Conformational features of the human U2-U6 snRNA complex

Ravichandra Bachu

Graduate Center, City University of New York
Conformational features of the human U2-U6 snRNA complex

by

Ravichandra Bachu

A dissertation submitted to the Graduate Faculty in Chemistry in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2014
This manuscript has been read and accepted for the Graduate Faculty in Chemistry in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Professor Nancy L. Greenbaum

Date

Chair of the Examining Committee

Professor Maria Tamargo

Date

Executive Officer

Professor Nancy L. Greenbaum
Professor Dixie J. Goss
Professor Lesley Davenport

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK
ABSTRACT

Conformational features of the human U2-U6 snRNA complex

by

Ravichandra Bachu

Mentor: Prof. Nancy L. Greenbaum

The splicing of precursor messenger (pre-m) RNA, during which noncoding intervening sequences are excised and flanking coding regions ligated, is an integral reaction of gene expression. In eukaryotes, it is carried out by a dynamic RNA-protein complex called the spliceosome, in which five small nuclear (sn) RNA components are actively involved in recognition and chemical aspects of the process. A complex formed between U2 and U6 snRNAs is implicated in the chemistry of pre-mRNA splicing. The catalytic activity of the U2-U6 snRNA complex is dependent on the presence of Mg$^{2+}$ ions, and the complex has been shown to have several specifically bound Mg$^{2+}$ binding sites in vitro. The overall goal of this research is to characterize the conformational changes of the human U2-U6 snRNA complex upon addition of Mg$^{2+}$. In order to pursue this question, we attempted to characterize the lowest energy structure of the complex in the absence of spliceosomal proteins using a combination of biophysical and biochemical techniques in the solution state. We first used enzymatic structure probing to evaluate the secondary structural fold of protein-free human U2-U6 snRNA complex. Cleavage patterns resulting from probing reactions were consistent with formation of four stem regions surrounding the junction, therefore favoring the four-helix model consistent with previous results of in vivo studies of the human U2-U6 snRNA complex. However, $^{19}$F
NMR studies from our laboratory also identified a lesser fraction (up to 14%) of a three-helix conformation. Upon addition of up to 100 mM Mg\(^{2+}\), a slight increase in cleavage by enzymes specific for both single-stranded and double-stranded regions was observed at the junction region, suggesting that this region is becoming more accessible, perhaps because of an increase in the fraction of the three-helix conformation. Analytical ultracentrifugation studies revealed that the Stokes radius of the RNA complex decreased slightly from 31.3 Å to 27.9 Å in the presence of 100 mM Mg\(^{2+}\), suggesting a slight compaction of the tertiary structure in the presence of divalent metal ions. Hydroxyl radical footprinting experiments on this complex showed signs of increased protection in some areas near and more distant from the junction upon addition of Mg\(^{2+}\), suggesting a change in three-dimensional conformation. Therefore, it appears that Mg\(^{2+}\) induces a small three-dimensional conformational change on human U2-U6 snRNA complex. In order to build a three-dimensional model for the four-helix conformation, we designed a mutant that favors the formation of four-helix conformation and performed SAXS experiments on it. The preliminary SAXS studies suggest that the human U2-U6 snRNA complex and the mutant complex may also be amenable to further study by SAXS. These results act as a good starting point to characterize further the overall global conformation of human U2-U6 snRNA complex and effects of spliceosomal proteins on it.
To my Parents
Acknowledgements

I take this opportunity to express my sincerest thanks and gratitude to my mentor Nancy L. Greenbaum for her excellent guidance and her immeasurable patience in introducing me to the RNA structural biology field. I am grateful to Prof. Dixie J. Goss and Prof. Lesley Davenport for their valuable suggestion from time to time and for serving on my thesis committee.

I also extend my deepest gratitude to Prof. Michael Brenowitz, Prof. Alain Laederach, Prof. Stefano Rivella and Prof. Lois Pollack for their successful collaboration with various projects, accommodating me in their labs and training me with necessary techniques. I especially would like to thank Dr. Joerg Schlatterer for training me with various techniques and providing his expert guidance and suggestions in every step both professionally and personally. I sincerely appreciate your point of view.

A big thanks to all my lab mates past and present: Dr. Manjula Kolipaka, Dr. Milena Popovic, Dr. Faqing Yuan, Caijie Zhao, He Huang, Huong Chu and my buddy William Perea. I thank the lab members from my collaborators Dr. Laura Breda, Dr. Suzette Pabit for their help. I also thank all the undergraduate & Masters students who worked with me as a team and helping me out in tough times: Liwen chen, Yuriy Khanin, Daniel Shimonov, Steven Cajamarca, Nauman Abid, Gaddiel Rodriguez, Anthony Ventimiglia and Shawn Adams.

Finally, I thank my family and friends who have continuously supported me over these years. Special thanks to my siblings Sindhuri and Madhuri for their love and support. Last but not least I would like to thank my best friend and my wife Sohoni Das Sharma for her love and encouragement during all these years.
Table of Contents

ABSTRACT ............................................................................................................................... iv
List of Tables ........................................................................................................................... x
List of Figures .......................................................................................................................... xi
List of Abbreviations and symbols ......................................................................................... xvi

CHAPTER 1 ................................................................................................................................ 1

INTRODUCTION ...................................................................................................................... 1
1.1. Introduction: ..................................................................................................................... 1
1.2. Nucleic acids: .................................................................................................................. 3
1.4. RNA as a catalyst: Ribozymes ....................................................................................... 10
1.5. Pre-mRNA splicing: ....................................................................................................... 11
1.6. Mechanism of splicing: ................................................................................................. 14
1.7. Spliceosome: Structure and dynamics: .......................................................................... 16
1.8. Parallels between the spliceosome and group II introns: ............................................. 19
1.9. Role of snRNAs in catalysis: is the spliceosome a ribozyme? ...................................... 23
1.10. Role of proteins in pre-mRNA splicing: ....................................................................... 25
1.11. Conformation of the U2-U6 snRNA complex: ............................................................ 27

CHAPTER 2 .............................................................................................................................. 31

SECONDARY STRUCTURE PROBING OF THE HUMAN U2-U6 snRNA COMPLEX 31
2.1. Introduction: ..................................................................................................................... 31
2.2. Materials and methods: ................................................................................................ 35
  2.2.1. RNA design and synthesis: ....................................................................................... 35
  2.2.2. Design of the mutant: .............................................................................................. 36
  2.2.3. Enzymatic structure probing: .................................................................................. 38
  2.2.4. Principle and background: ...................................................................................... 39
  2.2.5. Enzymatic probes: .................................................................................................... 40
  2.2.6. Chemical probes: ...................................................................................................... 40
  2.2.7. Enzymatic probing Method: .................................................................................... 42
2.3. Results: ............................................................................................................................ 43
  2.3.1. Enzymatic structure probing: .................................................................................. 43
  2.3.2. Effect of divalent metal ions on secondary structure: ............................................. 49
2.4. Discussion: ...................................................................................................................... 53

CHAPTER 3 ................................................................................................................................ 59

TERTIARY STRUCTURAL FEATURES OF THE HUMAN U2-U6 snRNA COMPLEX. ......... 59
3.1 Introduction: ...................................................................................................................... 59
3.2. Materials and methods: ................................................................................................ 61
  3.2.1 Analytical ultracentrifugation (AUC) ......................................................................... 62
  3.2.2. Hydroxyl Radical Footprinting: ............................................................................... 63
  3.2.3. SAXS: ....................................................................................................................... 65
3.3. Results: ............................................................................................................................ 68
  3.3.1. Analytical Ultracentrifugation: ................................................................................ 68
  3.3.2. Hydroxyl Radical footprinting: ............................................................................... 70
  3.3.3. SAXS: ....................................................................................................................... 77
3.4. Discussion: .......................................................................................................................... 85

CHAPTER-4 .................................................................................................................................. 93

CONCLUSIONS AND FUTURE DIRECTIONS ........................................................................ 93

4.1. Conclusions: .......................................................................................................................... 93
   4.1.1 Introduction: ....................................................................................................................... 93
   4.1.2 Secondary Structural fold: ............................................................................................... 94
   4.1.3 Tertiary conformational features: .................................................................................... 95

4.2 Future directions: .................................................................................................................. 98

Supplementary Data ................................................................................................................. 100

References: ............................................................................................................................... 107
List of Tables

Table 2.1. Various enzymatic and chemical probes that are generally used for RNA structure probing, their molecular weight and specificities ….41

Table 3.1. Results of analytical ultracentrifugation measurements on the human U2-U6 snRNA complex ….70

Table 3.2. Results of SAXS experiments of the human U2-U6 snRNA complex and the complex between mutated U2 and U6 snRNA ….79
List of Figures

**Figure 1.1.** The central dogma of molecular biology ....2

**Figure 1.2.** Common bases found in Nucleic acids. ....4

**Figure 1.3.** Common secondary structure motifs present in RNA. ....6

**Figure 1.4.** Regular base pairs found in RNA double helices. ....7

**Figure 1.5.** Processing of pre-mRNA in eukaryotes ....12

**Figure 1.6.** Basic modes of alternative splicing ....13

**Figure 1.7.** Figure shows universally conserved nucleotides in the introns. And the consensus sequences near the 5’ splice site, branch site and 3’ splice site. ....15
Figure 1.8. Schematic representation of two-step transesterification reaction catalyzed by spliceosome. ....17

Figure 1.9. Schematic representation of assembly of mammalian spliceosome. ....18

Figure 1.10. Cartoon showing the structural similarities between various domains of group II introns and U snRNAs of spliceosome. ....20

Figure 1.11. Figure shows the structural and sequence similarities of Domain V of group II intron with U6 ISL of yeast and human U6 snRNA. ....22

Figure 1.12. A) Three helix junction model: proposed secondary structure model for yeast U2-U6 snRNA complex B) Four-helix junction: proposed secondary structure model for human U2-U6 snRNA complex. ....28

Figure 2.1. Models for the conformation of the protein-free human (h) U2-U6 snRNA complex. ....32
Figure 2.2. Sequence and proposed secondary structure of a mutation favoring the four-helix model.

Figure 2.3. Representative sequencing gel exhibiting cleavage patterns of $^{32}$P-labeled (*) hU2 and *hU6 snRNA paired with respective counter strands subjected to enzymatic probing.

Figure 2.4. Pairing of RNA oligomers representing hU2 and U6 snRNA analyzed by non-denaturing gel electrophoresis.

Figure 2.5. A & B. Normalized cleavage intensities of complex between *hU6 and hU2 snRNAs and the complex between *hU2 and hU6 snRNAs following reaction with RNase V1.

Figure 2.6. A & B. Normalized cleavage intensities of complex between *hU6 and hU2 snRNAs and the complex between *hU2 and hU6 snRNAs following reaction with RNases A and T1.

Figure 3.1. Schematic representation of hydroxyl radical footprinting experiment.
**Figure 3.2.** The figure shows $g(s^*)$ versus $s^*$ distribution curve for the human U2-U6 snRNA complex with increasing concentrations of $\text{Mg}^{2+}$.

**Figure 3.3.** Representative sequencing gel exhibiting cleavage patterns of $^{32}\text{P}$-labeled (*) hU6 paired with unlabeled hU2 strand, subjected to hydroxyl radical cleavage.

**Figure 3.4.** Thermal/color plots showing cleavage intensities of complex between *hU2 and hU6 snRNAs following reaction with hydroxyl radicals.

**Figure 3.5.** Normalized cleavage intensities of complex between *hU6 and hU2 snRNAs and the complex between *hU2 and hU6 snRNAs following reaction with hydroxyl radicals were mapped onto the major structural fold i.e. four-helix conformation.

**Figure 3.6.** A & B. Figure shows $P(r)$ profiles from the wild type (A) and mutant (B) samples.

**Figure 3.7.** A&B. Figure shows the normalized SAXS profile of the WT (aI2), Mutant (bI2) and the alternate minor conformation presumably three-helix profile.
**Figure 3.8.** The figure shows shape reconstructions of WT (Red) and Mutant that favors four-helix conformation (Grey) ....84

**Figure 3.9.** The figure shows the proposed fold of the human U2-U6 snRNA complex based on preliminary computational analysis. ....90

**Supplementary material**

**Figure 1.** Ribonuclease cleavage patterns for the U2-U6 snRNA complex ....101

**Figure 2.** Representative sequencing gel for the *hU2+hU6 complex and the mutant *hu2+hU6 complex subjected to partial hydrolysis ....104

**Figure 3.** Thermal/color plots showing cleavage intensities of complex between *hU6 and hU2 complex, following the reaction with hydroxyl radicals ....105
List of Abbreviations and symbols

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Å</td>
<td>Angstrom, $10^{-10}$ meters</td>
</tr>
<tr>
<td>A</td>
<td>Adenosine</td>
</tr>
<tr>
<td>3'-ss</td>
<td>3'-splice site</td>
</tr>
<tr>
<td>³²P</td>
<td>Radioactive phosphorous-32</td>
</tr>
<tr>
<td>5'-ss</td>
<td>5'-splice site</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine triphosphate</td>
</tr>
<tr>
<td>AUC</td>
<td>Analytical ultracentrifugation</td>
</tr>
<tr>
<td>C</td>
<td>Cytidine</td>
</tr>
<tr>
<td>DEPC</td>
<td>Diethyl pyrocarbonate</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylene diamine tetraacetic acid</td>
</tr>
<tr>
<td>FRET</td>
<td>Förster Resonance Energy Transfer</td>
</tr>
<tr>
<td>G</td>
<td>Guanine</td>
</tr>
<tr>
<td>ISL</td>
<td>Intramolecular stem loop</td>
</tr>
<tr>
<td>mL</td>
<td>$10^{-3}$ liters</td>
</tr>
<tr>
<td>mM</td>
<td>$10^{-3}$ molar</td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger RNA</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance</td>
</tr>
<tr>
<td>OH</td>
<td>Hydroxyl</td>
</tr>
<tr>
<td>PAGE</td>
<td>Polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>Pre-mRNA</td>
<td>Precursor messenger RNA</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>RNP</td>
<td>Ribonucleo protein</td>
</tr>
<tr>
<td>T</td>
<td>Thymidine</td>
</tr>
<tr>
<td>tRNA</td>
<td>Transfer RNA</td>
</tr>
<tr>
<td>Tris</td>
<td>Tris (hydroxymethyl) aminomethane, buffer</td>
</tr>
<tr>
<td>UV</td>
<td>Ultraviolet</td>
</tr>
<tr>
<td>U</td>
<td>Uridine</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

1.1. Introduction:

The central dogma of molecular biology states that the role of Ribonucleic Acid (RNA) is to transfer the genetic information from Deoxyribonucleic Acid (DNA) to proteins, the workhorse molecules of the cell (Crick, 1970; Crick, 1958) (Figure 1.1). However, recent developments in the field of biology show that there are numerous exceptions and additions to this paradigm and that RNA maintains many essential roles in addition to acting as an important intermediary role in protein synthesis. Noncoding (nc) RNAs, RNAs that do not encode a protein, which were previously regarded as transcripts of “junk” DNA, may play crucial roles in gene expression (Mattick, 2005). These ncRNAs maintain several essential functions inside the cell, which are still being catalogued. The ability of RNA to perform both informational and catalytic functions, and their diverse regulatory roles has led to an “RNA world” hypothesis, i.e. RNA may have preceded both DNA and proteins in evolution in both coding and functional roles (Gilbert, 1986).

One important class of ncRNAs is the small nuclear (sn) RNAs of the eukaryotic spliceosome, which play a critical role in splicing of precursor messenger (pre-m) RNAs. In order for these RNAs to have such functional versatility, they must be able to interact with each other to form dynamic complexes and to fold into specific structures. The objective of this research is to characterize the conformational changes of the U2-U6 snRNA complex of the human spliceosome, the nuclear ribonucleoprotein assembly that mediates the catalysis of pre mRNA splicing. In order to address this question we have used a wide variety of techniques such
Figure 1.1. The central dogma of molecular biology (Crick, 1958). This is described as “DNA makes RNA makes Protein”
as enzymatic probing, analytical ultracentrifugation (AUC), hydroxyl radical footprinting, and small angle X-ray scattering (SAXS). Our results suggest that the human U2-U6 snRNA complex predominantly forms a four-way junction structure with formation of U2 Stem I, and that Mg$^{2+}$ has little effect on either the secondary or tertiary structure of the complex.

Before going into the specifics it is important to understand the building blocks of RNA i.e. nucleic acids their structure and properties.

1.2. Nucleic acids

Nucleic acids can be subdivided into two categories: Deoxyribo Nucleic Acid (DNA) and Ribo Nucleic Acid (RNA). Both DNA and RNA consist of long chains of repeating units called nucleotides. Each nucleotide is made up of three components a nitrogenous base, a ribose sugar, and an inorganic phosphate group.

Bases: There are four different nitrogenous bases in RNA and DNA; each of them is a derivative of one of two parent compounds, purine or pyrimidine (Figure 1.2 A). The naturally occurring pyrimidines are cytosine (C), uracil (U) found in RNA, and thymine (T) in DNA. The naturally occurring purines both in DNA and RNA are adenine (A) and guanine (G). The structure and native numbering of each base is shown in the Figure 1.2 A.

Sugars: There are two types of sugars present in nucleic acids, D-ribose in RNA and 2-deoxy D-ribose present in DNA. In both cases, sugars form a five-membered ring (pentose) called furanose and are numbered as shown in Figure 1.2 B. The numbering system 1´, 2´, 3´ and so on helps distinguish them from the base’s carbon atoms. The main difference between the sugars in DNA and RNA is the presence of hydroxyl group at 2´ position in RNA; in DNA a hydrogen atom is present at the same position. This structural difference will be reflected in their role in hydrolysis; specifically, RNA is more prone to hydrolysis because of the 2´-hydroxyl
Figure 1.2. A) Common bases found in Nucleic acids are shown adenine (A), guanine (G), cytosine (C), uracil (U) and Thymine (T). They are derivatives of either a purine or pyrimidine. B) Sugars present in Nucleic acids Ribose and Deoxyribose. C) Guanine is connected to a ribose sugar moiety forming Guanosine nucleoside.
group. This structural difference will be reflected in their role in hydrolysis; specifically, RNA is more prone to hydrolysis because of the 2′-hydroxyl group. The presence of 2′-OH in RNA also provides additional properties, including metal ion interaction and catalysis.

