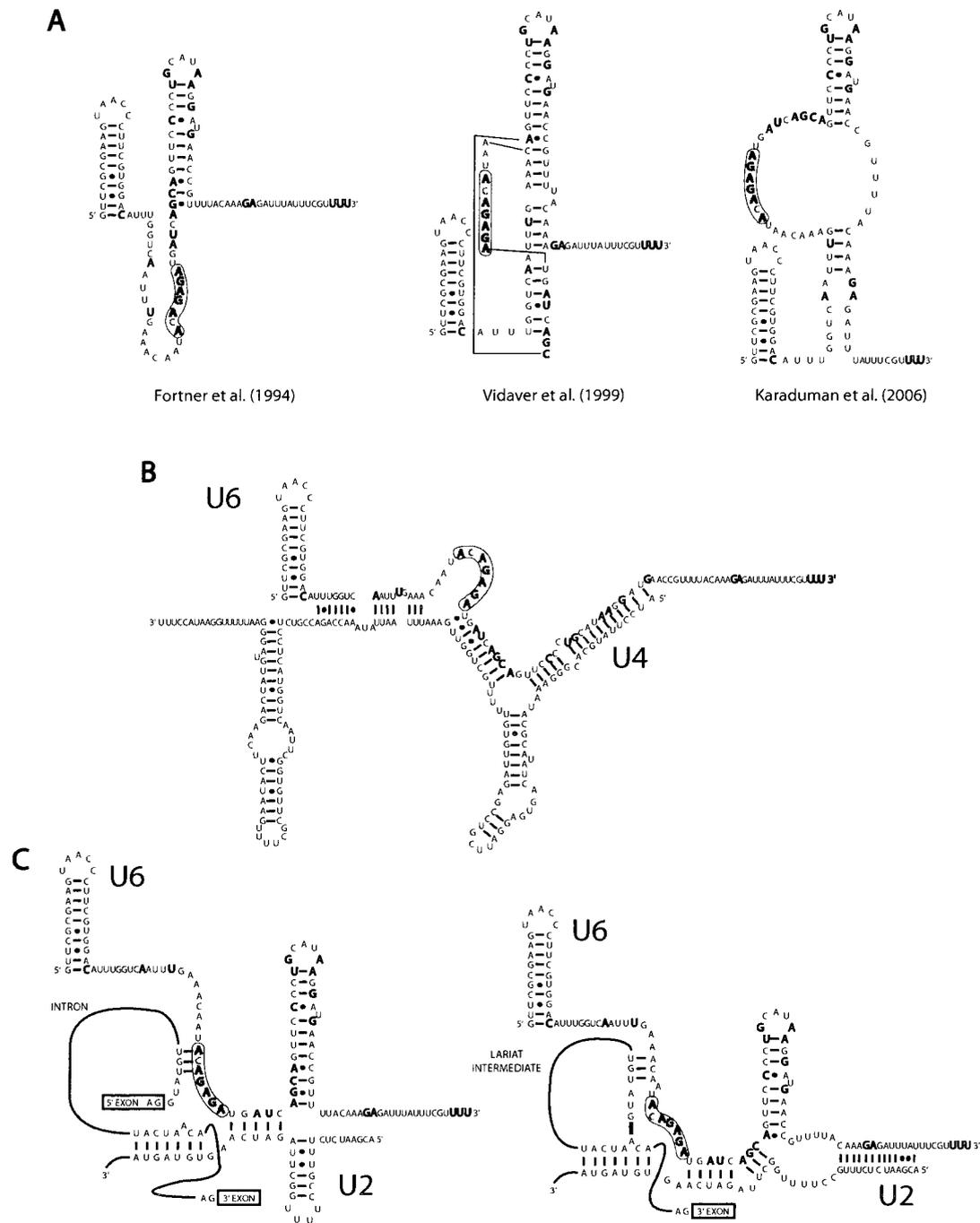


Figure 2. U6 snRNA conformational re-arrangements throughout the splicing cycle. Proposed U6 snRNA secondary structures in free U6 snRNP (A), U4U6 di-snRNA (B), and the first and second step catalytic spliceosome conformations (C). Invariant nucleotides are shown in bold and the ACAGAGA sequence is shaded in gray.

sequence base paired in a structure that resembles U2-U6 helix I or an extended U6 intramolecular stem/loop (ISL), both of which have been proposed to form near the active site of the spliceosome at different times throughout splicing (Peebles et al. 1995, Madhani & Guthrie 1992, Sashital et al. 2004). The AGC triad has been shown to be essential for catalysis of self-splicing introns and the essential metal ion binding at a bulge in domain V of group II self-splicing introns and in the U6 3' ISL further extend the parallels between these systems (Peebles et al. 1995, Yean et al. 2000). These observations suggest that nuclear pre-mRNA splicing may in fact be catalyzed by RNA through a mechanism similar to that of group II introns (Madhani & Guthrie 1992, Peebles et al. 1995). Like pre-mRNA splicing, group II self splicing is achieved through two transesterification reactions, resulting in the excision of a lariat intron that is formed through a 2'-5' phosphodiester linkage (Peebles et al. 1986, van der Veen et al. 1986, Cech 1986).

U6 snRNA is found in three major forms throughout the splicing cycle: intramolecularly base paired in a free U6 snRNP, base paired extensively to U4 snRNA in the U4/U6 di-snRNP, and base paired to the pre-mRNA transcript and U2 snRNA in activated spliceosomes (Fig. 2, Brow & Guthrie 1988, Datta & Weiner 1991, Wu & Manely 1991, Madhani & Guthrie 1992, Field & Friesen 1996). These interactions are mutually exclusive and may reflect temporal regulation of structural re-arrangements during spliceosome assembly and activation since interactions between U6, U2, and the pre-mRNA transcript within the catalytically competent spliceosome may juxtapose important sequence elements for RNA catalyzed splicing (Fig. 2, Madhani & Guthrie

1992, Johnson & Abelson 2001). Thus it may be critical to ensure that one interaction is correctly established prior to engagement in a second interaction.

1.4 U6 snRNA in Free U6 snRNP

In its free U6 snRNP form, U6 snRNA is thought to interact with only a small complement of proteins: the precursor RNA processing protein, Prp24, and a group of seven small proteins, known as the Lsm complex (Shannon & Guthrie 1991, Cooper et al. 1995, Séraphin 1995, Stevens et al. 2001, Karaduman et al. 2006, Karaduman et al. 2008). Homologs of Prp24 have been identified in a number of organisms and the strong conservation of repeated RNA recognition motifs (RRMs) and a stretch of 12 highly conserved amino acids at the C-terminus implies that the function has been conserved (Bell et al. 2002, Rader & Guthrie 2002). Indeed, both Prp24 and its human homolog, SART3/p110^{nrb}, have been shown to be required under U4/U6 di-snRNP recycling conditions (Raghunathan & Guthrie 1998, Bell et al. 2002), and the highly conserved C-terminal domain has been shown to be involved in promoting formation of this complex in yeast (Rader & Guthrie 2002).

The Lsm proteins are thought to form a doughnut shaped complex that interacts with the uridine rich 3' end of U6 snRNA (Achsel et al. 1999, Vidal et al. 1999, Ryan et al. 2002). This complex is required for U6 snRNA stability and has been proposed to play a role in U6 snRNP conformational re-arrangements (Cooper et al. 1995, Mayes et al. 1999). Further, the entirety of this complex has recently been shown to be required for nuclear localization of free U6 snRNP (Spiller et al. 2007). The presence of the Lsm complex appears to enhance Prp24 binding, suggesting that protein association with U6

snRNA to form free U6 snRNP is a co-operative binding event (Ryan et al. 2002, Rader & Guthrie 2002).

Although it will be some time before high resolution structural information for free U6 snRNP is obtained, low resolution electron microscope images are now available (Karaduman et al. 2008). These images have revealed the relative positioning of the two large protein domains and suggest that the Lsm complex interacts with U6 in a fixed orientation relative to Prp24 (Karaduman et al. 2008). Unfortunately it is not possible to visualize U6 snRNA at such low resolution, and there is still some disagreement surrounding the RNA interactions within free U6 snRNP (Ryan et al. 2002, Karaduman et al. 2006, McManus et al. 2007). Understanding the structure of U6 snRNA in the free snRNP will provide insight into the mechanism of U4/U6 di-snRNP formation and ultimately the activation of U6 snRNA for splicing.

Several models of *S. cerevisiae* U6 snRNA secondary structure in free U6 snRNP have been proposed over the last two decades (Fortner et al. 1994, Brow & Vidaver 1995, Vidaver et al. 1999, Karaduman et al. 2006). The first widely considered proposal consisted of a 5' stem/loop, which was phylogenetically conserved, and two other stem structures: the central stem/loop and the 3' ISL (Brow & Guthrie 1988, Fortner et al. 1994). The latter two structures were based on the presence of 3' exonuclease degraded U6 species of length 60 and 90 nucleotides found in yeast sub-cellular extract (Eschenlauer et al. 1993). Fortner et al. (1994) attributed this observation to 3' exonuclease activity that defined the 3' border of the two stem structures. While the 3'

ISL is phylogenetically conserved, the central stem/loop is not and was regarded as phylogenetically unproven (Fortner et al. 1994).

Phylogenetic, genetic, and chemical structure probing experiments have provided strong support for both the 5' stem/loop and 3' ISL and consequently these structures have been proposed in all other models of yeast and mammalian secondary structure (Brow & Guthrie 1988, Fortner et al. 1994, Jandrositz & Guthrie 1995, Wolff & Bindereif 1993, Wolff & Bindereif 1995, Karaduman et al. 2006, McManus et al. 2007). In contrast, the stretch of 35 nucleotides located between these stem structures has undergone considerable remodeling ranging from a pseudoknot structure that later evolved into a telestem proposed by Brow and colleagues (Brow & Vidaver 1995, Vidaver et al. 1999), to completely unstructured in order to accommodate binding of Prp24 as proposed by Lührmann and colleagues (Karaduman et al. 2006).

In contrast to the yeast models, the mammalian model had remained unchanged since its first proposal in 1980 until recently. The initial version of mammalian U6 was much more extensively base paired throughout the central region of the molecule with an asymmetric bulge present about mid-way through an extended 3' stem/loop (Epstein et al. 1980). Karaduman et al. (2006) disrupted much of this intramolecular network to propose a mammalian model that is conserved between yeast and mammals, arguing that the high level of sequence conservation should be reflected through similar secondary structures (Karaduman et al. 2006). Like the yeast model, these researchers suggest that the Prp24 homolog binds U6 snRNA at this large bulge.

1.5 Motivation for a New Model of U6 snRNA Secondary Structure

In spite of extensive phylogenetic and structural information, there is no widely accepted secondary structure model of U6 snRNA in free U6 snRNP, let alone any understanding of its three dimensional shape. More importantly, these models do not offer compelling explanations of how U6 becomes incorporated into the spliceosome, nor do they explain the functional significance of the large structural re-arrangements in U6 throughout splicing. In particular, they do not address the function of transient association with U4 snRNA, an absolute requirement for U6 snRNA assembly into functional spliceosomes. Further, many of the experimental techniques employed to study U6 snRNA secondary structure have generated ambiguous results that have largely been interpreted in a manner consistent with the accepted model of the time without exploration of other possible interpretations of the data. The main objective of the work described in this thesis was to generate a new model of U6 snRNA secondary structure that is compatible with existing data, and offers some explanation of the functional relevance of the large structural re-arrangements in U6 snRNA during spliceosome assembly and activation, particularly in the function of the transient U4/U6 interaction.

Chapter Two – Introduction to the Dunn-Rader Model of U6 snRNA Structure in Free U6 snRNP

Two main objectives in our manual re-modeling of U6 snRNA were i) to minimize the extent of unstructured regions of RNA, and ii) to sequester the catalytically important ACAGAGA sequence within the RNA interaction network. Such sequestration of this potentially catalytic element may prevent U6 from engaging in premature interactions with pre-mRNA transcripts until signaled to do so through an induced conformational change. This chapter introduces the Dunn-Rader model of U6 snRNA secondary structure and assesses the quality of the structure by mapping data from the literature to the model, as well as to previously proposed models for comparison.

2.1 The Dunn-Rader Model of U6 snRNA Secondary Structure

In addition to a 5' stem/loop that has been proposed in all other models of U6 snRNA secondary structure, the Dunn-Rader model predicts three additional helical segments which meet at a three-way junction: the central stem, stem/loop A, and stem/loop B (Fig. 3). While the base pair composition of the central stem is very similar to that proposed by Karaduman et al. (2006), the presence of two additional stems is unique to the Dunn-Rader model. Stem/loops A and B greatly reduce the number of unstructured nucleotides by incorporating residues that have previously been proposed to reside in a large asymmetric bulge into these stems (Fig. 2A, 3). Importantly, the highly conserved ACAGAGA sequence is predicted to base pair within stem A in the Dunn-Rader model while all other models predict that this sequence lies in a large unstructured region comprising up to 23 unpaired nucleotides (Karaduman et al. 2006). The significance of this is discussed in chapter six.

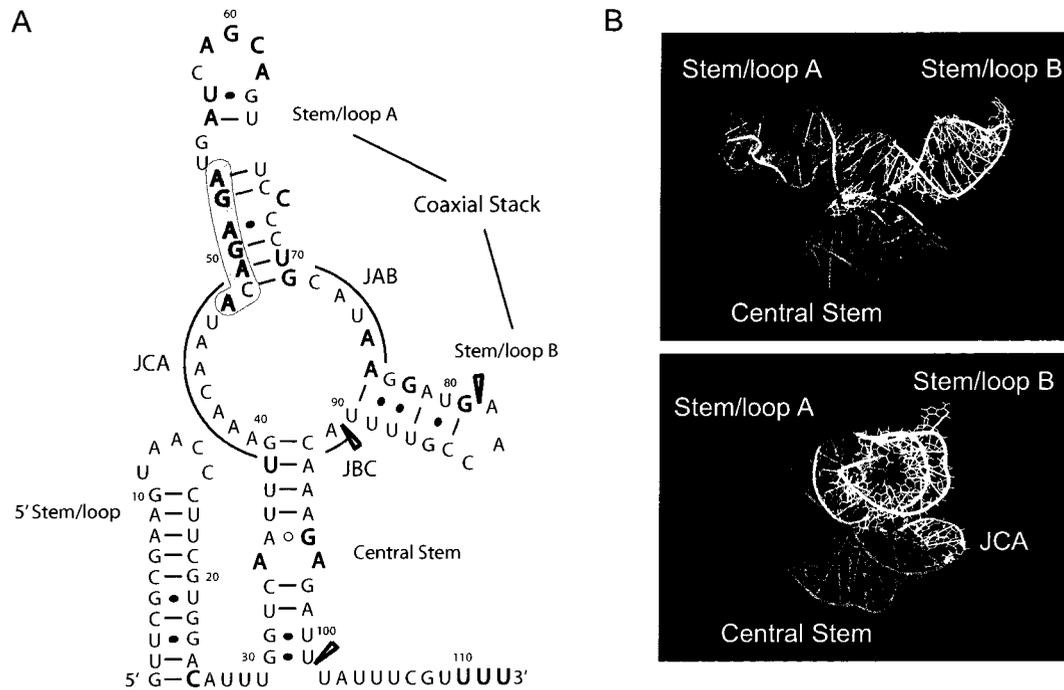


Figure 3. The Dunn-Rader model of U6 snRNA structure in the free U6 snRNP.
(A) U6 snRNA secondary structure. Invariant nucleotides are shown in bold and the ACAGAGA sequence is shaded in gray. The 3' end of 3' exonuclease degradation products are indicated with wedges. The junction between stem/loops A and B (JAB), stem/loop B and the central stem (JBC), and the central stem and stem/loop A (JCA) are indicated. **(B)** Three-dimensional representation of the proposed three-way junction from the front (top) and side (bottom): stem/loop A (red), stem/loop B (yellow), the central stem (blue), and the single-stranded junctions (green).

Three-way junctions are common RNA folds characterized by the convergence of three helical segments at a single junction, where two of these helices tend to stack three-dimensionally in a nearly coaxial fashion (Lescoute & Westhof 2006). The helices are connected by no more than three single stranded RNA segments, the length of which dictates which two of the three helices likely stack and the relative orientation of the third helix to the stack (Lescoute & Westhof 2006). The Dunn-Rader model predicts single stranded linking segments of lengths eight, four, and one nucleotide for the junction

between the central stem and stem/loop A (JCA), stem/loop A and stem/loop B (JAB), and stem/loop B and the central stem (JBC) respectively (Fig. 3). Since JCA is larger than JBC, the Lescoute and Westhof (2006) classification scheme suggests that stem/loops A and B may stack coaxially while the central stem bends in toward stem/loop A.

2.2 Three Dimensional Modeling of U6 snRNA

2.2.1 Methods

To generate a theoretical three-dimensional representation of our proposed base pairing interactions, the core U6 snRNA nucleotide sequence consisting of the *S. cerevisiae* residues 30 to 101 (excluding two nucleotides in JAB) was loaded into the RNA modeling program MC-Sym (<http://www.major.ircic.ca/MC-Pipeline>) along with our secondary structure constraints (Parisien & Major 2008). Three-dimensional models were viewed and manipulated in MacPyMOL (DeLano 2008).

2.2.2 U6 snRNA in Three Dimensions

MC-Sym is an RNA modeling program that is capable of generating a three-dimensional model from a primary nucleotide sequence, which can be supplied with or without secondary structure constraints (Parisien & Major 2008). Small RNA cycles, known as nucleotide cyclic motifs (NCMs), have been extracted from RNA crystal structures to generate a database of RNA tertiary structure building blocks. Since adjacent NCMs share a common border, a three dimensional model can be generated by overlapping the common edge of NCMs selected from the database. Thus an entire three-dimensional model can be produced through the step-wise addition of NCMs to a growing chain of building blocks.

To model our U6 snRNA secondary structure in three dimensions, we supplied MC-Sym with a 5' and 3' truncated U6 snRNA construct consisting of nucleotides 30-101 along with our secondary structure constraints. This construct contains the central stem and stem/loop A and B regions of U6 snRNA and will be referred to as the core domain. While MC-sym generated a number of potential structures, many of these were rejected for one of two reasons: either the path of the theoretical backbone contained bond angles that would produce substantial steric clashing, or the structure of the three-way junction did not agree with predictions based on empirical data collected from high resolution X-ray crystal structures. However, as predicted from the lengths of the single stranded junctions connecting the helices, MC-Sym did predict a model with a relatively compact three dimensional structure in which stem/loops A and B stack coaxially, and the central stem bends in toward stem/loop A (Fig. 3). The compact structure was made possible by wrapping JCA around the backside of the molecule (Fig. 3). This model was free of severe steric clashing and was in agreement with empirically derived predictions.

A second model generated by MC-Sym takes on a much more open conformation where the base pairing has been maintained, however twisting the central stem, in combination with the flexibility of the eight nucleotide single stranded junction, un-tucks this helix from folding up toward stem/loop A (Fig. 4). Karaduman et al. (2006, 2008) have provided evidence to suggest that binding of U6 snRNA associated proteins induces a conformational change that results in a much more open structure in the presence of Prp24. These authors accommodated this observation by simply disrupting base pairing of their extensively base paired protein-free U6 snRNA model to create a large

asymmetric bulge in which Prp24 could bind. The three dimensional models depicted here suggest that upon protein binding, U6 snRNA structure may open, however this conformational change does not require disruption of the base pairing interactions.

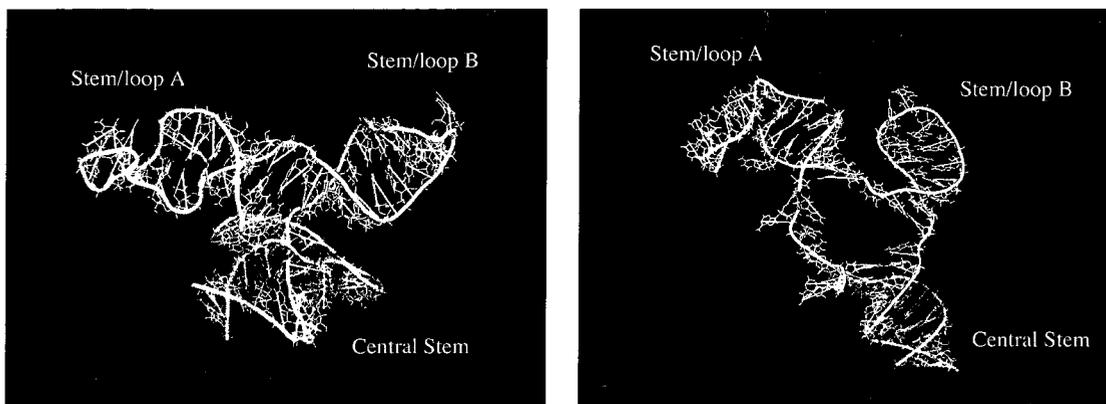


Figure 4. U6 snRNA can fold into a closed (left) and open (right) conformation without disrupting the base pairing interactions proposed in the Dunn-Rader model. The 5' end is blue and the 3' end is red.

To assess the overall flexibility of U6 snRNA, 2D Total Coherence Transfer Spectroscopy (TOCSY) Nuclear Magnetic Resonance (NMR) was performed in collaboration with Dr. Philip Johnson at York University. The TOCSY-NMR spectrum obtained indicated that there were nine or ten pyrimidines within the core domain of U6 that reside in unstructured regions of the molecule (Fig. 5). While the Karaduman et al. (2006) model of U6 snRNA secondary structure in the free U6 snRNP predicts sixteen pyrimidines in flexible, single-stranded regions of the core domain, the Dunn-Rader and Vidaver et al. (1999) models predict only nine and ten respectively. Thus this experiment revealed that the core domain of U6 snRNA was not as unstructured as Karaduman et al. (2006) propose for the protein bound U6. However, since the TOCSY-NMR experiment was performed on *in vitro* transcribed U6 in the absence of proteins, it is more

appropriate to consider the structure proposed by Karaduman et al. (2006) in the absence of protein. Since this structure also predicts nine unstructured pyrimidines, the TOCSY-NMR experiment was not capable of distinguishing between the models.

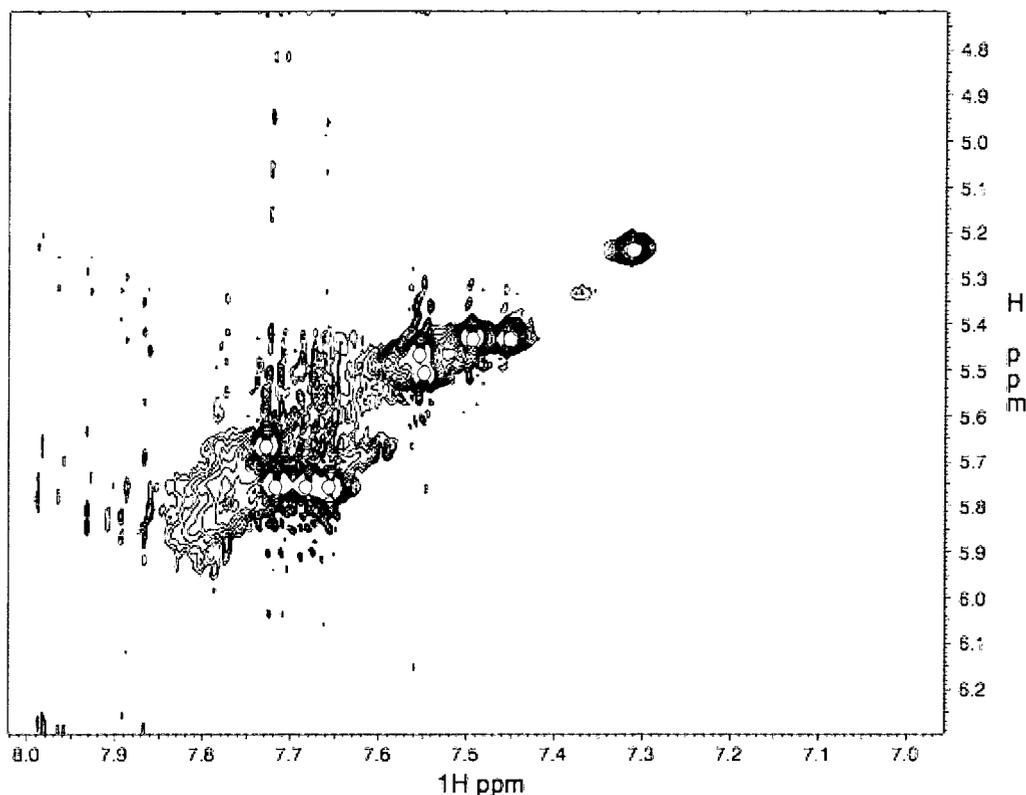


Figure 5. 2D TOCSY NMR spectrum of *in vitro* transcribed U6 snRNA.
Chemical shifts are given on the x and y axes and mobile pyrimidines are indicated in red.

2.3 Support for the Dunn-Rader Model from the Literature

An important first step in testing our new model of U6 snRNA secondary structure was to assess the compatibility with experimental results obtained previously by other labs. Extensive chemical structure probing and genetic techniques have been employed to study U6 snRNA conformation and function. Unfortunately, many of these experimental methods have generated results with ambiguous interpretations, and while

we do not dispute the authenticity of these data, we do question whether the data support various structural elements within the context of a free U6 snRNP as opposed to some other U6 containing complex that forms throughout the splicing cycle. Interestingly, much of the data present in the literature can be explained at least as well by the Dunn-Rader model, and in some cases more thoroughly, than either the Vidaver et al. (1999) or Karaduman et al. (2006) models. Here we present a brief analysis of such data analyzed in the context of each of the proposed free U6 snRNA conformations in order to assess the quality of each model based on known experimental information.

2.3.1 Chemical Modification of Accessible U6 snRNA Bases

U6 snRNA secondary structure, both in the absence and presence of protein, has been tested through chemical structure probing by several labs (Fortner et al. 1994, Jandrositz & Guthrie 1995, Karaduman et al. 2006). The principle behind these experiments is that various reagents can chemically modify nucleotides that have accessible Watson/Crick base pairing positions (Stern et al. 1988). During subsequent primer extension reactions, the activity of reverse transcriptase is blocked when the enzyme encounters such a modification, resulting in the production of truncated molecules that can be separated on a gel to identify the modified nucleotides. Thus chemical structure probing can be used to distinguish between nucleotides that are involved in canonical Watson/Crick base pairing, which are strongly protected from chemical modification, and those that are not, which become strongly modified.

While these experiments can be very informative, it is important to acknowledge that the experimental conditions greatly influence the results, and more importantly, that

the assignment of the level of modification to each nucleotide is a subjective process. Interpretation of the results is complicated even further by dynamic processes such as helix breathing where modification signals are weaker and repeated experiments may provide conflicting results. In addition, the protection of a nucleotide from chemical modification in the absence of proteins suggests that the nucleotide is base paired; however, no information is provided as to which nucleotide it is base paired with. Thus the results are ambiguous and should be interpreted with caution since they may support multiple secondary structure predictions as illustrated below.

The first chemical structure probing of *S. cerevisiae* U6 snRNA was performed *in vivo* by allowing the chemical reagent to diffuse into the cells (Fortner et al. 1994). Consequently the chemical modification pattern obtained was representative of all possible U6 snRNA conformations throughout its lifetime and was therefore not considered here. In contrast, Jandrositz and Guthrie (1995) and Karaduman et al. (2006) applied chemical structure probing to U6 snRNPs partially purified on a glycerol gradient or TAP-tag purified respectively. In addition, these two groups compared the snRNP modification pattern to protein-free U6 obtained either through hot phenol: chloroform extraction of RNA from peak U6 snRNP-containing glycerol gradient fractions, or from *in vitro* transcribed U6 snRNA respectively.

We have mapped the Jandrositz and Guthrie (1995) and Karaduman et al. (2006) chemical modification data in the absence and presence of protein to each of the three competing models of U6 snRNA secondary structure (Fig. 6, 7). Note that while the Vidaver et al. (1999) (VB) and Dunn-Rader (DR) models were considered for both the

RNA and RNP data, Karaduman et al. (2006) (KL) have proposed different structures in the absence and presence of protein and thus we consider each structure where appropriate. Both sets of data were consistent with the 5' stem/loop proposed in all models and therefore only residues following the 5' stem/loop region were examined here. Since it is not clear whether weak chemical modifications represent nucleotides that are base paired or single stranded, we have limited our analysis to only nucleotides for which there was a strong protection or strong modification in both sets of data. This reduced our analysis to only 13 residues in protein-free U6 snRNA and 22 residues in U6 snRNA in free U6 snRNPs.

Of the thirteen nucleotides that met our criteria, eight were consistent with the VB model, 11 were consistent with the DR model, and 12 were consistent with the KL model (Fig 6). It should be noted that Karaduman et al. (2006) have proposed a U•U mismatch as part of their extended stem structure and that such a mismatch would be expected to disrupt the canonical form of the helix (Leontis et al. 2002). Since both of these nucleotides were strongly protected, the number of nucleotides consistent with the data may be artificially high for the KL model if this interaction does not actually occur.

The U6 snRNP data set contained one nucleotide out of 22 where the strong modification from the Jandrositz and Guthrie (1995) data was not consistent with the strong protection observed by Karaduman et al. (2006). Of the remaining 21 nucleotides, 16 positions were consistent with the VB model, 16 were consistent with the DR model, and 13 were consistent with the KL model. Three of these nucleotides, A40, A41, and A42, were strongly modified in the protein free data set and strongly protected in the U6

Figure 6. Chemical modification of naked U6 snRNA and U6 snRNA in U6 snRNPs. (A) Data from Jandrositz & Guthrie (1995) and Karaduman et al. (2006) are given on the left and right of each grid respectively. Strongly modified residues are shown in a black box, weakly modified residues in gray, strongly protected residues in white, and residues for which there is no information are replaced by an X. The middle three columns represent the predicted modification (black) or protection (white) pattern for the Vidaver et al. (1999) (V), Dunn-Rader (D), and Karaduman et al. (2006) (K), models of U6 snRNA secondary structure. Residues that UV cross-link to Prp24, the Lsm complex, or both are indicated by one, two, or three asterisks respectively (Karaduman et al. 2006, 2008). Residues protected by hydroxyl radical cleavage are shown by small (weak protection) or large (strong protection) circles (Karaduman et al. 2006) or triangles (Ghetti & Abelson 1995).

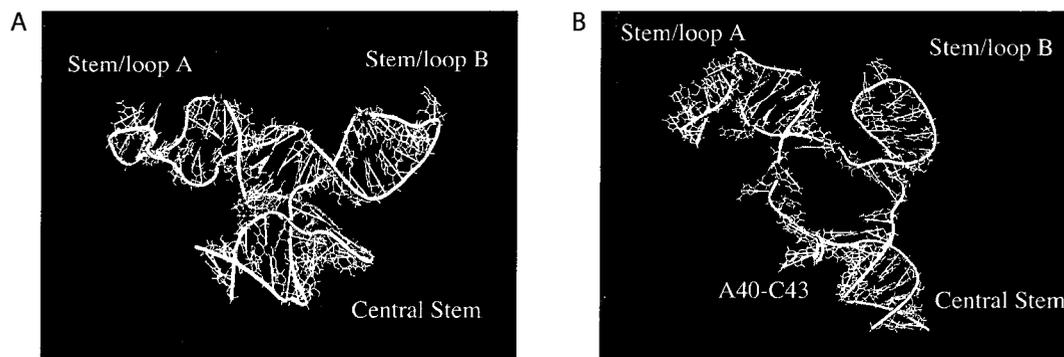


Figure 7. The Karaduman et al. (2006) chemical modification data for *in vitro* transcribed U6 snRNA and U6 snRNPs mapped to the proposed closed (A) and open (B) Dunn-Rader models. Strongly modified residues are red, weakly modified are orange, strongly protected are yellow, and no information are green.

snRNP data set. Further, treatment of free U6 snRNPs with SDS and proteinase K eliminated the strong protection of these residues (Jandrositz & Guthrie 1995). These observations, combined with strong protection of the RNA backbone from hydroxyl radical cleavage, suggest that residues A40-A42 are in fact single stranded and that protection in the snRNP was due to the presence of protein (Fig. 6, Jandrositz & Guthrie

1995, Karaduman et al. 2006). This increases the consistency between the snRNP data and the KL and DR models to 16 and 19 respectively.

In summary, while each of the models under consideration predicted dramatically different base pairing interactions, the DR model was no less consistent with chemical modification data than either the VB or KL models. In fact, while the consistency between the data and the DR model was approximately the same as that for the KL model when considering protein-free U6 snRNA, consistency between the DR model and the U6 snRNP data was slightly higher than the VB and KL models, providing additional support for the DR model.

2.3.2 Compensatory Base Mutations

One region of U6 snRNA secondary structure that has been thoroughly tested through compensatory base mutation is the telestem structure proposed by Vidaver et al. (1999). This structure can be divided into the upper and lower telestem, which are composed of base pairing between nucleotides 40-43 with 86-89, and 36-39 with 92-95 respectively. Ryan et al. (2002) showed through a series of compensatory base mutations that disruption of the lower telestem inhibited Prp24 binding to synthetic U6 snRNA *in vitro*, and that Prp24 binding could be restored by restoring base pairing in the lower telestem. In contrast, restoration of base pairing in the upper telestem did not restore Prp24 binding. While tri-nucleotide substitution of the 5' strand of the upper telestem inhibited Prp24 binding, tri-nucleotide substitution of the 3' strand of the upper telestem did not affect Prp24 binding, tri-snRNP assembly, or pre-mRNA splicing. Thus Ryan et al. (2002) concluded that base pairing in the upper telestem does not occur during

spliceosome assembly or splicing and that the lower telestem and residues 40-42 are important for Prp24 binding *in vitro*.

In a slightly more elaborate *in vivo* experiment, Vidaver et al. (1999) drew essentially the same conclusion as Ryan et al. (2002). These authors had created a cold sensitive mutant, U6-A62G, and isolated *cis*-acting suppressors in a previous publication (Fortner et al. 1994). Many of these suppressors fell into the telestem region and were predicted to suppress the cold sensitive phenotype of A62G, which was thought to hyperstabilize the 3' ISL (see below), by destabilizing the telestem. Vidaver et al. (1999) reasoned that if this truly was the mechanism of suppression, then introducing a mutation that restored base pairing in the telestem would result in reversion back to a cold sensitive growth phenotype. Consistent with this hypothesis, restoring base pairing at positions 36 and 37 with positions 94 and 95 in the lower telestem reverted the growth phenotype. However, restoration of base pairing between positions 41 and 42 with positions 87 and 88, and between residues 40 and 89 in the upper telestem did not revert the growth phenotype. Thus Vidaver et al. (1999) were able to find *in vivo* support for the lower telestem, but not the upper telestem.

Both the Karaduman et al. (2006) and Dunn-Rader models of U6 snRNA secondary structure propose the lower telestem as the upper half of the central stem, and do not incorporate residues 40-42 into any base pairing interactions. Thus both of these models have maintained the stem structure and adjacent single stranded nucleotides required for Prp24 binding. These structural features are consistent with the strong protection of nucleotides in the 5' strand of the upper central stem as well as with the

strong modification of residues 40-42 in the absence of Prp24, and the strong protection of these residues in the presence of Prp24 (Fig. 6, Jandrositz & Guthrie 1995, Karaduman et al. 2006). Further, the entire 5' strand of the central stem and residues 40-42 are strongly protected from hydroxyl radical cleavage, suggesting that a protein component of the U6 snRNP does indeed bind at this location (Fig. 6, Ghetti et al. 1995, Karaduman et al. 2006). We conclude that the compensatory analyses described above provide support for at least the upper half of the central stem, but not for the upper telestem proposed in the VB model.

2.3.3 The U6 snRNA 3' Intramolecular Stem/loop (3' ISL)

Previously proposed models of U6 snRNA secondary structure in free U6 snRNP suggest that a universally proposed 3' intramolecular stem structure must unwind to allow for interaction with U4 snRNA during spliceosome formation (Fig. 2, Epstein et al. 1980, Fortner et al. 1994, Vidaver et al. 1999, Ryan et al. 2002, Karaduman et al. 2006). This structure is then thought to reform in active spliceosomes where it coordinates a metal ion that has been shown to be important for the splicing reactions (Madhani & Guthrie 1992, Fortner et al. 1994, Yean et al. 2000). Hydroxyl radical cleavage of portions of the 3' ISL that are consistent with the turn of a helix places this stem structure in proximity of other important catalytic elements within the active spliceosome (Rhode et al. 2006).

Support for the 3' ISL in the free U6 snRNP comes largely from a genetic analysis of two mutants, A62G and A79G, which were predicted to hyperstabilize the yeast 3' ISL and resulted in the predicted cold sensitivity and U4/U6 di-snRNP assembly defect expected for stem hyperstabilization (Fortner et al. 1994, McManus et al. 2007). The U4/

U6 assembly defect, which was also present at 30C where growth was wild type, could be overcome by over-expressing U4, however, the cold sensitive phenotype persisted even when the U4/U6 assembly defect had been alleviated (Fortner et al. 1994). If the cold sensitive growth phenotype were due to a shortage of U4/U6, as predicted by the authors, then restoring U4/U6 levels would allow the yeast to grow in the cold, which was not observed. These observations suggest that destabilization of the 3' ISL is important at a later stage in spliceosome assembly and activation, or during the splicing reactions. Since destabilization of the 3' ISL would be required for interaction with U4 snRNA to generate the di-snRNP particle, these results argue against the presence of the stem in the free U6 snRNP.

Similar mutational analyses conducted in a mammalian system also argue strongly for the importance of the 3' ISL following, but not preceding, U4/U6 interaction. Disruption of the C62-G72 base pairing interaction with either a C62G or a G72C point mutation resulted in wild type U4/U6 levels and spliceosome assembly, however splicing was reduced to less than 10% (Wolff & Bindereif 1993). When the double mutation C62G/G72C, which inverted the base pair interaction and should therefore not affect the stability of the 3' ISL, was introduced, U4/U6 and spliceosome assembly levels were again comparable to wild type, however splicing had increased to between 10% and 50% (Wolff & Bindereif 1993). This observation suggested that formation of the C62-G72 base pairing interaction was important in the context of the active spliceosome (Wolff & Bindereif 1993).

To our knowledge there are currently no unambiguous data available to suggest that the 3' ISL forms in the free U6 snRNP. While the genetic data discussed above highlighted the importance of the formation of this structure in the active spliceosome, it neither supported nor disproved the existence of the structure in free U6 snRNP. Chemical structure probing of U6 snRNA in free U6 snRNPs was just as consistent with the formation of the 3' ISL as they were with the replacement of this structure with stem/loops A and B as proposed in the Dunn-Rader model (see above). Thus the available genetic data at this time are fully consistent with the base pairing interactions proposed in the Dunn-Rader model.

2.3.4 Genetic Suppression Analysis – A62G and A79G

Suppressors of the cold sensitive growth phenotype exhibited by both the A62G and A79G point mutations have been proposed to exert their effect through two different mechanisms: destabilization of the 3' ISL, or destabilization of the telestem or telestem/Prp24 interactions (Fortner et al. 1994, Vidaver et al. 1999, McManus et al. 2007). We agree with these modes of suppression, however we suggest that suppressors that act through 3' ISL destabilization do so in the context of the active spliceosome rather than in free U6 snRNPs (Fig. 8). Destabilization of this structure may be required during active site structural re-arrangement between splicing steps and/or for spliceosome disassembly. Thus hyperstabilizing the 3' ISL with the original A62G or A79G mutations may stall the active spliceosome and prevent recycling of the spliceosomal components, slowing growth of these strains at low temperature.