**Nucleoside:** The C1′ of the ribose/deoxyribose sugar is linked to a nitrogenous base (at the N1 position in pyrimidines and the N9 position in purines) via a β-N-glycosidic bond to form corresponding nucleosides (Figure 1.2 C). Nucleosides of A, G, U and C in RNA are called adenosine, guanosine, uridine and cytidine. Similarly, nucleosides A, G, T, and C in DNA are called deoxyadenosine, deoxyguanosine, deoxycytidine and deoxythymidine.

**Nucleotide:** Nucleotides are formed when 5′carbon of a nucleoside is connected to a phosphate group by an ester bond. Depending on the number of phosphate groups a nucleoside is attached to, nucleotides are termed as nucleotide monophosphate (attached to one phosphate group), diphosphate (two phosphate groups) and triphosphate (three phosphate groups) (1.2 D). The 3′-hydroxyl group of one nucleotide is linked to the 5′ hydroxyl of the following nucleotide through a phosphate group to form the backbone of the nucleic acids, a process that makes nucleic acid chains directional (i.e. they have 5′ → 3′ directionality). The nitrogenous bases, which are connected to the sugar groups, participate in hydrogen bonding with other bases, forming various secondary structure motifs of the nucleic acid (Figure 1.3) In the canonical Watson-Crick base pairing, A forms two hydrogen bonds with U, and G forms three hydrogen bonds with C. In addition to these canonical base pairs there are several non-canonical base pairs that have been found in RNA structures, the most common of which include the G•U wobble pair, the sheared G•A pair, the reverse Hoogsteen pair, and the G•A imino pair (Figure 1.4).
Figure 1.3. Common secondary structure motifs present in RNA. Adapted from (Chastain and Tinoco, 1991)
Figure 1.4. Common base pairs found in RNA double helices. Dashed lines represent hydrogen bonds. Both canonical (Watson-Crick) and non-canonical base pairs are shown.
1.3. RNA functional versatility:

Broadly, RNA molecules have two different roles inside the cell: 1) it can act as a information carrier from DNA to protein; and 2) it can fold into specific three-dimensional structures and interact with other biological molecules to perform particular functions inside the cell. There are three different types of RNAs that convert the genetic code of DNA into polypeptides: 1) mRNA molecules, which carry the protein building information from DNA to ribosomes (the molecular machines that catalyze protein synthesis); 2) ribosomal (r)RNA, which constitutes a major part of the ribosome; 3) transfer (t)RNA, an adapter molecule that transports appropriate amino acids into the ribosome to incorporate them into the newly synthesized protein.

The second class of RNAs, which maintain several functions inside the cell without being translated into a protein, are called as noncoding (nc)RNAs. There are several classes of ncRNAs exist in eukaryotic cells, these include snRNAs, micro (mi)RNAs, small interfering (si)RNAs, small nucleolar (sno)RNAs, riboswitches and catalytic RNAs. Each class of the RNAs has a specific function inside the cell.

**snRNAs:** Uridine rich U RNAs present inside the nucleus comes under the category of snRNAs. These snRNAs namely U1, U2, U4, U5 and U6 in association with several other proteins form ribonucleoprotein (RNP) complexes, which play an important role in the catalysis of pre-mRNA splicing.

**snoRNAs:** snoRNAs are a class of ncRNAs located inside the nucleolus, which vary in length from 70-250 nucleotides. The primary role of the snoRNAs is to guide chemical modifications of rRNAs, tRNAs and snRNAs (Eliceiri, 1999). snoRNAs can be subdivided into two classes, the C/D box snoRNAs that are associated with site-specific 2’-O-ribose methylations to rRNAs
(Kiss-Laszlo et al., 1996) and the H/ACA box snoRNAs which are associated with pseudouridylation of rRNAs (Ganot et al., 1997).

**miRNAs:** miRNAs are short RNA molecules that are approximately 22-26 nucleotides in length, they exist in many organisms such as the nematode *C. elegans*, flies, mice and humans (He and Hannon, 2004; Lee et al., 1993; Wightman et al., 1993). These small regulatory molecules play an important role in post-transcriptional gene silencing. miRNAs inhibit gene expression by translation repression mechanism; more specifically, they bind to the 3’ untranslated region of their target mRNA through imperfect base pairing, thus preventing the expression of functional protein or degrading the target mRNA.

**siRNAs:** siRNAs are a class of short double stranded RNA molecules averaging 20-25 nucleotides in length. Like miRNAs, the siRNAs also play an important role in RNA interference (RNAi) pathway, in which these RNA molecules target the 3’UTR of specific mRNA thereby repressing the translation. It is generally believed that RNAi is related with anti viral defense mechanism, transposon silencing, and post-transcriptional gene silencing in higher eukaryotes (Carthew and Sontheimer, 2009). Currently siRNA technology is used as a powerful tool by many researchers for the suppression of specific genes in mammals.

**Riboswitches:** A riboswitch is a part of a mRNA molecule that binds to a metabolite or metal ion and regulates the gene expression by forming altered structures in response to ligand binding (Tucker and Breaker, 2005). A riboswitch consists of two domains, the aptamer domain that acts as a sensor that specifically binds to a ligand, and the expression platform, which toggles between two different secondary structures in response to the ligand binding by aptamer domain, thus effecting the gene expression.
1.4. RNA as a catalyst: Ribozymes

For a long time it was thought that only proteins acted as biological catalysts, but in a seminal discovery in 1982 Thomas Cech and co-workers for the first time demonstrated that a single intron of large ribosomal RNA of Tetrahymena thermophila could cleave and splice itself in the absence of any external proteins or energy source (Kruger et al., 1982). One year later Sidney Altman and co-workers at Yale University showed that the RNA component of ribonuclease (RNase) P in the bacterium Escherichia coli can carry out the processing of precursor tRNA in the absence of any protein in vitro. In 1989 both Cech and Altman were awarded the Nobel Prize in chemistry for their influential work showing the catalytic role of RNA. RNA molecules with catalytic activity are called ribozymes, which stands for ribonucleic acid enzyme. Ribozymes fold into specific three-dimensional structures in order to perform efficient catalysis. Most ribozymes are metalloenzymes, requiring divalent metal ions (Mg$^{2+}$) in the active site to attain catalytically active structure. Depending upon the size and mechanism, naturally occurring ribozymes are classified into two categories, small and large ribozymes. Large ribozymes include RNA component of RNase P, group I and group II introns, with a size of few hundred nucleotides to 3000 nucleotides. Several small ribozymes were discovered, which include the hammerhead ribozyme (O'Keefe et al., 1996), the hairpin ribozyme (Buzayan et al., 1986), the hepatitis delta virus (HDV) ribozyme (Kuo et al., 1988), the varkud satellite (VS) ribozyme (Saville and Collins, 1990), and the glmS riboswitch ribozyme (Winkler et al., 2004). The size of the catalytic region of small ribozymes, which generally reside within the RNA genome of a virus, varies from 40 nucleotides to 154 nucleotides.
1.5. Pre-mRNA splicing:

One of the important tenets of the central dogma of biology is that one gene encodes one protein, which is very much true for bacteria and other prokaryotes. However, in eukaryotes the situation is more complicated, and most eukaryotic genes are expressed as precursor messenger (pre-m) RNAs that contain short coding regions (exons) interrupted by long stretches of nucleotides that do not encode a protein (introns). After being transcribed, these pre-mRNAs undergo extensive modifications inside the nucleus to become mature messenger (m) RNA that can be transported into cytoplasm for protein translation. These modifications include 5′-capping with 7-methyl guanosine, 3′-polyadenylation, and pre-mRNA splicing, which involves removal of introns and ligation of exons (Figure 1.5). Introns make it possible for eukaryotic cells to evolve new genes by a process called exon duplication using non-homologous recombination within the intron regions. Another advantage of having introns is that one gene can encode multiple proteins via alternative splicing (Black, 2003). Alternative splicing is a process in which exons of one gene are reconnected in different ways during splicing i.e. some exons are constitutively spliced and some exons are left out as cassette exons, which produces multiple isoforms of proteins. There are several modes of alternative splicing, which includes exon skipping, mutually exclusive exons, alternative donor site, alternative acceptor site, and intron retention modes (Figure 1.6). This process enhances the proteomic diversity of higher eukaryotes; recent deep sequencing studies have shown that more than 95% of human multi-exon genes are alternatively spliced (Pan et al., 2008). Exploring the process of pre-mRNA splicing at the molecular level is not only important for understanding gene expression, but is also of enormous medical relevance, because errors occurring in pre-mRNA splicing results in many cancers and neurodegenerative diseases (Licatalosi and Darnell, 2006).
Figure 1.5. Processing of pre-mRNA in eukaryotes includes 5’-capping, 3’ poly adenylation and splicing of introns, which involves removal of intervening sequences called introns and connecting the exons.
Figure 1.6. Basic modes of alternative splicing 1) Exon skipping: In this mode an exon termed as cassette exon may be spliced out or retained 2) Mutually exclusive exons: Any one of the two exons are retained in the final transcript but not both 3) Alternative 5’ donor sites: an alternative donor site is used in the splicing 4) Alternative 3’ acceptor sites: an alternative acceptor site is used 5) Intron retention: An intron may be spliced out or retained.
Introns are spliced out from pre-mRNA molecules by cleavage at conserved sequence called splice sites, conserved sequences at the 5´ and 3´ ends of the introns. The first nucleotide of the intron is usually called as splice donor/5´ splice site, in most cases the first two nucleotides of the intron are defined by GU. Another important sequence occurs 18 to 40 nucleotides upstream from the 3´ end of the intron, which is called the branch site and always contains an adenine; in higher eukaryotes the branch site is followed by a polypyr imidine tract. The very end (3´ side) of the intron is called the 3´-splice site/ splice acceptor, and is typically defined by a dinucleotide AG. The above-mentioned nucleotides are universally conserved. In addition to these universally conserved nucleotides there are consensus sequences in all introns, which will help spliceosome recognize the splice sites with absolute accuracy (Figure 1.7).

Splicing requires absolute precision because even a single nucleotide addition or deletion will shift the reading frame of mRNA, which results in coding of aberrant proteins, thus having adverse consequences. Several factors like splicing enhancers and silencers in conjunction with SR proteins, hnRNP proteins and spliceosomal machinery are responsible in recognizing precise intron exon boundaries and consensus sequences for splicing (Ward and Cooper, 2010). RNA-RNA, RNA-protein and protein-protein interaction play an important role during the catalysis of pre-mRNA splicing (Wahl et al., 2009).

1.6. Mechanism of splicing:

The pre-mRNA splicing reaction proceeds in two consecutive trans-esterification reactions, in the first step the 5´ phosphate of the first nucleotide of the intron is attacked by the 2´-OH of the branch site Adenosine, forming a lariat intermediate. In the second step, the last nucleotide of the intron at the 3´ splice site is attacked by the newly formed 3´-OH of the exon
Figure 1.7. Figure shows universally conserved nucleotides in the introns. And the consensus sequences near the 5’ splice site, branch site and 3’ splice site. The height of the alphabet in the motif logo signifies the conserved nature of the nucleotide. Adapted from (Lim and Burge, 2001)
by a nucleophilic attack thus joining the exons and releasing the intron lariat, which is subsequently degraded (Figure 1.8).

1.7. **Spliceosome: Structure and dynamics:**

Pre-mRNA splicing in eukaryotes is catalyzed by a large complex in the cell called spliceosome, arguably the most complicated and dynamic macromolecular machine. The human spliceosome is comprised of five recyclable small nuclear ribonucleoprotein (snRNPs) U1, U2, U4, U5 and U6 and numerous (~100 proteins in yeast, and >300 proteins in humans) non-snRNP proteins (Stark and Luhrmann, 2006). Each snRNP is made of a small nuclear RNA bound to a unique set of proteins as well as a shared common set of 7 sm proteins. The sm proteins B/B’, D1, D2, D3, E, F, and G proteins form a ring like structure that interacts with highly conserved uridyl-rich sequence called sm site of snRNAs, which forms the core snRNP domain (Kambach et al., 1999). The process of pre-mRNA splicing is carried out by spliceosome in a highly orchestrated way, which involves several ordered RNA-RNA, RNA-protein and protein-protein interactions. Figure 1.9 represents assembly and rearrangements of spliceosome at various steps of splicing. In the first step of spliceosome assembly, U1 snRNA base pairs with the 5′ splice site (SS) forming early (E) complex. In the second step U2 snRNA base pairs with the branch site (BS) region, forming U2-BS duplex in which the branched adenosine is in bulged out conformation, creating the pre-spliceosome (complex-A). In the next step the U4/U6.U5 tri-snRNP in which U4 and U6 snRNAs are extensively base paired with each other, is incorporated into the spliceosome forming mature spliceosome (Complex B1). At this stage major rearrangements in RNA-RNA and RNA-protein interaction occur. At first the U4/U6 interaction is disrupted and the U4 snRNA is ejected out. The U6 snRNA base pairs with the 5′ SS, replacing the U1 snRNA. In addition to these rearrangements, an extensive base pairing
Figure 1.8. Schematic of two-step transesterification reaction catalyzed by spliceosome. In the first step 2’ OH group of branch point Adenosine performs a nucleophilic attack on 5’ splice site thus freeing 3’ OH group of exon and resulting in the formation of lariat intron. In the second step the newly formed 3’ OH group attacks on 3’ splice sites joining both the exons. Adapted from (Brow, 2002)
Figure 1.9. Schematic representation of assembly of mammalian spliceosome. At first U1snRNP binds to the 5’ splice site to form E-complex. Next, U2 snRNP binds to the branch site to form A-complex. In the next step U4, U5 and U6 tri snRNP binds to form B-complex. Major rearrangements take place in the spliceosome at this stage, U1 and U4 snRNPs are ejected out to form B^{act} complex. At this stage, first step of splicing is catalyzed followed by the second step. The components of spliceosome disassemble for the next cycle. Adapted from (Will and Luhrmann 2010, CSHL press)
interactions are formed between U6 and U2 snRNA, which bring the 5’SS and BS closer, for the first step of splicing. The first step of splicing then occurs followed by the second step, after which the individual components are dissociated and recycled for the subsequent splicing reactions.

1.8. **Parallels between the spliceosome and group II introns:**

   Group II introns are a large class of self-catalytic ribozymes *i.e.* although their catalytic rates are enhanced by interactions with proteins *in situ*, they can catalyze the splicing reaction under high salt concentrations *in vitro* in the absence of any proteins (Michel et al., 1982). Both spliceosome and group II introns catalyze the splicing reaction through the identical two-step mechanism with similar stereochemistry (Dayie and Padgett, 2008). Moreover the intermediates and products produced during the splicing reaction by them are similar (Jacquier, 1990). As with group II introns, both steps of the splicing reaction catalyzed by the spliceosome are reversible (Tseng and Cheng, 2008). Phosphorothioate substitution analysis at specific sites of group II introns and the spliceosome indicated that both the reactions require divalent metal ions for catalytic activity and that the catalytically important metal ions have similar arrangement at the active site for both the reactions (Chanfreau and Jacquier, 1994; Gordon et al., 2000a, b; Maschhoff and Padgett, 1992; Sontheimer et al., 1999a; Sontheimer et al., 1997; Toor et al., 2008). In addition to these mechanistic similarities, there are significant parallels in the sequence and secondary structures between some domains of group II introns and snRNAs (Figure 1.10). The secondary structure of group II introns is highly conserved and is characterized by six domains named domain I to VI, which fold into a complicated tertiary structure (Keating et al., 2010; Toor et al., 2008). The primary sequence and overall secondary
Figure 1.10. Cartoon showing the structural similarities between various domains of group II introns and U snRNAs of spliceosome. Adapted from (Sharp, 1991).
structure of domain V, the catalytically important and most conserved domain of group II introns, has close resemblance with the U6 snRNA intramolecular stem loop (Sashital et al., 2004a; Sigel et al., 2004)(Figure 1.11). The functional equivalence of these two structures have been shown by domain swapping experiments, where domain V of group II intron is shown to replace the U6 intramolecular stem loop (ISL) in an in vivo splicing assay, which further substantiates the similarities between these two structures (Shukla and Padgett, 2002).

The interaction between U2 snRNA and the branch site of introns is similar to domain VI of the group II intron. In both cases the branch site adenosine residue, which acts as a nucleophile for the first step of splicing, is positioned in a bulged conformation by a phylogenetically conserved psuedouridine residue in U2 snRNA, the kind of architecture required for optimal catalysis (Chu et al., 1998; Greenbaum and Newby, 2002; Newby and Greenbaum, 2003).

The binding between exon binding site (EBS) 1, 2 and 3 of domain I of group II introns with exonic sequences is similar to that of interaction between U5 snRNA and exons. Domain swapping experiments in which the domain I of group II introns was replaced by U5 also confirmed their functional equivalence (Hetzer et al., 1997). Another subdomain of group II intron that has a functional counterpart in U1 snRNA is a subdomain of domain I (ε′), which plays an important role in the recognition of 5´splice site. These extensive mechanistic parallels between reactions catalyzed by the spliceosome and the self-splicing group II intron (Sontheimer et al., 1999b; Valadkhan and Manley, 2002b; Weiner, 1993), as well as sequence and structural similarities between functionally analogous RNA sequences in the two systems (Gordon et al., 2000b; Keating et al., 2010), led to the hypothesis that the spliceosomal snRNAs
**Figure 1.11.** Figure shows the structural and sequence similarities of Domain V of group II intron with U6 ISL of yeast and human U6 snRNA. All of them have similar structure i.e two short helices separated by an asymmetric bulge which binds to a catalytically important metal ion. All the structures contain highly conserved AGC triad, which interacts with catalytically critical divalent metal ions. Adapted from (Valadkhan, 2010)
and group II introns might share evolutionary ancestors and thus, may also share a catalytic role.

1.9. Role of snRNAs in catalysis: is the spliceosome a ribozyme?

One of the longstanding debates in the field of splicing is whether the spliceosome is a ribozyme. There are several lines of evidence supporting the possible role of RNA mediated catalysis in splicing. The arguments that are in favor of RNA mediated catalysis include: 1) snRNAs are absolutely required for splicing activity; moreover, certain mutations and backbone substitutions have severe effects on catalysis, both in the first and second steps of splicing (Fabrizio and Abelson, 1990a, b, 1992; Madhani and Guthrie, 1992; Mcpheeters and Abelson, 1992b); and 2) several intermolecular structures formed between U2, U5, and U6 snRNAs and pre-mRNA have close resemblance to the structures formed by domains of group II introns (Keating et al., 2010). All these studies fueled the hypothesis that the spliceosome itself were derived from the group II self-splicing introns (Cech, 1986; Jacquier, 1990; Sharp, 1985).

Spliceosome purification at various stages of its assembly have shown that out of the five snRNAs, U1 and U4 snRNAs leave the spliceosomal assembly before the time of catalytic activation; therefore, it is safe to assume that they are not part of the catalytic core of the spliceosome required for catalysis. The conserved loop of U5 snRNA interacts with the exons before second step of splicing, properly positioning them for the reaction. However, the U5 snRNA is shown to be unessential for both the steps of splicing in mammals (O'Keefe and Newman, 1998; O'Keefe et al., 1996; Segault et al., 1999). These results suggest that U2 and U6 are the only snRNAs that are required for both steps of splicing, which might form at least a part of catalytic core of spliceosome.

U6 snRNA is the functional counterpart of domain V, the catalytic domain of group II introns. Two important and highly conserved regions in U6 snRNA are the ACAGAGA box and
the AGC triad; these regions are also shown to bind functionally important metal ions (Huppler et al., 2002a; Sontheimer et al., 1997; Yean et al., 2000a; Yuan et al., 2007b). Mutagenesis analysis performed by several groups on these regions have shown that, the mutations in these regions block both the steps of splicing reaction (Datta and Weiner, 1993; Hilliker and Staley, 2004; Lesser and Guthrie, 1993a; Wolff et al., 1994). Another indication that snRNAs mediate catalysis is there metal ion dependent pathway of splicing (Sontheimer et al., 1997). Phosphorothioate backbone substitution studies in several positions of U6 snRNA are not compatible with splicing, specifically at position 74 of human U6 snRNA the phosphate that is substituted was shown to be involved in direct contact with the catalytically important divalent metal ion (Valadkhan and Manley, 2002a; Yean et al., 2000a; Yuan et al., 2007b). NMR studies have shown that the U6 ISL has intrinsic capability of binding to a divalent metal ion (Huppler et al., 2002a). These studies using metal ions suggest that U6 RNA is capable of binding catalytically important metal ions, and thus potentially use them for catalysis.

The recently solved crystal structure of a self-splicing group II intron (Toor et al., 2008) has shown that the active site of the group II intron is formed by the AGC triad, an asymmetric internal bulge of domain V and a short purine-rich sequence between domains II and III, a functional equivalent to the ACAGAGA box in U6 snRNA. Finally, the most compelling evidence that supports the RNA mediated catalysis is that reactions similar to splicing are performed in vitro by protein free U2, U6 snRNA fragments, albeit with low rate and efficiency (Lee et al., 2010; Valadkhan and Manley, 2001; Valadkhan et al., 2009a; Valadkhan et al., 2007) leading to the conclusion that cofactors are required for splicing of pre-mRNA inside the cell.
1.10. Role of proteins in pre-mRNA splicing:

On the other side of the story, there is a growing body of evidences supporting the idea that protein comprises a part of spliceosome’s active site. Although genetic analysis has demonstrated the requirement for dozens of spliceosomal proteins in catalysis, the most likely candidate to play a critical role in splicing is Prp8, which is the largest and highly conserved protein in the entire spliceosome (Hodges et al., 1995). Prp8 is located at the heart of the spliceosome and extensively interacts with crucial sequences required for splicing, such as the 5′ splice site (Siatecka et al., 1999; Teigelkamp et al., 1995a), branch site (Collins and Guthrie, 1999) and 3′ splice site of pre-mRNA (Teigelkamp et al., 1995a; Teigelkamp et al., 1995b) and to the U6 (Vidal et al., 1999) and U5 (Dix et al., 1998) snRNAs. In addition to these interactions with RNA fragments, Prp8 is implicated in interacting with several proteins in spliceosome including a helicase Brr2, which is required for splicing catalysis and unwinding helices formed between snRNA components during the course of the reaction (Laggerbauer et al., 1998; Raghunathan and Guthrie, 1998). Recent studies by Query and co-workers have suggested that the spliceosome undergoes significant change in conformation between the two steps of splicing and that Prp8 plays a pivotal role (Konarska et al., 2006). Mutational analysis of Prp8 identified certain group of mutants that improve the second step of splicing by inhibiting the first step (Query and Konarska, 2004b) and vice versa (Liu et al., 2007b).

Taken together, these data the authors suggested a two-step model for spliceosome during catalysis i.e. spliceosome adapts two different conformations for two steps of splicing. The first conformation favors the first step of splicing inhibiting the second and the second conformation supports the second step of splicing while constraining the first step. This model is reinforced by recent studies from Tseng and Cheng, who have shown that both the steps of splicing are
reversible under certain salt conditions (Tseng and Cheng, 2008). All these studies have shown that the Prp8 protein is indispensable for the splicing reaction; in order to characterize further the function of Prp8, several researchers have made a major effort to solve high-resolution structure of the protein, however this is not successful until now.

Three independent research groups recently solved crystal structures of a domain located near the carboxyl terminal of human/yeast Prp8, which contains the amino acids that interact with the 5’ splice site of the pre-mRNA (Pena et al., 2008; Ritchie et al., 2008; Yang et al., 2008). Surprisingly, they have all discovered that the subdomain fold is similar to that of the RNase H domain, with a five-stranded β-sheet surrounded by three α-helices. Even though the overall fold of the domain is similar to RNase H, some of the key catalytic residues are missing at the active site. Ritchie and co-workers have shown that the fragment of protein crystallized can interact with model RNAs that contain the sequences of U2, U6 snRNAs and 5´ splice site (Ritchie et al., 2008). Furthermore, single point mutations in the RNase H-like domain of Prp8 were detrimental to cell growth. All the groups made an attempt to determine if the domain binds to any divalent metal ions; however, no metal ion binding was observed when the crystals were grown in Mg$^{2+}$ concentrations as high as 200 mM. Taken together, these data suggest the RNase H-like domain of Prp8 protein may play an important role in forming at least a part of the active site with other residues of the protein and/or RNA.

Irrespective of whether a spliceosome is a ribozyme or not the snRNAs in particular, U2 and U6 snRNAs play a critical role in catalysis. Also, proteins are required (as cofactors) for assembly of spliceosome and to coordinate several rearrangements between the RNA fragments during the course of the reaction.
1.11. **Conformation of the U2-U6 snRNA complex:**

The spliceosomal U2 and U6 snRNAs are indispensable for splicing and they are evolutionarily most conserved of the snRNA molecules. In addition, the complex is implicated in catalytic activity, which depends upon several specifically bound Mg$^{2+}$ ions (Huppler et al., 2002b; Sontheimer et al., 1997; Valadkhan and Manley, 2002b; Yean et al., 2000b; Yuan et al., 2007a), one of which is located within the highly conserved AGC triad of U6 snRNA. Catalytic activity, albeit at a low rate and yield, by the protein-free human U2-U6 snRNA complex underscores the importance of the snRNA components in the splicing reaction (Valadkhan and Manley, 2001; Valadkhan et al., 2009b; Valadkhan et al., 2007). All these studies suggest the importance of U2 and U6 snRNAs in the catalysis of pre-mRNA splicing, in order for efficient catalysis U2 and U6 snRNAs should interact and fold into a specific 3-D structure.

The U2 and U6 snRNA molecules are shown to form extensive intra- and intermolecular base pairing sequences, which are thought to act as scaffolds in positioning both the 5′ splice site and the branch site region of the intron for the first cleavage reaction (Black et al., 1985; Lesser and Guthrie, 1993b; Parker et al., 1987; Wachtel and Manley, 2009). The distance between the bound metal ions in the secondary structure of the U2-U6 snRNA complex (ACAGAGA loop, AGC triad and U74) (highlighted yellow in Figure1.12. B) implies the need for folding and probably for conformational change during the splicing reaction. Therefore it is important to characterize the precise nature of interaction between U2 and U6 snRNAs and also the conformational changes of the U2-U6 snRNA complex during the course of the catalysis reaction. The overall goal of this research is to characterize the conformational changes of U2-U6 snRNA complex upon addition of Mg$^{2+}$ ions. In order to pursue this, it is necessary to
Figure 1.12. A) Three helix junction model: proposed secondary structure model for yeast U2-U6 snRNA complex (Madhani and Guthrie, 1992). Adapted from (Burke et al., 2012) B) Four-helix junction: proposed secondary structure model for human U2-U6 snRNA complex, and also protein free yeast U2-U6 complex (Sashital et al., 2004a; Sun and Manley, 1995a). Adapted from (Valadkhan, 2010)
characterize ground state (i.e. lowest energy) structure of the complex. The structure of the U2–U6 pairing has been a matter of debate for years. Early in vivo genetic studies on yeast U2-U6 snRNA complex support a three-helix structure in which the highly conserved AGC triad in U6 forms three base pairs with U2 (Madhani and Guthrie, 1992). NMR studies on the truncated version of the yeast U2-U6 snRNA complex have shown that in the absence of proteins and Mg$^{2+}$ AGC triad forms base pairs with U6 nucleotides (Sashital et al., 2004a) (Figure 1.12 B). However, a recent model of the protein-free yeast U2-U6 snRNA complex derived from a combination of solution NMR, small angle x-ray scattering, and computer modeling of a sequence in which an additional four basepairs of the native sequence in Helix II were included indicated formation of a three-helix junction structure consistent with the conformation identified in cellular studies of yeast (Burke et al., 2012). Single molecule studies have identified multiple conformations of U2-U6snRNA complex depending upon the concentration of Mg$^{2+}$ ion concentration( Guo et al., 2009). There is no clear consensus on the secondary structure of human U2-U6 snRNA complex. Interactions identified by genetic suppression assays were consistent with formation of an extended U6 snRNA ISL and formation of U2 Stem I (Sun and Manley, 1995a) (Figure 1.12 B), indicating that interactions characteristic of the four-helix junction are required for active splicing. Their data also suggested that Helix Ib and the three-helix structure (Figure 1.12 A) are not required for splicing activity. Even though the mechanism of splicing is identical in yeast and in humans (Moore and Sharp, 1993), it is possible that the conformation of human and yeast U2-U6 snRNA complex might be different.

The goal of our experiments is to probe the lowest energy/ground state structure of human U2-U6 snRNA complex and the effect of Mg$^{2+}$ ions on the structure.
1.12. **Specific aims:**

In order to analyze conformational changes associated with catalytic activity, it is essential to know the lowest energy conformation i.e. in the absence of metal ion, or protein components that may induce structural changes. To address this question we used biochemical structure probing, solution NMR, hydroxyl radical footprinting, analytical ultracentrifugation and small angle x-ray scattering studies. Our results are consistent with the formation of a four-helix junction characterized by presence of U2 Stem I, rather than the three-helix structure, as the predominant fold. Mg$^{2+}$ has very little effect on the secondary structure and tertiary structure of the complex.
CHAPTER 2

SECONDARY STRUCTURE PROBING OF THE HUMAN U2-U6 snRNA COMPLEX

2.1. Introduction:

The U2-U6 snRNA complex plays a critical role in pre-mRNA splicing. In order to perform this role efficiently, the U2-U6 snRNA complex must fold into a specific structure, which involves both secondary (base pairing) and tertiary structural components. In this chapter, I will discuss and analyze the secondary structural features of human U2-U6 snRNA complex. I have used enzymatic structure probing experiments to characterize the secondary structural features of this complex. Before discussing the results, I will briefly explain what is known about the secondary structure of the U2-U6 snRNA complex and why it is important to characterize the structural features of the complex.

It has been shown that there is extensive inter- and intramolecular base pairing interaction between the U2 and U6 snRNAs. The precise nature of the interactions between the snRNAs is a matter of debate over the years, with two different secondary structural models proposed (Figure 2.1). Genetic studies in the yeast *Saccharomyces cerevisiae* support the importance of formation of Helix Ib, in which the AGC triad forms three intermolecular base pairs with U2 snRNA and U2 Stem I is opened, forming a three-helix structure (Figure 2.1.B) (Madhani and Guthrie, 1992). This model was reinforced by mutational studies that implicated the requirements of Helix Ib in at least one (Hilliker et al., 2007), and possibly both (Mefford and Staley, 2009), cleavage steps.
Figure 2.1. Models for the conformation of the protein-free human (h) U2-U6 snRNA complex. (A) Four-helix junction model proposed by Sun and Manley (1995). (B) Three-helix model adapted from the model proposed by Madhani and Guthrie (1992) for the yeast U2-U6 snRNA complex. The sequences shown are fragments of hU2 and hU6 snRNAs with several changes made to the native sequences to increase the transcription yield and pairing efficiency as specified in Materials and Methods. For $^{19}$F NMR studies, C13 of U2 snRNA was substituted with 5-fluoro-cytidine (5-$^{19}$F-C; red nucleotide).
Additionally, results of cross-linking assays in yeast identified tertiary interactions that could only exist in the three-helix model (Ryan et al., 2004).

In contrast, NMR investigation of a truncated complex suggested a different structural model, in which the U6 ISL is extended to include the AGC triad and U2 snRNA forms intramolecular Stem I, therefore creating a four-helix junction (Figure 2.1 A). In addition to it, no spectral evidence was reported for an alternative conformation (Sashital et al., 2004b). This finding is consistent with data of McPheeters and Abelson, which had demonstrated that the presence of certain base pairs in Stem I was important for splicing in yeast (McPheeters and Abelson, 1992a). However, a recent model of the protein-free yeast U2-U6 snRNA complex derived from a combination of solution NMR, small angle x-ray scattering, and computer modeling of a sequence in which an additional four base pairs of the native sequence in Helix II were included indicated formation of a three-helix junction structure consistent with the conformation identified in cellular studies of yeast (Burke et al., 2012). Single molecule studies on protein-free yeast U2-U6 snRNA complex identified at least three distinct states, including the three-way and four-way junction structures in equilibrium with each other, in the presence of Mg$^{2+}$ (Guo et al., 2009), suggesting the propensity for ion-dependent conformational change. Computational studies from Cao and Chen have predicted the presence of more than one conformation of yeast U2-U6 snRNA complex, depending upon the sequences used for the study (Cao and Chen, 2006).

The situation appears to be different in case of human U2-U6 snRNA complex: interactions identified by genetic suppression assays were consistent with formation of an extended U6 snRNA ISL and formation of U2 Stem I (Sun and Manley, 1995b), indicating that interactions characteristic of the four-helix junction are required for active splicing. Their data
also suggested that Helix Ib and the three-helix structure are not required for splicing activity. The proposed four-helix junction model is consistent with NMR studies by Sashital and coworkers on yeast U2-U6 snRNA. All evidence indicates that the chemical mechanism of splicing is identical in yeast and human systems (Moore and Sharp, 1993), however because of the difference in sequences and variable resistance to the mutations of the yeast and human U2-U6 snRNA complex, it is possible that the human and yeast U2-U6 snRNA complexes adopt different folds in the protein-free state. The computational study by Cao and Chen also supported the formation of a four-way junction model as the lowest energy conformation of human U2-U6 snRNA complex in the protein free state (Cao and Chen, 2006).

In order to measure the conformational changes of the U2-U6 snRNA complex during the course of the reaction, it is important to characterize the ground state conformation of human U2-U6 snRNA complex i.e. in its protein free state. Both the current models (Figure 2.1) proposed comprises U2-U6 snRNA intermolecular helices I and II as well as a U6 intramolecular stem loop (U6 ISL); the features that vary include subhelices of Helix I, as well as lengths of the other helices and appearance or absence of an intramolecular U2 snRNA Stem I. Importantly, the models depict the genetically conserved AGC triad in different pairing environments, which may have an impact on folding and function. The presence or absence of helix Ib plays a major role in orientation of several functionally critical regions like evolutionarily conserved ACAGAGA loop, branch point binding sequence of U2 snRNA, the AGC triad and U6 ISL. These conserved regions also contain a functionally important metal ion-binding site (Huppler et al., 2002a; Yean et al., 2000a; Yuan et al., 2007b), therefore it is important to measure the ground state conformation of the human U2-U6 snRNA complex. The goal of this chapter is to probe the secondary structure of the human U2-U6 snRNA complex in the absence of proteins and introns,
and to characterize the effect of divalent metal (Mg$^{2+}$) ions on the secondary structure. We have used enzymatic structure probing technique to address the question. The results obtained are consistent with the formation of a four-helix junction characterized by the presence of U2 Stem I, rather than the three-helix structure, as the predominant fold. However, NMR studies (performed by Caijie Zhao from our laboratory) also identified a small but significant fraction of the U2-U6 snRNA complex forming an alternative conformation, which may be the three-helix structure in equilibrium with the major conformation.