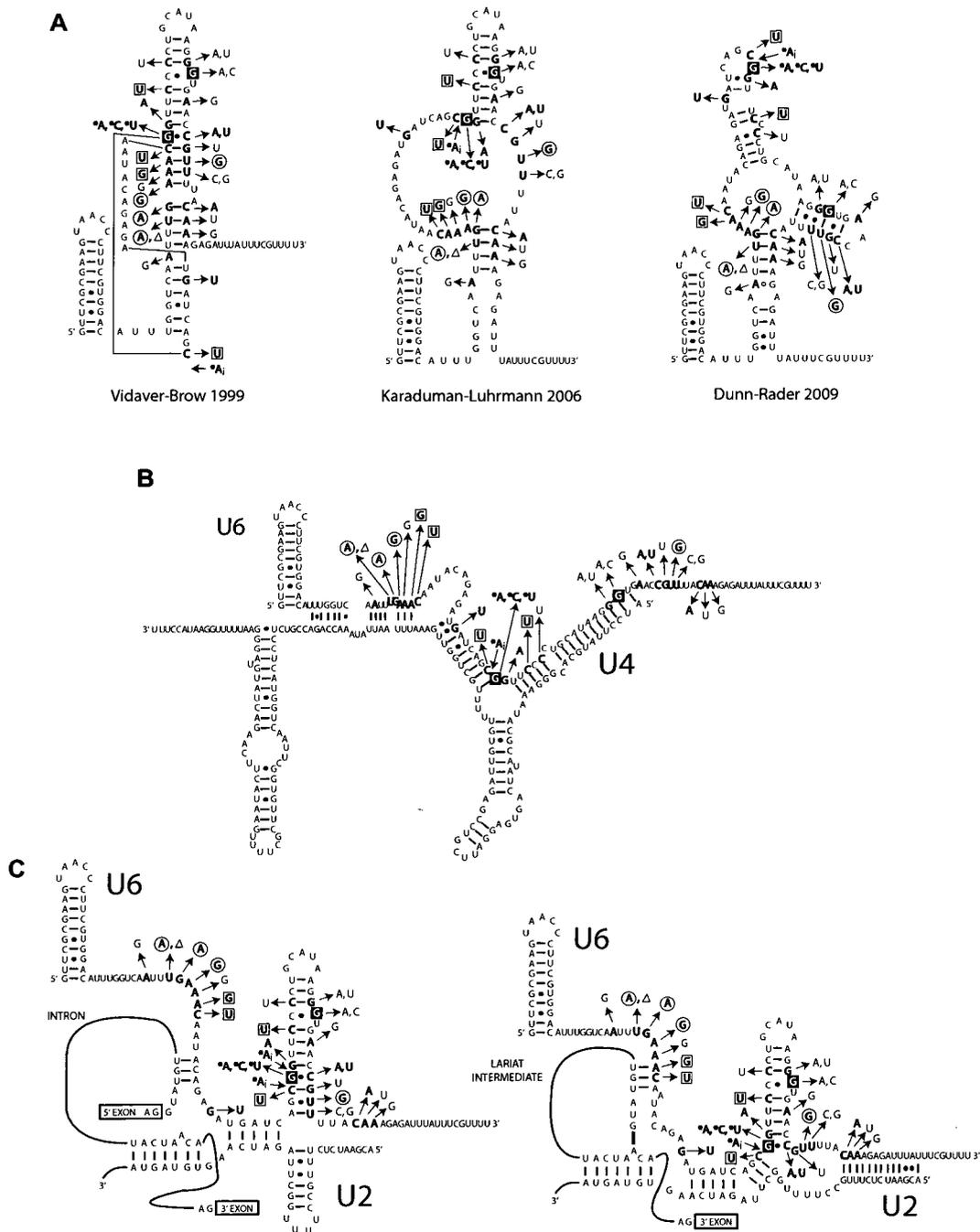


Figure 8. A62G and A79G suppressor mutations mapped onto U6 snRNA secondary structures in the free U6 snRNP (A), di-snRNP (B), and active spliceosome (C). The A62G and A79G mutations are indicated in a black box. Suppressor mutations are indicated as follows: Bold: suppress A62G, regular font: suppress A79G, bold with an asterisk: suppress A62G but has no affect on A79G, circled bold: suppress both A62G and A79G, square bold: suppress A62G but enhance A79G.

In the context of the Dunn-Rader model of U6 snRNA secondary structure, A62G may increase the stability of stem/loop A by creating an additional base pair at the top of the stem while A79G would have a relatively mild affect on stem/loop B stability (Fig. 8). Many of the suppressors of both mutations map to the Prp24 binding site located at the upper central stem and residues 40-43, and may act through central stem or central stem/Prp24 destabilization as proposed by Fortner et al. (1994) and Vidaver et al. (1999). In most cases, suppressors that map outside of this region may provide a slight destabilizing affect on stem/loop A or B. Interestingly, many of the A79G suppressors cluster in the stem/loop B region very close to the original A79G mutation (Fig. 8). Nevertheless, many of these suppressors likely exert their effect at a later stage in the splicing cycle in the context of the 3' ISL.

If the sole effect of the A62G and A79G mutations were to hyperstabilize the 3' ISL, then suppressors of one mutation would be expected to also suppress the other. While this was true for four of the A62G suppressors, four others suppressed A62G and had no affect when combined with A79G, and four others suppressed A62G but actually enhanced the A79G phenotype (Fig. 8, McManus et al. 2007). Thus something more is going on with one or both of these mutations beyond 3' ISL hyperstabilization. In the context of the active spliceosome, protonation of A79 in the C67•A79 mismatch facilitates coordination of a catalytically important Mg^{2+} in the 3' ISL (Huppler et al. 2002). A to C is the only known sequence variation at position 79, and is also the only mutation that would allow for similar protonation and maintenance of almost perfect isostericity with the C67•A79 mismatch (Huppler et al. 2002). Interestingly, the only two

A79G suppressors identified at position 79 were reversion back to an A, or mutation to a C (McManus et al. 2007).

2.3.5 Exonuclease Degradation of U6 snRNA

The 3' end of U6 snRNA is generally protected from 3' exonuclease activity *in vivo* as a result of binding of the Lsm complex to this region. However, U6 snRNA added to sub-cellular or splicing extracts is typically not protected from this activity and the length of degradation products can provide some insight into the U6 snRNA secondary structure. 3' exonuclease activity is thought to occur progressively from the 3' end of the molecule to the 3' end of a stem structure where nuclease activity becomes blocked (Eschenlauer et al. 1993). Full length U6 transcribed in yeast sub-cellular extract is routinely degraded to a 90-nucleotide fragment, suggesting that a stem structure ends with a base pair containing residue 90 (Margottin et al. 1991, Eschenlauer et al. 1993).

This observation is not consistent with any of the proposed U6 snRNA secondary structures in free U6 snRNP, the di-snRNA, or the active spliceosome. Residue 90 is predicted to either lie in a single stranded region, where nuclease activity should continue, or in the middle of a stem structure, where nuclease activity should be blocked several nucleotides earlier, in all but the Dunn-Rader model where nucleotide 90 is found at the 3' edge of stem/loop B (Fig. 2, 3). However, in order for exonuclease activity to continue through to stem/loop B, the central stem would have to be disrupted. It is possible that the *in vitro* transcription assay produces U6 transcripts that have not yet fully folded, thus disruption of the long-range central stem interaction, if it has formed at

all, may allow for degradation up to the first stably formed stem, which may be stem/loop B.

Ryan et al. (2002) have also observed 3' degradation products of synthetic full length and 3' truncated versions of U6 snRNA reconstituted into U6 snRNPs in U6 snRNA-depleted splicing extract. Full length U6 snRNA was typically degraded to 101 nucleotides while shortened versions consisting of residues 1-94 and 1-91 were degraded to 90 nucleotides, and even shorter versions containing residues 1-88, 1-86, and 1-84 were degraded to 81 nucleotides (Ryan et al. 2002). Degradation of full length U6 to 101 nucleotides is consistent with degradation continuing up to the 3' edge of U2/U6 helix II as suggested by Ryan et al. (2002), but is also consistent with degradation up to the 3' edge of the intramolecular U6 snRNA central stem, proposed in both the Karaduman et al. (2006) and Dunn-Rader models (Fig. 2, 3). Degradation of the 1-94 and 1-91 constructs is consistent with degradation to the 3' edge of stem/loop B in the Dunn-Rader model, since much of the central stem base pairing region has been eliminated in these constructs, allowing for exonuclease activity to proceed to stem/loop B (Fig. 3). A degradation product of 81 nucleotides is not consistent with any of the proposed U6 snRNA secondary structures except for U6 snRNA in the U4/U6 di-snRNP (Fig. 2, Ryan et al. 2002).

Chapter Three – Structure Probing by Oligonucleotide Accessibility

Accessible regions of RNA are capable of annealing to short complementary oligonucleotides while regions of RNA that are involved in RNA/RNA or RNA/protein interactions are not. The accessibility of segments of an RNA molecule to a complementary oligonucleotide can therefore indicate a great deal about the overall secondary structure of the molecule. These types of experiments have been employed both in the absence and presence of proteins as well as in both human and yeast splicing systems (Black & Steitz 1986, Blencowe et al. 1989, Fabrizio et al. 1989, Li & Brow 1993, Brow & Vidaver 1995). Here we have tested the Dunn-Rader model of U6 snRNA secondary structure by designing oligonucleotides that were complementary to *S. cerevisiae* U6 snRNA and examining the ability or inability of these oligonucleotides to anneal to yeast U6 snRNA in a protein-free system.

3.1 Materials and Methods

Cold Phenol:Chloroform Total RNA Extraction

One milliliter of wild type W303 yeast culture, grown overnight to an OD₅₉₅ between one and two, was spun down in a microcentrifuge tube in an Eppendorf Centrifuge 5417C at 13,000 rpm for one minute at room temperature. The cell pellet was washed twice with cold, sterile dH₂O prior to resuspension in 300µL chilled, filtered RNA extraction buffer (50mM Tris-HCl pH 7.5, 100mM NaCl, 10mM EDTA). Acid washed, baked 0.5mm Zirconia/Silica beads (BioSpec Products Inc.) were added to the 200µL mark and tubes were vortexed for one minute on the maximum setting. Following a five minute incubation on ice, the tubes were vortexed for an additional minute before

adding 300 μ L chilled RNA extraction buffer [60 μ L 10% SDS, and 400 μ L acid phenol: chloroform (5:1, pH 4) (Ambion)]. The samples were then vortexed for one minute on the highest setting and centrifuged in an Eppendorf Centrifuge 5415D at 13,200 rpm for two minutes at 4C. The aqueous phase was transferred to a tube containing 500 μ L cold acid phenol: chloroform (5:1, pH 4) (Ambion), was vortexed for one minute, and centrifuged for two minutes as before. A third phenol: chloroform (5:1, pH 4) (Ambion) extraction was performed followed by an extraction with 500 μ L chloroform (Sigma) as described above. The aqueous phase from the chloroform extraction was transferred to a clean eppendorf tube where 40 μ L 3M sodium acetate and 1mL 100% ethanol was added. Samples were cooled at -80C for at least 20 minutes. Precipitated RNA was pelleted by centrifugation in an Eppendorf Centrifuge 5415D at 13,200 rpm for 20 minutes at 4C. The pellets were washed with 70% ethanol and allowed to air dry for 5-10 minutes prior to resuspension in 30 μ L 10mM Tris-HCl, pH 7.5. The final nucleic acid concentration was determined by Nanodrop.

5' End Labeling of DNA Oligonucleotides

Fifty picomoles of DNA oligonucleotide were incubated with 2.5 μ L of 10X T4 Polynucleotide Kinase Buffer (NEB), 20 units of T4 Polynucleotide Kinase (PNK) (NEB), and 3 μ L of [γ -³²P]-ATP (PerkinElmer) in a final reaction volume of 25 μ L. The reactions were incubated for one hour at 37C followed by heat inactivation of the T4 PNK for twenty minutes at 65C. The reaction was diluted to 50 μ L with dH₂O and unincorporated [γ -³²P]-ATP was removed using either an Illustra MicroSpin G25 Column

or an illustra ProbeQuant G50 Micro Column (GE Healthcare) according to the manufacturer's instructions.

Oligonucleotide Accessibility Experiments

Twenty micrograms of total RNA was incubated with 1 μ L (approximately 1 pmol) of 5' end labeled DNA oligonucleotide and 1 μ L of Hybridization Buffer [150mM NaCl, 50mM Tris-HCl pH 7.5, 1mM EDTA] at either 4C or 70C for 40 minutes. Reactions that were pre-incubated with unlabeled oligonucleotide were first heat treated at 70C for 5 minutes in the presence of 1 μ M unlabeled oligonucleotide and were then allowed to slow cool to room temperature for 15 minutes prior to incubation with the labeled oligonucleotide for 20 minutes. All reactions were moved to ice prior to gel loading and each reaction was loaded into a pre-chilled 9% non-denaturing polyacrylamide gel and the gel was run at 250V for 1 hour. The gel was then exposed to a phosphor-imager screen at -80C and the resulting autoradiogram was visualized and quantified with a Cyclone phosphor imager and OptiQuant Software[©] (Packard Instruments).

Temperature course experiments with oSDR464 were conducted by incubating 2 μ L of 5' end labeled oSDR464 with 100 pmol of *in vitro* transcribed U6 snRNA (Ambion MEGAshortscript Kit) for 15 minutes at 4C, 10C, 19C, 40C, and 70C. An additional heat treatment was performed in which the incubation was placed at 70C for 5 minutes and then slow cooled to room temperature for 10 minutes. Each reaction was loaded into a pre-chilled 9% non-denaturing polyacrylamide gel and the gel was run at 250V for 1 hour 20 minutes. The gel was exposed to a phosphor imager screen and visualized as above.

Oligonucleotide GC content and melting temperature (T_m) were calculated using the Integrated DNA Technologies (IDT) SciTools Oligo Analyzer 3.1 at <http://www.idtdna.com> (Owczarzy et al. 2008).

Table 1. DNA Oligonucleotide Sequence, GC Content, and Melting Temperature

Oligo Number	Sequence	Length	% GC	T _m (C)
oSDR180	5'-AAAACGAAATAAATCTCTTTG-3'	21nts	23.8	44.8
oSDR461	5'-ATCCTTATGCAGGGGAAC-3'	18nts	50	51.6
oSDR464	5'-CATCTCTGTAT-3'	11nts	36.4	24.9
oSDR465	5'-GATCATCTCTGTATTG-3'	16nts	37.5	39.5
oSDR466	5'-AATCTCTTTGTAAAAC-3'	16nts	25	36.8
oSDR467	5'-TTGTTTCAAATTGACC-3'	16nts	31.3	41.1
oSDR494	5'-CTTCGCGAAC-3'	10nts	60	34.4

3.2 Results and Discussion

The Dunn-Rader model of U6 snRNA secondary structure was tested using an oligonucleotide (oligo) accessibility experiment with several oligos that were designed to be complementary to specific regions of U6 snRNA sequence (Fig. 9). The oligos were 5' end labeled in order to detect annealing to U6 snRNA, which would generate a complex that migrated through a non-denaturing polyacrylamide gel with a slower mobility than the oligo alone. Thus we expected to observe a shift in the mobility of oligos that were predicted to anneal to single stranded regions of U6 upon incubation with yeast total RNA, and no shift for oligos designed to anneal to regions of U6 predicted to take part in base pairing interactions. Since the oligos were incubated with total RNA under non-denaturing conditions, some oligos were also predicted to anneal to

U6 in the U4/U6 complex and therefore we also expected to see oligo shifts of an even larger U4/U6/oligo complex running more slowly than the U6/oligo complex.

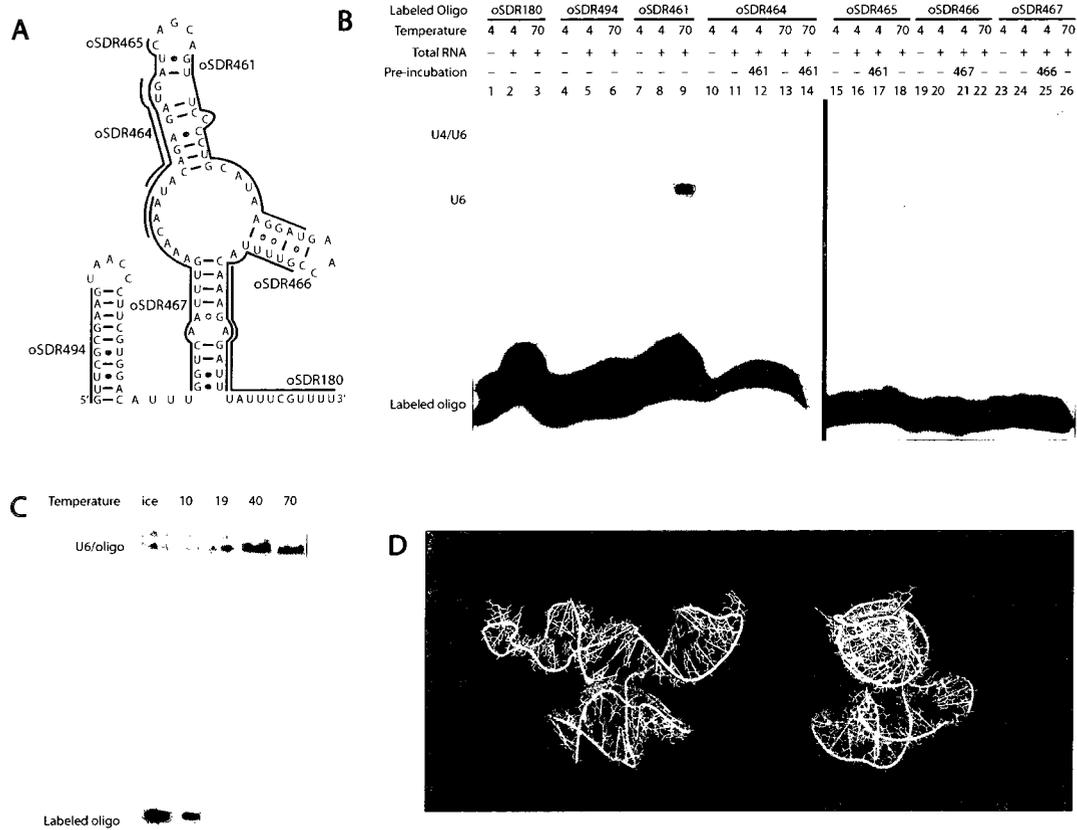


Figure 9. U6 snRNA secondary structure probing by oligonucleotide accessibility. (A) The annealing sites of complementary DNA oligonucleotides are indicated on the Dunn-Rader model of U6 snRNA secondary structure. (B) Oligonucleotide accessibility of U6 snRNA in cold phenol: chloroform extracted yeast total RNA. Reaction conditions are indicated above each lane. (C) Temperature course of oSDR464 annealing with *in vitro* transcribe U6 snRNA. (D) oSDR466 (blue) and oSDR467 (red) annealing sites mapped onto the MC-Sym three-dimensional model of U6 snRNA secondary structure.

oSDR494 was designed to be complementary to the 5' strand of the proposed 5' stem/loop and was used here to visualize inefficient annealing since all evidence suggests that this structure does exist in free U6 and U4/U6 (Fig. 9). oSDR494 did not anneal to free U6 snRNA or the U4/U6 di-snRNA complex at either 4C or following heating at 70C (Fig. 9). This result was not surprising since the estimated melting temperature of the 5' stem/loop is approximately 95C. Even if this structure became destabilized enough at 70C for some annealing of oSDR494, the much stronger intramolecular base pairing interaction would probably out-compete the U6/oligo interaction, knocking the oligo off of the RNA to regenerate the hairpin. The absence of oSDR494 annealing is consistent with the existence of a 5' stem/loop as predicted by all proposed models of U6 snRNA secondary structure.

Sequestering the ACAGAGA region in a base pairing interaction is unique to the Dunn-Rader model. All other models of U6 snRNA in free U6 snRNP, as well as an extensively base paired model for protein-free U6 snRNA proposed by Karaduman et al. (2006), predict that this sequence is located in a bulge and should therefore be accessible to a labeled oligo. oSDR464 was designed to be complementary to an eleven nucleotide stretch containing the ACAGAGA sequence and was used here to distinguish between our model, where it would not be expected to anneal, and all other proposed models where it would be expected to anneal. Consistent with only the Dunn-Rader model, oSDR464 did not efficiently anneal to free U6 at 4C, however this oligo also did not anneal any more stably to U6 when samples were heat-treated (Fig. 9). We expected that heat treatment would result in destabilization of the overall U6 snRNA structure,

allowing for efficient annealing of the labeled oligo. However, since oSDR464 was capable of annealing to *in vitro* transcribed U6 snRNA, it was clear that the oligo should have annealed to the ACAGAGA sequence had it been accessible as proposed by all other models (Fig. 9).

In contrast to U6 in total RNA, annealing of oSDR464 to *in vitro* transcribed U6 occurred even at low temperatures, and increased with increasing temperature. One explanation for the discrepancy in behavior of the oligo when incubated with yeast total RNA compared to *in vitro* transcribed U6 is that the RNAs may not fold into the same structure. *In vivo*, U6 snRNA is bound by the Lsm complex, which is important for U6 snRNA stability, and Prp24, which has been proposed to stabilize base pairing of the long range interaction between nucleotides 35-39 and 92-96 (Vidaver et al. 1999, Cooper et al. 1995). Conducting the RNA extraction on ice stabilized base pairing even when the proteins were removed by phenol: chloroform extraction. In the absence of these proteins, as was the case for *in vitro* transcribed U6, the long-range central stem interaction may not form, leading to an overall less stable U6 structure. Thus while the *in vitro* transcribed U6 may still sequester the ACAGAGA sequence through base pairing in stem/loop A, oSDR464 may disrupt stem/loop A base pairing more easily in the absence of longer range RNA/RNA interactions.

While the oSDR464 annealing experiments were supportive of only the Dunn-Rader model, the results were somewhat dissatisfying in that we were unable to observe annealing to U6 following heat treatment when incubated with yeast total RNA. We therefore designed a second oligo, oSDR465, which also spanned the ACAGAGA

sequence but consisted of a two and three nucleotide extension at the 3' and 5' ends of oSRD464 respectively. This oligo also did not anneal to free U6 efficiently at 4C, but did stably associate with U6 following heat treatment at 70C (Fig. 9, compare lanes 16 and 18).

Since the melting temperature of the U4/U6 di-snRNA complex is approximately 53C (Brow & Guthrie 1988), the free U6 snRNA signal following heat treatment represents both the free U6 snRNA as well as U6 snRNA that has been released from the U4/U6 di-snRNA complex. Ideally, if a region of U6 snRNA is accessible in both free U6 and the U4/U6 di-snRNA, then the sum of these signals at 4C should be approximately equal to the U6 signal at 70C. If the annealing site is inaccessible, then annealing at low temperature will be relatively inefficient while annealing following denaturation will be much more efficient, resulting in a much greater U6 signal in heat-treated samples than the combined U6 and U4/U6 signals at low temperatures.

The oSDR465/free U6 signal was approximately five-fold stronger than the combined U4/U6 and free U6 signals at 4C, indicating that approximately 80% of the available U6 was inaccessible at low temperatures. While the Dunn-Rader model predicts that this annealing site is inaccessible, all other models of yeast secondary structure predict this region to be single stranded. Further, strong chemical modification in the di-snRNP indicates that this annealing site is fully accessible in the U4/U6 complex (Jandrositz & Guthrie 1995). Thus the inefficient annealing to free U6 at 4C and the five-fold increase in signal upon heating strongly support the Dunn-Rader model where heat treatment of total RNA would disrupt base pairing in stem/loop A, releasing the

ACAGAGA sequence and allowing for efficient annealing between oSDR465 and U6. All other proposed models of U6 snRNA secondary structure should have resulted in much more stable association of the oligo at low temperatures.

To test the prediction that disruption of stem/loop A would expose the ACAGAGA sequence to allow for stable association of complementary oligos, we attempted to open the structure by pre-incubating the total RNA preparation with unlabeled oSDR461 prior to incubation with either labeled oSDR464 or oSDR465. Labeled oSDR461 alone did not efficiently anneal to U6 unless heat treated, thus pre-incubations with the unlabeled oligo were heated to 70C and slow cooled to allow annealing prior to the addition of oSDR464 or oSDR465 (Fig. 9, compare lanes 8 and 9). Unfortunately, these experiments were inconclusive since the short heat treatment resulted in some destabilization of the U4/U6 species and it was consequently not clear whether the increase in intensity of the free U6 signal was the result of increased annealing of open free U6, or if the increase was the result of increased free U6 levels due to the destabilization of U4/U6 base pairing.

oSDR180, also known as U6 6D in the literature, has been used extensively by many splicing labs as a probe for Northern blot and solution hybridization since it anneals well to the U6 3' tail in both free U6 snRNA and U4/U6 di-snRNA (Li & Brow 1993). When the sample was heat treated at 70C, a much stronger free U6 signal was observed (Fig 9, lane 3). Some of this signal increase was a result of U4/U6 disassembly due to the high temperature, however the U6 signal at 70C was still approximately two-fold higher than the combined U4/U6 and U6 intensities at 4C.

The increased free U6 signal upon heating suggests that the region of U6 that is complementary to oSDR180 was moderately inaccessible at lower temperatures. While all models of yeast U6 snRNA secondary structure proposed prior to the Karaduman et al. (2006) and Dunn-Rader models predict the oSDR180 annealing site to be single stranded, the central stem in the Karaduman et al. (2006) and Dunn-Rader models predict half of this annealing site to be occluded through base pairing (Fig. 9). Thus a possible explanation for the two-fold increase in signal at 70C is that at lower temperatures oSDR180 was capable of annealing to the 3' half of the U6 annealing site, but that this interaction, with an estimated melting temperature of approximately 22C, was relatively unstable and may not have allowed for efficient annealing. Upon heating, denaturation of the central stem may have released the remaining half of the oSDR180 annealing site, allowing for much more stable association where the oligo/U6 complex had an estimated melting temperature of approximately 45C.

Interestingly, the annealing pattern of oSDR467 matched that observed for oSDR180 where some annealing occurred at 4C and an approximately two-fold increase in annealing was observed upon heat treatment (Fig. 9). oSDR467 was designed to anneal to the opposite strand of the central stem, and like oSDR180, extended into a single stranded region 3' of the stem. The annealing pattern observed for oSDR467 can be explained in the same way as that observed for oSDR180; that is, at low temperatures some oligo annealing was observed because the single stranded region of U6 could serve as a nucleation site for annealing, however the interaction was through so few residues that it was relatively unstable. Heating the RNA would have disrupted base pairing in the

central stem, allowing for more stable association of the oligo. The identical annealing patterns for similarly constructed oligos designed to anneal to opposite strands of the same stem offer strong support for the presence of the stem structure.

In contrast to oSDR180 and oSDR467, oSDR466, which was complementary to the same strand of the central stem as oSDR180 but extended to the 5' side of the stem into the 3' strand of stem/loop B, did not efficiently anneal to U6 at 4C or 70C (Fig. 9). This result was not too surprising since this annealing site in U6 was predicted to be base paired through both the central stem and stem/loop B, and further, when the annealing site was located in the three-dimensionally modeled structure, it became evident that portions of this region of U6 may be buried within the overall U6 structure (Fig. 9). The stem/loop B hairpin itself is quite stable, with an estimated melting temperature above 40C and in the context of the overall three-dimensional structure, may be even higher. Thus cooling on ice following the heat treatment at 70C may result in rapid regeneration of stem/loop B, resulting in the destabilization of any remaining oSDR466/U6 interactions.

The difference in annealing observed for oSDR180 and oSDR466 brings into question the interpretation of the results of a similar experiment conducted by Brow and Vidaver (1995) using HeLa cell nuclear extract. In this experiment an oligo that was complementary to nucleotides 85-99 of the human U6 snRNA sequence, hU63', showed the same inability to anneal to U6 as oSDR466, even when heat-treated to 90C and slow cooled. This observation has long since been interpreted as support for the proposal that yeast and mammalian U6 snRNA fold into different secondary structures, however, the

hU63' annealing site did not extend all the way through to the 3' end of human U6 like oSDR180 in yeast, but instead left seven nucleotides 3' of the annealing site. oSDR466, while spanning much of the same region of yeast U6 as oSDR180, also did not extend all the way through to the 3' end of the molecule. We have shown here that although oSDR466 was only shifted from the 3' end of the molecule by eleven nucleotides, this shift was enough to generate dramatically different annealing behaviors between the two oligos. We suggest that oSDR466 is a more appropriate candidate for comparison to hU63' since the annealing sites are more similar than that of oSDR180.

oSDR466 and oSDR467 were designed to anneal to similar regions of yeast U6 snRNA as two oligos designed by Brow and Vidaver (1995) to anneal to human U6, hU63' and hU6cen2 respectively. While hU63' did not anneal to free U6, hU6cen2 strongly associated with both free U6 and U4/U6. The authors attributed this discrepancy in annealing to an unexplained ability of hU6cen2 to more easily invade the proposed stem structure. However, the similar annealing capabilities observed here for the equivalent oligos in yeast, oSDR466 and oSDR467 respectively, suggest that the human and yeast structures are much more similar in this region than previously thought. Thus we propose that like oSDR466, hU63' was complementary to a region of human U6 that was largely inaccessible while hU6cen2 annealed to a region of human U6 that was at least partially accessible, much like oSDR467. This is consistent with the latest human U6 snRNA secondary structure proposed by Karaduman et al. (2006) where a central stem, similar in composition to the yeast central stem, has been proposed for human U6.

Interestingly, annealing of hU63' to free U6 snRNA can be stimulated by first pre-incubating the HeLa cell extract with unlabeled hU6cen2 (Brow & Vidaver 1995). We conducted the same experiment using yeast total RNA pre-incubated with unlabeled oSDR467 to determine whether or not annealing of oSDR466 could be similarly encouraged. Although the oSDR466 signals were often quite weak, quantification of the bands revealed similar annealing to U4/U6 and an approximately 1.5 fold increase in annealing to free U6 (Fig. 9). In contrast, pre-incubation of yeast total RNA with oSDR466 prior to incubation with labeled oSDR467 did not stimulate annealing of oSDR467 to either free U6 or U4/U6 (Fig. 9).

In conclusion, all of the oligos tested supported the Dunn-Rader model with the most important of those being the ACAGAGA complement, oSDR465, whose inability to anneal at low temperatures was only consistent with the Dunn-Rader model. The two-fold increase in annealing of heat-treated incubations of oSDR180 and oSDR467 was consistent with partial occlusion of the complementary U6 sequence through base pairing in the central stem at lower temperatures. The inefficient annealing of oSDR466 at both 4C and in heat-treated incubations was consistent with the complement of this oligo being both base paired and buried within the U6 structure. The comparable behavior of oSDR466 and oSDR467 to two oligos designed to anneal to human U6 snRNA suggested that the yeast and human U6 snRNA secondary structures are more similar than previously thought.

Chapter Four – Genetic Support for the Lower Central Stem

The central stem was proposed by Karaduman et al. (2006) and is therefore not unique to the Dunn-Rader model of U6 snRNA secondary structure. Nevertheless, this structural element should be tested experimentally since evidence for its existence is still lacking. Protection of residues from chemical modifying reagents in the 5' strand of the stem suggests that these residues are in fact base paired, however chemical modification data for the 3' strand of the stem is unavailable since the primer extension oligonucleotide anneals to U6 snRNA in this region (Jandrositz & Guthrie 1995, Karaduman et al. 2006). Further, the upper half of the central stem has been supported by compensatory mutation, however the lower half has not been subjected to genetic analysis (Vidaver et al. 1999, Ryan et al. 2002). To test the lower half of the stem, we designed three double mutations in the bottom of the central stem and subsequently evaluated the *in vivo* effect of the mutations on cell growth, as well as U4/U6 assembly.

4.1 Materials and Methods

Generation of Mutant SNR6

Point mutations were introduced into *SNR6* via PCR-mediated site directed mutagenesis of pSR85 (pUC + 1.2 Kb EcoRI-PstI genomic fragment containing *S. cerevisiae SNR6*) using 2.5 units of Pfu Turbo AD DNA polymerase (Stratagene). PCRs were treated with 20 units of DpnI (NEB) for 7 hours to ensure complete digestion of wild type *SNR6* template plasmid. DpnI digests were subsequently ethanol precipitated and transformed into RbCl₂ competent DH5 α by standard methods. Transformations were plated onto LB plates containing 0.05mg/mL carbenicillin and incubated overnight at

37C. A single colony of each mutant was selected and grown in liquid LB containing 0.05mg/mL carbenicillin overnight at 37C on a shaker at 200rpm. Plasmid DNA was collected from each strain through alkaline lysis and approximately 20µg was subsequently subjected to restriction enzyme digest with 20 units each of EcoRI (NEB) and SphI (NEB) for 3 hours at 37C. The *SNR6* containing EcoRI-SphI fragment was inserted into the yeast vector pSE358, which carries a TRP1 marker, in a standard ligation reaction containing 400 units of T4 DNA ligase (NEB) and a 3:1 molar ratio of insert to vector. Ligation reactions were incubated for 1.5 hours at 37C, heat inactivated for 20 minutes at 65C, and subsequently transformed into RbCl₂ competent DH5α, plated on LB – carbenicillin plates, and incubated as above. A single colony of each mutant was grown up in liquid culture as above and the plasmid was subsequently isolated by Qiagen® plasmid mini-prep according to the Qiagen mini-prep handbook. Two micrograms of each plasmid in a total volume of 20µL was sent to UNBC sequencing facilities to ensure the presence of the appropriate mutations.

PCR Primer Pairs:

U6 – G30U/G31U

oSDR551 FWD: 5'-CCCTTCGTGGACATTTTTTCAATTTGAAACAATAC-3'

oSDR552 REV: 5'-GTATTGTTTCAAATTGAAAAATGTCCACGAAGGG-3'

U6 – U100C/U101C

oSDR547 FWD: 5'-CCGTTTTACAAAGAGACCTATTTTCGTTTTTTTTTTATC-3'

oSDR548 REV: 5'-GATAAAAAAAAAAACGAAATAGGTCTCTTTGTAAAACGG-3'

U6 – U100A/U101A

oSDR549 FWD: 5'-CCGTTTTACAAAGAGAAATATTTTCGTTTTTTTTTTATC-3'

oSDR550 REV: 5'-GATAAAAAAAAAAACGAAATATTTCTCTTTGTAAAACGG-3'

Yeast Transformations and Plasmid Shuffle

Mutant *SNR6* sequences were introduced into the yeast strain YHM1 (genomic deletion of *SNR6* covered by a wild type *SNR6* carried on a URA3 marked YCp50 plasmid) by the plasmid shuffle technique. Approximately 15 μ L of YHM1 was collected from an overnight plate patch and resuspended in 100 μ L of TEL [10mM Tris pH 7.5, 3.33mM EDTA, 100mM lithium acetate]. Forty micrograms of sheared salmon sperm DNA, approximately 500ng of Qiagen® prepped mutant plasmid, and 1mL of PEG-TEL [10mM Tris pH 7.5, 3.33mM EDTA, 100mM lithium acetate, 40% PEG-4000] were added to the cells and mixed well. Following an overnight incubation at room temperature, the transformations were heat shocked at 42C for 10 minutes and then centrifuged for 30 seconds at 3000rpm. As much PEG-TEL was removed as possible and cells were resuspended in 200 μ L of TE [10mM Tris pH 7.5, 3.33mM EDTA]. The transformations were plated on trp (-) media and incubated at 30C for three days. A single colony of each mutant was selected and streaked onto fresh trp (-) media and incubated at 30C for three days. These plates were then replica plated in the order trp (-), ura (-), 5-FOA, and YPD, and all plates were incubated at 30C for three days. A single colony of each mutant that grew on trp (-), 5-FOA and YPD, but not ura (-) was selected and grown overnight to saturation at 30C in 3mL of liquid YPD. Glycerol stocks of each mutant were made by mixing one part of cell culture to two parts of 50% glycerol. Stocks were snap frozen in liquid nitrogen and stored at -80C.

Growth Phenotypes by Dot Dilution

Each U6 mutant yeast strain was grown up in 3mL of liquid YPD to an OD₅₉₅ between 2.0 and 3.0. A five-fold dilution series of eight dilutions was generated with the most concentrated of the set having a concentration of 1 OD/mL. Dilution series were set up in 96 well plates and were plated onto three different YPD plates with a frogger. The plates were then incubated at 18C, 30C, or 37C for three days.

Preparation of Yeast Total RNA

Each U6 mutant containing yeast strain was grown up from the glycerol stock on YPD plates at 30C for three days. A single colony of each was then grown overnight in liquid YPD at 30C to an OD₅₉₅ between 1.5 and 2.5. One milliliter of culture was collected for cold phenol: chloroform total RNA extraction (see chapter 3). Two 500µL aliquots of culture were collected and spun down for temperature shift. One pellet was resuspended in 100µL YPD pre-incubated to 18C while the other was resuspended in 100µL YPD pre-incubated to 37C. The resuspended cells were used to inoculate 2mL of YPD that had been pre-incubated at 18 or 37C. Temperature shifted cultures were incubated at the appropriate temperature for three hours and then 1mL was collected for cold phenol: chloroform total RNA extraction as before.

Northern Blots

Ten micrograms of total RNA was loaded onto a pre-chilled 4.5% non-denaturing polyacrylamide gel and samples were subjected to electrophoresis in the cold room for 1.5 hours at 250V. The RNA was transferred to Amersham Hybond-N⁺ nylon transfer membrane (GE Healthcare) for 30 minutes at 450mA in a semi-dry electro-blotter (Owl

Panther Hep-3) and was subsequently cross-linked to the membrane in a UV Stratalinker 1800 (Stratagene) with 120,000 J of ultraviolet radiation. The membrane was blocked with 7mL Rapid-hyb Buffer (GE Healthcare) for 45 minutes at 65C, and blocking continued while the temperature was slowly lowered to approximately 30C (approximately 30 minutes). 5' end labeled U4 probe oligonucleotide (U4 14B) was added and allowed to incubate with the membrane for one hour at 30C. The membrane was washed with 7mL Wash Solution [0.2% SDS, 6X SSC] four times: the first for two minutes, the second for 20 minutes, and the remaining two washes for 40 minutes. The membrane was exposed to a phosphor imager screen overnight and the resulting autoradiogram was visualized and quantified with a Cyclone phosphor imager and OptiQuant Software (Packard Instruments). Membranes were stripped with boiling Northern Strip Solution (0.1X SSC, 0.1% SDS, 40mM Tris pH 7.5) twice for 10 minutes and were re-probed with 5' end labeled U6 probe oligonucleotide (U6 6D) as described above.