2.2. Materials and methods:

2.2.1. RNA design and synthesis:

Based upon proposed models for secondary structural folds of the yeast (Burke et al., 2012; Madhani and Guthrie, 1992; Sashital et al., 2004b) and the human U2-U6 snRNA complexes (Sun and Manley, 1995b), RNA fragments representing the regions of human U2 and U6 snRNA sequences were designed with several modifications to minimize formation of undesirable self-paired complexes (Figure 2.1). Specifically, we replaced the hairpin loop of U6 snRNA ISL, GCGCA, with the yeast sequence GCAUA, changed U9 of the U2 strand to A9 to form a complementary pair with U89 of the U6 strand, and truncated the 3′ and 5′ sequences of U6 and U2, respectively, so that helix III comprised 9 base pairs. In addition, two guanosine residues were added to the 5′ end of each strand for efficient in vitro transcription, as well as cytidine residues on the 3′ termini. In order to verify if the mutations have any effect on secondary structure, we have used secondary structure prediction software such as m-Fold (Zuker, 2003). The sequences containing the mutations are submitted into the software with the
wild type sequence as a control. The output results of the software suggested that none of these mutations were likely to induce any change in the secondary structure of the complex. In addition, the mutations are not situated in the junction region of the complex, which has different structure in the two major models. The relative free energies for formation of secondary structures were estimated using m-Fold (Zuker, 2003).

2.2.2. Design of the mutant:

To validate the results obtained with the wild-type U2-U6 snRNA complex (see results section for details), we have designed mutants that will stabilize the formation of Stem I, thus favoring the formation of four-helix. In total, we have designed two mutants: (1) extension of U2 Stem I mutant by addition of two extra G-C/C-G base pairs to increase thermal stability of the stem (2) substitution of the UUUU tetra-loop sequence of Stem I to the hyperstable sequence UUCG, and mutation of the top pair of Stem I from G-C to C-G (which would inhibit the formation of Helix Ib) (Figure-2.2). Both of these mutations favor the formation of Stem I and the four-helix structure. Both hU6 and hU2 RNAs, as well as the mutant sequences with out 19F-substituted nucleotides, were transcribed from synthetic double-stranded DNA templates (Integrated DNA Technologies) using T7 RNA polymerase expressed and purified in the laboratory. Transcribed RNA was PAGE-purified, eluted using an electroeluter, precipitated, washed with a suitable buffer using a Centricon filter, dried, and resuspended to the final concentration.

In order to characterize intermolecular pairing and exclude significant contributions of self-paired U2 or U6 snRNA, we measured migration of individual and paired strands on a nondenaturing gel. Equal amounts of purified U2 and U6 snRNA strands were heated to 70°C
Figure 2.2. Sequence and proposed secondary structure of a mutation favoring the four-helix model. In this mutant, the UUUU tetraloop of U2 snRNA was mutated to a hyperstable UUCG loop and the top base pair in Stem I was mutated from G-C to C-G in order to disfavor the formation of Helix Ib. The mutated nucleotides from the original sequences are in blue and the 5-^{19}F-C substitution is in red.
for 3 min in a buffer containing 50 mM Tris, 100 mM NaCl, pH 7.5, and cooled at room temperature for 30–45 minutes. Samples of U2 and U6 strands alone were treated equivalently and observed as controls. The reaction mixtures were loaded onto a 12% nondenaturing gel and electrophoresed at 100 mV for 4 h at 4°C. Following staining by Nuclisstain, we observed a single band in the lane with combined strands, which migrated more slowly than individual U6 or U2 strands, and which we attributed to the U2-U6 snRNA complex (Fig. 1G). The pairing experiment was repeated in various buffers that were used for the enzymatic structure probing and spectroscopic experiments, with the same result. In all cases, we saw essentially complete pairing of the U2 and U6 snRNA strands, with no evidence of unpaired or self-paired U2 or U6 snRNA strands. For the mutant sequences, we verified pairing of the two strands by the method outlined above.

2.2.3. Enzymatic structure probing:

RNA molecules can fold into distinct three-dimensional architectures to perform wide range of cellular functions. Precise studying of the secondary and tertiary structures of these RNA molecules will help us understand the diverse functional role of RNAs. High-resolution structures of hundreds of RNAs and RNA protein complexes were solved using X-ray crystallography or solution state NMR. However these techniques require large amounts of purified sample and are restricted by the size of the molecules and the ability to form a single, in some cases crystallizable, conformation. In case of nucleic acids, secondary structure *i.e.* the base-pairing pattern, also gives important information about the biological functions of the RNAs. Because of this ability of RNA to form biologically important secondary structures, there
are several researchers attempting to develop the techniques that probe the secondary structures of RNA molecules (Wilkinson et al., 2006; Zuker, 2003). These techniques include both computational (e.g., RNA structure prediction software and tools such as m-fold) and experimental methods. In this work, ribonucleases were utilized to probe the RNA secondary features.

2.2.4. Principle and background:

To understand the technical details of the methods used and to analyze the validity of the results, it is important to understand the basics of the structure probing technique. At first the RNA of interest is transcribed in vitro or extracted from cells, purified, and folded in the required buffer. The purified RNA is treated with a reagent (organic molecule, enzyme or metal ion) that modifies the RNA at a specific position, characteristic of the reagent. The modification can result in strand scission of the RNA molecule or in the formation of a covalent adduct between the RNA molecule and the reagent being used. These modifications are performed under “single hit” kinetics i.e. less than one “hit” per molecule. The strand scission or modification may alter the secondary structure of the RNA thereby inducing secondary cuts that are not specific to the modifying reagent. That is why it is important to perform all hydrolysis reactions under single hit kinetics. There are two different paths to identify the position of these modifications on the RNA molecules. The first method is using end-labeled RNA molecules; this method is only used with the reagents that can cause the cleavage of RNA upon reacting. The second method involves reverse transcription using complementary primers. This method is used for reagents that result in the strand cleavage or modification, in either case the modification is detected as a stop in primer extension. The choice of which method to use depends upon the length of the RNA molecule that is being probed and by the nature of the nucleotide positions. If the RNA is above
150 nucleotides in length we can use reverse transcription method to characterize the secondary structure.

There are wide varieties of RNases and chemical probes that are specific to single stranded and structured regions. Different chemical and enzymatic probes, their characteristics and specificities are shown in the table below (Table 2.1). Both chemical and enzymatic probes have certain advantages and disadvantages in structure probing of RNAs in solutions.

2.2.5. **Enzymatic probes:**

Nucleases as shown in the above table are bulky, and are sensitive to steric hindrance *i.e.* sometimes they cannot access certain tightly sequestered regions of folded RNAs. Each nuclease has optimum conditions for hydrolysis, however when we are probing the RNA structure in a particular buffer with various enzymes we might have to use sub-optimal conditions, which are different for different enzymes. However, if the RNA that is being probed is small (~100 nucleotides or less), RNases act as very good probes that characterize the secondary structure information at a single nucleotide resolution.

2.2.6. **Chemical probes:**

Unlike enzymatic probes, chemical probes are small and are therefore less sensitive to steric hindrances. The chemical probes modify the RNA by forming a covalent adducts instead of strand scission and this modification is detected by reverse transcription using complementary primers. Because of the technical difficulties during reverse transcription and subsequent electrophoresis methods, information for the first 35 nucleotides and last 20 nucleotides of the
<table>
<thead>
<tr>
<th>Probes</th>
<th>Mol. Weight</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNase T1</td>
<td>11,000</td>
<td>unpaired G</td>
</tr>
<tr>
<td>RNase U2</td>
<td>12,490</td>
<td>unpaired A&gt;G</td>
</tr>
<tr>
<td>RNase CL3</td>
<td>16,800</td>
<td>unpaired C&gt;&gt;A&gt;U</td>
</tr>
<tr>
<td>RNase T2</td>
<td>36,000</td>
<td>unpaired N</td>
</tr>
<tr>
<td>nuclease S1</td>
<td>32,000</td>
<td>unpaired N</td>
</tr>
<tr>
<td><em>N.c. nuclease</em></td>
<td>55,000</td>
<td>unpaired N</td>
</tr>
<tr>
<td>RNase V1</td>
<td>15,900</td>
<td>paired or stacked N</td>
</tr>
<tr>
<td>DMS</td>
<td>126</td>
<td>N3-C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N1-A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N7-G</td>
</tr>
<tr>
<td>DEPC</td>
<td>174</td>
<td>N7-A</td>
</tr>
<tr>
<td>CMCT</td>
<td>424</td>
<td>N3-U</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N1-G</td>
</tr>
<tr>
<td>kethoxal</td>
<td>148</td>
<td>N1-G, N2-G</td>
</tr>
<tr>
<td>bisulfite</td>
<td>104</td>
<td>unpaired C---&gt;U</td>
</tr>
<tr>
<td>ENU</td>
<td>117</td>
<td>Phosphates</td>
</tr>
<tr>
<td>MPE-Fe(II)</td>
<td>780</td>
<td>paired N</td>
</tr>
</tbody>
</table>

**Table 2.1.** Table shows various enzymatic and chemical probes that are generally used for RNA structure probing, their molecular weight and specificities. Adapted from (Ehresmann et al., 1987)
RNA is lost. This makes it difficult to use chemical probes on smaller RNA constructs. For this reason, I have opted to use enzymes to probe the secondary structure of U2-U6 RNA construct.

2.2.7. Enzymatic probing Method:

Individual U6 and U2 snRNA fragments were dephosphorylated at the 5′ termini using Antarctic phosphatase (New England Biolabs) and labeled with γ-32P-ATP using T4 polynucleotide kinase (New England Biolabs). Labeled RNA was purified on a 12% denaturing gel, eluted by a crush-and-soak method, precipitated by ethanol, dried and resuspended in 50 mM Tris, 100 mM NaCl pH7.5 for all the RNase A and RNase T1, with 1 mM MgCl₂ for assays with RNase V1 (minimal [MgCl₂] requirement for enzyme activity). Each labeled strand was paired with a 1.5 times excess of unlabeled partner strand using the above-mentioned protocol. To assess pairing in the low concentrations generally used for 32P-labeled samples, an aliquot of paired samples (with nanomolar concentrations of either labeled U2 or U6 strand) was electrophoresed on a nondenaturing gel against labeled individual strands. From quantification of paired vs. unpaired strands, we observed that ~100% of each labeled strand was in a paired form. The resulting complex was individually subjected to partial cleavage by ribonucleases RNase A, T1 (both from Ambion Inc.), and RNase V1 (Pierce Milwaukee). Enzymatic probing assays were performed essentially by protocols accompanying the enzymes, using tRNA as carrier and identifying conditions that resulted in less than one “hit” per labeled molecule. For each experiment, reference guanosine ladders were prepared by incubation with RNase T1 (Fermentas Inc,) under semi-denaturing conditions, as were alkaline hydrolysis ladders. To test the effect of high [Mg^{2+}] on secondary conformation, MgCl₂ was added to a final concentration of 100 mM to the folded RNA and equilibrated at room temperature for 30 min. The complex was exposed to various enzymes as described earlier. The samples obtained were electrophoresed on a
denaturing gel against T1 and alkaline hydrolysis ladders (Figure-2.3). The gel was exposed to a phosphor screen for 12-16 hours (overnight) at 4 °C and scanned by a phosphorimager (STORM scanner from GE health care). The resulting picture was analyzed with SAFA (Laederach et al., 2008), which calculates the density of each band. The band densities were normalized to the total radioactivity per lane, and the resulting values were then plotted against the sequence of the RNA to give the cleavage profiles for various enzymes (See supplementary data). Cleavage intensities were mapped onto the two possible secondary structural folds of the complex.

2.3. Results:

To investigate the ground state secondary structure of the hU2-U6 snRNA complex in solution, we used enzymatic probing on the complexes formed by U2 and U6 snRNA fragments representing the functionally important sequences. Strand pairing to form a single complex with a stoichiometry of 1:1 was confirmed by nondenaturing gel electrophoresis under the conditions similar to those used for the different experiments (Figure-2.4) (details in Materials and Methods).

2.3.1. Enzymatic structure probing:

RNase V1 preferentially cleaves at regions that are double stranded or highly stacked. The products of the RNase V1 digestion conducted at 1 mM MgCl₂ were resolved and quantified to generate cleavage profiles (representative gel in Figure-2.3, and cleavage profiles in supplementary data). We observed greatest cleavage of the labeled hU2 strand (*hU2) strand at nucleotides 9-14, 19-21, and 26-28. *hU6 exhibited the most cleavage at nucleotides 44-52, 56-58, 62-64, 70-72, and 77-80. Very short fragments were not visible due to gel artifacts, however these sequences are predicted to be double stranded in either model.
Figure 2.3. Representative sequencing gel exhibiting cleavage patterns of $^{32}$P-labeled (*) hU2 and *hU6 snRNA paired with respective counter strands subjected to enzymatic probing by ribonucleases RNase A, T1 (see Materials and Methods for experimental details). Lanes are labeled on top with the ribonuclease used and the concentration of Mg$^{2+}$. 
**Figure 2.4.** Pairing of RNA oligomers representing hU2 and U6 snRNA analyzed by non-denaturing gel electrophoresis. Lanes 1 and 2 show relative migration of fragments representing U6 and U2 snRNA fragments, respectively; lane 3 demonstrates retarded migration upon annealing of the two snRNA fragments. All samples were electrophoresed in a single gel; lanes were cut to delete lanes of unrelated samples.
short fragments were not visible due to gel artifacts; however these sequences are predicted to be double stranded in either model. Formation of the stems associated with Helix III and Helix II was verified by characteristic imino-imino and imino-amino patterns in $^1$H NMR spectra performed by a fellow Ph.D. student in the Greenbaum group, Caijie Zhao.

The cleavage profile was mapped onto schematic secondary structural models corresponding to the four-helix and three-helix folds (Figure-2.5). The cleavage pattern for the U6 strand is consistent with expected formation of the U6 intramolecular stem loop (ISL), a common feature in both models. In addition, the intense cleavage patterns of several residues in the segment G12-C21 of U2 snRNA suggest formation of a short stem corresponding to the position of the proposed U2 snRNA Stem I, a prominent feature of the four-helix junction model.

We probed the single-stranded regions by reaction with RNase A (unpaired C and U) and RNase T1 (unpaired G). Cleavage profiles (Supplementary Data) indicate the greatest intensities for *hU2* at nucleotides 15-18, 21-24, and 28-36, with essentially no cleavage observed at nucleotides 12-14, 19-20, and the most intense cleavage for *hU6* at nucleotides 44-49, 52-55, 64-69, and 81-86. These locations, mapped onto secondary structural models (Figure-2.6) coincide with the expected unpaired residues in the hU6 ISL and the ACAGAGA loop, features that are common to both models. Other cleavage sites correspond to regions whose secondary structure differs in the two models: for example, when mapped onto the schematic figures of the two models, nucleotides 12-14, 19-20 of U2 snRNA map to a segment anticipated to be single stranded in the three-helix structure but double stranded in in the four-helix model. The lack of extensive cleavage in these regions, along with evidence for formation of Stem Loop I from the RNase V1 cleavage patterns, implies the formation of a four-helix junction by the U2-U6 snRNA complex under these conditions.
Figure 2.5 A & B. Normalized cleavage intensities of complex between *hU6 and hU2 snRNAs and the complex between *hU2 and hU6 snRNAs following reaction with RNase V1 were mapped onto the two possible folds (A: four-helix model; B: three-helix model) of the U2-U6 snRNA complex. Cleavage by RNase V1 at the corresponding nucleotides is represented by squares with open, gray, and black squares corresponding to low, medium and high intensity.
Figure 2.6 A & B. Normalized cleavage intensities of complex between \( ^*hU6 \) and \( hU2 \) snRNAs and the complex between \( ^*hU2 \) and \( hU6 \) snRNAs following reaction with RNases \( A \) and \( T1 \) were mapped onto the two folds. Cleavage by RNase \( A \) and RNase \( T1 \) at corresponding nucleotides is represented by open, gray, and black circles or triangles, respectively, corresponding to low, medium and high cleavage by the ribonuclease. Nucleotides in grey represent those for which information was not collected because of gel artifacts.
2.3.2. Effect of divalent metal ions on secondary structure:

To investigate whether high Mg\(^{2+}\) concentrations affect the secondary structure of the human U2-U6 snRNA complex, we repeated the experiments described above in the presence of 100 mM MgCl\(_2\) under otherwise identical buffer conditions. This concentration of Mg\(^{2+}\) is substantially higher than cellular concentrations. However, such high amounts of Mg\(^{2+}\) ion concentration have been used to study the equilibrium folding of different RNA molecules \textit{in vitro} (Schlatterer and Brenowitz, 2009; Schlatterer et al., 2008) and the dynamics of conformational change in the yeast U2-U6 snRNA complex (Guo et al., 2009). The added Mg\(^{2+}\) ions screen the electrostatic repulsion and may therefore generate an ensemble of secondary structure components that can interact with each other (Takamoto et al., 2004b), thereby facilitating tertiary contact formations that may otherwise only be favored in the presence of spliceosomal proteins.

After performing the cleavage reactions, we separated the products by denaturing PAGE. We analyzed the gel image with the SAFA software, and normalized the data relative to the total radioactivity per lane (Schlatterer and Brenowitz, 2009; Takamoto et al., 2004a). The relative cleavage intensities were plotted against the sequence of the radioactively labeled RNA strand as before. Results showed that the overall cleavage profiles of the hU2+hU6 complex were consistent with the formation of the four-helix conformation, as in the case without added Mg\(^{2+}\). However, we observed several changes in the presence of Mg\(^{2+}\) ions.

We first used RNase V1, which cuts after double stranded and stacked nucleotides. Cleavage profiles of RNase V1 on the *hU6+hU2 complex indicated an increase in cleavage intensity for the nucleotides 78-80 from medium to high intensity. These nucleotides correspond to the bottom part of the U6 ISL on 3' side. The increased cleavage intensity is consistent with
increased thermal stability of the U6 ISL in the presence of high concentrations of Mg$^{2+}$. On the other hand, we observed a decrease in cleavage intensity of nucleotides 45 to 51 of *hu6 from medium to low intensity (Figure 1B-Supplementaty information). These nucleotides correspond to the ACGAGA loop region, anticipated to be single stranded in both the folds. The decreased intensity of these nucleotides may be the result of: 1) decreased stacking of these single-stranded nucleotides; or 2) decreased accessibility of these nucleotides with increased Mg$^{2+}$ ion concentration. However, we did not observe the decreased accessibility with other enzymes since the other enzymes are smaller in size than RNase V1 (Ehresmann et al., 1987).