4.2 Results and Discussion

Genetic manipulations have been used extensively to test predicted base pairing interactions throughout splicing, providing strong support for interactions between U6 and other splicing snRNAs as well as the U6 3' ISL and lower telestem (Madhani & Guthrie 1992, Lesser & Guthrie 1993, Fortner et al. 1994, Vidaver et al. 1999, Ryan et al. 2002). Mutations that increase the stability of a stem are thought to generate a cold sensitive growth phenotype since lowering the temperature further stabilizes the already hyperstabilized structure. Consequently the stem may be more difficult to unwind for

subsequent structural re-arrangements. Mutations that destabilize a stem are thought to exhibit a heat sensitive growth phenotype since increasing the temperature is predicted to destabilize the structure further, hindering its formation *in vivo*. The double mutations constructed here were designed to either hyperstabilize (U100C/U101C) or destabilize (G30U/G31U and U100A/U101A) the bottom two central stem G•U wobbles and were therefore predicted to generate a cold or heat sensitive growth phenotype respectively.

To test the central stem mutations *in vivo*, a yeast strain carrying only the mutant U6 snRNA gene, *SNR6*, was constructed to ensure that any phenotype observed was the result of introducing the mutant sequence. The genomic copy of *SNR6* was replaced by *LEU2* in the yeast strain employed in this experiment, and since U6 snRNA is an essential gene product, a wild type copy of *SNR6* carried on a *URA3* marked plasmid was present to cover the genomic deletion. Mutant *SNR6* sequences were introduced by plasmid shuffling where the mutant sequences, carried on a *TRP1* marked yeast vector, were transformed into the yeast strain (Fig. 10). The presence of *TRP1* allowed these yeast strains to grow in media lacking tryptophan and therefore selective pressure to take up the *TRP1* marked plasmid, and consequently the mutant *SNR6*, was applied by growing the strain on trp (-) plates (Fig. 10). No selective pressure was applied to maintain the wild type *SNR6* carried on the *URA3* marked plasmid and this plasmid was eventually lost. The inability to grow on media lacking uracil and the ability to grow on media containing 5-Fluoroorotic Acid (5-FOA), a chemical that in the presence of the *URA3* gene is converted to fluorodeoxyuridine, which is toxic to yeast, confirmed that this plasmid had been successfully shuffled out (Fig. 10).

The growth phenotype of the mutant U6 snRNA strains were tested by dot dilution on YPD plates incubated at 18C, 30C, and 37C for three days. All yeast strains tested, including the wild type control and a strain serving as a control for cold sensitivity, U6-A62G, showed wild type growth at 30C (Fig. 11). U100A/U101A also showed wild type growth at 18C and 37C while U100C/U101C showed wild type growth at 37C and severe inhibition of growth at 18C (Fig. 11). Interestingly, G30U/G31U did not exhibit wild type growth at elevated or lowered temperatures. A weak growth defect at 37C and a severe growth defect at 18C were observed for this strain (Fig. 11).

The Karaduman et al. (2006) and the Dunn-Rader models of U6 snRNA secondary structure predict that the U100A/U101A and G30U/G31U mutations would destabilize the central stem. Such destabilization may be expected to push the equilibrium between free U6 and U4/U6 toward U4/U6 formation, and since this is the direction that spliceosome assembly proceeds, it was not clear if driving formation of the di-snRNA complex would generate a detectable growth phenotype at elevated temperatures. In fact, the U100A/U101A mutation grew as well as the wild type strain at 37C, suggesting that maintenance of the lower portion of the stem may not be critical for growth if the central stem in fact exists. Additionally, this result indicated that mutating nucleotides U100 and U101 to adenosines did not interfere with any RNA-RNA or RNA-protein interactions that may be important for progression through splicing.

In contrast, the weak heat sensitivity of the G30U/G31U mutant observed at 37C was consistent with central stem destabilization. However, since the G30U/G31U and U100A/U101A mutants were predicted to destabilize the same base pairing interactions,

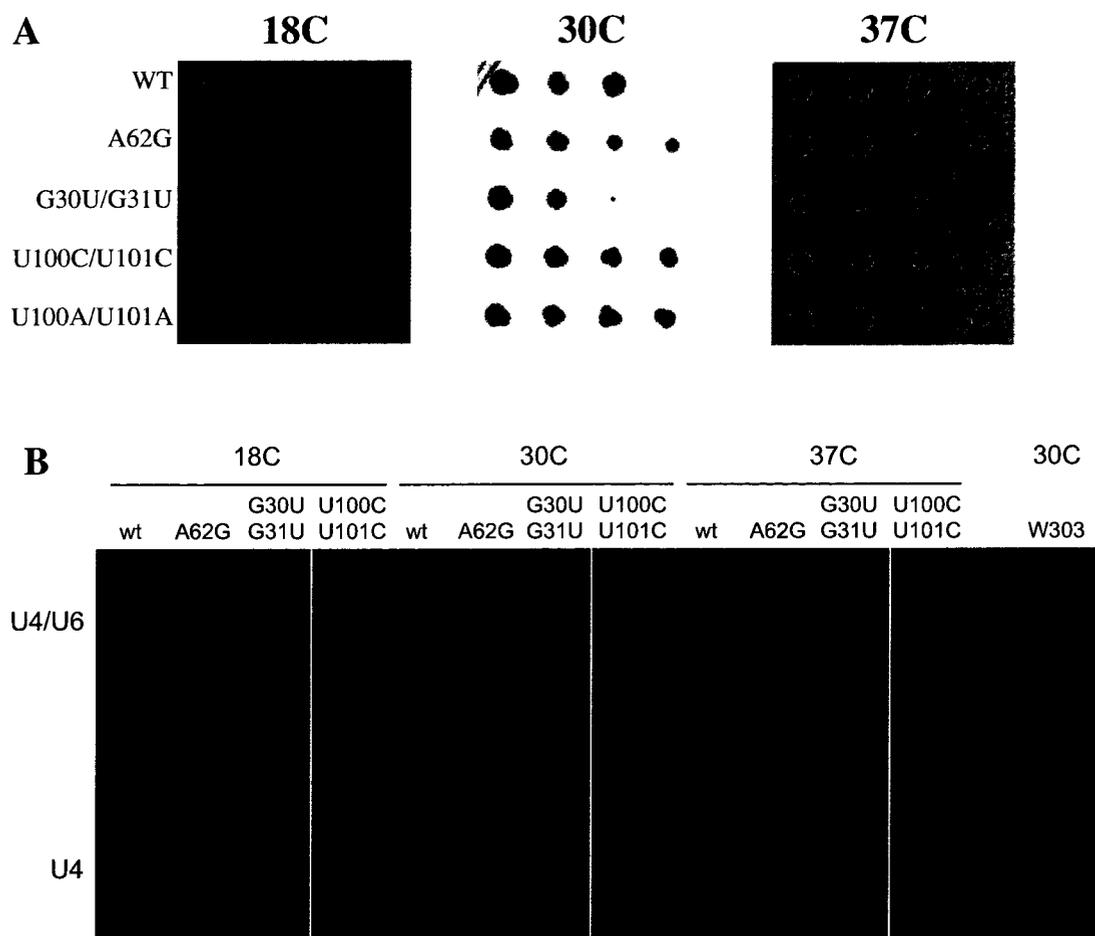


Figure 11. Growth phenotypes and U4/U6 assembly phenotypes of the central stem double mutants, a wild type control, and a control for cold sensitivity, U6-A62G. (A) U6 mutant growth phenotypes were tested by serial dot dilution on YPD plates that were incubated for three days at the temperature indicated. (B) U4/U6 assembly Northern blot probed for U4 snRNA. Total RNA was extracted from the central stem mutants after shifting to the temperature indicated and was separated on a 4.5% non-denaturing polyacrylamide gel.

the slightly different growth phenotypes exhibited by these mutants suggest that either the proposed base pairing interaction does not form *in vivo*, or that G30 and G31 are involved in some additional interaction(s) that is also disrupted by the double mutation. While we cannot rule out the first interpretation, we favor the latter since growth of this mutant

strain was severely inhibited at 18C and total RNA extracted from this strain grown at 18C, 30C, and 37C revealed a defect in U4/U6 assembly (Fig. 11). The U4/U6 assembly defect was evident by the increased levels of free U4 snRNA and diminished levels of U4/U6 in this strain compared to a wild type strain. These observations suggest that mutation of these nucleotides either increases the stability of free U6 or shifts the equilibrium toward free U6 by destabilizing interactions with U4.

The G30U/G31U mutation may destabilize a third phylogenetically conserved helix in U4/U6 that has been proposed at positions 35 to 42 in yeast U6 snRNA (Jakab et al. 1997). Introduction of a tri-nucleotide bulge in U4 snRNA could potentially extend base pairing by an additional eight base pairs to generate a relatively stable helix composed of fourteen base pairs with a di- and tri-nucleotide bulge in the U6 and U4 strands of the helix respectively. This lengthened helix would extend through both G30 and G31, which would be predicted to base pair with U4-C80 and C81. Thus the G30U/G31U mutation may result in a U4/U6 assembly defect by destabilizing this U4/U6 interaction, preventing efficient annealing of U4 and U6 during di-snRNP formation. Strong protection of G30, G31, and C33 from chemical modification in the U4/U6 di-snRNP suggests that this helix may form in yeast, however it is also possible that the protection of these nucleotides may be due to the presence of proteins in the di-snRNP particle (Jandrositz & Guthrie 1995).

An alternative explanation for the unexpected G30U/G31U phenotypes is that mutation of these residues may increase the stability of U6 snRNA/protein interactions in free U6 snRNP. G30 has been UV cross-linked to Lsm2 of the Lsm complex and

neighboring residues U28 and U29 have been UV cross-linked to both Prp24 and Lsm2 (Karaduman et al. 2006, 2008). Thus it is possible that introducing additional uridines may provide a stronger binding site for one or more of these proteins. The Lsm complex in particular is known to bind the uridine rich 3' tail of U6 snRNA and the proximity of this natural Lsm binding site to the lower central stem, where we have introduced additional uridine residues, may result in aberrant Lsm binding. Strengthening the interaction between U6 snRNA and its associated proteins may account for the severe cold sensitive growth phenotype and U4/U6 assembly defect exhibited by this mutation.

It is interesting to note that the U4/U6 assembly defect of G30U/G31U was present at 30C where the strain exhibited wild type growth. A similar result was observed for U6-A62G where a U4/U6 assembly defect was present at 18C and 30C while growth was severely inhibited or comparable to wild type at these temperatures respectively (Fortner et al. 1994). Additionally we have shown that although the U4/U6 assembly defect of A62G persisted at 37C, growth was also comparable to wild type at this elevated temperature (Fig. 11). These observations demonstrate that yeast growth phenotypes are not sensitive enough to detect all molecular disruptions, and further, that the mechanisms governing the generation of growth phenotypes are not well understood at this time.

In contrast to the destabilizing mutations, U100C/U101C was predicted to hyperstabilize the central stem, which was consistent with the observed severe cold sensitivity at 18C. The central stem is expected to destabilize at some point throughout the splicing cycle to accommodate interactions between U2 and U6 to form U2/U6 helix

II. However, at low temperatures where central stem destabilization may be less efficient, U2/U6 helix II formation may be impeded. This interpretation is consistent with the observation that the U100C/U101C mutation did not result in a defect in U4/U6 assembly, suggesting that the associated growth defect is generated following U4/U6 di-snRNP formation (Fig. 11). Since the U4/U6 interaction region of U6 lies outside of the central stem, disruption of the central stem at an early stage in spliceosome assembly may not be important, however disruption of the central stem following U4/U6 association may be a critical step in U6 activation in functional spliceosomes.

Human U6 snRNA has been proposed to carry a long-range intramolecular interaction of the 3' tail and an internal region of U6 through to the di-snRNP where RNase V1 cleavages on the 5' strand of the helix lend some support to this proposed base pairing interaction (Brow & Vidaver 1995, Mougin et al. 2002). In fact, Brow and Vidaver (1995) proposed that U2/U6 helix II formation actually stabilizes the U4/U6 interaction by antagonizing reformation of the U6 long-range intramolecular interaction. Thus it is possible that the proposed central stem does not disassemble until U6 and U2 snRNA come into contact, at which time disassembly of the central stem may allow for formation of both the third U4/U6 helix and U2/U6 helix II.

In conclusion, we have shown that the double mutations generated in the lower half of the central stem produce growth phenotypes and U4/U6 assembly defects consistent with either disruption or hyperstabilization of this stem. Surprisingly, however, mutations in the predicted central stem region also give rise to unexpected phenotypes that cannot be easily explained by existing models. These phenotypes suggest that the

central stem, while forming in the free U6 snRNP, may persist through to the di- and tri-snRNP complexes until U6 engages in interactions with U2 snRNA. Much of this hypothesis should be tested in a more rigorous manner with additional mutational analyses, including compensatory mutations and isolation and identification of both *cis* and *trans* acting suppressors of the observed growth phenotypes. Additionally, pull down assays could be performed to assess the level of U6 snRNA complexed with the U6 snRNP-associated proteins to determine whether or not introduction of these mutations has created additional protein recognition sites as proposed.

Chapter Five – Co-variation of U6 snRNA Secondary Structure and Prp24 Homolog Architecture

RNA sequence can change throughout evolution without dramatically altering the RNA secondary or tertiary structure, as long as mutations that occur in regions important to the overall structure are compensated for by additional mutation(s) (Leontis et al. 2002). This principle has opened up an entire field of RNA structure prediction based on the idea that nucleotide co-variation at a base pair position confirms the existence of that base pair, and that co-variation at only a few positions within a putative helix can confirm the existence of that helix (Woese et al. 1983, Gutell et al. 1994). As a testament to the power of comparative sequence analysis, approximately 97% of the predicted 16S and 23S rRNA base pairs have since been confirmed in high resolution crystal structures (Gutell et al. 2002).

Early comparative sequence analyses were strictly based on co-variations between the standard Watson-Crick base pair interactions, G-C, A-U, and G•U (Woese et al. 1983). It is now clear that many non-canonical base pairs occupy the same or very similar geometries as the standard base pairs and can thus substitute without substantial perturbation of the RNA structure (Leontis et al. 2002). This chapter examines the conservation of the Dunn-Rader model of U6 snRNA secondary structure throughout evolution by analyzing sequence co-variations within U6 snRNA from more than 70 different organisms, including U6atac from the minor spliceosome of plants and animals. This analysis revealed a pattern to some of the differences in U6 snRNA sequence and may represent an evolving secondary structure. Interestingly, these patterned changes in

U6 snRNA sequence and/or secondary structure are accompanied by structured differences in the protein domain architecture of Prp24 homologs.

5.1 Materials and Methods

U6 snRNA and U6atac sequences were obtained from the NCBI website using the BLAST algorithm and from the Rfam database (Altschul et al. 1990, Gardner et al. 2009). Multiple sequence alignments were generated with CLUSTALW in the software program MEGA 4.0 and the alignment was subsequently optimized by manual inspection (Thompson et al. 1994, Tamura et al. 2007). NCBI accession numbers are given for each sequence in figure 12. U6 snRNA evolutionary trees were constructed in MEGA 4.0 under bootstrap conditions of 1000 replicates where both transitions and transversions were allowed to occur at a uniform rate.

Prp24 homologs were obtained from the NCBI website using the BLAST algorithm and protein domains were identified with the Simple Modular Architecture Research Tool (SMART) at <http://smart.embl-heidelberg.de/> in normal mode with default settings (Schultz et al. 1998). Homologs used in our analysis are listed below along with a description of the BLAST hit and the NCBI accession number.

Table 2. Putative Prp24 Homologs

Organism Name	Description	NCBI Accession #
<i>S. cerevisiae</i>	Prp24	AY723858
<i>V. polyspora</i>	Hypothetical proteinKpol_1069p8	XP_001643044
<i>K. lactis</i>	Unnamed protein product	XP_454447
<i>M. griseae</i>	70-15 hypothetical protein MGG_04727	XP_362282
<i>S. pombe</i>	RNA-binding protein Prp24	NP_596721
<i>A. fumigatus</i>	Af293 pre-mRNA splicing factor Prp24	XP_749317

Organism Name	Description	NCBI Accession #
<i>N. crassa</i>	OR74A hypothetical protein NCU09048	XP_958552
<i>A. capsulatus</i>	NAm1 conserved hypothetical protein	XP_001536207
<i>P. anserina</i>	Unnamed protein product	XP_001903653
<i>D. hansenii</i>	CBS767 hypothetical protein DEHA0F08250g	XP_460681
<i>L. elongisporus</i>	NRRL YB-4239 hypothetical protein LELG_00207	XP_001527687
<i>Y. lipolytica</i>	Hypothetical protein	XP_504884
<i>P. stipitis</i>	CBS 6054 U4/U6 snRNA-associated splicing factor	XP_001384763
<i>P. nodorum</i>	SN15 hypothetical protein SNOG_11646	EAT80690
<i>C. albicans</i>	SC5314 hypothetical protein CaO19.10095	XP_718326
<i>H. sapiens</i>	Squamous cell carcinoma antigen recognized by T cells	NP_055521
<i>R. norvegicus</i>	Squamous cell carcinoma antigen recognized by T cells	NP_001100626
<i>D. melanogaster</i>	RNA binding protein	CAA75535
<i>O. sativa</i>	Os08g0113200	NP_001060838
<i>C. elegans</i>	Protein B0035.12, partially confirmed by transcript evidence	CAA97405

5.2 Phylogenetic Conservation of U6 snRNA Secondary Structure

We have generated a comprehensive alignment of U6 snRNA sequences from 72 different organisms in order to identify positions of genetic co-variation in our proposed stem structures (Fig. 12). Although these sequences span all eukaryotic kingdoms, more than half of the sequences collected are from fungal organisms while the remaining sequences include representatives from other unicellular organisms, the euglenozoa, through to the most complex plants and animals. Analysis of the U6 snRNA sequence from such a diverse group of organisms should allow for the identification of subtle

changes in sequence between closely related species and more pronounced changes in sequence between more distantly related species.

The integrity of the 5' stem/loop has been tested phylogenetically and multiple instances of genetic co-variation support the existence of this structure (Brow & Guthrie 1988). Our much more extensive sequence alignment provided many more examples of co-variation in the 5' stem/loop, lending further support to this structural feature (Fig. 12). While the ability to form the stem has been conserved throughout evolution, the length of the stem varies from three to ten base pairs and the loop varies from a highly stable UUCG tetra-loop in most species to either a penta-loop or a hexa-loop in some fungal species. A bulged residue is present in many stems, however the location of the bulge and whether it is present on the 5' or 3' strand of the stem does not appear to be well conserved.

The proposed central stem is strongly supported by genetic co-variation and the overall length of this stem appears to also be well conserved with only small variations in the length of the lower portion of the proposed structure (Fig. 12). True co-variations have been identified at four of the nine base pair positions: the U-A second from the top of the *S. cerevisiae* central stem is replaced by a G-C in the euglenozoa species while the U-A directly below has undergone complete conversion to A-U in all non-fungal species (Fig. 12). In the lower portion of the stem, the C-G directly below the bulge is A-U in plants and the G•U second from the bottom has completely inverted to U•G in the non-fungal species (Fig. 12). Examples of sequence variation that provide weak support for a

base pair, that is G•U to G-C or A-U, can also be found at almost every position in this proposed stem structure (Fig. 12).

Interestingly three invariant nucleotides, A34, G96, and A97 (numbered as in *S. cerevisiae*), cluster at the central stem bulge where both adenosines are bulged and G96 is engaged in an A-G mismatch in all but the euglenozoa where the guanosine is predicted to form a C-G base pair. A34 has not been shown to interact with any other RNA or protein throughout splicing, nor has it been proposed to do so. G96 and A97 have been shown to interact genetically with U2 snRNA through the formation of helix II, but have not been proposed to interact with any other RNA or protein (Datta & Weiner 1991, Wu & Manley 1991, Field & Friesen 1996).

The ability to form helix II is conserved in yeast and human, however phenotypes associated with disruption of this stem in human and yeast systems are surprisingly different. In humans, mutation or complete deletion of U2 or U6 residues in the helix II region is detrimental *in vivo* and abolishes splicing *in vitro*, while similar mutations in yeast have no observable effect *in vivo* (Datta & Weiner 1991, Field & Friesen 1996). This finding suggests that these nucleotides may have been strictly conserved for some other function and clustering in the central stem suggests that the nucleotides may function in the context of the free snRNP, perhaps mediating tertiary RNA-RNA interactions or serving as a protein recognition element. We intend to explore these possibilities through a mutational analysis of these residues.

In contrast to both the 5' stem/loop and the central stem, stem/loop A provides little opportunity to identify genetic co-variation due to the high level of sequence

conservation in this region (Fig. 12). The majority of residues residing in the loop and the 5' strand of the stem are invariant in our sequence alignment, however sequence variations in the 3' strand of the stem do provide some weak support for the proposed stem structure. Much of this support comes from the fungi since the more complex plants and animals included in our alignment have identical sequences throughout the entire proposed stem/loop structure to that of the *Aspergillus* species and other closely related fungi (Fig. 12). Both proposed G-C base pairs within the *S. cerevisiae* stem are supported by a C to U mutation, creating a G•U interaction in fungal species such as *D. hansenii*, *L. elongisporus*, and *P. stipitis* (Fig. 12). The C-G base pair at the base of the stem is supported with a single C to U change in *C. fasciculata*, the only confirmed ACAGAGA variant, to create a U•G interaction at this location as well (Fig. 12, Xu et al. 1994).

One stretch of nucleotides in the 3' strand of stem/loop A is particularly interesting since there appears to be three major sequence variants and one minor variant in this region. In the *Saccharomyces* species and other closely related fungi, positions 63-65 (numbers as in *S. cerevisiae* sequence) consist of a GUU sequence while this stretch of nucleotides is UUG in fungal species such as *D. hansenii*, *L. elongisporus*, and *P. stipitis*. The third major variant, UGG, is present in fungi including the *Aspergillus* species as well as all plant and animal sequences present in our alignment (Fig. 12). A less common UUU sequence has been identified in the *Candida* species *albicans* and *tropicalis*, although *Candida glabrata* is found to have a GUU sequence. These variants change the base pairing pattern in the upper half of the proposed stem structure slightly, reducing the bulge in the 5' strand of the stem to a single nucleotide while the invariant A

at position 62 may become base paired to reduce the proposed penta-loop to a tetra-loop in all GUU variants (Fig. 12).

GUU to UGG, the dominant variant, creates tandem A-G mismatches beside a bulged nucleotide in the 5' strand of stem/loop A. The C1'-C1' distance in a standard *cis* Watson-Crick/Watson-Crick base pair is 10.3Å and the only orientation of an A-G mismatch that would maintain a near canonical helical form is a *cis* interaction between the adenosine Watson-Crick edge and the guanosine Hoogsteen edge, generating a C1'-C1' distance of 10.5Å (Leontis et al. 2002). This interaction has been observed at a resolution of 1.9Å and requires protonation of the adenine N1 position in order to generate a base pairing interaction mediated through two hydrogen bonds (Leontis et al. 2002). All other A-G interactions generate C1'-C1' distances that range between 9.0Å in a *cis* Watson-Crick/Sugar edge interaction to 12.5Å in a *cis* Watson-Crick/Watson-Crick interaction, resulting in distortion of the helical form (Leontis et al. 2002). Consequently, A-G mismatches are typically found at the ends of helices or beside internal bulges within a helix where the altered C1'-C1' distance can be accommodated without disruption of the helix (Elgavish et al. 2001). This is consistent with the Dunn-Rader model where the tandem A-G mismatches are located next to a bulged guanosine.

Given the slight structural changes predicted for the GUU variants, we modeled the UGG variant with MC-Sym in order to determine first if MC-Sym could generate a stem structure under our secondary structure constraints, second, if a three-dimensional model could be produced, how it would differ from the GUU containing stem structure, and third, what form of base pairing interaction the A-G mismatches may be involved in

and how the helix accommodates such mismatches. Surprisingly MC-Sym is capable of generating similar three-dimensional structures for both the GUU and UGG variants despite the introduction of non-canonical interactions in the upper half of stem/loop A (Fig. 13). Both A-G mismatches in the modeled UGG variant are of the type *cis* Watson-Crick/Watson-Crick and the positioning of the tandem A-G mismatch twists the stem in such a way as to accommodate the unpaired guanosine on the inside of the helix where it may contribute to helix stability through base stacking interactions.

Sequence variation support for stem/loop B is limited to a subset of fungal species including the *Saccharomyces* species as well as the unicellular euglenozoa. The *S. cerevisiae* A79-U87 base pair is an A•C mismatch in a number of fungal species, and this A•C mismatch is completely converted to a C•A mismatch in the euglenozoa (Fig. 12). Several examples of G•U to G-C or A-U are also present in this proposed stem along with G•U to G-A variation. Surprisingly, despite an insertion of two or three unpaired nucleotides in the 3' strand of the stem, MC-Sym was able to produce a similar three-dimensional structure for human stem/loop B as for yeast (Fig. 13). However, stem B in mammals is probably less stable than that of yeast since the mammalian stem is predicted to contain non-canonical base pairing in addition to the unpaired residues (Fig. 12, 13). A less stable stem may accommodate binding of the mammalian Prp24 homolog which is approximately three times larger than yeast Prp24 (see below, Shannon & Guthrie 1991, Bell et al. 2002).

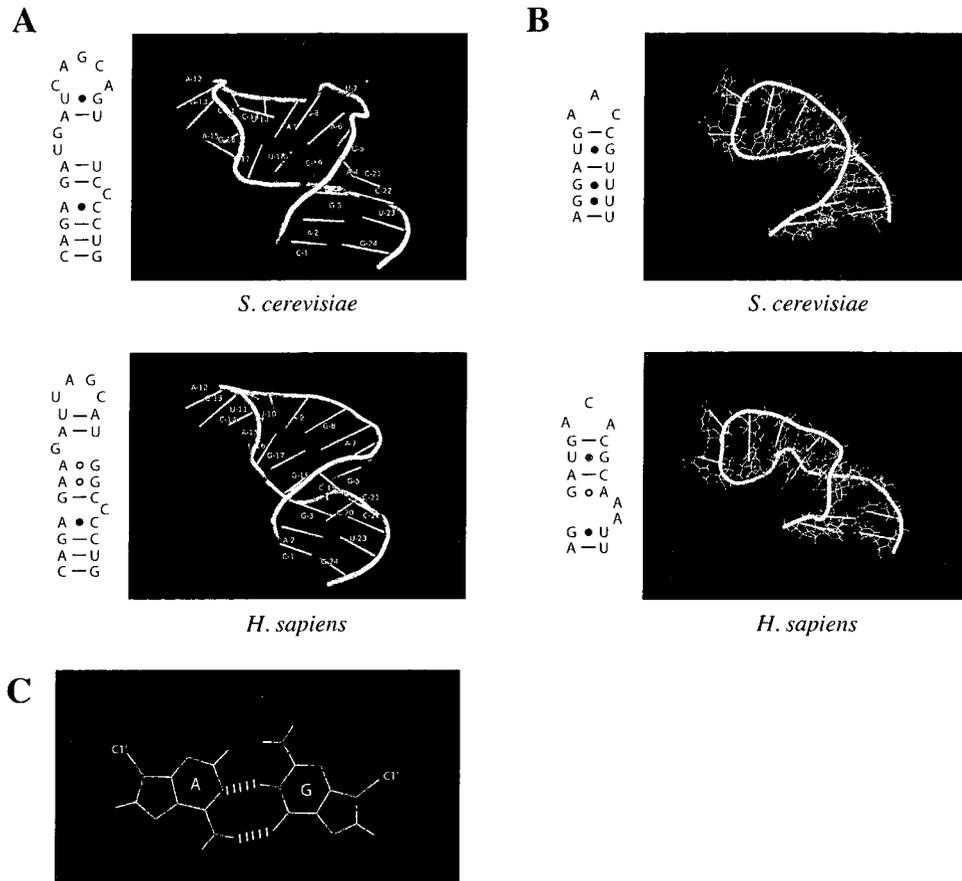


Figure 13. *S. cerevisiae* and *H. sapiens* stem/loop A (A) and B (B) can fold into similar three-dimensional structures. Both A-G mismatches in the *H. sapiens* stem/loop A are of the type *cis* Watson/Crick Watson/Crick (C).

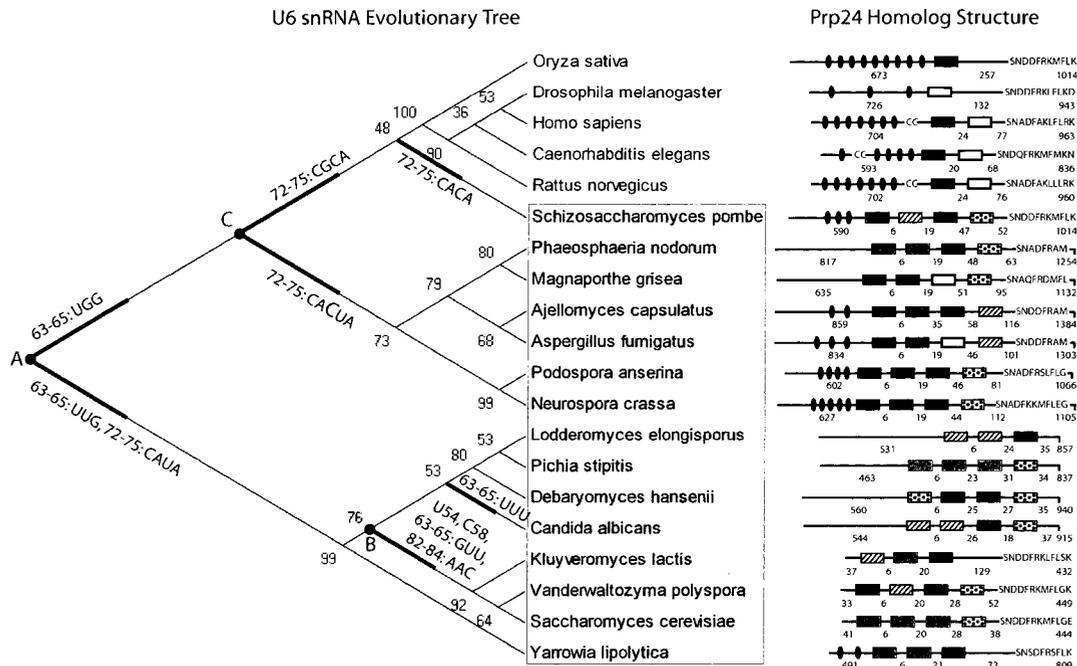
Interestingly U6atac from the minor spliceosome of plants and animals has the ability to generate the same structure as our proposed U6 model with the exception that the 5' stem/loop is replaced by a 3' stem/loop that is not present in U6 (Fig. 12). Sequence variations throughout the proposed U6atac central stem, stem/loop A, and to a lesser extent stem/loop B, provide further support for these structural elements. The ability of U6atac to fold into a similar structure to U6 is encouraging since SART3/p110^{nrb} is known to function in both the major and minor spliceosomes, binding to both

U6 snRNA and U6atac in humans (Damianov et al. 2004). Thus U6 snRNA and U6atac likely fold into similar structures, at least in the SART3/p110^{nrb} binding region, to allow for efficient binding of two different RNAs by the same protein.

5.3 Evolutionary Relationships Based on U6 snRNA Sequence and Secondary Structure

Despite the unusually high level of sequence conservation, our comparative sequence analysis of U6 snRNA revealed a number of distinct sequence variations that may contribute to differences in overall structure. To determine if these variations reflect an evolving U6 snRNA secondary structure, we first assessed the evolutionary relationship between organisms based on U6 snRNA sequence by constructing a neighbor joining evolutionary tree with an alignment of twenty representative U6 snRNA sequences from our original sequence alignment (Fig. 14). A bootstrap consensus tree was generated with one thousand replicates providing a level of confidence in the branching pattern generated. Although many of the nodes in our tree have a bootstrap value less than 75, we are confident in the three major clades predicted since three other tree building methods, minimum evolution, maximum parsimony, and unweighted pair-group with arithmetic means (UPGMA), all generated very similar overall topologies with identical branching of the three major clades (data not shown).

The evolutionary tree produced can be described on two levels: changes that occur in the linear nucleotide sequence and potential changes that may occur in the proposed secondary structure (Fig. 14). At the level of primary sequence, the twenty organisms included in our tree segregated into three major clades in groups of six or eight organisms. These clades were largely defined by the GUU variations discussed



U6 snRNA Secondary Structure Characteristics

- A. Up: 5-7 bps in 5' Stem/loop, 2 nucleotide insertion and non-canonical bps in Stem B
Down: 8-10 bps in 5' Stem/loop, Strong potential to form Stem B with 6 canonical bps
Both directions: Stem/loop A - 1 nucleotide bulge in stem and tetra-loop
Central Stem - Symmetric single nucleotide bulge
- B. Down: Stem/loop A - 2 nucleotide bulge in stem and penta-loop
- C. Up: Central Stem - Asymmetric multiple nucleotide bulge

Figure 14. Subtle changes in U6 snRNA sequence are accompanied by substantial changes in Prp24 homolog architecture. A neighbor joining evolutionary tree constructed from 20 representative U6 snRNA sequences is given on the left along with bootstrap values (1000 replicates) at the nodes. Fungal organisms are shaded. Notable differences in U6 snRNA sequence are indicated along the major branches. Corresponding changes in U6 snRNA secondary structure in the context of the Dunn-Rader model are listed below the tree and located on the tree as points A, B, and C. Prp24 homolog structure is given to the right of each organism name. Common protein domains identified by SMART are indicated as follows: coiled coils (CC), HAT domains (black ovals), RRM (rectangles with E-values 10^0 - 10^{-3} (dotted), 10^{-4} - 10^{-7} (hatched), 10^{-8} - 10^{-11} (gray), 10^{-12} - 10^{-19} (black), $<10^{-20}$ (white)). The amino acid sequence of the Lsm interaction domain is given at the C-terminus. The number on the right of the homolog schematic is the length in amino acids and the number on the left is the number of amino acids before the first RRM. All other values indicate the number of amino acids between RRMs or between the C-terminal RRM and the Lsm interaction domain.

previously and the sequence of a 4 to 5 nucleotide segment consisting of residues 72-75 in *S. cerevisiae*. Also to note, U6 snRNAs from the clade containing *S. cerevisiae*, *K. lactis* and *V. polyspora* contain a U at position 54, C at position 58, and an AAC sequence at positions 82-84. The positions corresponding to 54 and 58 are A and U, respectively, in U6 snRNAs from all other organisms included in our original sequence alignment, and the segment of RNA corresponding to positions 82-84 was ACA in all sequences analyzed, excluding the euglezoa, *Schizosaccharomyces* species and *C. albicans* and *C. tropicalis*. Additional changes in sequence in the 5' and 3' region of U6 were also present, however these mutations appeared to be much more variable and did not generate consistent grouping patterns.

A much more interesting story developed when the evolutionary tree was analyzed from the perspective of secondary structure in the context of the Dunn-Rader model where it was clear that the three major clades corresponded to three notable differences in U6 snRNA secondary structure (Fig. 14). The most dramatic evolutionary event occurred at the most ancient common ancestor where U6 in all organisms on the upper branch of our evolutionary tree contained a 5' stem/loop with 5-7 base pairs in the stem and a weaker potential to form stem/loop B. U6 snRNA from organisms on the lower branch contained 5' stem/loops with 8-10 base pairs and a strong potential to form stem/loop B through six canonical base pairs. A second evolutionary event occurring at a more recent common ancestor grouped the plants, worms, arthropods, and mammals apart from all fungal species, with the exception of *S. pombe* and likely other *Schizosaccharomyces* species, and may correspond to an increase in the number of

nucleotides residing in an asymmetric bulge in the middle of the central stem (Fig. 14). All other species included in our tree have the ability to generate a central stem with a single nucleotide symmetric bulge. The clade consisting of eight fungal organisms can further be divided into two clades when the slight change in stem/loop A structure described earlier is considered.

5.4 Subtle Changes in U6 snRNA Sequence Correlate with Dramatic Changes in Prp24 Homolog Architecture

It was surprising to find in our U6 snRNA sequence alignment that groups of three or four consecutive nucleotides in U6 tended to vary in only a small number of distinct patterns. While these sequence variations were relatively subtle, they may represent slight changes in the overall U6 snRNA secondary structure throughout evolution. If the RNA structure has evolved over time, then it is possible that the protein domain architecture of the RNA-associated proteins may have evolved with the RNA. In fact, Prp24 and its human homolog have both been characterized and have been shown to differ dramatically in both size and protein domain composition (Shannon & Guthrie 1991, Bell et al. 2002, Rader & Guthrie 2002). To further assess the differences in Prp24 architecture and to correlate these to potential changes in U6 snRNA structure, we obtained the Prp24 homolog sequence from each of the organisms included in our evolutionary tree and assessed the protein domain organization with the Simple Modular Architecture Research Tool (SMART).

SMART is capable of identifying common protein domains including the RNA Recognition Motif (RRM) and Half-A-Tetratricopeptide (HAT) repeat domains previously identified in Prp24 and SART3/p110^{nrB} (Preker & Keller 1998, Bell et al.

2002, Rader & Guthrie 2002). RRMs bind RNA across a four strand-sheet while HAT domains, which are typically found in tandem, are involved in facilitating protein-protein interactions (Auweter et al. 2006, Preker & Keller 1998, Medenbach et al. 2004). Additionally, SMART has been programmed to recognize a C-terminal motif (CTM) that is unique to Prp24 and its homologs. The CTM has been shown to interact with the Lsm complex in yeast U6 snRNPs where it is involved in promoting U4/U6 di-snRNP formation (Rader & Guthrie 2002). SMART not only identifies these domains, but also provides an E-value, which gives an indication of how close the identified sequence conforms to the consensus for a particular protein domain; the smaller the E-value the less likely that the sequence would occur by chance.

The Prp24 homologs analyzed can be classified into four major groups depending on the types of protein domains present, the number of repeated RRMs, the presence of a CTM, and the overall length of the putative protein (Fig. 14). Group one, consisting of homologs from the U6 snRNA clade containing *S. cerevisiae*, *V. polyspora*, and *K. lactis*, are characterized by three or four RRMs, the presence of a CTM, and a total length ranging from 432 to 449 amino acids (Fig. 14). The fourth RRM present in *S. cerevisiae* and *V. polyspora* is quite degenerate with assigned E-values of 10^0 and 10^{-2} respectively. Some sequence similarity between the amino acid sequence following RRM 3 of the *K. lactis* homolog and RRM 4 of the *S. cerevisiae* and *V. polyspora* homologs was apparent upon manual inspection. Thus the *K. lactis* homolog likely also contains a highly degenerate fourth RRM for which the E-value exceeded the threshold and consequently was not detected by SMART.

Group two consisted of Prp24 homologs from the U6 snRNA clade containing *L. elongisporus*, *P. stipitis*, *D. hansenii*, and *C. albicans* (Fig. 14). These proteins contain three or four RRM, no CTM, and a long N-terminal addition preceding the first RRM that drastically increases the length of the protein to between 837 and 940 amino acids. Like group one, the fourth RRM was relatively degenerate and when detected by SMART, was assigned an E-value greater than or equal to 10^{-3} . SMART did not detect any protein domains in the N-terminal domain of these homologs.