The cleavage patterns achieved with RNase T1, which cleaves after single stranded Gs, were very similar to the pattern achieved in the absence of added Mg$^{2+}$. However, the cleavage profile of *hU6 with RNase A, which cleaves after single stranded C and U, indicated decreased cleavage intensity of essentially all the nucleotides that exhibited medium or high intensity cleavage in the absence of Mg$^{2+}$, e.g. the nucleotides 52, 55, 66, 68, 81 (Figure-1D-supplementary information). While the reason for this change is not clear, it may be related to an increase in background cleavage of other nucleotides as a result of contamination, which would decrease the normalized cleavage intensities of the nucleotides described above. Overall the cleavage profiles of the U2-U6 snRNA complex in the presence of Mg$^{2+}$ ions are still in agreement with formation of four-helix conformation.

Enzymatic probing is a qualitative technique, which means that only relative cleavage intensity was assayed in these studies; therefore it is not possible to exclude the presence of a minor fraction of the alternative conformation. So a separate set of studies to characterize the distribution of the conformations were performed by a colleague Caijie Zhao from our laboratory. She has applied another technique $^{19}$F NMR, to determine whether a minor conformational
fraction was present and if so, determine whether the two alternative secondary structures are in
equilibrium with each other. She chose the nucleotide C13 of the U2 snRNA strand for the single
5-^{19}F-C residue because it would be expected to reside in a double-stranded region (Stem I) in a
four-helix model (Fig. 2.1 A) or in a single-stranded region in a three-helix model (Fig. 2.1 B).
By comparison with separate measurements of 5-^{19}F- C in control fragments, she observed a
large (approx. 86% of total), sharp peak at -167.8 ppm, suggesting that the 5-^{19}F-C resides
predominantly in a double-stranded environment, and consistent with formation of Stem I and
the four-helix structure (Fig. 2.1 A). A smaller (~14%), broad composite peak with a center in
the chemical shift range for unpaired 5-^{19}F-C (-165ppm), suggesting that the 5-^{19}F-C resides
predominantly in a single-stranded environment was tentatively attributed to several three-helix
conformer(s). Therefore, these results reinforce our enzymatic structure probing results and also
suggest that the two proposed models in the protein free human U2-U6 snRNA complex are in a
dynamic equilibrium. Addition of 5 mM Mg^{2+} shifted the relative ratio from 86:14% to 83:17%,
suggesting a slight tendency to stabilize the three-helix form (Zhao et al., 2013).

In order to characterize the overall (both secondary and tertiary) structural features of the
major conformation i.e. the four-helix conformation we designed a mutant to favor the formation
of Stem-I. We changed the hairpin loop sequence of the putative Stem I of U2 snRNA from
UUUU to the hyperstable UUCG sequence, which is likely to enhance the thermal stability of
Stem I, and the top base pair of Stem I from G-C to C-G, which we predicted would disfavor
formation of Helix Ib (a key feature of the three-helix model), both of which would be expected
to favor the four-helix structure (Figure-2.2)

Enzymatic probing experiments were performed on the mutant complex (see materials
and methods) to probe the overall secondary structure of the complex. Cleavage patterns
following incubation of the mutant with the UUCG-substituted Stem I with RNase V1 were very similar to those measured for the wild-type U2-U6 snRNA sequence. However, based upon visual comparison of the intensity of cleavage patterns, those regions associated with marked cleavage of the wild-type complex demonstrated a small increase in intensity in the mutant sequence for *hU2 (nucleotides 9–14 and 26–28) compared to regions expected to be common to both models. Cleavage patterns of the mutant sequence by RNase A and RNase T1 were also very similar to those of the wild type, although with a relatively small decrease in the cleavage intensity of the nucleotides 13–19 in the mutant with the UUCG Stem I with respect to invariant regions (Figure-2-Supplementary information). These data suggest that the mutations designed to favor formation of Stem I and the four-stemmed structure do not result in global conformational change, but that nucleotides in the region of U2 Stem I are less likely to be single-stranded in the mutant sequences, i.e., that formation of a four-stemmed conformation was, indeed, favored.

Caijie Zhao a colleague in my laboratory has performed $^{19}$F-NMR experiments on the mutant complex to characterize the distribution of the complex. For this mutated complex, she observed no peak at the single strand chemical shift. These findings suggest that the mutant complex forms essentially ~100% stem-I and four-helix conformation.

Taken together, our data from enzymatic probing suggest that the U2-U6 snRNA complex adopts the four-way junction structure as the dominant species in solution. The enzymatic probing results in presence of up to 100mM Mg$^{2+}$ ions suggest that, the divalent metal ions do not show any significant effect on the secondary structure of the complex. Our data from $^{19}$F NMR studies suggest that U2-U6 snRNA complex is in a dynamic equilibrium with four-way junction structure being a predominant conformation (86.3%) and three-way junction structure as a minor conformation (13.7%). In the presence of Mg$^{2+}$ the percentage of the three-way junction
structure increased from 13.7% to 16.6%, suggesting a minor shift in the equilibrium in presence of divalent metal ions. The mutant designed to favor the formation of Stem I and four-helix conformation did not result in global conformation, but that nucleotides in the region of Stem I are less likely to be single stranded in mutant sequence, i.e., that formation of a four-helix conformation was indeed favored.

2.4. Discussion:

Enzymatic structure probing results are more consistent with the formation of U2 snRNA Stem I in the predominant lowest energy conformation of the human U2-U6 snRNA complex, consistent with formation of a four-helix junction by the protein-free RNA complex. This conformation agrees with the functional model proposed by Sun and Manley (1995) based on results of genetic experiments in human cells and is analogous to the conclusion reached by Sashital et al. (Sashital et al., 2004b) from NMR studies of protein-free yeast RNA fragments. The enzymatic structure probing results are contrary to another secondary structural model proposed for the U2-U6 snRNA complex that is a three way junction model. Features of this model have been proposed for a protein-free yeast complex (Burke et al., 2012) and during the splicing process of yeast in situ (Hilliker and Staley, 2004; Madhani and Guthrie, 1992; Mefford and Staley, 2009).

There is a growing body of evidence suggesting that the spliceosome exists in at least two distinct conformational states during the course of the splicing activity. Opposite stereochemistry and different substrates associated with the two splicing reactions suggests the possibility of different catalytic centers for each of the two cleavage steps (Moore and Sharp, 1993) which in turn may require conformational rearrangement within the U2-U6 snRNA complex. Accordingly,
Query and Konarska proposed a “two-state” model for the two steps of splicing in which dual conformations are in equilibrium (Liu et al., 2007a; Query and Konarska, 2004a).

In support of the role of different sequences asserting an effect on the relative energetics of folding, computational studies by Cao and Chen (2006) indicated a propensity for conformational heterogeneity of protein-free U2-U6 snRNA complexes from both yeast and human sequences with a truncated Helix I/III, with a different balance observed for the different sequences. In each case, secondary structure formation depended on coaxial stacking of stems; as a result of the small sequence differences in the region of the junctions; different distributions were predicted for the two sequences. For yeast, they found that folds approximating the three-helix (containing Helix 1b) and four-helix junction models were most prevalent. Although some of the anticipated folds depended upon the heavily truncated stem regions in their calculations, the appearance of multiple low-energy folds may reflect the potential for rearrangement of the junction region during splicing activity. Investigating the ground state conformation of cell-free U2-U6 snRNA complexes, in which coaxial stacking patterns of RNA helices are not modulated by proteins, provides an advantage in characterization of conformational change associated with metal ions and spliceosomal proteins.

Structural rearrangement of individual protein-free yeast U2-U6 snRNA complexes into at least three distinct states in the presence of Mg$^{2+}$ was shown by fluorescence resonance energy transfer experiments (Guo et al., 2009), suggesting the propensity for ion-dependent conformational change. Specifically, two major conformations attributed to four-helix and three-helix models, respectively, with an obligatory intermediate. In the absence of Mg$^{2+}$, these authors observed a very predominant fraction that they attributed to the four-helix conformation, with no representation of the low-FRET state they attribute to the three-helix conformation.
Sequence differences in the human U2-U6 snRNA complex were associated with a different distribution of folds from that found in yeast by computational studies, with 85% predicted to form a four-helix junction and approximately 5% each in three alternative structures; however, each of the alternative folds would be favored only in the presence of the highly truncated Helix I/III in the human U2 and U6 snRNA sequences tested (Cao and Chen, 2006).

Using U2 and U6 snRNA sequences including more complete Helix I and III stems, we employed enzymatic structure probing technique to probe the conformation of the complex in solution. Enzymatic structure mapping, a commonly used technique for identification of single- and double-stranded regions of RNA molecules (Ehresmann et al., 1987), was most consistent with formation of a four-helix junction. However, these data are averages that cannot reliably detect minor populations of alternative conformations and are not absolute in their specificity for structure. For example, RNase V1 cleaves not only after base paired nucleotides but also stacked single stranded regions, and enzymes specific for single strand regions may not have access to certain nucleotides (Lowman and Draper, 1986). Thus these data do not exclude the possibility of another minor coexisting conformation.

In order to study the possibility of existence of a minor conformation a separate set of 19F-NMR studies were performed by Caijie Zhao (from Greenbaum’s lab). She exploited different chemical shifts of 5-19F-incorporated pyrimidine nucleotide, in either a single- or double-stranded region of RNA, which has previously been used to probe the secondary structure of RNA (Arnold and Fisher, 2000; Chu et al., 1992; Gollnick et al., 1986; Hammann et al., 2001; Hennig et al., 2007; Horowitz et al., 1977; Kanyo et al., 1996; Marshall and Smith, 1977; Olejniczak et al., 2002; Puffer et al., 2009; Sahasrabudhe and Gmeiner, 1997) to demonstrate the presence of an alternative conformation of U2-U6 snRNA complex conclusively. The advantage
of observing the $^{19}$F nucleus by NMR is the broader range and the greater sensitivity of fluorine chemical shifts in response to the local environment as compared with those of hydrogen because the fluorine nucleus is surrounded by nine electrons in molecules vs. a single electron in hydrogen. We specifically targeted C13 of U2 snRNA, which resides within the helix of Stem I in the four-helix junction model but would be single-stranded otherwise. Although this is also an ensemble approach, it allows us to quantify distribution between populations for a specific state.

Identification of a very dominant resonance peak at the chemical shift value corresponding to 5-$^{19}$F-cytidine in a double-stranded Stem I reinforces results of the enzymatic probing experiments, specifically inclusion of U2 snRNA Stem I in the lowest energy state conformation. However, the presence of the lesser peak corresponding to the single-stranded 5-$^{19}$F-cytidine suggests that the two proposed conformations are in dynamic equilibrium. Two observations are consistent with the conclusion that the two peaks reflect alternative conformations in a dynamic equilibrium on a relatively slow time scale: 1) the chemical shift difference between the dominant peak and the lesser peak is essentially the same as what we observed from the short oligomer controls, which indicates slow chemical exchange between conformations; and 2) the fraction of $^{19}$F peak representing the single-stranded environment is less at higher temperatures; both of these findings are contrary to what would be expected if the unpaired fraction were due only to thermal melting of Stem I. This result therefore suggests a concerted conformational change, perhaps to form Helix 1b. A study of the dynamics of chemical exchange between the two conformations will be reported elsewhere. Taken together our results show that the protein-free U2-U6 snRNA complex adopts a four-way junction structure as a predominant fold with a small fraction of three-way junction structure.
Several lines of investigation support conformational rearrangement between the two steps of splicing experimentally. For example, Tseng and Cheng (2008) demonstrated that both catalytic steps of splicing are reversible, which suggests the possibility of conformational rearrangement (Tseng and Cheng, 2008). Also, Prp8 assists substrate repositioning by altering the equilibrium between the two steps (Liu et al., 2007a; Query and Konarska, 2004a). Prp16p-dependent opening and closing of Helix I was demonstrated (Mefford and Staley 2009). In agreement with their conclusions, results of site-directed hydroxyl radical cleavage have shown an alteration in the spatial relationship between U6 snRNA ISL and the ACAGAGA loop between the two steps of splicing (Rhode et al., 2006b).

*In vitro* experiments also provide evidence for conformational rearrangement upon addition of Mg$^{2+}$. Butcher and coworkers observed a ~9% decrease in the radius of gyration at 2 mM Mg$^{2+}$ for the yeast complex assayed by small angle x-ray scattering (Burke et al., 2012). However, a notably greater change was reported from single molecule FRET studies (Guo et al., 2009), in which a large shift in the fraction of the three-helix conformation from the four-helix junction conformation was observed in the presence of 10 mM Mg$^{2+}$. We also investigated whether Mg$^{2+}$ induced conformational change in the human sequence as was found by Guo et al. (2009) in the yeast sequence. $^{19}$F NMR experiments indicated a small increase in the three-helix structure, and sedimentation velocity data suggested a small amount of compaction of the three-dimensional structure in the presence of Mg$^{2+}$; these changes may occur in the protein-free system as a result of altered patterns of coaxial stacking of stems upon interaction with the metal ion.

As noted by Cao and Chen (2006), it is possible that sequence differences may have some impact on the energetics of the two complexes. For example, Stem I of yeast has six base pairs,
compared with three in human U2 snRNA; there is a G·A pair in the human U6 snRNA ISL (human), compared with Watson-Crick pairing in the lower stem of the yeast ISL. Also, the catalytic AGC triad has different base pairing partners for both intra- and intermolecular pairing in yeast and human, which could alter metal ion binding properties. Although the identical chemical mechanism of splicing in the two systems was suggested (Moore and Sharp, 1993), it is not quite clear how much of the sequence difference is involved with the difference of conformational distribution and response to Mg$^{2+}$ between the two systems. Whether the difference in conformational distribution predicted between yeast and human U2-U6 snRNA complexes by computational methods extend to differential response to Mg$^{2+}$ and/or in cellular systems is not yet clear.

In summary, our data show that the U2-U6 snRNA complex in its ground state forms predominantly four-way junction structure (86%), with a minor conformation (~14%) as a three-way junction structure. Divalent metal ions (Mg$^{2+}$) have very little effect on secondary structure of the U2-U6 snRNA complex decreasing the percentage of major conformation in the distribution to 83%.
CHAPTER 3

TERTIARY STRUCTURAL FEATURES OF THE HUMAN U2-U6 snRNA COMPLEX.

3.1 Introduction:

Major structural changes are proposed to occur in the spliceosome during its catalytic activation, culminating in the pre-mRNA splicing reaction (Brow, 2002). The changes in snRNA conformation have been studied at the secondary level of RNA–RNA interactions in both human and yeast spliceosomes (Guo et al., 2009; Hilliker and Staley, 2004; Madhani and Guthrie, 1992; Mefford and Staley, 2009; Sashital et al., 2004a; Sun and Manley, 1995a; Zhao et al., 2013). Functional and structural studies suggest two different conformational models of the U2-U6 snRNA complex for yeast and human complexes, respectively. Genetic studies in yeast support the importance of Helix Ib formation, in which the AGC triad forms three intermolecular base pairs with U2 snRNA, and the U2 Stem I opens to form a three-helix structure (Hilliker et al., 2007; Hilliker and Staley, 2004; Madhani and Guthrie, 1992; Mefford and Staley, 2009)(Figure 2.1.). NMR studies on the yeast U2-U6 snRNA complex supported formation of Stem I and therefore the four-helix conformation in the protein-free state (Sashital et al., 2004a). However, a subsequent and comprehensive set of structural studies on the yeast complex from the same group using solution NMR, small angle x-ray scattering (SAXS), and computational modeling define a three-helix conformation with no evidence of an alternative fold (i.e. the three-helix fold) suggested by genetic studies (Burke et al., 2012). Single molecule studies of a truncated yeast U2-U6 snRNA complex provided evidence for Mg$^{2+}$ mediated interconversion between two folds attributed to four-helix and three-helix conformations.
In contrast, genetic studies on the U2-U6 snRNA complex in human cells implicate a four-helix conformation (Sun and Manley, 1995a). In the previous chapter, enzymatic structure probing and $^1$H and $^{19}$F NMR techniques were used to characterize the secondary structural fold of protein-free human U2-U6 snRNA complex. Data were consistent with formation of the four-helix junction structure as a predominant conformation (Sun and Manley, 1995a), with $^{19}$F NMR data also identifying a fraction (~14% at 25 °C) of a three-helix conformation (Madhani and Guthrie, 1992) (See chapter 2 for more details). Each of this secondary structure may be associated with an alternative three-dimensional structure in the region of junction. The next step is to investigate the tertiary conformational features of the human U2-U6 snRNA complex.

The crystal structure of a group II intron has revealed that the regions analogous to the AGC triad, ACAGAGA loop and U74 in human U2-U6 snRNA complex converge to form the active site. Thus, it is conceivable that these regions may be juxtaposed in the 3D structure of U2-U6 snRNA complex of the activated spliceosome to form an active site.

Based upon their NMR, SAXS and molecular modeling studies, Butcher and coworkers have proposed a model for the three-dimensional fold of the yeast U2-U6 snRNA complex (Burke et al., 2012) that adopts a three-helix junction structure that forms an extended “Y” shape. From their model, the authors noted that the essential features of the complex, including the U80 metal ion-binding site in the U6 ISL, AGC triad, and pre-mRNA recognition sites are located on one face of the complex. They concluded that this fold would orient substrate and other cofactors during the splicing reaction. Interestingly, the authors did not find any evidence of an alternative four-helix conformation as observed by their earlier studies (Sashital et al., 2004a) or in our studies of the human U2-U6 snRNA complex (Zhao et al., 2013). Additionally, in contrast with
results of single-molecule FRET studies by Rueda and coworkers they did not observe any significant change upon addition of Mg$^{2+}$ (Guo et al., 2009).

All evidence suggests that the chemical mechanism of splicing is identical in yeast and human systems (Moore and Sharp, 1993). It is intriguing to note that the yeast and human complexes appear to make use of a different structural context to catalyze identical mechanism. Therefore it is important to characterize the tertiary structural features of the human U2-U6 snRNA complex. The goal of this chapter is to investigate the tertiary conformational features of the human U2-U6 snRNA complex and analyze the effect of Mg$^{2+}$ ions on the global structure. This may help us build models and analyze if the metal ion binding sites are brought into close proximity for catalytic activity. Toward this goal, we have used a combination of analytical ultracentrifugation (AUC), hydroxyl radical footprinting and small angle x-ray scattering (SAXS) to study the overall three-dimensional structural features of the human U2-U6 snRNA complex.