Group three homologs belong to organisms including *P. nodorum*, *M. grisea*, *A. capsulatus*, *A. fumigatus*, *P. anserina*, and *N. crassa* and are characterized again by four RRM and the presence of a CTM (Fig. 14). In addition, a long N-terminal extension, in some cases containing multiple HAT domains, and a C-terminal extension following the CTM, which ranges from a couple of amino acids up to thirty, increases the size of these proteins to between 1066 and 1384 amino acids.

All Prp24 homologs from non-fungal species belong to group four and are characterized by only one RRM in plants and arthropods or two RRM in all other species including worms and mammals (Fig. 14). These homologs also contain a CTM and an N-terminal extension consisting of multiple HAT domains and in some cases a coiled coil region. Like group two and three, these homologs are relatively large, ranging in size from 836 to 1014 amino acids.

While we expected the protein architecture of the Prp24 homologs collected for this analysis to differ dramatically based on the already known differences between Prp24 and SART3/p110^{rb}, it was surprising to find that the homolog domain architecture could

be classified into four groups which corresponded perfectly to the four U6 snRNA clades identified in the U6 snRNA evolutionary tree (Fig. 14). It is not at all clear how the differences in U6 snRNA sequence or structure may be related to differences in Prp24 homolog architecture, and this link will likely not be known in detail until U6 snRNP crystal structures from several different organisms become available. However, the increase in protein size appears to correspond to a decrease in U6 snRNA stem/loop B stability in more complex organisms (Fig. 13). Complete disruption of stem/loop B may be required to accommodate binding of such a large protein in many of the organisms examined in our analysis.

Sequence alignments indicate that RRM 1 and 2 of Prp24 are most like RRM 1 and 2 of SART3/p110^{nrb} respectively, and mutational analyses and binding assays have shown that these RRMs of Prp24 are responsible for high affinity binding to U6 snRNA (Kwan & Brow 2005, Bae et al. 2007). The function of the third and fourth RRMs in Prp24 is not clear, however RRM 4 does not conform well to the RRM consensus sequence and further has been suggested to have at least partial redundancy with RRM 1 (Rader & Guthrie 2002). In contrast to the other RRMs, RRM 3 has an acidic surface and this observation has led to the proposal that RRM 3 may not be important for binding U6 snRNA, but may instead be important for binding some other ligand (Bae et al. 2007).

While the Prp24 homolog from most fungal species consisted of four RRMs, the large N-terminal domain preceding the first RRM is absent in only a small number of fungal species. Medenbach et al. (2004) have shown that the HAT domain of SART3/p110^{nrb} interacts specifically with the C-terminus of the U4/U6 90K protein and they state

that this region of 90K is conserved in the yeast U4/U6 protein Prp3. Thus it is possible that protein-protein interactions that are mediated through the HAT domain in the N-terminal region of most of the species examined here are mediated through RRM 3 or 4 in a small number of fungal species. Alternatively, Rader and Guthrie (2002) suggested that the function of the N-terminal domain might be supplied by some other polypeptide in organisms such as *S. cerevisiae* where the N-terminal domain is missing.

In summary, our analysis of U6 snRNA sequence and structure has identified slight structural changes in U6 snRNA, which may correlate with dramatic changes in Prp24 size or protein domain architecture. Since Prp24 and U6 snRNA are the main constituents of U6 snRNPs, it is unlikely that the structure of one would change without co-variation in the other to accommodate these changes. The most pronounced difference in Prp24 homolog structure is the absence of a large N-terminal domain in a small number of fungal species. The N-terminal domain appears to have been lost in this subset of fungi since this domain is present in ancestors common to these organisms and all other organisms analyzed. In order to understand the co-variations observed here, a much more extensive evolutionary analysis should be performed on a genome-wide scale.

Chapter Six – The Dunn-Rader Model of U6 snRNA Secondary Structure Reveals the Primary Function of U4 snRNA in Pre-mRNA Splicing

The focus of this thesis was to develop and experimentally test a new model of U6 snRNA secondary structure that would better address a number of questions concerning U6 snRNA activation for splicing than previously proposed models. In particular, our goal was to generate a model that would explain the functional relevance of the large U6 snRNA structural re-arrangements that are thought to occur throughout spliceosome assembly and activation, with a key interest in addressing the function of transient association with U4 snRNA in early stages of spliceosome assembly. However, in producing this new model, we have actually stumbled upon something much larger: the primary function of U4 snRNA in nuclear pre-mRNA splicing.

U4 snRNA is an essential gene product, however its function until now has remained largely unknown since it is not tightly associated with the active spliceosome and may even dissociate from the spliceosome following U4/U6 unwinding (Siliciano, et al. 1987, Yean & Lin 1991). Brow and Guthrie (1989) proposed that U4 snRNA might act as an anti-sense negative regulator of U6 snRNA activity, masking catalytically important residues while U6 adopts a conformation that is more favorable for its incorporation into the spliceosome. Further, Stevens et al. (2001) suggest that the sub-stoichiometric levels of U4 snRNP, relative to other splicing snRNPs, implies that U4 is a limiting factor in spliceosome assembly. Thus, interaction with U4 might serve as a regulatory point for inclusion of U6 into functional spliceosomes. While these proposals address the function

of U4 snRNA, they do not explain the large structural re-arrangements required in U6 snRNA to accommodate the U4/U6 interaction.

The Dunn-Rader model of U6 snRNA secondary structure is consistent with the proposed roles of U4 snRNA/P in splicing, however, we suggest that U4 snRNA performs an additional role as an activator of U6 prior to its role as a negative regulator in the di-snRNP. All other proposed yeast and mammalian models of U6 secondary structure in U6 snRNP predict that the U4/U6 interaction domain base pairs with itself in the 3' ISL, a structure that reforms in the catalytically active spliceosome following dissociation of U4. Our model proposes instead that this region of U6 base pairs with regions of U6 that are 5' and 3' of the interaction domain. Strikingly, the ACAGAGA sequence becomes inaccessible in our model through base pairing with the U4/U6 interaction domain in U6 snRNPs. We propose that base pairing between U4 and U6 is essential to release the ACAGAGA sequence from the U6 snRNA intramolecular interaction so that it can subsequently interact with the 5' splice site of a pre-mRNA transcript. Thus U4 snRNA can be viewed as an allosteric activator of U6 snRNA that induces a riboswitch-like conformational change in U6 (Fig. 15).

The mechanism of U4/U6 di-snRNP formation is not yet understood, however annealing of the U4 and U6 snRNAs appears to be facilitated by the U6-snRNP specific protein Prp24 (Shannon & Guthrie 1991, Ghetti *et al.* 1995, Raghunathan & Guthrie 1998, Rader & Guthrie 2002). Chemical modification/interference experiments of mammalian U6 snRNA suggest that nucleotides 65-70 as well as the invariant AGCA tetra-nucleotide at positions 53-56 are important for initiating base pairing between

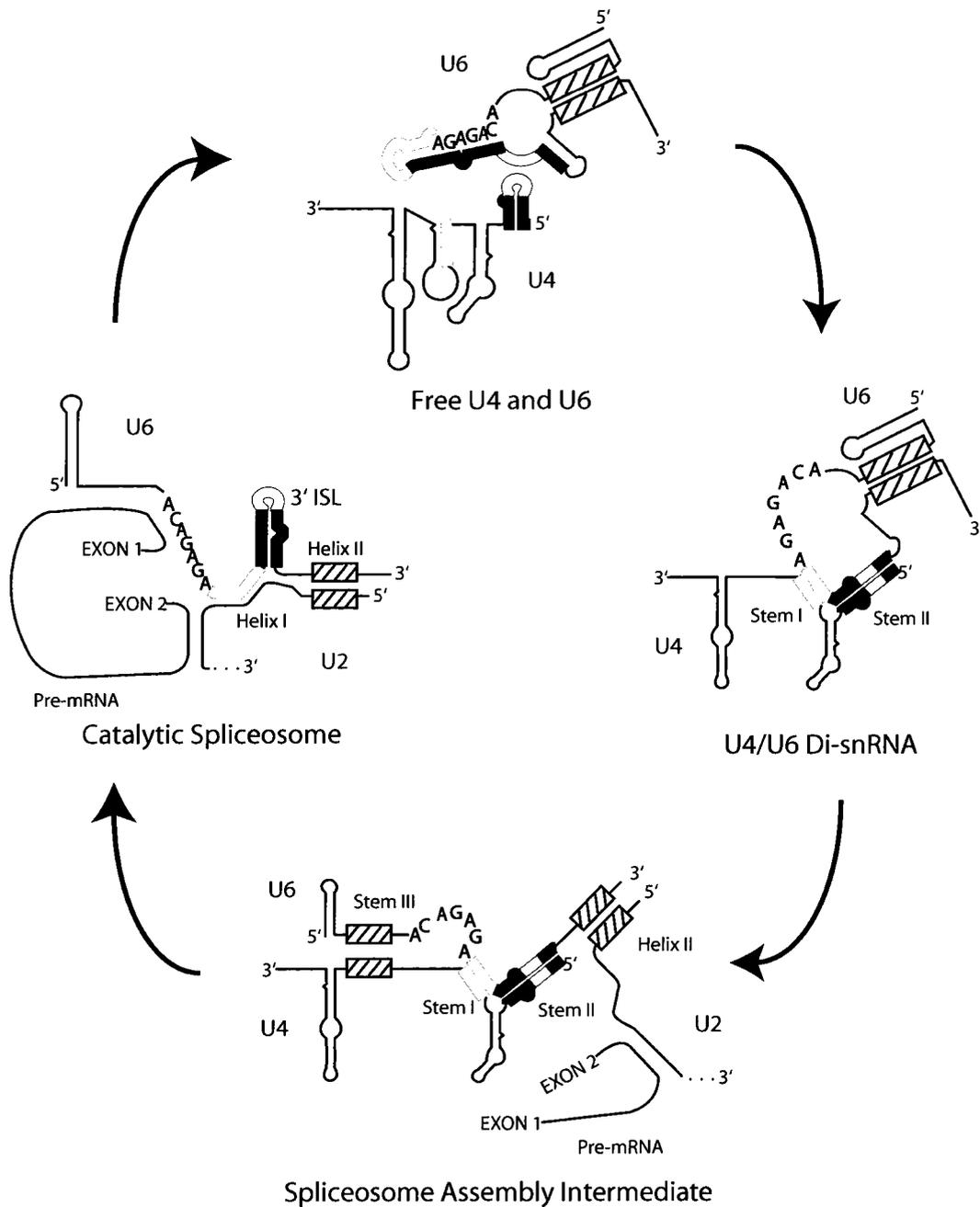


Figure 15. U6 snRNA is tightly regulated through riboswitch-like conformational changes during spliceosome assembly and activation. Mutually exclusive interactions between U6, U4, and U2 snRNA are shown where regions of RNA that interact with each other are indicated in various shades or hatched boxes.

U4 and U6 (Wolff & Bindereif 1993). These regions correspond to the loop of stem/loop A and the single stranded junction connecting stem/loops A and B in the Dunn-Rader model. Thus Prp24 likely interacts with U6 snRNA in such a way as to present these nucleotides to U4 snRNA. Following the initiation of base pairing, extension of the intermolecular interaction would result in the complete disruption of stem/loops A and B of U6, releasing the ACAGAGA sequence from intramolecular base pairing in stem/loop A.

Once in the di-snRNP, U4 may perform its second function as a negative regulator of allosterically activated U6 as proposed by Brow and Guthrie (1989) (Fig. 15). According to our model, the only U6 present in the cell with an exposed ACAGAGA sequence would be found in the di-snRNP in which it is present in an equimolar ratio to U4. Upon formation of the di-snRNA, the ACAGAGA sequence of U6 would become exposed and capable of interacting with the pre-mRNA, while the U2/U6 helix I and U6 3' ISL regions of U6 would be engaged in interactions with U4, preventing premature formation of these functionally important structures (Fig. 15). The ACAGAGA sequence is predicted to lie outside of the U4/U6 interaction domain and strong chemical modification of this sequence in U4/U6 di-snRNPs suggests that the ACAGAGA sequence is in fact accessible in the context of this complex (Brow and Guthrie 1988; Jandrositz and Guthrie 1995).

Cross-links observed between U4 snRNA and the 5' splice site in early stages of spliceosome formation imply that a proof-reading mechanism may be in place to ensure that correct base pairing between U6 and the 5' splice site is established prior to U4/U6

disassembly (Johnson and Abelson 2001). While this proof-reading is taking place, the U4/U6 interaction may be stabilized through the formation of U2/U6 helix II as proposed by Brow and Vidaver (1995) (Fig. 15). Phenotypes produced by our central stem mutations suggest that disruption of the U6 central stem, which is mutually exclusive with U2/U6 helix II, occurs after di-snRNP assembly, and therefore may occur concomitantly with U2/U6 helix II formation (Fig. 11, 15). Dissociation of U4 following establishment of the correct U6/pre-mRNA interaction would then allow for the formation of U2/U6 helix I and the U6 3' ISL within the catalytically active spliceosome (Fig. 15).

As illustrated in the preceding paragraphs, U6 snRNA is involved in a number of interactions throughout the splicing cycle. The mutual exclusivity of these interactions suggest that spliceosome assembly and activation are tightly regulated through an intricately timed exchange of base pairing partners. We have proposed that U4 snRNA plays a key role in these timing events through both the allosteric activation and negative regulation of U6 snRNA. While we have focused on regulation of U6, other splicing snRNAs are likely regulated in a similar fashion. Indeed, the accepted model of U4 snRNA secondary structure in free U4 snRNPs suggests that intramolecular base pairing in U4 must be disrupted to allow for interaction with U6 snRNA (Krol et al. 1981).

Chapter Seven – Future Directions and Concluding Remarks

U6 snRNA has been studied extensively in the context of the catalytically active spliceosome where U6 is thought to play a direct role throughout the splicing reactions. Much less work has been done to understand the structure and function of U6 snRNA in the catalytically inactive free U6 snRNP; however, it is the structure of the RNA, and the free snRNP as a whole, that will reveal important mechanistic requirements of the spliceosomal assembly and activation process. Unfortunately, work on the free U6 snRNP is limited to a few structure probing studies aiming to elucidate the intramolecular base pairing interactions and to define the sites of protein binding (Fortner et al. 1994, Jandrositz & Guthrie 1995, Karaduman et al. 2006). Here we propose new directions to explore the U6 snRNP structure and activation process.

7.1 U6 snRNP Structure Determination

To our knowledge, only one lab in the world is working toward understanding the three-dimensional structure of free U6 snRNPs, and a recent electron microscope image from their lab demonstrates that the splicing community is getting closer to understanding the snRNP in three-dimensions (Karaduman et al. 2008). While the electron microscope image reveals the relative orientation of the protein domains in the snRNP, U6 snRNA cannot be distinguished at all. It is possible that while atomic resolution structures may not be attainable for some time, Small Angle X-ray Scattering (SAXS) might produce middle range resolution structures of the snRNP. RNA tends to scatter X-rays more strongly than proteins, thus, while the individual RNA atoms may not be resolvable, the

electron density of the RNA molecule may provide some details of how U6 snRNA interacts with its associated proteins in the free U6 snRNP (Putnam et al. 2007).

7.2 Genetic Studies

The mutations that we have generated in the lower central stem, while generating phenotypes that were consistent with predictions of the model, also produced some unexpected phenotypes. Further mutation of these base pairs, specifically compensatory mutation of G30U/G31U to restore the G•U wobble pairs at the base of the stem, will provide further insight into the importance of these base pairing interactions. Restoration of the G•U wobble through the introduction of a U100G/U101G double mutation would be expected to suppress the weak heat sensitive growth phenotype if this phenotype is due to destabilization of the lower central stem. However, compensatory mutation of G30U/G31U would not be expected to rescue the severe cold sensitivity observed for this double mutation since we suspect that the cold sensitivity results either from the disruption of an interaction between these residues and U4 snRNA in the putative stem III, disruption of an interaction with a protein component in the di-snRNP, or through the stabilization of an interaction with a U6 snRNP specific protein.

To uncover potential G30 and G31 interaction partners, the double mutation should be subjected to suppression analysis to identify both *cis* and *trans* acting suppressors of the cold sensitivity exhibited by the G30U/G31U strain. While suppression of the growth defect represents a genetic interaction, which does not necessarily correspond to a direct physical interaction, elucidating the genetic interaction network will provide some insight into the functional roles of G30 and G31. It is

important to follow-up the phenotypes exhibited by this strain since this region of U6 snRNA, which lies far to the 5' side of the highly studied, highly conserved middle third of U6 snRNA, has been left largely unexplored.

Additional mutational analyses can be performed to test the Dunn-Rader stem/loop A and B interactions, however extreme care must be taken in both designing these experiments and interpreting the results due to the large number of known interactions involving these residues. For example, nucleotides that base pair with the ACAGAGA sequence in the free U6 snRNP base pair with U4 in the di-snRNP and with another region of U6 in active spliceosomes. Thus it is critical to compensate for the mutation under study in these other known interactions to ensure that any phenotype observed is not the result of interfering with these other interactions.

7.3 Structure Probing

RNA secondary structure can be probed using nucleases that cleave the RNA in specific structural environments. For example, RNase A will cleave 3' of single stranded C's and U's, RNase T1 will cleave 3' of single stranded G's, and RNase V1 will cleave double stranded regions. Thus nuclease cleavage of U6 snRNA in U6 snRNPs and deproteinized U6 snRNPs would reveal valuable information about base pairing interactions within the snRNP and would strengthen the chemical modification data collected by other labs. Additionally, flexible regions of RNA could be identified by subjecting U6 snRNA to in-line probing where the innate instability of the RNA results in cleavage of the phosphodiester backbone only when the attacking nucleophile and leaving group are positioned for in-line attack (Soukup & Breaker 1999).

7.4 U6 snRNA Structural Re-arrangements in Pre-mRNA Splicing

The major hypothesis that comes from the work presented here is that base pairing between U4 snRNA and U6 snRNA allosterically activates U6 by releasing the ACAGAGA sequence from intramolecular base pairing in the free U6 snRNP. We propose that this is just the beginning of a cascade of riboswitch-like conformational changes that occur throughout pre-mRNA splicing. Very little is known about when these structural re-arrangements occur relative to one another, however deciphering the precise timing of these exchanges will undoubtedly provide valuable insight into the process of generating catalytically competent spliceosomes.

The incorporation of fluorescent dyes at strategic positions in one or more of the splicing snRNAs could reveal information about when various intra- and intermolecular interactions form and break. Briefly, energy released from one dye molecule that has been excited at its characteristic wavelength can excite a second dye with a lower excitation wavelength if the second dye is in close enough proximity to the first dye. Monitoring the release of energy from the second dye will reveal whether or not the dyes are close enough for fluorescence resonance energy transfer (FRET) to occur; if the dyes are close enough, FRET will be relatively efficient, however if the dyes are separated by too great a distance, FRET will be relatively inefficient. Thus, the combination of choosing an appropriate pair of dyes and placing those dyes at informative locations within the snRNAs may be a powerful technique to establish the correct cycle of snRNA interactions, particularly interactions involving U6 snRNA.

7.5 Concluding Remarks

The experiments presented in this document, combined with data collected from the literature, have provided strong support for the Dunn-Rader model of U6 snRNA secondary structure in free U6 snRNPs. In particular, the inaccessibility of the ACAGAGA sequence to complementary oligonucleotides was consistent only with the Dunn-Rader model where this sequence is sequestered within an intramolecular base pairing interaction. Since the ACAGAGA sequence recognizes and base pairs to the 5' splice site of pre-mRNA transcripts early in spliceosome assembly, some mechanism must exist to disrupt the intramolecular ACAGAGA interaction. We propose that U4 snRNA is responsible for inducing a riboswitch-like conformational change in U6 that releases the ACAGAGA sequence from intramolecular interactions in U6 snRNP, allowing this sequence to subsequently engage in intermolecular interactions with the 5' splice site of the pre-mRNA transcript.

Works Cited

- Abelson J. (2008) Is the spliceosome a ribonucleoprotein enzyme? *Nat Struct Mol Biol.* **15**:1235-1237.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. (1990) Basic local alignment search tool. *J Mol Biol.* **215**:403-410.
- Achsel T, Brahm H, Kastner B, Bachi A, Wilm M, Lührmann R. (1999) A Doughnut-shaped heteromer of human Sm-like proteins binds to the 3'-end of U6 snRNA, thereby facilitating U4/U6 duplex formation *in vitro*. *EMBO J.* **18**:5789-5802.
- Auweter SD, Oberstrass FC, Allain FH-T. (2006) Sequence-specific binding of single-stranded RNA: is there a code for recognition? *Nucleic Acids Res.* **34**:4943-4959.
- Bae E, Reiter NJ, Bingman CA, Kwan SS, Lee D, Phillips GN Jr, Butcher SE, Brow DA. (2007) Structure and interactions of the first three RNA recognition motifs of splicing factor Prp24. *J Mol Biol.* **367**:1447-1458.
- Bell M, Schreiner S, Damianov A, Reddy R, Bindereif A. (2002) p110, a novel human U6 snRNP protein and U4/U6 snRNP recycling factor. *EMBO J.* **21**:2724-2735.
- Berget SM, Moore C, Sharp PA. (1977) Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc Natl Aca. Sci USA.* **74**:3171-3175.
- Bindereif A, Green MR. (1987) An ordered pathway of snRNP binding during mammalian pre-mRNA splicing complex assembly. *EMBO J.* **6**:2415-2424.
- Black DL, Chabot B, Steitz JA. (1985) U2 as well as U1 small nuclear ribonucleoproteins are involved in premessenger RNA splicing. *Cell.* **42**:737-750.
- Black DL, Steitz JA. (1986) Pre-mRNA splicing *in vitro* requires intact U4/U6 small nuclear ribonucleoprotein. *Cell.* **46**:697-704.
- Blencowe BJ, Sproat BS, Ryder U, Barabino S, Lamond AI. (1989) Antisense probing of the human U4/U6 snRNP with biotinylated 2'-OMe RNA oligonucleotides. *Cell.* **59**:531-539.
- Bon E, Casaregola S, Blandin G, Llorente B, Neuvéglise C, Munsterkötter M, Guldener U, Mewes H-W, Van Heldon J, Dujon B, Gaillardin C. (2003) Molecular evolution of eukaryotic genomes: hemiascomycetous yeast spliceosomal introns. *Nucleic Acids Res.* **31**:1121-1135.

- Brody E, Abelson J. (1985) The “Spliceosome”: Yeast pre-messenger RNA associates with A 40S complex in a splicing-dependent reaction. *Science*. **228**:963-967.
- Brow DA, Guthrie C. (1988) Spliceosomal RNA U6 is remarkably conserved from yeast to mammals. *Nature*. **334**:213-218.
- Brow DA, Guthrie C. 1989. Splicing a spliceosomal RNA. *Nature*. **337**:14-15.
- Brow DA, Vidaver RM. (1995) An element in human U6 snRNA destabilizes the U4/U6 spliceosomal RNA complex. *RNA* **1**:122-131.
- Cech TR. (1986) The generality of self-splicing RNA: Relationship to nuclear mRNA splicing. *Cell*. **44**:207-210.
- Cheng, S-C, Abelson, John. (1987) Spliceosome assembly in yeast. *Genes Dev*. **1**:1014-1027.
- Collins CA, Guthrie C. (2000) The question remains: is the spliceosome a ribozyme? *Nat Struct Biol*. **7**:850-854.
- Cooper M, Johnston LH, Beggs JD. (1995) Identification and characterization of Uss1p (Sdb23p): a novel U6 snRNA-associated protein with significant similarity to core proteins of small nuclear ribonucleoproteins. *EMBO J*. **14**:2066-2075.
- Damianov A, Schreiner S, Bindereif A. (2004) Recycling of the U12-type spliceosome requires a component of the U6atac snRNP. *Mol Cell Biol*. **24**:1700-1708.
- Datta B, Weiner AM. (1991) Genetic evidence for base-pairing between U2 and U6 snRNA in mammalian mRNA splicing. *Nature*. **352**:821-824.
- DeLano WL. The PyMOL Molecular Graphics System (2008). DeLano Scientific, Palo Alto, California, USA. <http://www.pymol.org>.
- Elgavish T, Cannone JJ, Lee JC, Harvey SC, Gutell RR. (2001) AA.AG@Helix.Ends: A:A and A:G Base-pairs at the ends of 16S and 23S rRNA helices. *J Mol Biol*. **310**:735-753.
- Epstein P, Reddy R, Henning D, Busch H. (1980) The nucleotide sequence of nuclear U6 (4.7S) RNA. *J Biol Chem*. **255**:8901-1906.
- Eschenlauer JB, Kaiser MW, Gerlach VL, Brow DA. (1993) Architecture of a yeast U6 snRNA gene promoter. *Mol Cell Biol*. **13**:3015-3026.

- Fabrizio P, Abelson J. (1990) Two domains of yeast U6 small nuclear RNA required for both steps of nuclear precursor messenger RNA splicing. *Science*. **250**:404-409.
- Fabrizio P, McPheeters DS, Abelson J. (1989) In vitro assembly of yeast U6 snRNP: a functional assay. *Genes Dev*. **3**:2137-2150.
- Field DJ, Friesen JD. (1996) Functionally redundant interactions between U2 and U6 spliceosomal snRNAs. *Genes Dev*. **10**:489-501.
- Fortner DM, Troy RG, Brow DA. (1994) A stem/loop in U6 RNA defines a conformational switch required for pre-mRNA splicing. *Genes Dev*. **8**:221-233.
- Gardner PP, Daub J, Tate JG, Nawrocki EP, Kolbe DL, Lindgreen S, Wilkinson AC, Finn RD, Griffiths-Jones S, Eddy SR, Bateman A. (2009) Rfam: updates to the RNA families database. *Nucleic Acids Res*. **37**:D136-D140.
- Ghetti A, Company M, 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.
- Grabowski PJ, Seiler SR, Sharp PA. (1985) A multicomponent complex is involved in the splicing of messenger RNA precursors. *Cell*. **42**:345-353.
- Gutell R, Larsen N, Woese CR. (1994) Lessons from an evolving rRNA: 16S and 23S rRNA structures from a comparative perspective. *Microbiol Rev*. **58**:10-26.
- Gutell RR, Lee JC, Cannone JJ. (2002) The accuracy of ribosomal RNA comparative structure models. *Curr Opin Struct Biol*. **12**:301-310.
- Guthrie C, Patterson B. (1988) Spliceosomal snRNAs. *Annu Rev Genet*. **22**:387-419.
- Guthrie C. (1991) Messenger RNA splicing in yeast: clues to why the spliceosome is a ribonucleoprotein. *Science*. **253**:157-163.
- Hilliker AK, Staley JP. (2004) Multiple functions for the invariant triad of U6 snRNA. *RNA*. **10**:921-928.
- Hilliker AK, Mefford MA, Staley JP. (2007) U2 toggles iteratively between the stem IIa and stem IIc conformations to promote pre-mRNA splicing. *Genes Dev*. **21**:821-834.
- Huppler A, Nikstad LJ, Allmann AM, Brow DA, Butcher SE. (2002) Metal binding and base ionization in the U6 RNA intramolecular stem-loop structure. *Nature*. **9**:431-435.

- Irimia M, Penny D, Roy SW. (2007) Co-evolution of genomic intron number and splice sites. *Trends Genet.* **23**:321-325.
- Jandrositz A, Guthrie C. (1995) Evidence for a Prp24 binding site in U6 snRNA and in a putative intermediate in the annealing of U6 and U4 snRNAs. *EMBO J.* **14**:820-832.
- Jakab G, Mougin A, Kis M, Pollak T, Antal M, Branlant C, Solymosy F. (1997) *Chlamydomonas* U2, U4, and U6 snRNAs. An evolutionary conserved putative third interaction between U4 and U6 snRNAs which has a counterpart in the U4atac-U6atac snRNA duplex. *Biochimie.* **79**:387-395.
- Johnson TL, Abelson J. (2001) Characterization of U4 and U6 interactions with the 5' splice site using a *S. cerevisiae* in vitro trans-splicing system. *Genes Dev.* **15**:1957-1970.
- Jurica MS, Moore MJ. (2003) Pre-mRNA splicing: awash in a sea of proteins. *Mol Cell.* **12**:5-14.
- Kandels-Lewis S, Séraphin B. (1993) Roles of U6 snRNA in 5' splice site selection. *Science.* **262**:2035-2039.
- Karaduman R, Fabrizio P, Hartmuth K, Urlaub H, Lührmann R. (2006) RNA structure and RNA-protein interactions in purified yeast U6 snRNPs. *J Mol Biol.* **356**:1248-1262.
- Karaduman R, Dube P, Stark H, Fabrizio P, Kastner B, Lührmann R. (2008) Structure of yeast U6 snRNPs: arrangement of Prp24 and the LSM complex as revealed by electron microscopy. *RNA.* **14**:2528-2537.
- Kim CH, Abelson J. (1996) Site-specific crosslinks of yeast U6 snRNA to the pre-mRNA near the 5' splice site. *RNA.* **2**:995-1010.
- Konarska MM, Grabowski PJ, Padgett RA, Sharp PA. (1985) Characterization of the branch site in lariat RNAs produced by splicing of mRNA precursors. *Nature.* **313**:552-557.
- Konarska MM, Sharp PA. (1987) Interaction between small nuclear ribonucleoprotein particles in formation of spliceosomes. *Cell.* **49**:763-774.
- Krol A, Branlant C, Lazar E, Gallinaro H, Jacob M. (1981) Primary and secondary structures of chicken, rat and man nuclear U4 RNAs. Homologies with U1 and U5 RNAs. *Nucleic Acids Res.* **9**:2699-2716.
- Kwan SS, Brow DA. (2005) The N- and C-terminal RNA recognition motifs of splicing factor Prp24 have distinct functions in U6 RNA binding. *RNA.* **11**:808-820.

- Leontis NB, Stombaugh J, Westhof E. (2002) The non-Watson-Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res.* **30**:3497-3531.
- Lescoute A, Westhof E. (2006) Topology of three-way junctions in folded RNAs. *RNA.* **12**:83-93.
- Lesser, CF, Guthrie C. (1993) Mutations in U6 snRNA that alter splice site specificity: implications for the active site. *Science.* **262**:1982-1988.
- Li Z, Brow DA. (1993) A rapid assay for quantitative detection of specific RNAs. *Nucleic Acids Res.* **21**:4645-4646.
- Madhani HD, Bordonne R, Guthrie C. (1990) Multiple roles for U6 snRNA in the splicing pathway. *Genes Dev.* **4**:2264-2277.
- Madhani HD, Guthrie C. (1992) A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. *Cell.* **71**:803-817.
- Margottin F, Dujardin G, Gerard M, Egly J-M, Huet J, Sentenac A. (1991) Participation of the TATA factor in transcription of the yeast U6 gene by RNA polymerase C. *Science.* **251**:424-426.
- Maschhoff K, Padgett R. (1993) The stereochemical course of the first step of pre-mRNA splicing. *Nucleic Acids Res.* **21**:5456-5462.
- Mayes AE, Verdone L, Legrain P, Beggs JD. (1999) Characterization of Sm-like proteins in yeast and their associations with U6 snRNA. *EMBO J.* **18**:4321-4331.
- McManus JC, Schwartz ML, Butcher SE, Brow DA. (2007) A dynamic bulge in the U6 RNA internal stem-loop functions in spliceosome assembly and activation. *RNA.* **13**:1-14.
- Medenbach J, Schreiner S, Liu S, Lührmann R, Bindereif A. (2004) Human U4/U6 snRNP recycling factor p110: mutational analysis reveals the function of the tetratricopeptide repeat domain in recycling. *Mol Cell Biol.* **24**:7392-7401.
- Moore MJ, Sharp PA. (1993) Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. *Nature.* **365**:364-368.
- Mougin A, Gottschalk A, Fabrizio P, Lührmann R, Branlant C. (2002) Direct probing of RNA structure and RNA-protein interactions in purified HeLa cell's and yeast spliceosomal U4/U6.U5 tri-snRNP particles. *J Mol Biol.* **317**:631-649.

Ohi MD, Vander Kooi CW, Rosenberg JA, Ren L, Hirsch JP, Chazin WJ, Walz T, Gould KL. (2005) Structural and functional analysis of essential pre-mRNA splicing factor Prp19p. *Mol Cell Biol.* **25**:451-460.

Owczarzy R, Tataurov AV, Wu Y, Manthey JA, McQuisten KA, Almabrazi HG, Pedersen KF, Lin Y, Garretson J, McEntaggart NO, Sailor CA, Dawson RB, Peek AS. (2008) IDT SciTools: a suite for analysis and design of nucleic acid oligomers. *Nucleic Acids Res.* **36**:W163-W169.

Padgett RA, Konarska MM, Grabowski PJ, Hardy SF, Sharp PA. (1984) Lariat RNAs as intermediates and products in the splicing of messenger RNA precursors. *Science.* **225**:898-903.

Parisien M, Major F. (2008) The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature.* **452**:51-55.

Parker R, Siliciano PG, Guthrie C. (1987) Recognition of the TACTAAC box during mRNA splicing in yeast involves base pairing to the U2-like snRNA. *Cell.* **49**:229-239.

Peebles CL, Perlman PS, Mecklenburg KL, Petrillo ML, Tabor JH, Jarrell KA, Cheng HL. (1986) A self-splicing RNA excises an intron lariat. *Cell.* **44**:213-223.

Peebles C, Zhang M, Perlman P, Fanzen J. (1995) Catalytically critical nucleotides in domain 5 of a group II intron. *Proc Natl Acad Sci USA.* **92**:4422-4426.

Perriman RJ, Ares M Jr. (2007) Rearrangement of competing U2 RNA helices within the spliceosome promotes multiple steps in splicing. *Genes Dev.* **21**:811-820.

Preker PJ, Keller W. (1998) The HAT helix, a repetitive motif implicated in RNA processing. *Trends Biochem Sci.* **23**:15-16.

Punam CD, Hammel M, Hura GL, Tainer JA. (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q Rev Biophys.* **40**:191-285.

Rader SD, Guthrie C. (2002) A conserved Lsm-interaction motif in Prp24 required for efficient U4/U6 di-snRNP formation. *RNA.* **8**:1378-1392.

Raghuathan PL, Guthrie C. (1998) A spliceosomal recycling factor that reanneals U4 and U6 small nuclear ribonucleoprotein particles. *Science.* **279**:857-860.

Raghuathan PL, Guthrie C. (1998b) RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2. *Curr Biol.* **8**:847-855.

Rhode BM, Hartmuth K, Westhof E, Lührmann R. (2006) Proximity of conserved U6 and U2 snRNA elements to the 5' splice site region in activated spliceosomes. *EMBO J.* **25**:2475-2486.

Ryan DE, Stevens SW, Abelson J. (2002) The 5' and 3' domains of yeast U6 snRNA: Lsm proteins facilitate binding of Prp24 protein to the U6 telestem region. *RNA.* **8**:1011-1033.

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

Séraphin, B. (1995) Sm and Sm-like proteins belong to a large family: identification of proteins of the U6 as well as the U1, U2, U4, and U5 snRNPs. *EMBO J.* **14**:2089-2098.

Schultz J, Milpetz F, Bork P, Ponting CP. (1998) SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci USA.* **95**:5857-5864.

Shannon KW, Guthrie C. (1991) Suppressors of a novel U4 snRNA mutation define a novel U6 snRNP protein with RNA-binding motifs. *Genes Dev.* **5**:773-785.

Siliciano PG, Brow DA, Roiha H, Guthrie C. (1987) An essential snRNA from *S. cerevisiae* has properties predicted for U4, including interaction with a U6-like snRNA. *Cell.* **50**:585-592.

Soukup GA, Breaker RR. (1999) Relationship between internucleotide linkage geometry and the stability of RNA. *RNA.* **5**:1308-1325.

Spiller MP, Boon K-L Reijns, MAM, Beggs JD. (2007) The Lsm2-8cComplex determines nuclear localization of the spliceosomal U6 snRNA. *Nucleic Acids Res.* **35**:923-929.

Staley JP, Guthrie C. (1999) An RNA switch at the 5' splice site requires ATP and the DEAD box protein Prp28p. *Mol Cell.* **3**:55-64.

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

Stevens SW, Barta I, Ge HY, Moore RE, Young MK, Lee TD, Abelson J. (2001) Biochemical and genetic analyses of the U5, U6, and U4/U6.U5 small nuclear ribonucleoproteins from *Saccharomyces cerevisiae*. *RNA.* **7**:1543-1553.

- Stevens SW, Ryan DE, Ge HY, Moore RE, Young MK, Lee TD, Abelson J. (2002) Composition and functional characterization of the yeast spliceosomal penta-snRNP. *Mol Cell*. **9**:31-44.
- Tamura K, Dudley J, Nei M, Kumar S. (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. *Mol Biol Evol*. **24**:1596-1599.
- Thompson JD, Higgins DG, Gibson TJ. (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. **22**:4673-4680.
- Valadkhan S. (2007) The spliceosome: a ribozyme at heart? *Biol Chem*. **388**:693-697.
- van der Veen R, Arnberg AC, van der Horst G, Bonen L, Tabak HF, Grivell LA. (1986) Excised group II Introns in yeast mitochondria are lariats and can be formed by self-splicing in vitro. *Cell*. **44**:225-234.
- Vidal VP, Verdone L, Mayes AE, Beggs JD. (1999) Characterization of U6 snRNA-protein interactions. *RNA*. **5**:1470-1481.
- Vidaver RM, Fortner D, M., Loos-Austin LS, Brow DA. (1999) Multiple functions of *Saccharomyces cerevisiae* splicing protein Prp24 in U6 RNA structural rearrangements. *Genetics*. **153**:1205-1218.
- Wasserman DA, Steitz JA. (1992) Interaction of small nuclear RNAs with precursor messenger RNA during in vitro splicing. *Science*. **257**:1918-1925.
- Woese CR, Gutell R, Gupta R, Noller HF. (1983) Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acids. *Microbiology Reviews*. **47**:621-669.
- Wolff T, Bindereif A. (1993) Conformational changes of U6 RNA during the spliceosomal cycle: an intramolecular helix is essential both for initiating the U4-U6 interaction and for the first step of splicing. *Genes Dev*. **7**:1377-1389.
- Wolff T, Menssen R, Hammel J, Bindereif A. (1994) Splicing function of mammalian U6 small nuclear RNA: conserved positions in central domain and helix I are essential during the first and second step of pre-mRNA splicing. *Proc Natl Acad Sci USA*. **91**:903-907.
- Wolff T, Bindereif A. (1995) Mutational analysis of human U6 RNA: stabilizing the intramolecular helix blocks the spliceosomal assembly pathway. *Biochimica et Biophysica Acta*. **1263**:39-44.

Wu J, Manley, JL. (1991) Base pairing between U2 and U6 snRNAs is necessary for splicing of a mammalian pre-mRNA. *Nature*. **352**:818-821.

Xu G-L, Wieland B, Bindereif A. (1994) trans-spliceosomal U6 RNAs of *Crithidia fasciculata* and *Leptomonas seymouri*: deviation from the conserved ACAGAG sequence and potential base pairing with spliced leader RNA. *Mol Cell Biol*. **14**:4565-4570.

Yean S-L, Lin R-J. (1991) U4 small nuclear RNA dissociates from yeast spliceosomes and does not participate in the subsequent splicing reaction. *Mol Cell Biol*. **11**:5571-5577.

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