Preliminary data acquired from SAXS measurements suggest the likelihood that the two observed secondary conformations are associated with different three-dimensional profiles; results from AUC are consistent with small amount of compaction upon addition of MgCl$_2$ and OH· footprinting results support protection of certain regions of the complex consistent with altered arrangement of helices and tertiary structural fold that may be more compact upon addition of MgCl$_2$. We speculate that these latter changes may correlate with 1) Mg$^{2+}$-dependent partial shift from the four-helix to the three-helix structure. 2) A direct compaction of the tertiary fold of the four-helix conformation.

3.2. Materials and methods:
3.2.1 Analytical ultracentrifugation (AUC)

AUC has been used extensively to study global conformational changes of RNA molecules (Costantino and Kieft, 2005; Takamoto et al., 2002). It uses relatively small amounts of RNA and reports precisely, based on fundamental hydrodynamic principles, the size and shape of RNA molecules in solution. There is no necessity of labeling the RNA with external radioactive or fluorescent label as we use optical absorption properties of RNA molecules in the UV range (~ 260 nm).

When a homogenous solution of RNA molecules in an enclosed compartment is subjected to a centrifugal force, the solute particles move towards the bottom of the cell, thus establishing a solvent-solution boundary. The diffusion forces acting on the molecules tend to redistribute the concentration gradient formed at the boundary in different directions. Here in we have used sedimentation velocity experiments, in which the samples are centrifuged at high velocities so that the centrifugal force dominates the diffusion forces resulting in the movement of bulk solute to the bottom of the compartment. Monitoring the rate of movement of the solvent-solution boundary can help us analyze the size, shape, and conformational homogeneity of the macromolecular solute under a given set of solution conditions.

Equimolar amounts of U2 and U6 snRNA strands were resuspended in a buffer containing 50 mM Tris-HCl, 100 mM NaCl at pH 7.5, to a final absorbance value ~1 OD, heated to 70 °C, and cooled for 15-30 min. AUC experiments and their analysis were performed by Prof. Michael Brenowitz at AECOM on samples transcribed, purified, and previously analyzed for pairing by me. The sequences used for the native U2 and U6 snRNAs are identical to those used in Chapter 2. MgCl₂ was added to a final concentration of 5 mM or 100 mM at least 30 min before loading the samples into centrifuge cells. Sedimentation velocity experiments were
performed using a Beckman XL-I analytical ultracentrifuge at 20 °C in double-sector cells loaded into a Ti-60 rotor and centrifuged at 30,000 rpm as described (Mitra, 2009). Sedimentation boundaries were fit using DCDT+ (Philo, 2000, 2006) to determine the sedimentation and diffusion coefficients; resulting values were normalized to standard conditions using the buffer density calculated using SEDNTERP (http://www.jphilo.mailway.com/download.htm) with hydration and partial specific volume values of 0.59 and 0.53 cm³/g, respectively (Mitra, 2009). The Stokes radius (R_H) and axial ratio (a/b) were calculated using SEDNTERP as described (Mitra, 2009).

3.2.2. Hydroxyl Radical Footprinting:

Individual U2 and U6 snRNA fragments were dephosphorylated at the 5’ terminus using Antarctic phosphatase (New England Biolabs) and labeled with γ-[^32]P-ATP using T4 polynucleotide kinase (New England Biolabs). Labeled RNA was purified on a 12% denaturing gel, eluted by a crush-and-soak method, precipitated by ethanol, dried, and resuspended in 10 mM sodium cacodylate, 0.1 mM EDTA, and 100 mM NaCl, pH 7.4. The components of some buffers quench OH radicals, so it is important to choose the buffers carefully.

Each labeled strand was paired with a 1.5x excess of unlabeled partner, heated to 90 °C for 2 min, and cooled for 15-30 min at room temperature. To assess pairing, an aliquot of paired samples (with nanomolar concentrations of either labeled U2 or U6 strand) was electrophoresed on a nondenaturing gel against labeled individual strands. From quantification of paired vs. unpaired strands, we observed that effectively 100% of each labeled strand was in a paired form. We have repeated this experiment with a constant amount of labeled strand with an increase in concentrations of the unlabeled counter strand; we observed essentially complete pairing at all
the concentrations. This suggests that, under the experimental conditions that used in these experiments, there was ~100% complex formation.

Reaction aliquots of U2-U6 RNA complex in a buffered solution were prepared so that each reaction tube contains enough radioactively labeled RNA for visualization and quantitation of the bands of the reaction products (typically 30,000 cpm/ul of RNA solution). To investigate the divalent metal ion (Mg$^{2+}$) mediated equilibrium folding of RNA complex, 100 mM MgCl$_2$ is added to the complex in otherwise identical buffer conditions. The solutions were equilibrated at room temperature for 1 hour. The cleavage reaction was initiated by addition of freshly prepared Fenton reaction mix, which contains (Fe(NH$_4$)$_2$(SO$_4$)$_2$, EDTA and sodium L-ascorbate to a final concentration of 0.1 mM, 0.14 mM, and 6.94 mM, respectively). Hydroxyl radicals were generated by following Fenton reaction:

$$\text{Fe}(II) - \text{EDTA} + H_2O_2 \rightarrow \text{Fe}(III) - \text{EDTA} + \bullet OH + OH^-$$

It is important to identify the optimal concentration of OH radical needed for the cleavage reaction to occur under single hit kinetics. The optimal OH radical cleavage condition was found by a “dose response” experiment. For this experiment, the RNA sample was prepared as described above and aliquoted into several tubes containing the reaction mix in question. Then Fenton reaction was performed such that the concentration of any one of the Fenton reagents increases from one test tube to another. Hydroxyl radical footprinting assays were performed identifying conditions that resulted in less than one “hit” per labeled molecule.

A negative control was also included in which the sample was not treated with the Fenton reaction mix in otherwise identical conditions was resuspended in the denaturing gel loading dye.
For each experiment, reference guanosine ladders were prepared by incubation with RNase T1 (Fermentas, Inc.) under semidenaturing conditions. The samples were electrophoresed on a denaturing gel against a T1 digestion ladder. The gel was exposed to a phosphor screen for 12–16 h at 4 °C and scanned by a phosphorimager (STORM scanner from GE Health Care). The resulting gel image was analyzed with SAFA (Laederach et al. 2008), which calculates the density of each band. The band densities were normalized, and the resulting values were used to generate a thermal plot, which represents an electronic version of the gel (Figure 3.1). For a more detailed protocol please refer to the following video article (Bachu et al., 2011).

3.2.3. SAXS:

In order to exclude the excess of an unpaired strand and to obtain a sample containing essentially 100% paired U2-U6 complex, equimolar amounts of U2 and U6 snRNA strands were resuspended in a buffer containing 50 mM Tris-HCl, 100 mM NaCl at pH 7.5, heated to 70 °C, and cooled for 15-30 min. The paired complex was loaded onto a 12 % non-denaturing preparative gel and electrophoresed at 4 °C (cold room). The paired complex was visualized with UV radiation and excised from the gel. The complex was eluted out of the gel by electroelution at 4 °C inside the cold room. The buffer was replaced to 20 mM MOPS, 100 mM NaCl, 100 mM EDTA with a pH of 7.1 using Amicon ultra filters and stored at 4°C prior to the SAXS experiments. To characterize tertiary structural features of the lower-abundance alternate conformation, we designed a mutant with similar length shown to form ~100% four-helix structure by favoring the formation of Stem I. The mutant includes substitution of the UUUU tetra-loop sequence of Stem I to the hyperstable sequence UUCG, and mutation of the top pair of Stem I from G-C to C-G (which would inhibit the formation of Helix Ib). These mutations favor the formation of Stem I and the four-helix structure, with out changing the overall electron
Figure 3.1. Schematic representation of a hydroxyl radical footprinting experiment. A) Dephosphorylation and 5'-end labeling of RNA with $^{32}$P. B) Purification of $^{32}$P-labeled RNA on a denaturing polyacrylamide gel. C) Excision of RNA band, subsequent RNA extraction, and ethanol precipitation. D) Prefolding and folding of RNA. E) Addition of freshly prepared Fenton reaction mixture to generate the hydroxyl radicals. F) RNA fragment separation by denaturing polyacrylamide gel electrophoresis. G) Quantitation of RNA fragments by SAFA software.
density (this is the same mutant that was used in chapter -2). The above-described procedure was repeated with U6 and mutant U2 RNA complex, in order to obtain 100% paired complex. The samples were moved to room temperature and millimolar amounts of MgCl$_2$ in the same buffer were added at least 20 minutes prior to the SAXS measurements. The SAXS data were collected at the G1 Station of the Cornell High Energy Synchrotron Source (CHESS) by Dr. Suzette Pabit and Prof. Lois Pollack of Cornell University. Measurements were done in 0, 2, 5, and 10 mM MgCl$_2$ at varying RNA concentrations to check for structure factor contributions to the SAXS profiles. We used 0.5 and 1 mg/mL for all MgCl$_2$ concentrations and also measured 2 and 2.8 mg/mL for samples in 2 and 5 mM MgCl$_2$.

For SAXS measurements, the RNA solutions were loaded into 1-mm quartz capillary (Hampton Research, Aliso Viejo, CA) and oscillated back and forth to reduce radiation damage due to full beam exposure. Data from matching buffer solutions (with the same concentrations of MgCl$_2$ as the samples) were taken before and after each RNA sample for background subtraction. For each sample and buffer solution, at least 8 scattering images were taken with 30-second exposures using a low-noise area detector (Pilatus 100 K, Dectris, Baden, Switzerland). A pin diode beam stop was used to measure the transmitted intensity for normalization. The flight path from the capillary to the detector, a distance of 1.6 m, was kept under vacuum. The images were integrated for conversion into scattering intensity as a function of $q$ where $q = (4\pi/l)\sin(2q)/2$, 2$q$ is the scattering angle and $l$ is the x-ray wavelength. For this experiment, the $q$ range was $0.015 < q < 0.262$ Å$^{-1}$. The scattering intensity profiles were obtained by subtracting the background intensity (average scattering profile from the buffer solutions taken immediately before and after the samples) from the sample intensity. Drs. Pabit and Pollack used the software BioXTAS RAW (http://sourceforge.net/projects/bioxtasraw/) and MATLAB for processing the SAXS
data. *Ab initio* structure calculations for these samples were done using DAMMIF (Franke and Svergun, 2009), DAMAVER (Volkov and Svergun, 2003), and SUPCOMB (Kozin and Svergun, 2001) developed by the Svergun group.

### 3.3. Results:

In order to make an attempt to build a 3D model of human U2-U6 snRNA complex and the two conformations it was shown to form, we used a combination of AUC, Hydroxyl radical footprinting and SAXS on the complexes formed by U2 and U6 snRNA fragments representing the functionally important sequences and the mutant sequence favoring the four-helix fold. Strand pairing to form a single complex with a stoichiometry of 1:1 was confirmed by nondenaturing gel electrophoresis under the conditions similar to those used for the different experiments.

#### 3.3.1. Analytical Ultracentrifugation:

To detect changes in the global conformation of the human U2-U6 snRNA complex as a function of Mg$^{2+}$ concentration, sedimentation velocity measurements were obtained using AUC. U2 and U6 snRNA fragments were annealed and folded in buffer containing 0, 5, and 100 mM Mg$^{2+}$, and the sedimentation and diffusion coefficients ($S_{20, w}$ and $D_{20, w}$) were determined (See Materials and Methods for details). Each sample is characterized by a single symmetric peak in the $g$ ($s^*$) versus $s^*$ distributions, consistent with formation of unique or almost identical global conformation (Figure 3.2). Our previous 19F-NMR studies demonstrated the formation of two secondary structural folds. Based on this observation, we would have anticipated a major and minor peak.
Figure 3.2. The figure shows $g(s^*)$ versus $s^*$ distribution curve for the human U2-U6 snRNA complex with increasing concentrations of Mg$^{2+}$. 
Interestingly, we did not observe any evidence of an alternate peak or a small shoulder to the major peak. These results suggest that the two conformations have similar sedimentation properties. (See discussion for more details). The Stokes radius ($R_H$) and the axial ratio (a/b) for human U2-U6 snRNA complex in the absence of Mg$^{2+}$ were calculated to be 31.3 Å and 8.2 (Table 3.1).

<table>
<thead>
<tr>
<th>$\text{Mg}^{2+}$ (mM)</th>
<th>$S_{20,w}$ (S)</th>
<th>$D_{20,w}$ (F)</th>
<th>$R_H$ (Å)</th>
<th>a/b</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>4.452 ± 0.003</td>
<td>6.26 ± 0.03</td>
<td>31.3</td>
<td>8.2</td>
</tr>
<tr>
<td>5</td>
<td>4.532 ± 0.005</td>
<td>6.49 ± 0.03</td>
<td>30.7</td>
<td>7.7</td>
</tr>
<tr>
<td>100</td>
<td>4.996 ± 0.005</td>
<td>7.29 ± 0.03</td>
<td>27.9</td>
<td>5.6</td>
</tr>
</tbody>
</table>

Table 3.1. Results of analytical ultracentrifugation measurements of the human U2-U6 snRNA complex. The table shows the standard sedimentation ($S_{20,w}$) and diffusion constant ($D_{20,w}$) of human U2-U6 snRNA complex at 0, 5 and 100 mM Mg$^{2+}$. The Stokes Radius ($R_H$) and axial ratio (a/b) calculated using software SEDNTERP are also shown.

These values decreased to 30.7 Å/7.7 and 27.9 Å/5.5 in the presence of 5 and 100mM Mg$^{2+}$, respectively, suggesting that Mg$^{2+}$ induces a slightly more compact and less asymmetric tertiary structure. In addition to the shift that Mg$^{2+}$ ions create from four-helix to three-helix conformation, they may bring about a tertiary change that affects the orientation of the stems not just the structure of the junction. This is consistent with the observed decrease of Stokes radius and axial ratio of human U2-U6 snRNA complex.

3.3.2. Hydroxyl Radical footprinting:

Hydroxyl radicals ($\cdot$OH) act as unique probes of the structure of nucleic acids due to their
high reactivity, small size (Tullius and Dombroski, 1986) and non-specific cleavage of nucleic acid backbone. When used as a footprinting probe, •OH map the solvent accessible surface of the phosphodiester backbone of DNA (Tullius and Dombroski, 1986) and RNA (Celander and Cech, 1990) with as fine as single nucleotide resolution. Here in, we have used this technique to characterize the tertiary structural features of U2-U6 snRNA complex at a single nucleotide resolution. RNA molecules are devoid of any tertiary interactions in conditions of low ionic strength or in the absence of divalent metal ions. We monitored the solvent accessibility of RNA molecules as a function of increasing concentration of Mg$^{2+}$, which stabilizes the RNA tertiary interactions. The U2-U6 snRNA complex was folded in separate tubes with increasing concentrations of Mg$^{2+}$ ions (starting from 0mM to 100mM) under otherwise identical buffer conditions, and samples were subjected to cleavage by hydroxyl radicals generated by Fenton chemistry (see Materials and Methods for details). The cleavage products from the hydroxyl radical footprinting assay were resolved on a sequencing gel (Figure 3.3), quantified and analyzed using the software SAFA (Das et al., 2005). Thermal/Color plot was generated using SAFA, which represents an electronic version of the gel (Figure 3.4). Each pixel in the thermal plot represents a band on the gel.

The cleavage profiles were generated by plotting normalized cleavage intensity against the nucleotide sequence. In the absence of Mg$^{2+}$, we observed that the cleavage intensities of hU2+hU6 complex were in the range of 0.012 to 0.045, suggesting that some nucleotides are more protected than others. This is a small RNA complex and we would not expect it to have a stable tertiary structure, i.e. areas that are deeply buried and have low cleavage intensity. The
Figure 3.3. Representative sequencing gel exhibiting cleavage patterns of $^{32}$P-labeled (*) hU6 paired with unlabeled hU2 strand, subjected to hydroxyl radical cleavage (see Materials and Methods for experimental details). Lanes are labeled on top with the controls and the gradient of Mg$^{2+}$. 
Figure 3.4. Thermal/color plots showing cleavage intensities of complex between *hU2 and hU6 snRNAs following reaction with hydroxyl radicals with increasing concentration of Mg$^{2+}$ ions.
cleavage profiles observed here are consistent with that notion. Upon addition of Mg$^{2+}$, we noticed small changes in the cleavage intensities of some nucleotides, suggesting that some nucleotides are becoming more exposed or protected.

For *hU2, we observed that the cleavage intensity decreased consistently for the nucleotides 11-14, 19-22 and 36-37 and 39-41, suggesting that they are becoming more protected at increasing Mg$^{2+}$ concentrations. On the other hand, cleavage intensity for nucleotides 15-18 and 33 increased (Figure 3.4), consistent with their becoming increasingly accessible at increasing Mg$^{2+}$ concentrations. It is important to note that the increase/decrease that we observed is 5-15% suggesting, these are small changes that would be consistent with a dynamic system rather than formation of a stable alternative structure. Nucleotides 11-14, 19-22 are close to the junction in Stem 1, these nucleotides are being protected, perhaps as a result of change in the orientation of helices at the junction. However, the nucleotides 36-37 and 39-41 are distant from the junction, in Helix-III. Increased protection of these nucleotides suggests formation of new tertiary contact. On the other hand, nucleotides with increased cleavage intensity, i.e. 15-18, correspond to the loop of Stem 1, and nucleotide 33 corresponds to a nucleotide in the ACAGAGA loop.

In contrast, for the *hU6+hU2 complex, the changes in the cleavage intensities are inconsistent. As shown in Figure 3 (supplementary information), we did not observe consistent increase/decrease in the cleavage intensity (Figure-3-Supplementary information), making it difficult to analyze clear trends. However, several regions suggested a trend toward increased/decreased cleavage for certain ranges of Mg$^{2+}$ concentration. Among the nucleotides showing trends toward increasing protection upon addition of Mg$^{2+}$ were 44-46, 54-59, 66-67, 70-72 and 88-90. Nucleotides 44-46 correspond to the ACAGAGA loop, 66-67 and 70-72 are on
the top part of U6 ISL on the 3’ side and nucleotides 88-90 are close to the junction on helix-II. In contrast, we observed increased cleavage intensity for nucleotides 60-65 and 73-86. Nucleotides 60-65 are on the top half of U6 USL on the 5’ side, nucleotides 73-86 are on the bottom half of the U6 ISL on the 3’ side. We note that for the nucleotides 54-58 and 70-72 the protections are observed at concentrations of Mg$^{2+}$ from 60 to 100 mM, while for other nucleotides the protection was observed only at concentrations less than 60 mM Mg$^{2+}$.

These nucleotides were mapped onto the schematic secondary structural model corresponding to four-helix junction (Figure 3.5). We noticed that the protected and exposed nucleotides on the U6 ISL fall on the two sides of the helix suggesting that one side of the ISL is being protected and other side of the helix is being exposed. Also the increased protection on the helix-III nucleotides suggests that these two helices might come into close proximity. These results are consistent with the hypothesis made by Butcher and his coworkers that the U6 ISL and helix-III of four-helix conformation serve as a scaffold to juxtapose the catalytically important metal binding site in U6 ISL and the 5’splice site (Sashita et al., 2004a). Four-helix junctions that fold into two coaxially stacked helices in anti parallel arrangement are common in nucleic acids, and some of these have been shown to undergo Mg$^{2+}$-dependent conformational change (Hohng et al., 2004; Lilley, 1998). Site-directed hydroxyl radical experiments indicate that the nucleotides upstream to ACAGAGA loop and the residues close to metal binding pocket in the U6 ISL are in close proximity in a folded U2-U6 structure (Rhode et al., 2006a). However, single molecule FRET studies on the yeast U2-U6 snRNA complex suggest that the distance between the U6 ISL and helix-III increases with increasing Mg$^{2+}$ ion concentration (Guo et al., 2009). Overall the •OH foot printing results are consistent with previously proposed model in
Figure 3.5. Normalized cleavage intensities of complex between *hU6 and hU2 snRNAs and the complex between *hU2 and hU6 snRNAs following reaction with hydroxyl radicals were mapped onto the major structural fold i.e. four-helix conformation. The nucleotides highlighted in the red are those whose solvent accessibility increased in the presence of Mg\(^{2+}\) and nucleotides highlighted in blue are those whose solvent accessibility decreased. The *hU6 RNA data is less consistent and therefore it might be associated with some ambiguity.
which U6-ISL and helix-III might come into close proximity to act as scaffolding for the splicing reaction (Sashital et al., 2004a).

3.3.3. SAXS:

**Structure Factor Effects**

In order to understand the relative positions of the helices and the tertiary structural features of U2-U6 snRNA complex, we used SAXS. SAXS has proven to be a valuable biophysical tool to derive information about the global shape of RNA molecules (Lipfert and Doniach, 2007; Lipfert et al., 2008; Schlatterer and Brenowitz, 2009). In an ideal case, SAXS samples are dilute, monodisperse, and display no interparticle interactions (Svergun and Koch, 2003). Increasing the concentration of the SAXS samples improves the strength of the signal to noise. However, high concentrations of biological molecules are prone to interparticle interaction effects that affect the SAXS profiles (e.g. aggregation or dimer formation). In the absence of interparticle interactions, SAXS profiles from samples of different concentrations under the same solution conditions should overlap fully.

In nucleic acids, interparticle interaction effects are further complicated by the salt concentration of the surrounding buffer solution. Since nucleic acids are negatively charged, inadequate salt concentrations in the buffer solution can lead to inefficient nucleic acid charge screening that manifests as interparticle repulsive interactions in the SAXS profiles. For example, a moderate concentration of DNA in low ionic strength solution would exhibit interparticle repulsion peaks in the SAXS profiles, which is indicated by a downturn in the low q region (i.e. at \( q < 0.5 \text{ Å}^{-1} \)) in the SAXS profile (Qiu et al., 2007). On the other hand, higher ionic strengths or the presence of even low quantities of divalent ions in the solution can induce interparticle attraction leading to intermolecular association, a signature of which is an upturn in the low q
region of the SAXS profile (Pabit et al., 2009). Thus, the first step in any SAXS analysis of a nucleic acid system is to check for interparticle effects at all solution conditions used in the measurements. From scattering profiles acquired from samples at different RNA and buffer concentrations, we determined that 2 mg/mL RNA in 2 mM MgCl$_2$ yields samples with minimal inter-particle interactions.

**P(r) analysis:**

Another way to assess the interparticle association and the size of the molecules being investigated is to look at the pair distance distribution function, P(r). In SAXS, the P(r) function is a histogram that describes interatomic distances within the molecule. As the function describes the set of all paired distances within the structure, a small change in the relative position of few atoms can manifest as detectable changes in the P(r) distribution. Thus this is a useful tool to detect conformational changes and interparticle interactions leading to aggregation. The P(r) function is also used to calculate the correct maximum dimension of the molecule ($d_{max}$). When there is a molecule undergoing conformational change, it also changes the mass distribution around its center of gravity, which manifests as change in radius of gyration, $R_g$. Thus we can compare the $R_g$ of a macromolecule from two different conditions and characterize the condition specific conformational changes associated with it.

The results of the analysis as applied to the SAXS profiles with the best signal-to-noise ratios are displayed in Table 3.2 below:
<table>
<thead>
<tr>
<th>Sample Name</th>
<th>Variant</th>
<th>Concentration (mg/mL)</th>
<th>[Mg$^{2+}$] (mM)</th>
<th>$R_g$ (Å)</th>
<th>$d_{max}$ (Å)</th>
<th>$d_{max}$ range (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>an1</td>
<td>Wild type</td>
<td>1</td>
<td>0</td>
<td>35.4 ± 0.7</td>
<td>115.5</td>
<td>[110.8, 123.2]</td>
</tr>
<tr>
<td>al2</td>
<td>Wild type</td>
<td>2</td>
<td>2</td>
<td>33.64 ± 0.12</td>
<td>119.4</td>
<td>[115.9, 123.8]</td>
</tr>
<tr>
<td>al3</td>
<td>Wild type</td>
<td>2.8</td>
<td>2</td>
<td>32.68 ± 0.06</td>
<td>105.4</td>
<td>[102.9, 109.0]</td>
</tr>
<tr>
<td>am2</td>
<td>Wild type</td>
<td>2</td>
<td>5</td>
<td>36.2 ± 0.3</td>
<td>147.5</td>
<td>[143.1, 154.0]</td>
</tr>
<tr>
<td>am3</td>
<td>Wild type</td>
<td>2.8</td>
<td>5</td>
<td>36.1 ± 0.2</td>
<td>148.7</td>
<td>[145.1, 153.5]</td>
</tr>
<tr>
<td>ah1</td>
<td>Wild type</td>
<td>1</td>
<td>10</td>
<td>38.6 ± 0.3</td>
<td>149.2</td>
<td>[145.9, 154.2]</td>
</tr>
<tr>
<td>bn1</td>
<td>Mutant</td>
<td>1</td>
<td>0</td>
<td>35.07 ± 0.16</td>
<td>110.6</td>
<td>[106.8, 117.6]</td>
</tr>
<tr>
<td>bl2</td>
<td>Mutant</td>
<td>2</td>
<td>2</td>
<td>33.8 ± 0.2</td>
<td>136.1</td>
<td>[129.5, 142.3]</td>
</tr>
<tr>
<td>bl3</td>
<td>Mutant</td>
<td>2.8</td>
<td>2</td>
<td>32.44 ± 0.07</td>
<td>105.0</td>
<td>[101.6, 109.9]</td>
</tr>
<tr>
<td>bm2</td>
<td>Mutant</td>
<td>2</td>
<td>5</td>
<td>36.6 ± 0.3</td>
<td>150.8</td>
<td>[146.3, 156.3]</td>
</tr>
<tr>
<td>bml3</td>
<td>Mutant</td>
<td>2.8</td>
<td>5</td>
<td>36.6 ± 0.2</td>
<td>151.0</td>
<td>[146.6, 157.7]</td>
</tr>
<tr>
<td>bh1</td>
<td>Mutant</td>
<td>1</td>
<td>10</td>
<td>39.2 ± 0.3</td>
<td>156.7</td>
<td>[152.4, 161.8]</td>
</tr>
</tbody>
</table>

**Table 3.2.** Results of SAXS experiments of the human U2-U6 snRNA complex and the complex between mutated U2 and U6 snRNA. The table shows the sample name on the first column, which is a short hand representation of the samples, its concentrations and the conditions used. The table also shows Radius of gyration ($R_g$) values at various concentrations of the complex and MgCl$_2$. It also represents maximum possible diameter of the complex $d_{max}$.
Figure 3.6 A, B compare the $P(r)$ profiles for the wild type and mutant complexes. The $P(r)$ follows the same trend in both sets, which suggests that both the complexes have similar global conformation. We have noticed that the RNA in buffer only (BLUE, no MgCl$_2$) exhibited more extended interatomic distances than samples in MgCl$_2$, which suggests that there is interparticle repulsion under these conditions. At 10 mM MgCl$_2$ (CYAN), the SAXS profiles show long tails in the P(r) curves, which is a characteristic of interparticle attraction and signs of aggregation. Samples in 5 mM MgCl$_2$, also show these long tailed curves. At 2 mM MgCl$_2$, we did not notice any extended interatomic distances nor did we notice long tailed curves suggesting that the interparticle repulsions and aggregation were minimal under these conditions. Thus, the combination of structure factor and $P(r)$ analysis suggest that samples in 2 mM MgCl$_2$ should be used to compare wild type and the mutant data (GREEN).

Preliminary SAXS data of the wild type and the mutant human U2-U6 snRNA complex, in conjunction with ab initio modeling, suggest a well-defined, non-globular conformation. Kratky profiles for the wild type and the mutant conformations exhibit multiple peaks showing that all conformations are non-globular. The $P(r)$ shows two peaks; one near 20 Å, likely corresponding to the A-form helical width, and the other near 40 Å, likely corresponding to helical lengths. These values are similar to the values found by Burke et al. for the yeast U2-U6 snRNA complex (Burke et al., 2012).

We have previously shown that the mutant sample is homogenous and forms the four-helix conformation only, while the wild type sample is a heterogeneous mixture containing ~86% four-helix conformation and rest hypothesized to be a three-helix conformation (Zhao et al., 2013). In an attempt to generate the SAXS profile for the minor conformation (presumably
Figure 3.6 A & B. Figure shows $P(r)$ profiles from the wild type (A) and mutant (B) samples. The short hand representation of the samples and conditions is shown on the top right hand side.
the three-helix), we subtracted 86% mutant from the wild type profile in the 2 mg/mL, 2 mM MgCl₂ sample. We were able to do the subtraction because the wild type and mutant data were collected at same concentration and conditions. The resulting normalized SAXS profiles are shown in Figure 3.7 A&B.

The Kratky profile still exhibits multiple peaks showing that the minor conformation is also non-globular. In order to compare the P(r) profiles of the four-helix conformation and the minor conformation (presumably the three-helix one) we have used GNOM software to generate corresponding profiles. The P(r) still shows 2 peaks, one near 20 Å, which most likely corresponds to the A-form helical width. The other peak near 40 Å may correspond to helical lengths. Interestingly the 20-Å peak is more pronounced in the mutant four-helix structure than the alternate three-helix conformation, likely due to extra stability of the helix in the four-helix mutant. Additional to that the P(r) of the three-helix conformation persists at higher values, e.g. between 70-90 Å, which could indicate that the three-helix conformation is more extended. And P(r) of four-helix conformation shows more short-range pair distances, which agree with a four-helix less extended model.

We then reconstructed three-dimensional profiles by DAMMIF in fast mode to calculate 20 dummy atom models for the four-helix mutant and alternative conformer and averaged using DAMAVER (Figure 3.8). The corresponding shape reconstructions were aligned using SUPCOMB and compared to that of the U2-U6 yeast snRNA complex modeled by the Butcher group (Burke et al., 2012). The normalized spatial discrepancy (NSD) in the four-helix and three-helix conformation are 0.722 and 0.824 respectively. These values suggest that the models did not align well. A helix is missing in the reconstructed profiles of three-helix and four-helix structures compared to the yeast U2-U6 snRNA structure (Burke et al., 2012). Nevertheless,
Figure 3.7 A&B. Figure shows the normalized SAXS profile of the WT (aI2), Mutant (bI2) and the alternate minor conformation presumably three-helix profile generated by subtracting 86% mutant profile from the WT (A). Corresponding Kratky profiles of the WT, mutant and alternate three-helix conformation are shown in B.
Figure 3.8. The figure shows shape reconstructions of WT (Red) and Mutant that favors four-helix conformation (Grey) from 20 dummy atom models calculated using SAXS profiles. It also shows the superposition of WT and mutant models with previously solved yeast U2-U6 snRNA model (2LKR.pdb). (Burke et al., 2012)
the P(r) profiles generated for the model are very similar to that of the yeast model, therefore we hypothesize that the missing helix is a result of using different parameters for the 3-D reconstruction. Our next step is to try different parameters in order to build the four-helix model. In addition ab-initio modeling from SAXS data alone could be ambiguous therefore we can also use the constraints obtained from other experiments like FRET. We will also generate a series of four-helix mutants of the human U2-U6 snRNA complex with and without the ACAGAGA loop and with Helix I and Helix II (in separate constructs) extended by 10-12 bp to assist in identification of individual helices through added envelope density (Burke et al., 2012). Once we clarify the locations of different helices, we will attempt to measure changes with different concentrations of Mg$^{2+}$, if possible, to analyze its effect on orientation.

3.4. Discussion:

In this chapter the tertiary structural features of both the WT and mutant U2-U6 snRNA complex favoring the formation of four-helix conformation were studied using AUC, OH radical foot printing and SAXS. The results will provide a foundation to build 3D models in which metal ion binding sites might be brought into close proximity for catalytic activity in at least one of the conformations studied.

The results from the AUC studies suggest that the human U2-U6 snRNA undergoes small amount of compaction in the presence of up to 100 mM Mg$^{2+}$ ions. We observed a single peak in sedimentation coefficient distribution curve, which is an indicator of a completely homogenous sample. This result is contradictory to the data obtained from 19F-NMR studies, which suggest that the human U2-U6 snRNA complex adapts at least two different conformations. The reason for this inconsistency in the results might be explained by the fact that the 19F-NMR and AUC
techniques measure different parameters. AUC can measure the hydrodynamic radius of a macromolecule, so it cannot differentiate between two conformations of a macromolecule with almost similar radius. On the other hand 19F-NMR can easily quantify the heterogeneity in the secondary structural profile. AUC is a very sensitive technique and we would expect to see ~14% of the minor conformation as a separate peak in the sedimentation coefficient distribution curve if there was a difference of hydrodynamic radius between the major and minor secondary folds. However, we did not observe any alternate minor peak suggesting that the two conformations observed earlier might have similar hydrodynamic properties. We hypothesize that upon addition of Mg\(^{2+}\) the orientation of the helices in the major conformer might have altered slightly thus reducing the measured hydrodynamic radius of the complex. The compaction observed (~2%) here is very little compared to the compaction observed for *Tetrahymena thermophila* ribozyme (~35%) by Dr. Brenowitz and his coworkers (Takamoto et al., 2002). This relatively small compaction of human U2-U6 snRNA in comparison with *Tetrahymena thermophila* ribozyme suggests that the proposed rearrangement of the helices in the human U2-U6 snRNA complex is minute. Altered orientation of helices was measured using hydroxyl radical footprinting.

Hydroxyl radical footprinting technique used here measures the solvent accessibility of local regions at single nucleotide resolution. Thus, changes in the nucleic acid structure can be probed as a function of divalent metal ions. The hydroxyl radical footprinting experiment data show decreased solvent accessibility of the nucleotides close to as well as distal from the junction. The increased protection of the nucleotides close to the junction may be a result of stabilization of the junction region upon addition of divalent metal ions. The regions distal to the junction can be subdivided into two categories 1) the nucleotides around the metal ion binding
site in U6 ISL (i.e. U74) 2) The nucleotides of U2 snRNA pairing with A & C of ACAGAGA sequence in U6 snRNA (Figure 3.6). The increased protection observed in these regions may be a result of formation of long-range interactions between the U6 ISL and the stem-III. These results are consistent with the hypothesis by Butcher and coworkers that U6 ISL and helix-III of four-helix conformation serve as a scaffold to juxtapose the catalytically important metal binding site in U6 ISL and the 5’ splice site (Sashital et al., 2004a). Hydroxyl radical footprinting experiments by Rhode et al. similarly show an alteration in the spatial relationship between U6 snRNA ISL and the ACAGAGA loop between the two steps of splicing (Rhode et al. 2006) in the presence of spliceosomal proteins. Four-helix junctions that fold into two coaxially stacked helices in anti parallel arrangement are common in nucleic acids, and some of these have been shown to undergo Mg$^{2+}$ dependent conformational change (Hohng et al., 2004; Lilley, 1998). However, single molecule FRET studies on the yeast U2-U6 snRNA complex suggest that the distance between the U6 ISL and helix-III increases with increasing Mg$^{2+}$ ion concentration (Guo et al., 2009). Overall the •OH foot printing results are consistent with previously proposed model in which the human U2-U6 snRNA complex can undergo similar kind of rearrangement upon addition of Mg$^{2+}$ in the absence of spliceosomal proteins, which may explain the limited catalytic activity in its protein free state. However, as an ensemble technique that measures the average solvent accessibility of a particular nucleotide, hydroxyl radical footprinting cannot distinguish between less cleavage by the major conformation and full cleavage by the minor conformation.

Preliminary SAXS data of the wild type human U2-U6 snRNA complex, in conjunction with ab initio modeling, suggest a well-defined, non-globular conformation. Kratky profiles exhibit multiple peaks showing that both four-helix and three-helix conformations are non-globular (Figure 3.7). $P(r)$ of both the mutant and WT show two peaks; one near 20 Å, likely
corresponding to the A-form helical width, and the other near 40 Å, likely corresponding to helical lengths. These values are similar to the values found by Burke et al. for the yeast U2-U6 snRNA complex. Preliminary results on the 3-D reconstruction of the four-helix conformer indicate that the model generated is missing a helix compared to the yeast U2-U6 snRNA model. However, the P(r) profiles generated for the model are very similar to that of the yeast model, therefore we hypothesize that the missing helix is a result of using different parameters for the 3-D reconstruction. Our next step is to try different parameters in order to build the four-helix model. Ab initio modeling from SAXS data alone could sometimes be ambiguous. We will perform fluorescence resonance energy transfer (FRET) experiments on RNA fragments representing the U2-U6 snRNA complex to which organic FRET donor–acceptor dye pairs have been covalently attached to selected sites. Combining the distance constraints obtained from FRET experiments and solvent accessibility data from hydroxyl radical footprinting experiments we anticipate that we will be able to define a three-dimensional model for the four-helix conformer of the human U2-U6 snRNA complex, which can act as a good starting point to characterize the effects of protein on the complex. In order to clarify the location of different helices, we will also generate a series of four-helix mutants of the human U2-U6 snRNA complex with different helices extended by 10-12 bp and perform SAXS experiments. Once we clarify the locations of different helices, we will attempt to measure changes with different concentrations of Mg^{2+} to analyze its effect on orientation of the helices in the major conformation.

The data suggest the U2-U6 snRNA complex forms a heterogeneous mixture of at least two conformations. We were able to deconvolute the three-dimensional structure profiles of the four-helix and three-helix conformation that contribute to the heterogeneity of the complex. The
SAXS profiles obtained show substantial differences between the two structures. We hypothesize that the compaction that we have observed using AUC studies is a result of change in the arrangement of helices in the major conformation i.e. four-helix conformation. Taken together, the results from AUC and hydroxyl radical footprinting experiments suggest that the effect of Mg\(^{2+}\) ions on the tertiary conformational features of the four-helix junction is not significant. Recently Dr. Laing and Dr. Schlick analyzed existing RNA four-way junctions in terms of base pair interactions and 3D configurations and have identified nine broad junction families, which have characteristic helical configurations and stacking patterns (Laing and Schlick, 2009). Examining the sequence of human U2-U6 snRNA complex using the junction explorer software developed by Dr. Schlick, we observed that this might belong to the family called cX. The software takes into account coaxial stacking and the likelihood of stabilization by long-range interaction involving unpaired adenosine residues near the junction such as A-minor interactions. There is a GU wobble base pair near the end of helix-I. This is usually an indication of a perpendicular helix arrangement via a ribose-base interaction combined with p-interaction. There are also several adenine residues in the loop between ISL and helix-II, which indicates that there might be a crossing at the point of strand exchange stabilized by A-minor motifs. This family of junctions is rare and is so far observed only in 16S rRNA. The family is characterized by longer single strand elements between the helices and no coaxial stacking. The family cX has two pairs of helical arms arranged perpendicular to each other by helix packing interactions (Figure 3.9.). The “c” before the letter “X” in the family name denotes the crossing observed at the point of strand exchange.
Figure 3.9. The figure shows the proposed fold of the human U2-U6 snRNA complex based on preliminary computational analysis using Junction explorer software (Laing et al., 2012). The four-helix conformation belongs to a family called cX, in which four helices orient themselves perpendicular to each other.
However, we cannot exclude the possibility that the minor conformation i.e. the three-
helix conformation also undergoes conformational change. Such a hypothesis is consistent with
earlier studies from the Rueda group on yeast U2-U6 snRNA(Guo et al., 2009). Rueda and
coworkers made a similar observation, e.g. they found that the yeast U2-U6 snRNA complex
adapts at least three different conformations in the presence of Mg$^{2+}$ ions 1) a three-helix
conformation 2) four-helix conformation with U6 ISL and helix III close to each other 3) four-
helix conformation with U6 ISL and helix III distant to each other as observed by their smFRET
studies.

Overall, our results from AUC, hydroxyl radical experiment suggest that the human U2-
U6 snRNA complex may undergo a slight change in the orientation of helices upon addition of
Mg$^{2+}$, which results in the small amount of compaction observed. We constructed a mutant
conformation that favors formation of the four-helix conformation. Preliminary SAXS studies
suggest that the human U2-U6 snRNA complex and the mutant complex are amenable to further
study by SAXS and, with complementary structural parameters by other methods, are likely to
yield high quality models.

The splicing reaction mediated by the protein-free U2-U6 snRNA complex results in a
very low yield and a very slow rate of catalysis(Valadkhan and Manley, 2001; Valadkhan et al.,
2007). However, the intact spliceosome is dynamic and contains numerous proteins. The small
amount of compaction/conformational change that we have observed using AUC and hydroxyl
radical footprinting experiments shows very little change upon addition of Mg$^{2+}$ and (assuming
that addition of high concentrations of divalent metal ions helps to facility conformational
change analogous to changes seen upon interaction with spliceosomal proteins) concurs with the
low yield and slow catalytic rate of protein-free splicing reaction. It is possible that only a small
fraction of RNA underwent conformational changes, which was measured as an average by the equilibrium techniques that we have used. Thus, one of the future directions that we can take is to include protein factors to the study this project.
CHAPTER-4

CONCLUSIONS AND FUTURE DIRECTIONS

4.1. Conclusions:

4.1.1 Introduction:

The research described in this dissertation focuses on the secondary and tertiary structural features of the human U2-U6 snRNA complex in the absence of spliceosomal proteins and the effect of Mg\(^{2+}\) ions on those structural features. The complex formed between U2 and U6 snRNA is implicated in a critical role in the catalysis of pre-mRNA splicing. As discussed in the chapter-1, a very small amount of catalytic activity is performed by the protein-free complex formed between U2 and U6 snRNA in the presence of divalent metal ions (Mg\(^{2+}\)) alone, supporting the catalytic ability of these snRNAs (Jaladat et al., 2011; Valadkhan and Manley, 2001). It has also been shown that there are three catalytically important metal ion-binding sites at distant places of U2-U6 snRNA complex 1) the two G residues of invariant ACAGAGAGA loop 2) the A residue of invariant AGC triad and 3) at the U74 (Blad et al., 2005; Fabrizio and Abelson, 1992; Gordon et al., 2000b; Yu et al., 1995; Yuan et al., 2007b) (for a detailed discussion see chapter-1). Mechanistic parallels between reactions catalyzed by the spliceosome and the self-splicing group II intron (Sontheimer et al., 1999a; Valadkhan and Manley, 2002a; Weiner, 1993), as well as sequence and structural similarities between functionally analogous RNA sequences in the two systems(Gordon et al., 2000b; Keating et al., 2010), suggest that the metal ion-binding sites in the U2-U6 snRNA complex are brought into close proximity to each other (Steitz and Steitz, 1993), as observed in the Group II intron of Oceanobacillus iheyensis.
(Toor et al., 2008). The important role of these metal ions suggests that the ability of the U2-U6 snRNA complex to undergo conformational change is likely to be important. The overall goal of this research was to characterize these conformational changes of the human U2-U6 snRNA complex upon addition of Mg$^{2+}$ ions.

### 4.1.2 Secondary Structural fold:

As a first step, we attempted to characterize the lowest energy structure of the complex in the absence of spliceosomal proteins. The ground state secondary structural fold of the human U2-U6 snRNA complex was investigated by enzymatic structure probing technique (chapter-2). The results from the probing experiments are consistent with the formation of Stem I and therefore the four-helix conformation. Addition of up to 100 mM Mg$^{2+}$ ions did not alter the overall secondary structure of the complex, suggesting that the divalent metal ions has little effect on the secondary structure of the complex. Separate studies performed by Caijie Zhao using $^{19}$F-NMR technique complemented the results obtained from the enzymatic probing technique. Specifically, she also found that the U2-U6 snRNA complex forms predominantly four-helix conformation, and her approach permitted quantification of this conformer as $\sim$86% of the total, with an alternate (presumably three-helix) minor conformation, or conformations, comprising the remaining 14% at 25$^\circ$ C. Addition of up to 5 mM Mg$^{2+}$ ions decreased the percentage of four-helix conformation to $\sim$83%. Taken together, these data suggest that the human U2-U6 snRNA complex forms predominantly a four-helix conformation with some degree of heterogeneity. We have also successfully designed a mutant model of the human U2-U6 snRNA complex, which essentially forms $\sim$100% four-helix conformation. This mutant verified the identity of the four-helix fold and will be used for future study of the three-dimensional structural features of the major conformation i.e. four helix conformation.
These experiments for the first time show that the protein free human U2-U6 snRNA complex in vitro forms predominantly a four-helix conformation, consistent with the previous genetic studies, performed on human cells (Sun and Manley, 1995a). Therefore, we can say that the fragments used here representing the U2 and U6 snRNAs can form similar basepairing pattern in its protein free state. It is intriguing to note that the major conformation formed by the human U2-U6 complex is different than that of the yeast, even though the splicing mechanism in yeast and humans are identical. Since the genetic studies in human cells did not find any evidence of the three-helix conformation required for the splicing, it is difficult to state if the small fraction of alternate conformation observed here (presumably the three-helix conformation) is biologically relevant at certain point of splicing or an artifact of various mutations/substitutions that were introduced into the WT U2 and U6 snRNA sequences.

The small amount of catalytic activity in vitro displayed by the protein free human U2-U6 snRNA complex is strictly dependent on high concentrations of Mg\(^{2+}\), which might necessitate the formation of catalytically active conformation. Therefore, it is important to characterize the effect of Mg\(^{2+}\) on the conformation of the human U2-U6 snRNA complex. Addition of Mg\(^{2+}\) to the complex decreased the percentage of the predominant four-helix conformation from ~86% to ~83%, however the four-helix conformation was still predominant. Therefore we can speculate that the Mg\(^{2+}\) has very little effect on the basepairing pattern of the human U2-U6 snRNA complex.

4.1.3 Tertiary conformational features:

The tertiary conformational features of the human U2-U6 snRNA complex were investigated using AUC, hydroxyl radical footprinting studies and SAXS (Chapter 3). The results
of AUC experiments performed on the human U2-U6 snRNA complex suggest the formation of homogenous complex with unique hydrodynamic radius with no self-association. This observation is inconsistent with the results obtained from earlier $^{19}$F-NMR studies. This anomaly may be a result of measuring two different things with different experiments. AUC is an equilibrium technique that can measure the hydrodynamic radius of the molecule therefore it cannot differentiate between two populations of similar hydrodynamic properties. Whereas $^{19}$F-NMR can measure the heterogeneity of the complex using different chemical shift in single and double stranded regions. These results suggest that the hydrodynamic properties of both the four-helix and the alternate minor (presumably three-helix) conformation are similar. The decreased values of Stokes’ radius and axial ratio with increase in the concentration of Mg$^{2+}$ ions suggest slight compaction in the tertiary fold. The hydroxyl radical foot printing data identified local regions whose solvent accessibility decreased upon addition of Mg$^{2+}$ ions. The increased protection observed in these regions may be a result of slight alteration in the arrangement of the helices U6 ISL and the stem-III. We have generated preliminary 3D models for the WT human U2-U6 snRNA complex, which is a mixture of 86% four-helix conformation and 14% alternate conformation. We have also constructed 3D model for the mutant U2-U6 snRNA complex, which forms ~100% four-helix conformation which is predominant in the distribution. Since they are 86% similar we would expect them to have similar 3D profiles and we observed that both the models have similar 3D structure. However, more SAXS experiments have to be performed to identify the different helices observed in the 3D model.

Overall the results from AUC, hydroxyl radical foot printing and SAXS studies suggest that the complex undergoes very small change in presence of divalent metal ions. The results from the hydroxyl radical foot printing experiment support minor rearrangement in the
orientation of helices ISL and Helix-III. Hydroxyl radical footprinting experiments by Rhode et al. similarly show an alteration in the spatial relationship between U6 snRNA ISL and the ACAGAGA loop between the two steps of splicing (Rhode et al. 2006) in the presence of spliceosomal proteins. However, the conformational change that we have observed is small, probably because of lack of the spliceosomal proteins, which might act as scaffolding agents to facilitate the required conformational change. As discussed in chapter-3 the junction explorer software (http://datalab.njit.edu/bioinfo/coaxial/index.html) predicted that the human U2-U6 snRNA complex might belong to a category called cX family of junctions. The characteristic of this family is that the four helices orient in mutually perpendicular manner (Figure 3.9), with very little conformational change in presence of Mg$^{2+}$ ions, which is consistent with our observation. If the four-helix conformation formed by the U2-U6 snRNA complex belongs to this category, the orientation of the helices will be as shown in the figure 3.9, in which the U6-ISL is perpendicular to helix-III. And the orientation of ISL and stem-I are subject to change in the presence of Mg$^{2+}$, which bring U6-ISL closer to helix-III containing ACAGAGA sequence. This kind of arrangement can act as a scaffold to juxtapose a metal ion binding pocket in U6 ISL with the 5’ splice site and was earlier proposed by Prof. Butcher and his co-workers (Sashital et al., 2004a). This proposed model is similar to the recent 3-D model of yeast U2-U6 snRNA complex characterized by the same group in terms of facilitating the splicing mechanism, even though the yeast adopts the three-helix conformation (Burke et al., 2012). From their model, the authors noted that the essential features of the complex, including the U80 metal ion-binding site in the U6 ISL, AGC triad, and pre-mRNA recognition sites are located on one face of the complex. Therefore the human U2-U6 snRNA complex might undergo a small amount of
conformational change in the presence of divalent metal ions alone and that the spliceosomal proteins may play an important role in the proposed conformational change.

4.2 Future directions:

1) **Identification of the helices on the 3D model:** In order to analyze the orientation of the helices to characterize the three-dimensional model of the four-helix conformation, we need to first identify the stems in the preliminary model that was built. We will design the mutant models with extended helices II or III, these extended sequences appear as added electron density. By overlapping the previous mutant models with the extended mutant model we will be able locate the position of helices II, III and U6 ISL.

2) **Construct the model of three-helix conformation:** The three-helix conformer we observed in the wild type construct is only ~14% and it might have more than one structure, therefore it was difficult to obtain sufficient resolution for structural information. To characterize the tertiary structural features of the three-helix junction structure i.e. minor conformation, we will perform similar SAXS studies on a construct shown to favor the formation of the three-helix structure. This will help us build a model for an alternate three-helix conformation. We can compare both the models and characterize if the metal ion binding sites are brought into close proximity for the catalytic activity in at least one of the folds.

3) **To characterize the effect of spliceosomal proteins on the conformation of human U2-U6 snRNA complex using FRET:** Fluorescence Resonance energy transfer (FRET) is a well-established technique to study the distance information in biomolecules (Cardo et al., 2012; Lilley, 2004; McDowell et al., 2010). Members of the Greenbaum laboratory will examine the effect of spliceosomal proteins in facilitating folding of the human U2-U6 snRNA complex.
Changes in conformation upon addition of nuclear extract (which contains the full complement of spliceosomal proteins) will be measured by (FRET) experiments on RNA fragments representing the U2-U6 snRNA complex to which organic FRET donor–acceptor (Cy3-Cy5 or Cy3-Fluorescein) dye pairs have been covalently attached to selected sites. As the efficiency of energy transfer is strictly dependent on the distance, the experiments can characterize conformational changes in the presence of Mg$^{2+}$ and/or spliceosomal proteins.
Supplementary Data

Figure legend:

Figure 1: Ribonuclease cleavage patterns for the U2-U6 snRNA complex. Internally normalized cleavage intensities of each $^{32}$P-labeled (*) hU2 or hU6 strand (paired with an excess of unlabeled strand) were plotted against the sequence of RNA; the height of each peak represents the intensity of the band of that nucleotide in the sequencing gel. Each data point is an average of at least two repetitions.

(A) Cleavage profile diagram for *hU2 snRNA paired with hU6 snRNA followed by cleavage with RNaseV1 in the presence of 1 mM and 100 mM Mg$^{2+}$. (B) Cleavage profile diagram for hU2 snRNA paired with *hU6 snRNA followed by cleavage with RNaseV1 in the presence of 1 mM and 100 mM Mg$^{2+}$. (C) Cleavage profile diagram for the complex of *hU2 snRNA paired with hU6 snRNA after partial cleavage with RNase A with 0 mM and 100 mM Mg$^{2+}$. (D) Cleavage profile for hU2 snRNA paired with *hU6 snRNA after partial cleavage with RNase A with 0 mM and 100 mM Mg$^{2+}$. (E) Cleavage profile diagram for *hU2 snRNA paired with hU6 snRNA after partial cleavage with RNase T1 with 0 mM and 100 mM Mg$^{2+}$. (F) Cleavage profile diagram for hU2 snRNA paired with *hU6 snRNA after partial cleavage with RNase T1 with 0 mM and 100 mM Mg$^{2+}$. 
1. (A)
Figure 2. Representative sequencing gel for the *hU2+hU6 complex and the mutant *hu2+hU6 complex subjected to partial hydrolysis separately with RNase A, T1 and V1. Controls such as T1-digestion ladder and alkaline hydrolysis ladder were used as markers.
Figure 3. Thermal/color plots showing cleavage intensities of complex between *hU6 and hU2 complex, following the reaction with hydroxyl radicals with increasing concentration of Mg$^{2+}$ ions. The change is not as consistent as we have observed in *hU2.
Figure 4: Cleavage profile diagrams showing normalized cleavage intensities of *hU6+hU2 and *hU2+hU6 snRNA complex, following reaction with OH radicals at 0 mM Mg$^{2+}$ ions. We noticed cleavage intensity of some nucleotides is less than the other suggesting that these nucleotides are more protected.
References:


