Conformational Heterogeneity and Exchange within the Junction of the Human U2-U6 snRNA Complex Measured by 19F NMR Spectroscopy

Caijie Zhao

Graduate Center, City University of New York

How does access to this work benefit you? Let us know!

Follow this and additional works at: https://academicworks.cuny.edu/gc_etds

Recommended Citation
https://academicworks.cuny.edu/gc_etds/644

This Dissertation is brought to you by CUNY Academic Works. It has been accepted for inclusion in All Dissertations, Theses, and Capstone Projects by an authorized administrator of CUNY Academic Works. For more information, please contact deposit@gc.cuny.edu.
Conformational heterogeneity and exchange within the junction of the human U2-U6 snRNA complex measured by $^{19}$F NMR Spectroscopy

By

Caijie Zhao

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

2015
This manuscript has been read and accepted for the
Graduate Faculty in Biochemistry in satisfaction of the
dissertation requirement for the degree of Doctor of Philosophy.

Professor Nancy L. Greenbaum

Date

Chair of Examining Committee

Professor Richard S. Magliozzo

Date

Executive Officer

Professor Nancy L. Greenbaum

Professor Dixie J. Goss

Professor Frida E. Kleiman

Professor Lesley Davenport

Professor Ranajeet Ghose

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK
Abstract

Conformational heterogeneity and exchange within the junction of the human U2-U6 snRNA complex measured by $^{19}$F NMR Spectroscopy

By
Caijie Zhao

Advisor: Professor Nancy L. Greenbaum

Pre-mRNA splicing, the process by which non-coding sequences are removed and coding sequences are ligated, plays an important role in the maturation of precursor messenger RNA molecules prior to their translation into proteins. In eukaryotes, splicing is catalyzed by the spliceosome, a large and dynamic nuclear macromolecular “machine” that comprises five small nuclear (sn)RNA molecules and more than one hundred proteins. Among the RNA components, the complex formed between the U2 and U6 snRNA molecules is implicated in catalytic activity. In this dissertation, I have used $^{19}$F NMR techniques to characterize the conformation (“fold”) and dynamics of a protein-free biophysically tractable paired construct representing the human U2-U6 snRNA complex in which a single $^{19}$F-substituted nucleotide is incorporated in the U2 snRNA sequence. $^{19}$F spectra of the complex in the absence of Mg$^{2+}$ are consistent with formation of a four-helix junction structure as the predominant fold. However, spectra also
identify a lesser fraction (up to 14% at 25°C) of conformation that suggesting a collection of folds comprising three helices. In the presence of 5 mM Mg$^{2+}$, the fraction of the three-helix conformation increased to ~17%, suggesting a slight shift to the alternative conformation. The low energy cost of interconversion between different conformers ($\Delta G$ is approximately $-4.6$ kJ/mol for the interconversion) suggests the complex is very flexible and may be capable of undergoing spontaneous conformational rearrangement between conformers. Supporting this, $^{19}$F-$^{19}$F EXSY data provide the evidence of the chemical exchange between different conformations in a sub-second timescale. Taken together, these results indicate conformational heterogeneity and interconversion in the protein-free human U2-U6 snRNA complex, consistent with a model in which the RNA has sufficient conformational flexibility to facilitate rearrangements of the spliceosome between steps of splicing.
Acknowledgements

First of all, I would like to express my great gratitude and sincere respect to my mentor, Professor Nancy L. Greenbaum for all her guidance though my PhD work in her lab. She has always been extremely supportive and has constantly encouraged me whenever I felt not confident. I have always been inspired by her intelligence and hard-working.

I would like to thank Professor Dixie J. Goss, Professor Lesley Davenport, Professor Frida E. Kleiman and Professor Ranajeet Ghose for serving on my committee and their valuable guidance and interest in my education.

I especially would like to thank Professor Ranajeet Ghose, Dr. Clemens Anklin and Dr. Matthew Devany for their help with NMR training, data collection and analysis as well as trouble shooting. I am also grateful to Professor Saad Mneimneh and Professor Ruel Desamero for their guidance and help on collaborative projects.

I would like to express my thanks to all my labmates: Dr. Manjula Kolipaka, Dr. Milena Popovic, Dr. Faqing Yuan, Dr. Ravichandra Bachu, William Perea, He Huang and Huong Chu. I also thank all the undergraduates who worked with me as a team and helping me out in tough times: Alina Masters, Kulsuma Akhtar, Carolina Salguero and Sylwia Trzyna.

Finally I would like to thank my parents who have always been loving me and supporting me.
# Table of contents

Copyright page…………………………………………………………………………………………………II
Approval page…………………………………………………………………………………………………III
Abstract……………………………………………………………………………………………………IV
Acknowledgements………………………………………………………………………………………………VI
List of Tables……………………………………………………………………………………………………X
List of Figures…………………………………………………………………………………………………XI

## Chapter I  Introduction

1.1 Introduction……………………………………………………………………………………………1
1.2 Building Blocks of Nucleic Acid……………………………………………………………………2
1.3 Importance of RNA structure and dynamics in biological process…………………………9
1.4 Spliceosome and pre-mRNA splicing in eukaryotes .........................................................11
1.5 U2-U6 snRNA complex in spliceosome .........................................................................14
1.6 Use of NMR techniques in studies of RNA structure and dynamics ..............................15
1.7 Application of $^{19}$F NMR in biological systems.............................................................17
1.8 Summary of questions to be addressed...........................................................................18

## Chapter II  Secondary structural fold of the human U2-U6 snRNA complex

2.1 Introduction.....................................................................................................................20
2.2 Materials and Methods

2.2.1 Models for folding of the U2-U6 snRNA complex ......................................................25
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2.2 Formation of the U2-U6 snRNA complex</td>
<td>26</td>
</tr>
<tr>
<td>2.2.3 Design of the mutants</td>
<td>28</td>
</tr>
<tr>
<td>2.2.4 $^1$H-NMR spectra</td>
<td>35</td>
</tr>
<tr>
<td>2.2.5 $^{19}$F NMR studies</td>
<td>35</td>
</tr>
<tr>
<td>2.3 Results</td>
<td>37</td>
</tr>
<tr>
<td>2.4 Discussion</td>
<td>50</td>
</tr>
<tr>
<td>Chapter III  Probing chemical exchange between RNA conformers by $^{19}$F NMR spectroscopy</td>
<td></td>
</tr>
<tr>
<td>3.1 Introduction</td>
<td>58</td>
</tr>
<tr>
<td>3.2 Materials and Methods</td>
<td></td>
</tr>
<tr>
<td>3.2.1 Sample preparation</td>
<td>60</td>
</tr>
<tr>
<td>3.2.2 Acquisition parameters</td>
<td>61</td>
</tr>
<tr>
<td>3.3 Results</td>
<td>63</td>
</tr>
<tr>
<td>3.4 Discussion</td>
<td>73</td>
</tr>
<tr>
<td>Chapter IV  Conclusions and Future directions</td>
<td></td>
</tr>
<tr>
<td>4.1 Introduction</td>
<td></td>
</tr>
<tr>
<td>4.1.1 Fold of human U2-U6 snRNA complex</td>
<td>76</td>
</tr>
<tr>
<td>4.1.2 Effect of magnesium ion</td>
<td>78</td>
</tr>
<tr>
<td>4.1.3 Dynamics of human U2-U6 snRNA complex</td>
<td>80</td>
</tr>
<tr>
<td>4.2 Future directions</td>
<td></td>
</tr>
<tr>
<td>4.2.1 Characterize the relative orientation between Helix I and U6 ISL of human U2-U6 snRNA complex</td>
<td>81</td>
</tr>
<tr>
<td>4.2.2 Structural features of the junction region in human U2-U6 snRNA complex</td>
<td>83</td>
</tr>
<tr>
<td>4.2.3 Develop a three-dimensional model of human U2-U6 snRNA complex</td>
<td>83</td>
</tr>
</tbody>
</table>
4.2.4 Comparison of U2-U6 snRNA complex with U12-U6atac complex from minor spliceosome………………………………………………………………………..84

References………………………………………………………………………………...86
List of Tables

Table 1.1 Nitrogenous bases and the corresponding nucleotides in RNA or DNA…………….4

Table 3.1 Extracted parameters from 2D $^{19}$F-$^{19}$F EXSY experiments of the bi-stable sample…69
List of Figures

Figure 1.1 Structure of Ribose and 2-Deoxyribose .........................................................3
Figure 1.2 Demonstration of C2'-endo and C3'-endo sugar confirmation .........................3
Figure 1.3 Polymerization of two nucleotides to form single-stranded RNA ....................7
Figure 1.4 Common base pairs found in RNA .................................................................8
Figure 1.5 Examples of common secondary structure motifs in RNA .............................10
Figure 1.6 Schematic of the two-step transesterification reactions carried out by the Spliceosome ..............................................................12
Figure 1.7 Schematic of mammalian major spliceosome assembly and disassembly cycle ......13

Figure 2.1 Three-helix model of the protein-free human (h)U2-U6 snRNA complex ...........23
Figure 2.2 Four-helix junction model of the protein-free hU2-U6 snRNA complex ..........24
Figure 2.3 Demonstration of the position of 5-19F-C substitution in the two models of the protein-free human U2-U6 snRNA complex ..........................................................27
Figure 2.4 Unimolecular sample representing human U2-U6 snRNA complex ...............29
Figure 2.5 Sequence and proposed secondary structure of a mutation that favors the four-helix model (mutation 1) .................................................................30
Figure 2.6 Sequence and proposed secondary structure of a mutation favoring the four-helix model (mutation 2) .................................................................31
Figure 2.7 Sequence and proposed secondary structure of a mutation that favors the three-helix model (mutation 3) .................................................................32
Figure 2.8 Proposed four-helix junction model for mutated human U6 (mAGA) and U2 snRNA complex (mutation 4) .................................................................34
Figure 2.9 Pairing of RNA oligomers representing hU2 and U6 snRNA analyzed by non-denaturing gel electrophoresis .................................................................36
Figure 2.10 Aromatic-anomeric (base-H1') region of a NOESY spectrum of the unimolecular sample shown in Figure 2.4

Figure 2.11 One-dimensional $^{19}$F NMR spectra of control RNA oligomers

Figure 2.12 Imino region of one-dimensional $^1$H NMR spectrum of the double-stranded control RNA oligomer

Figure 2.13 One dimensional $^{19}$F NMR spectrum of human U2-U6 snRNA complex

Figure 2.14 (A) One dimensional $^{19}$F NMR spectrum of the extended U2 snRNA pairing with U6 snRNA (mutation 1). (B) One dimensional $^{19}$F NMR spectrum of the mutated UUCG tetra loop U2 snRNA pairing with U6 snRNA (mutation 2)

Figure 2.15 One dimensional $^{19}$F NMR spectra of a mutation that favors the three-helix model (mutation 3)

Figure 2.16 One dimensional $^1$H NMR spectra of human U2-U6 and U2-mAGA snRNA complex

Figure 2.17 One dimensional $^{19}$F NMR spectrum of human U2 and mAGA complex (mutation 4)

Figure 2.18 Overlaid one dimensional $^{19}$F NMR spectra of the human U2-U6 snRNA complex with (upper trace) and without (lower trace) 5mM Mg$^{2+}$

Figure 3.1 (A & B) Proposed secondary structures of the bi-stable sequence; (C) One-dimensional $^1$H spectrum of the bi-stable sample at 10°C

Figure 3.2 One-dimensional $^{19}$F spectrum of the bi-stable sample at 25°C

Figure 3.3 Two-dimensional $^{19}$F-$^{19}$F EXSY spectra of the bi-stable sample at mixing times of 2 ms, 10 ms, 100 ms and 700 ms

Figure 3.4 Plot of the intensities of exchange peaks vs. mixing times (2-760 ms)

Figure 3.5 Two-dimensional $^{19}$F-$^{19}$F EXSY spectra of human U2-U6 snRNA complex acquired at 600 MHz with mixing times of 100 ms, 200 ms, 400 ms and 500 ms
Chapter I

Introduction

1.1 Introduction

Nucleic acids and proteins are ubiquitous biological macromolecules that play essential roles in all living organisms. Compared with a wide range of functions mediated by proteins, such as catalyzing metabolic reactions (Northrop 1929), participating as building blocks of cytoskeleton (Wickstead and Gull 2011) and transporting other molecules across cell membranes (Alber et al. 2007), the principle functions of nucleic acids relate to the carrying, passage, and expression of genetic information (Crick 1970). Deoxyribonucleic acids (DNA) are responsible for encoding genetic information and ribonucleic acids (RNA), which are transcribed from DNA, function as the carriers of genetic information and specify the amino acid sequence of protein in gene expression. Those RNA molecules that carry the primary sequence of protein from DNA to ribosome are called messenger-RNA (mRNA). Other RNA molecules directly involved in protein synthesis include ribosomal (r)RNA and transfer (t)RNA. In addition, there are many non-coding RNA molecules that play important roles in catalysis of post-transcriptional processing and regulation of gene expression. Among this latter group are small nuclear RNA (snRNA) (Ohshima et al. 1981) and riboswitches (Nudler and Mironov 2004) etc. The work described in this dissertation studies the structure and dynamics of non-coding RNA molecules.
1.2 Building Blocks of Nucleic Acid

Both DNA and RNA are polymers of nucleotides. Each nucleotide has three components: a 5-carbon sugar, a nitrogenous base and a triphosphate group which loses a pyrophosphate group while polymerizing. There are two types of sugars, ribose and deoxyribose, both of which are pentose monosaccharides forming a five-membered ring (furanose) (Figure 1.1). The difference between them is that in deoxyribose, the hydroxyl group attached to the 2′ carbon is replaced by hydrogen. The presence of the chemically active hydroxyl group at 2′ position of ribose enables RNA molecules to catalyze nucleolytic cleavage of the phosphodiester, among them the splicing of precursor (pre)m-RNA molecules (discussed in details in the following section), while DNA cannot mediate such reactions. However, the high reactivity of the 2′-OH renders RNA chemically unstable under alkaline conditions.

The 2′-OH group also has an impact on the conformation of the sugar. In deoxyribose, the common conformation is the C2′-endo sugar pucker. While in ribose, because of the steric hindrance associated with 2′-OH, the C3′-endo sugar pucker is preferred (Figure 1.2).

Based on the parent compound they are derived, each nitrogenous base is classified as belonging to one of two groups, purine and pyrimidine. The purine molecules in DNA and RNA are the same: adenine (A) and guanine (G); for pyrimidines, cytosine (C) and uracil (U) are found in RNA and in DNA, thymine (T) replaces uracil. The structure of nitrogenous bases and the corresponding nucleotides is shown in Table 1.1.
Figure 1.1 Structure of Ribose (left) and 2-Deoxyribose (right).

Figure 1.2 Demonstration of C2'-endo and C3'-endo sugar confirmation.

(Figure from Julien KR, Sumita M, Chen PH, Laird-Offringa IA, Hoogstraten CG. (2008) Conformationally restricted nucleotides as a probe of structure-function relationships in RNA. RNA 14(8): 1632-1643.)
Table 1.1 Nitrogenous bases and the corresponding nucleotides in RNA or DNA

<table>
<thead>
<tr>
<th>Nitrogenous base</th>
<th>Nucleotide</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adenine</td>
<td>![Adenosine triphosphate (ATP)]</td>
</tr>
<tr>
<td>Guanine</td>
<td>![Guanosine triphosphate (GTP)]</td>
</tr>
<tr>
<td>Cytosine</td>
<td>![Cytidine triphosphate (CTP)]</td>
</tr>
</tbody>
</table>
Table 1.1 Contd

<table>
<thead>
<tr>
<th>Nitrogenous base</th>
<th>Nucleotide</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thymine</td>
<td>Deoxythymidine triphosphate (dTTP)</td>
</tr>
<tr>
<td>Uracil</td>
<td>Uridine triphosphate (UTP)</td>
</tr>
</tbody>
</table>
The nucleotide molecules form single-stranded polymers through the esterification reaction between the triphosphate attached to the 5’ carbon of the incoming nucleotide and the hydroxyl group of the 3’ carbon in the nucleotide to which it joins (Figure 1.3). During the esterification, a pyrophosphate molecule is released. In DNA molecules and in paired regions of RNA, two polynucleotide stands assemble to form a stable structure called double helix. The two stands are anti-parallel and held together through hydrogen bonds between complementary base pairs (Figure 1.4). The most common base pairs are Watson-Crick A-T (A-U in RNA) and G-C base pairs. Because of the geometry required by the hydrogen bonds inside the base pair, the angle between the two glycosidic bonds is approximate 120° for the narrower side and 240° for the wider side, thus creates the minor groove and major groove in double helix DNA or RNA structures. However, it is also possible to form “wobble” bases between G and U, or between U and U, as well as other non-canonical base pairs such as G•A. In such cases, the angles between the glycosidic bonds will differ from those found in the canonical pairings, resulting in variations of backbone conformation.
Figure 1.3 Polymerization of two nucleotides to form single-stranded RNA.

(Figure from http://en.wikipedia.org/wiki/Nucleic_acid_sequence)
Figure 1.4 Common base pairs found in RNA. Hydrogen bonds are represented in dashed lines.

Both Watson-Crick base pairs (AU and GC) and non-canonical base pair (GU wobble) are shown. (Figure from http://eternawiki.org/wiki/index.php5/Base_Pair)
1.3 Importance of RNA structures and dynamics in biological process

RNA is synthesized as a single-stranded polymer copied from a template strand of DNA. However, the single strand is likely to fold into secondary (base-paired) and sometimes tertiary structures (involving long range interactions) depending on its sequence and its role in biological process. Due to the great flexibility in backbone (six single bonds), RNA molecules are often very plastic and can undergo changes in base pairing patterns, thus forming alternative secondary and tertiary conformations. This results in the possibility of alternative secondary structure motifs in functional RNA molecules (Figure 1.5) and is critical for multiple roles of some noncoding RNA molecules in gene expression (Dethoff et al. 2012). For example, analysis of the chemical exchange rate for conformational change is important for understanding ribozyme-mediated catalysis as well as RNA-ligand interactions associated with activity of riboswitches (Batey 2012) and in drug design (Penchovsky and Stoilova 2013). Such rearrangements are also associated with the activity of the small nuclear (sn)RNA components of the eukaryotic spliceosome, which are dynamic and undergo major rearrangements throughout the splicing reaction (see ahead, section 1.4). In order to understand mechanisms of RNA-mediated activity, it is essential to probe the secondary structures of functional RNA under different conditions as well as kinetics and thermodynamics associated with conformational rearrangement between different conformers.
Figure 1.5 Examples of common secondary structure motifs in RNA.

(Figure from Nowakowski J, Tinoco I Jr. (1997) RNA Structure and Stability. Seminars in Virology 8(3): 153-165.)
1.4 Spliceosome and pre-mRNA splicing in eukaryotes

Pre-mRNA splicing is the process of excision of noncoding intervening sequences (introns) from eukaryotic pre-mRNA molecules and ligation of flanking coding regions (exons). It is a critical step in the maturation of nascent mRNA transcripts and in the generation of alternative products from polycistronic genes. This process involves two sequential transesterification reactions catalyzed by the spliceosome, a dynamic nuclear ribonucleoprotein complex that comprises five recyclable snRNA molecules and at least 70 small nuclear ribonucleoprotein particle (snRNP)-associated proteins and more than 100 non-snRNP proteins (Stark and Lührmann 2006). In the first transesterification reaction, the 2′OH of a conserved adenosine residue in the intron, called the branch site because of the branched lariat intermediate it forms, executes a nucleophilic attack at the 5′ splice site and results in formation of a free 3′OH. The second reaction is characterized by attack of this free 3′OH at the 3′ splice site, thus joining the two exons and release of the lariat intron (Figure 1.6).

Figure 1.7 illustrates the process of spliceosome assembly (Wahl et al. 2009). Specifically, the assembly starts with binding of U1 snRNP with the 5′ splice site of intron. Following the binding of U1 snRNP, U2 snRNA binds the branchpoint of the intron with the help of corresponding proteins, forming the prespliceosome (complex A). After the formation of complex A, the U4/U6 and U5 snRNPs are involved as a preassembled U4/U6.U5 tri-snRNP, leading to the formation of the complex B. Complex B then undergoes conformational and compositional rearrangements and releases U1 and U4 snRNA to become catalytically active. The activated spliceosome catalyzes the first transesterification reaction and results in complex C.
Figure 1.6 Schematic of the two-step transesterification reactions carried out by the spliceosome. (Figure from Brow DA (2002) Allosteric cascade of spliceosome activation. *Annu Rev Genet* **36**: 333-360.)
Figure 1.7 Schematic of mammalian major spliceosome assembly and disassembly cycle.

U1, U2, U4, U5 and U6 (the colored balls) represent the corresponding spliceosomal snRNPs.

(Figure is adapted from Wahl MC, Will CL, Lührmann R. (2009) The spliceosome: design principles of a dynamic RNP machine. *Cell* 136(4): 701-718.)
After the second reaction taking place, the spliceosome dissociates, which releases the U2, U5, and U6 snRNPs for the next round of splicing. (Wahl et al. 2009)

1.5 U2-U6 snRNA complex in spliceosome

Despite of more than a hundred of proteins involved (Stark and Lührmann 2006), the snRNA components of the spliceosome are implicated in key splicing roles, including recognition, pairing, and catalysis. Of the five snRNAs, only U2 and U6 snRNA are directly involved in binding with pre-mRNA in both splicing steps and participate in generation of the catalytic core (Fabrizio and Abelson 1990; Madhani and Guthrie 1992; McPheeters and Abelson 1992). The complex between U2 and U6 snRNA molecules, comprising extensive intra- and intermolecular base pairing sequences, assists in positioning both the 5’ splice site and the branch site region of the intron for the first cleavage reaction (Black et al. 1985; Parker et al. 1987; Lesser and Guthrie 1993).

In addition, the complex is implicated in catalytic activity, which depends upon several specifically bound Mg$^{2+}$ ions (Sontheimer et al. 1997; Yean et al. 2000; Huppler et al. 2002; Valadkhan and Manley 2002; Yuan et al. 2007), 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. 2007; Valadkhan et al. 2009). However, the Prp8 protein, an integral component of the U5 snRNP,
contains a RNase H domain near the C-terminus, suggesting that a protein component contributes to the reaction chemistry (Pena et al. 2008; Ritchie et al. 2008; Yang et al. 2008).

Mechanistic parallels between reactions catalyzed by the spliceosome and the self-splicing group II intron (Weiner 1993; Sontheimer et al. 1999; Valadkhan and Manley 2002), as well as sequence and structural similarities between functionally analogous RNA sequences in the two systems (Gordon et al. 2000; 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 noted in the Group II intron of *Oceanobacillus iheyensis* (Toor et al. 2008). The important role of these metal ions implies the need for a specific structural context for each site and for any interaction between them, which may differ for the two steps of splicing. Thus it is important to probe the secondary structure of the U2-U6 snRNA complex and conformational change associated with different steps of splicing. One goal of the work presented in this dissertation is to analyze the lowest energy conformation(s) of the protein-free U2-U6 snRNA complex and to examine the possibility of interconversion. The significance of this aim is to relate structural studies with functions of protein-free U2-U6 snRNA complex, thus to understand the mechanisms of spliceosome and its RNA components as ribozymes.

### 1.6 Use of NMR techniques in studies of RNA structure and dynamics

Solution state NMR has unique advantages in measuring the structure and dynamics of biological molecules such as protein and nuclear acids. Nuclei typically observed in NMR studies of RNA include $^1$H, $^{13}$C, $^{15}$N and $^{31}$P, which are either naturally present in high abundance ($^1$H, $^{31}$P) or can be enriched in RNA oligomers by transcription with labeled
nucleotide triphosphates ($^{13}$C, $^{15}$N). NMR approaches monitoring $^1$H, $^{13}$C, or $^{15}$N significantly improves the spectra overlap that is generally encountered in $^1$H-NMR spectra, thus making it possible to solve the structure of complicated RNA systems. The application of heteronuclear NMR experiments also addresses a broad range of detection time scales spanning picoseconds (Akke et al. 1997) to seconds (Lee et al. 2010), which enables probing of a many dynamic processes such as base flipping (Blad et al. 2005), interhelical motions in the interhelical twist (Zhang et al. 2007), sugar puckering transitions (Johnson and Hoogstraten 2008) and conformational rearrangement between conformers (Kloiber et al. 2011). In addition, specific NMR techniques have been developed that can focus on the entire structure and on particular sites within a molecule. The recent development of the non-uniform sampling technique, which enables rapid detection of multi-dimensional experiments, provides an excellent way for measuring structural rearrangements induced by binding of ligands dynamically (Schmieder et al. 1993; Tugarinov et al. 2005).

Although the application of heteronuclear NMR experiments significantly improves the spectra overlap, it may still be difficult to monitor individual nucleotides, particularly in the case of conformational heterogeneity. One way around this complication is observation of a molecule into which a single $^{13}$C- or $^{15}$N-labeled nucleotide has been incorporated by chemical synthesis (Kloiber et al. 2011). An unfortunate limitation of this approach is that the phosphoramidites are not commercially available thus requiring specialized chemical synthesis expertise. So it would be great advantages to develop such a commercially available approach that can be used to monitor the behavior of a single nucleotide in a complex system.
1.7 Application of $^{19}$F NMR in biological systems

Fluorine NMR has been used to probe the secondary structure of RNA for a long time (Horowitz et al. 1977; Marshall and Smith 1977; Gollnick et al. 1986; Chu et al. 1992; Kanyo et al. 1996; Sahasrabudhe and Gmeiner 1997; Arnold and Fisher 2000; Hammann et al. 2001; Olejniczak et al. 2002; Hennig et al. 2007; Puffer et al. 2009). The fluorine substituted bases include 5-fluorouracil (Arnold and Fisher 2000; Hammann et al. 2001; Olejniczak et al. 2002), 5-fluorocytosine (Puffer et al. 2009), 5-(trifluoromethyl)uracil (Gmeiner et al. 1991), 2-fluoroadenine (Scott et al. 2004). It can also be used to substitute the 2′-OH group in RNA molecules (Kreutz et al. 2005). Among the above molecules, the 5-F substituted pyrimidines and 2′-F substituted purine have been proved to not alter the thermostability of RNA molecules, and thus can be used in structural and dynamic studies of RNA complex (Kreutz et al. 2005; Puffer et al. 2009).

Compared to the traditional NMR experiments that observe $^1$H, the $^{19}$F nucleus is far more sensitive to its environment because it has 19 electrons, resulting in a greater chemical shift dispersion and environment-dependent change than observed for $^1$H, and thus simplifies spectral analysis. Substitution of a single $^{19}$F at the 5 position of a pyrimidine base by commercially available solid phase synthesis permits monitoring of the local environment of that nucleotide. NMR measurement of the $^{19}$F-labeled pyrimidine does not involve any two-bond H-F couplings, and three-bond H-F couplings are within the line width of the peak; thus no decoupling is necessary. It is noted that at higher field strength or for larger biomolecules, chemical shift anisotropy may contribute to peak broadening. However, the high sensitivity of the $^{19}$F nucleus (83.3%), which is similar to that of $^1$H and much greater than that of $^{13}$C (1.59%) or $^{15}$N (0.104%), makes it a good candidate for NMR observation. As a result, $^{19}$F NMR has been
widely used in the study of structure and dynamics of biological molecules, such as proteins (Larda et al. 2013) and nucleic acids (Kreutz et al. 2005; Puffer et al. 2009).

1.8 Summary of questions to be addressed

Research from Valadkhan and coworkers (Valadkhan and Manley 2001; Valadkhan et al. 2007; Valadkhan et al. 2009) revealed the catalytic activity by the protein-free human U2-U6 snRNA complex, which underscores the importance of the snRNA components in the splicing reaction. Comparison of the spliceosome with a mechanistic parallel, the self-splicing group II intron (Toor et al. 2008), suggests the folding of U2-U6 snRNA complex in order to bring the active sites into close proximity to each other (Steitz and Steitz 1993). Functional and structural studies suggest different conformational models of the U2-U6 snRNA complex. Each of the proposed folds comprises U2-U6 snRNA intermolecular Helices II and III 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. The famous models include the four-helix junction with the formation of U2 Stem I suggested by NMR investigation of a truncated yeast complex (Sashital et al. 2004) and the three-helix structure with Helix I brought up by genetic studies (Madhani and Guthrie 1992). Importantly, the two models depict the genetically conserved AGC triad in different pairing environments, which may have an impact on folding and function.

In order to understand the mechanisms of the splicing reactions, it is essential to probe the lowest energy conformation of the RNA complex in the absence of metal ion, RNA substrates, or protein components that may induce structural changes, as well as its ability for conformational
change of it. To address this, I used solution $^{19}$F NMR to investigate conformational features of a set of biophysically tractable constructs representing the human U2-U6 snRNA (and mutations thereof) \textit{in vitro} (sequence changes described in Chapter II). Results presented here 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. However, my results 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. These findings suggest that the energetic barrier for conformational change is low, consistent with the findings of Guo et al. (2009), thus suggesting the interconversion between the steps of splicing is possible. In order to verify the conformational interconversion between different conformers of the U2-U6 snRNA complex, I applied $^{19}$F-$^{19}$F EXSY experiments to the U2-U6 snRNA complex. The results confirmed the interconversion between different conformers in protein-free U2-U6 snRNA complex, suggesting that the RNA components alone are capable of rearrangements that may be important in catalysis of the splicing reactions.

The research presented in this dissertation illustrates studies of the structural and dynamic features of the junction region in the protein-free human U2-U6 snRNA complex. The flexible nature of this region has been speculated to assist in juxtaposing the critical elements during the splicing reaction. Thus this research provides the basis for understanding the mechanisms of spliceosome assembly and activity.
Chapter II

Secondary structural fold of the human U2-U6 snRNA complex

2.1 Introduction

As discussed in Chapter I, the complex formed between U2 and U6 snRNA is implicated in mediating catalysis in pre-mRNA splicing (Fabrizio & Abelson, 1990; Madhani & Guthrie, 1992; McPheeters & Abelson, 1992). Previous research reveals that the human U2-U6 snRNA complex binds several catalytically important metal ions at distant sites, implying the necessity of precise folding (Huppler et al., 2002; Sontheimer et al., 1997; Valadkhan & Manley, 2002; Yean et al., 2000; Yuan et al., 2007). Functional and structural studies suggest different conformational models of the U2-U6 snRNA complex. Each of the proposed folds comprises U2-U6 snRNA intermolecular Helices II and a U6 intramolecular stem loop (U6 ISL); the features that vary include sub-helices 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. A great deal of information was gained from the crystal structure of a spliced Group II intron (Toor et al. 2008), one of a large family of RNA transposable elements found in prokaryotes and eukaryotic organelles that perform the identical splicing reaction as that mediated by the spliceosome and which share a number of structural features in common with the U2-U6 snRNA complex. This structure reveals a tertiary interaction involving the CGC triad (analogous to AGC in the U2-U6 snRNA complex) with nucleotides from an inter-domain junction (analogous to GA from the ACAGAGA loop), an interaction that may be hindered by formation of Helix III. If a similar structure forms in the U2-U6 snRNA complex to facilitate splicing activity, the
precise position of the AGC triad will be important. Supporting this finding, a recent model derived from chemical structure probing provided evidence that Helix III does not form in human U2-U6 snRNA complex (Anokhina et al. 2013), which might be important for folding into a tertiary structure. These findings suggest that probing the structure of U2-U6 snRNA complex is necessary for understanding the active core of spliceosome.

Genetic studies in 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 (Madhani and Guthrie 1992). This model (Figure 2.1) was reinforced by mutational studies that implicated the requirements of the three-base-pair Helix Ib in at least one (Hilliker et al. 2007) and possibly both (Mefford and Staley 2009) cleavage steps. Additionally, results of cross-linking assays in yeast identified tertiary interactions only possible in the three-helix model (Ryan et al. 2004). On the other hand, findings by McPheeters and Abelson that formation of certain base pairs corresponding to Stem I was important for splicing in yeast suggests the possibility that the alternative four-helix model is active at some point (McPheeters and Abelson 1992). In human cells, however, interactions identified by genetic suppression assays were consistent only with formation of an extended U6 snRNA ISL and formation of U2 Stem I (Sun and Manley 1995) (Figure 2.2), characteristics of the four-helix model.

Biochemical studies by Moore and Sharp, using splicing substrates that contained a chiral phosphorothioate, demonstrated that both steps of splicing are strongly inhibited by one diastereomer, but not by the other, suggesting that the active site undergoes conformational change between steps (Moore and Sharp, 1993). Interestingly, most genetic studies in yeast support the formation of Helix Ib, associated with the three-helix model; in contrast, genetic
studies in human cells support the formation of the extended ISL and Stem I associated with the
four-helix model. Since all evidence indicates that the chemical mechanism of splicing is
identical in yeast and human systems (Moore and Sharp 1993), interconversion between these
two structures may be a possibility. However, no direct evidence of interconversion between
conformers has been shown for either system in in situ. It is thus tempting to speculate that the
differences in conformations observed in the yeast and human systems are related to differences
in experimental design, modulating effects of spliceosomal proteins, and differences in
energetics of stem formation by different RNA sequences, and/or observation of different events.

Conformational features of the U2-U6 snRNA complex have also been examined in vitro.
NMR investigation of a truncated yeast complex demonstrated that U6 ISL is extended to
include the AGC triad and U2 snRNA forms the intramolecular Stem I, therefore creating a four-
helix junction; no spectral evidence was reported for an alternative conformation (Sashital et al.
2004). However, a more 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). Thus it appears that the complex is capable of
forming multiple conformations in the region of the junction under different conditions.

Supporting these findings, Cao and Chen, using computational studies, have demonstrated
that both human and yeast U2-U6 complex can form multiple conformations (Cao and Chen
2006). The distribution between them can be affected by minor changes in the sequence,
presence of Mg$^{2+}$ and/or spliceosomal proteins. Consistent with this notion, single-molecule
Figure 2.1 Three-helix model of the protein-free human (h)U2-U6 snRNA complex 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.
Figure 2.2 Four-helix junction model of the protein-free hU2-U6 snRNA complex proposed by Sun and Manley (1995).
fluorescence data of a model yeast U2-U6 snRNA complex have given a strong evidence of interconversion between two folds attributed to four- and three-helix conformers that was highly dependent upon Mg\textsuperscript{2+} concentration (Guo et al. 2009).

In order to analyze conformational changes associated with catalytic activity, it is essential to know the lowest energy conformation in the absence of metal ion, RNA, or protein components that may induce structural changes. To address this question, solution NMR was used to investigate conformational features of a set of biophysically tractable constructs representing the human U2-U6 snRNA (and mutations thereof) in vitro (sequence changes described in the following Materials and Methods section). My results, as well as the biochemical structural probing data from Ravichandra Bachu in the Greenbaum laboratory, 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. However, the NMR results also identified a small but significant fraction of the U2-U6 snRNA complex forming an alternative family of conformations, which may represent the three-helix structure in equilibrium with the major conformation. These findings suggest that the energetic cost for conformational change is low and the junction region of the U2-U6 snRNA complex is quite flexible, which facilitates interconversion between the steps of splicing.

2.2 Materials and methods

2.2.1 Models for folding of the U2-U6 snRNA complex

Based upon proposed models for secondary structural folds of the yeast (Madhani and Guthrie 1992; Sashital et al. 2004; Burke et al. 2012) and the human (h)U2-U6 snRNA
complexes (Sun and Manley 1995), I generated proposed secondary structures representing the two alternative folds of the human U2-U6 snRNA complex: one including a four-helix junction in which the AGC triad is paired with opposing nucleotides at the base of the U6 ISL and the other describing a three-helix structure in which the AGC sequence pairs intermolecularly with U2 snRNA to form Helix Ib (Figures 2.1 and 2.2, respectively).

2.2.2 Formation of the U2-U6 snRNA complex

RNA fragments representing the pairing regions of human U2 and U6 snRNA sequences were designed with several modifications to minimize formation of undesirable self-paired complexes. Specifically, I replaced the hairpin loop of U6 snRNA ISL, GCGCA, with the yeast sequence GCAUA, changed the 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 (Figure 2.1). NMR evidence of formation of a U-U pair in a complex containing the native sequence in this region confirmed the formation of Helix II in either case (data not shown). In addition, two guanosines were added to the 5’ end of each strand for efficient in vitro transcription, as well as cytidines on the 3’ termini to optimize pairing. None of these changes were predicted by m-Fold (Zuker 2003) to induce any conformational change from the native fold. For 19F NMR experiments, a 5-fluoro-cytidine (5-19F-C) was introduced at U2 snRNA position C13 (Figure 2.3, red nucleotide, the 5-19F-C oligomers were purchased from Dharmaco). This site was chosen because it would reside within a double-stranded stem (the middle residue of the 5’ strand of Stem I) in the four-helix junction model or single-stranded in the three-helix structure (Zhao et al. 2013).
Figure 2.3 Demonstration of the position of $5^{-19}$F-C substitution in the two models of the protein-free human U2-U6 snRNA complex (Zhao et al. 2013).
I also designed three unimolecular samples; 1) in the first, by joining the U2 and U6 snRNA sequences 5’ to the U6 ISL regions and Helix I with a super stable ‘UUCG’ tetra loop; 2) in the second, by joining the U2 and U6 snRNA sequences 3’ to U6ISL and Helix II with UUCG; and 3) in the third, to focus on the junction region, I truncated the U6ISL to leave only 5 base pairs from the junction and the base pair near the loop, removed the ACAGAGA loop to facilitate the folding and to decrease the size (molecular mass) of the sample. Two extra G•C base pairs were added at the end of the sample for efficient in vitro transcription. The resulted unimolecular samples are approximate 70-nt. Figure 2.4 shows an example of the unimolecular sample combing described in example 3.

2.2.3 Design of the mutants

To validate the results obtained with the wild type junction region of the U2-U6 snRNA complex, I created mutants designed to stabilize or destabilize the formation of Stem I, thus favoring the formation of four-helix or three-helix structures respectively. In total I have designed three 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 (Figure 2.5) (Zhao et al. 2013); 2) substitution of the moderately thermostable UUUU tetraloop sequence of Stem I to the hyperstable sequence UUCG, and mutation of the top pair of Stem I from G-C to C-G to inhibit the formation of Helix Ib (Figure 2.6) (Zhao et al. 2013). Both of these mutations favor the formation of Stem I and the four-helix structure; 3) mutation of U2 snRNA G12 and G19 to A12/19 to disfavor the formation of Stem I, along with mutation of U6 snRNA G80C and C81G to disfavor the extended ISL and C61U to disfavor interactions between U6 ISL and the single-stranded loop in U2 snRNA. This latter set of mutations was designed to favor formation of the putative three-helix conformation (Figure 2.7) (Zhao et al. 2013).
Figure 2.4 Unimolecular sample representing human U2-U6 snRNA complex. The U6ISL was truncated and the ACAGAGA loop was removed. The two strands were connected by a super stable ‘UUCG’ tetra loop. And an ‘A’ (marked in red) was erroneously added to the loop sequence in this sample.
Figure 2.5 Sequence and proposed secondary structure of a mutation that favors the four-helix model (mutation 1) in which hU6 snRNA is paired with a mutated hU2 snRNA with two extra G-C/C-G base pairs to increase the stability of the Stem I. The extra G-C/C-G base pairs are in blue and the 5'-19F-C substitution is in red (Zhao et al. 2013).
**Figure 2.6** Sequence and proposed secondary structure of a mutation favoring the four-helix model (mutation 2). 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'-19F-C substitution is in red (Zhao et al. 2013).
Figure 2.7 (A) Sequence and proposed secondary structure of a mutation that favors the three-helix model (mutation 3). The mutated nucleotides from the original sequences are in blue and the 5'-19F-C substitution is in red. (B) Same sequence as (A) but with different 5'-19F-C substitution (red nucleotide) (Zhao et al. 2013).
In mutations 1 and 2, a 5'-19F-C was introduced at U2 snRNA position 14 and 13; in mutation 3, separate samples were created with 5'-19F-C in U2 snRNA positions 13 and 21, respectively.

To determine the effect of ACAGAGA loop on the conformation of the junction, I designed a mutated U6 snRNA sample so that the sequence forms complementary base pairs with U2 snRNA at the ACAGAGA loop region (mutation 4, Figure 2.8). The mutated U6 snRNA were paired with the same U2 snRNA with 5'-19F-C substitution as the original U6 snRNA sample (Zhao et al. 2014a).

Both hU6 and hU2 RNAs, as well as the mutant sequences without 19F-substituted nucleotides and the unimolecular constructs, 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, gel purified to separate transcription products, eluted from a gel slice of the desired full-length band using an electroeluter, precipitated, washed with a suitable buffer using 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, I measured migration of individual and paired strands on a non-denaturing gel. Equal amounts of purified U2 and U6 snRNA (or the U6 snRNA with mutated ACAGAGA region (mAGA) ) strands were heated to 70 °C for three minutes 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 on to a 12% non-denaturing gel and electrophoresed at
**Figure 2.8** Proposed four-helix junction model for mutated human U6 (mAGA) and U2 snRNA complex (mutation 4). The mutated nucleotides in U6 snRNA and their pairing partners in U2 snRNA were indicated in red. The substitution of $5'$-$^{19}$F-$C$ in U2 snRNA was also indicated in red (Zhao et al. 2014a).
100 mV for four hours at 4 °C. Following staining by Nuclistain (a colorimetric stain that interacts chemically and stoichiometrically with nucleic acid backbones), a single band was observed for each of the U6 samples alone, consistent with previous observations of self-folding (Zhao et al. 2013); the lane for the U2 sample showed diffuse folding (also previously observed), consistent with a relative lack of secondary structure. A single band was also observed in the lane with the combined strands, which migrated more slowly than individual U6 or U2 strands, and which was attributed to formation of the U2-U6 snRNA complex (Figure 2.9). I note that this approach does not provide sufficient resolution to distinguish between populations of the complex differing in local conformation in the junction region.

2.2.4 \(^1\)H NMR spectra

\(^1\)H NMR spectra were acquired on a Bruker Avance III 600 MHz spectrometer equipped with a cryoprobe. Quadrature detection was achieved using the States-TPPI method (Marion et al. 1989). Spectra were processed and assigned using Bruker Topspin and NMRPipe software (Delaglio et al. 1995). One-dimensional experiments were acquired with 3-9-19 watergate pulse sequence for water suppression. Two-dimensional spectra were apodized using a Gaussian function and zero filling was performed in both dimensions.

2.2.5 \(^19\)F NMR spectra

Two short oligomers containing 5-\(^19\)F-C in place of cytidine, 5’-UG\(^{5\text{F}}\)CUUUU-3′ and 5’-GG\(^{5\text{F}}\)CUUUUGGCCG-3′ were designed as controls in which the 5-\(^19\)F-C was predicted to be in a single- or double-stranded environment, respectively (Zhao et al. 2013). The 5-\(^19\)F-C oligomers including U2 snRNA and all the mutants were purchased from Dharmacon and deprotected according to their protocols. Unmodified sequences were prepared by \textit{in vitro}
Figure 2.9 Pairing of RNA oligomers representing hU2 and U6 snRNA analyzed by non-denaturing gel electrophoresis. Lanes 1, 3 and 6 show relative migration of fragments representing U6, mutated U6 (mAGA, mutation 4, described in Materials and methods) and U2 snRNA respectively; lane 2 and lane 4 shows annealing of the two strands with U2 snRNA (Zhao et al. 2013).
transcription using synthetic DNA templates and purified by gel electrophoresis as described above. The modified U2 snRNA was paired with a small excess (1.0:1.1) of the transcribed U6 snRNA by heating both strands to 70 °C and cooling slowly to room temperature as described above. The paired complex was then dried and resuspended in 95% H2O/5% 2H2O (Cambridge Isotope Laboratories) for NMR experiments. Pairing of the two strands was verified on a non-denaturing gel. NMR samples had a concentration of approximately 0.35 mM RNA in 5 mM NaPi, pH 6.5, 50 mM NaCl, and 0.1 mM EDTA.

19F NMR spectra with 1H-coupling were acquired on a Varian INOVA 500 MHz spectrometer equipped with a broadband probe. Acquisition parameters of 19F NMR experiments were as follows: spectrometer frequency 470.220 MHz, spectral width 61633.3Hz, 19F excitation pulse length 15 µs, number of scans 12000-24000, acquisition time 1.063 s, and relaxation delay 1.5 s. Data were processed by Varian VNMRJ software with a line broadening factor of 30 Hz. Spectra were referenced by external neat trifluoroacetic acid (-78.5ppm).

2.3 Results

My first attempt was to use solution NMR on truncated unimolecular samples representing human U2-U6 snRNA complex (represented sequence shown in Figure 2.4). The goal was to identify the resonances in the junction and then probe their structural and dynamic features. Although the complex was truncated to remove the ACAGAGA loop and several base pairs at U6 ISL, the size of the molecule is still relatively large, and likely too flexible, to obtain clear 1H NMR spectra. As a result, the relaxation of the signal is too fast and the overlap and line
Figure 2.10 Aromatic-anomeric (base-H1') region of a NOESY spectrum of the unimolecular sample shown in Figure 2.4. Spectrum was acquired in D$_2$O at 25°C with 250 ms mixing time.
broadening of the spectra is significant (Figure 2.10), which makes it very difficult to assign the
peaks and thus prompted an alternative approach.

Since Stem I only appears in the four-helix junction model, monitoring its presence or
absence provides a straightforward “handle” to distinguish between the two proposed
conformation models for the human U2-U6 snRNA complex. I thus pursued $^{19}$F NMR studies to
differentiate between the two proposed models. The fluorine chemical shift of a single 5-$^{19}$F-
cytidine (5-$^{19}$F-C) at position C13 of U2 snRNA was measured as a reporter for single- or
double-stranded status. This substituted nucleotide residue is base paired if Stem I forms or
single-stranded if Stem I does not form (Figure 2.3).

$^{19}$F spectra were acquired for two controls containing 5-$^{19}$F-C, one within a single stranded
fragment and the other as part of a stem in a short stem loop, each with a sequence context
matching that of the equivalent region of U2 snRNA, to provide reference spectra for the two
alternative conformations (Figure 2.11). The predominant $^{19}$F resonance peaks for the single 5-$^{19}$F-C in each oligomer were at -165.1 ppm and -167.4 ppm for the single- and double-stranded
controls, respectively, values that correspond with previous measurements (Puffer et al. 2009).
Small peaks (~5-7% of the total) at -167.1 and -165.8 ppm in the controls were noted, which
suggests some distribution between paired and unpaired states in each case (Zhao et al. 2013).

A spectrum of exchangeable $^1$H of the double-stranded control at 10 °C revealed four imino
protons in 12-15 ppm range, consistent with the anticipated four base pairs in the stem (Figure
2.12). A spectrum of exchangeable $^1$H for the single stranded control, however, did not reveal
any hydrogen bonded imino protons, suggesting that any protons involved in base pairing in the
single-stranded control were in rapid exchange with the solvent at 10 °C and did not form a
Figure 2.11 One-dimensional $^{19}$F NMR spectra of control RNA oligomers. (A) Single stranded control oligomer; (B) Double stranded control oligomer. The sequences were indicated by each spectrum and the position of $5'$-$^{19}$F-C substitution is shown in red (Zhao et al. 2013).
Figure 2.12 Imino region of one-dimensional $^1$H NMR spectrum of the double-stranded control RNA oligomer (5'-GG$^5$C CUUUGGGCG-3') at 15°C in 95%H$_2$O/ 5% $^2$H$_2$O. The sequence of the sample can be found in Figure 2.11B. Each of the pronounced peaks corresponds to one of the four anticipated Watson-Crick base pairs of the sequence; the very small peaks correspond to alternative base pairing schemes.
stable base-paired structure. It was also noted that there were two small peaks in the $^1$H spectrum of the double-stranded control, consistent with the conformational heterogeneity observed in the $^{19}$F spectrum.

The proton-coupled one dimensional $^{19}$F spectrum of the U2-U6 snRNA complex acquired at 25 °C displayed a large peak at approximately -167.8 ppm, suggesting that the 5-19F-C at position 13 is predominantly within a duplex (Figure 2.13) (Zhao et al. 2013). This peak appears to include a small small “shoulder” of a second peak with a very similar chemical shift (also in the range consistent with a double-stranded environment), suggesting conformational heterogeneity, perhaps associated with different orientations of the stem loop. However, there was also a very broad (apparently multi-component) peak with a center at approximately -165.4 ppm, with an area ~14% of the total, implying a family of conformations with single-stranded 5-19F-C at position 13 in this fraction. Whether these conformers represented in this broad peak correspond to the three-helix conformation proposed by Madhani and Guthrie (1992), dissociation of the intramolecular base pairing, or both, cannot be distinguished from these data alone.

In order to distinguish between the two alternatives and help identify the conformation associated with the single-stranded broad peak, I designed several mutations to favor either the four-helix or the three-helix structure, respectively. In the first mutation, formation of putative Stem I in U2 snRNA is strengthened by addition of two base pairs to either side of the 19FC-G base pair (Figure 2.5). Because the chemical shift of the 19F-substituted nucleotide is sensitive to the sequence context, the identity of the neighboring nucleotides was maintained. The one-dimensional $^{19}$F spectrum of this mutated complex displayed a major peak at approximately -167.5 ppm and a minor peak (~5%) centered at approximately -165.8 ppm (Figure 2.14A) (Zhao et al. 2013). The decrease in the area of the peak with single-stranded chemical shift from 14% to
Figure 2.13 One dimensional $^{19}$F NMR spectrum of human U2-U6 snRNA complex at 25°C in 95%H$_2$O/ 5% $^2$H$_2$O. The spectrum was acquired by Varian NOVA 500 MHz spectrometer and the acquisition parameters were stated in Materials and methods (Zhao et al. 2013).
~5%, significant narrowing of this peak as compared with the very broad (and perhaps combined) resonance of the wild type complex, as well as precise overlap with the single-stranded peak observed as the minor conformation of the stem-loop control, suggest that the downfield-shifted single-stranded peak of the mutant sequence with the extended Stem I is a different species from that seen in the one-dimensional $^{19}F$ spectrum of the wild type U2-U6 snRNA complex. Based upon similarities with the downfield peak in the double-stranded control, it is anticipated that this peak also represents a minor species associated with transient opening of Stem I, and that this species is also likely to contribute (as the upfield region) to the combined single-stranded broad peak observed in the spectrum of the wild type complex.

To investigate this possibility further, I designed and tested another mutant in which I 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 is predicted to 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 2.6). For this mutated complex, I observed no peak at the single strand chemical shift (Figure 2.14B) (Zhao et al. 2013). These findings support our hypothesis that at least some components of the broad peak observed in the wild type complex represent an alternative conformation and the broad peak does not simply represent “breathing” of Stem I or dissociation of the two strands in the complex. It is also noted that, in both mutations, the double-stranded peak is narrower than that of the wild type U2-U6 snRNA complex, which can be attributed to the increase of thermal stability in Stem I.

Formation of Helix Ib has not been previously demonstrated in the human U2-U6 snRNA complex. In order to verify the ability of the human complex to form a three-stemmed
Figure 2.14 (A) One dimensional $^{19}$F NMR spectrum of the extended U2 snRNA pairing with U6 snRNA (mutation 1). (B) One dimensional $^{19}$F NMR spectrum of the mutated UUCG tetra loop U2 snRNA pairing with U6 snRNA (mutation 2). Both spectra were acquired at 25°C in 95%H$_2$O/ 5% $^2$H$_2$O. Representation of a component corresponding to the proposed three-helix conformation is greatly diminished (A) or essentially nonexistent (B) for these mutant complexes (Zhao et al. 2013).
conformation analogous to that proposed in yeast, I also designed a mutated complex to favor Helix Ib formation. This was accomplished by disrupting base pair formation in the lower region of the U6 ISL (G80C, C81G of U6 snRNA) and Stem I (C12A, G19A) (Figure 2.7). Presence of an upfield-shifted $^{19}$F peak at -168.1 ppm (Figure 2.15A) is consistent with a double-stranded environment for the substituted nucleotide in the mutant complex shown in Figure 2.7A.

Moreover, presence of several downfield-shifted peaks in the range of -164.6 ppm (Figure 2.15B) implies a somewhat heterogeneous single-stranded environment for a substituted nucleotide in a complex with the same sequence mutations shown in Figure 2.7B. The calculated distribution of the three-helix conformation in Figure 2.7 was 43.1% by the double-stranded 5-$^{19}$F-C (the peak at -168.1 ppm in Figure 2.15A), consistent with the value of 42.3% by the single-stranded 5-$^{19}$F-C (the peaks at -163.1 and -164.6 ppm in Figure 2.15B). These findings suggest that a three-helix structure containing Helix Ib is stable and can be favored by disrupting U2 Stem I and lower U6 ISL.

In order to determine whether the large internal ACAGAGA loop has any impact on the distribution of paired vs. unpaired status of the nucleotides participating in Stem Loop I, I also acquired spectra of the mutated U2-U6 snRNA complex in which the ACAGAGA loop was replaced by a complementary stem (Figure 2.8). The well overlap of the 1D $^1$H NMR of the native sequence and the mutated U6 snRNA forming complementary base pairs with U2 snRNA in the ACAGAGA region, suggests that the two complexes have similar overall structure, which implicates the formation of helix I and III (Figure 2.16).

$^{19}$F spectra at 25 °C displayed a similar peak pattern as that of the wild type U2-U6 snRNA complex, i.e. a large double peak centered at -167.8 ppm with two resonance peaks near each
Figure 2.15 One dimensional $^{19}$F NMR spectrum of a mutation that favors the three-helix model (mutation 3). (A) Spectrum of a sample with 5-$^{19}$F-C substitution in double-stranded environment (Sequence in Figure 2.7A). (B) Spectrum of a sample with 5-$^{19}$F-C substitution in single-stranded environment (Sequence in Figure 2.7B). Both spectra were acquired at 25°C in 95%H$_2$O/ 5% $^2$H$_2$O. Appearance of multiple resonance peaks suggests considerable heterogeneity in single-stranded conformers.
Figure 2.16 One dimensional $^1$H NMR spectra of the human U2-U6 and U2-mAGA snRNA complex at 20°C in 95%H$_2$O/ 5% $^2$H$_2$O. The blue spectrum displayed the exchangeable imino protons in human U2-U6 snRNA complex and the red one is for U2-mAGA snRNA complex, as marked in the figure. The number of scans was 1200 and the relaxation delay was set to 2s. Similarities between the resonance peak locations of the imino protons for the two complexes implies that the presence or absence of the ACAGAGA loop was not associated with any notable difference in base pairing patterns, and therefore of global structure, of the two complexes.
Figure 2.17 One dimensional $^{19}$F NMR spectrum of human U2 and mAGA complex (mutation 4) at 25$^\circ$C in 95%H$_2$O/ 5% $^2$H$_2$O. The distribution of peak intensities was essentially as that seen for the native complex containing the ACAGAGA loop (Figure 2.13). However, the broad peak corresponding to the single-stranded region for the $^{19}$F-C appears slightly narrower in this spectrum; the significance of this observation is not clear (Zhao et al. 2014a).
other and a smaller but broad peak centered around -165.4 pm with a distribution of ~14.0% of total (Figure 2.17). The small “shoulder” within the double-stranded peak suggests conformational heterogeneity within the mutant complex. We also observed that the single-stranded peak is a little narrower than that of the wild type sequence, likely to be the result of decreased flexibility from the complementarity in the ACAGAGA loop. Although one may also anticipate that the double-stranded peak is narrower because of the increased thermal stability, I observed similar linewidth as that for the wild type sequence. These data imply that the conformation in the junction region is independent on the formation of ACAGAGA loop (Zhao et al. 2014a).

**Effect of Mg**\(^{2+}\). I also repeated the \(^{19}\)F NMR experiment in the presence of 5mM MgCl\(_2\). The result showed that both conformations were present and the fraction of single-stranded peak increased slightly to 17% at 25°C, suggesting a minor shift in distribution induced by Mg\(^{2+}\) (Figure 2.18) (Zhao et al. 2013). There is a shift for the peak of double-stranded 5-\(^{19}\)F-C, which may be due to changes in tertiary fold or simply be the electrostatic effects of Mg\(^{2+}\). It is also noted that the single-stranded peak in the sample with 5mM MgCl\(_2\) was a little sharper than the same complex without Mg\(^{2+}\), which is likely to be the stabilization effect from Mg\(^{2+}\) on the tree-helix conformations with single-stranded 5-\(^{19}\)F-C.

### 2.4 Discussion

Results of \(^{19}\)F NMR demonstrate 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
Figure 2.18 Overlaid one dimensional $^{19}$F NMR spectra of the human U2-U6 snRNA complex with (upper trace) and without (lower trace) 5mM Mg$^{2+}$ at 25°C in 95%H$_2$O/5% $^2$H$_2$O. Each spectrum was labeled in the figure. A slight increase in the fraction of a single-stranded component (from ~14% to ~17%) is observed in the presence of Mg$^{2+}$ (Zhao et al. 2013).
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. 2004) from
NMR studies of protein-free yeast RNA fragments. The NMR results also suggest that the
predominant four-helix junction is in equilibrium with a lesser fraction of the three-helix
conformation; features of this model have been proposed for a protein-free yeast complex in
vitro (Burke et al. 2012) and during the splicing process of yeast in situ (Madhani and Guthrie

Consistent with these different models, it is possible 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 suggest 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 (Query and Konarska 2004; Liu et al.
2007).

In addition, computational studies by Cao and Chen (2006) demonstrated a propensity for
conformational heterogeneity of protein-free U2-U6 snRNA complexes from both yeast and
human sequences with a truncated Helix I/III. They observed a different distribution for the yeast
and human sequences, which they have attributed to different sequences near the junction region.
The appearance of multiple low-energy folds may reflect the potential for rearrangement of the
junction region during splicing activity. Quantification of conformational heterogeneity 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 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; Karunatilaka and Rueda 2014), suggesting the propensity for ion-dependent conformational change. Specifically, two major conformations were attributed to four-helix and three-helix models, respectively, with an obligatory intermediate. The equilibrium distribution they found was very different from that predicted by Cao and Chen (2006) using the computational analysis, although this observation may be the result of differences in the lengths and sequences of the stems the two teams used, as well as for the differences in data obtained from experimental and statistical approaches, respectively.

The computational study also found a different distribution of folds for the human complex than for that of yeast, specifically favoring formation of a four-way junction, which they attributed to the sequence differences in the human complex junction (Cao and Chen 2006). It is noted that several minor conformations identified by their calculations were associated with the presence of highly truncated Helix I/III in the human U2 and U6 snRNA sequences tested.

My approach was to exploit different chemical shifts of 5-$^{19}$F-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 (Horowitz et al. 1977; Marshall and Smith 1977; Gollnick et al. 1986; Chu et al. 1992; Kanyo et al. 1996; Sahasrabudhe and Gmeiner 1997; Arnold and Fisher 2000; Hammann et al. 2001; Olejniczak et al. 2002; Hennig et al. 2007; Puffer et al. 2009) to demonstrate the presence of an alternative conformation of U2-U6 snRNA.
complex conclusively. The advantages of observing the $^{19}$F nucleus by NMR include (1) 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; (2) the 100% natural abundance of $^{19}$F nucleus and its very high NMR sensitivity, providing high signal to noise ratio in NMR experiments; (3) Because substitution of $^{19}$F nucleus is usually used to replace $^1$H in the molecule, $^{19}$F NMR experiments usually do not require water depression or $^1$H decoupling, which reduces the signal lost during NMR acquisition. Along with these advantages, C13 of U2 snRNA was specifically targeted, 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 allowed 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 suggests the formation of a four-helix junction as the predominate conformation, 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 there are alternative conformations present in the complex, which may be in equilibrium. NMR data verifying exchange between conformations in which the 5-$^{19}$F-cytidine is located in different environments strongly suggest a time scale of several hundreds of milliseconds, which is far slower than that expected for transient opening and closing of the base pairs (more discussion about the conformational exchange can be found in chapter III). $^{19}$F NMR data for the mutant 3 sequence, designed to favor formation of a three-helix model (Figure 2.7), indicated that the human U2-U6 snRNA complex is indeed capable of forming Helix Ib. Therefore, while it is likely that part of the broad peak observed for the original sequence
corresponds to “breathing”, my data are consistent with the ability to adopt the three-helix conformer. The low free energy difference between conformations suggests that the human U2-U6 snRNA complex adopts alternative structures under different conditions, perhaps stabilized by specific contacts favoring the formation or presentation of different active sites associated with the two steps of splicing.

Results from enzymatic structure mapping of the human U2 and U6 snRNA complex, performed by Ravichandra Bachu in the Greenbaum laboratory, were consistent with formation of a four-helix junction as the predominate conformation (Zhao et al. 2013). Dr. Bachu also observed some features that can be attributed to the alternative three-helix model, suggesting the possibility of equilibrium between the major fold and a minor coexisting conformation. However, a recent model derived from the chemical structure probing of human spliceosomal snRNAs in cells (Anokhina et al. 2013), suggests the three-helix structure is at the core of active spliceosome. Although some level of ambiguity is present in both structure probing data (although not noted in the paper by Anokhina et al.), it is possible that the different conformation observed is related to the binding of spliceosomal proteins, which may favor formation of the three-helix fold.

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 (Query and Konarska 2004; Liu et al. 2007). 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. 2006).

*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$^{2+}$. I also investigated whether Mg$^{2+}$ induced conformational change in the human sequence as was found by Guo et al. (2009) in the yeast sequence by $^{19}$F NMR. The data indicated a small increase in the three-helix structure, which may occur in the protein-free system as a result of formation specific interaction upon binding with the metal ion, as suggested by the crystal structure of Group II intron (Toor et al. 2008), thus stabilized the three-helix structure. However, the small increase I noted (from ~14% to ~17% of the total) is substantially less than the very great shift observed by FRET techniques by Guo et al. (Guo et al. 2009). As noted by the authors, part of the reason for the dramatic Mg$^{2+}$-dependent change observed by these authors may be the result of unraveling by the short and thermally unstable Helix III that may or may not be related to a change in junction structure.

Based upon the relative population of the two conformers obtained from the distribution of $^{19}$F NMR resonance peaks at 25 °C, I calculated a $\Delta G$ of -4.6 kJ/mol for formation of the four-helix structure from the three-helix structure in the absence of Mg$^{2+}$ and -4.0 kJ/mol in the presence of 5 mM Mg$^{2+}$, consistent with a very low interconversion cost that would favor the changes necessary to form different active sites for each of the two steps in the context of the active spliceosome. Since it is likely that some component of the broader peak is the result of
“breathing”, that is, transient opening and closing of the Stem I helix, assignment of the entire broad peak, which represents ~14% of the total, to the alternative conformation may be an overestimate. If, in fact, the alternative conformation only reflects a lesser population, then the actual ΔG value would be somewhat greater. Overall, the low free energy difference may facilitate conformational change between steps of splicing.

In summary, my data show that the human U2-U6 snRNA complex adopts two conformations with an apparent low free energy difference between conformations. Such a finding is energetically consistent with interconversion alternative structures under different conditions, perhaps stabilized by specific contacts favoring the formation or presentation of different active sites that facilitate the two steps of splicing.
Chapter III

Probing chemical exchange between RNA conformers by $^{19}$F NMR spectroscopy

3.1 Introduction

The ability of RNA molecules undergoing conformational interconversion between alternative folds is essential for their roles in multiple events in gene expression (Dethoff et al. 2012). Therefore, characterizing the kinetics and thermodynamics associated with conformational rearrangement of functional RNA will facilitate the study of RNA-mediated activity.

Dynamic motions observed for RNA molecules span a wide range of timescales (Bothe et al. 2011) from picoseconds (Akke et al. 1997) to hundreds of seconds (Zarrinkar and Williamson 1994). Rapid motions, including librations, interhelical motions, and base flipping, have a significant impact on binding small molecules, proteins, or other RNA molecules, e.g. Eldho et al. (2007). However, interconversion between alternative conformations, including those from lowest energy states to higher energy states and secondary structural transitions (Bothe et al. 2011), which typically have timescales in the millisecond to second range, are of great interest as they are directly implicated in biological processes such as ribozyme activity (Eldho and Dayie 2007).

Kinetics of the RNA folding landscape over a wide range of time scales has been analyzed by a number of experimental methods. Analysis of kinetics of the forward reaction induced by external stimuli such as light (Wenter et al. 2005) or addition of an ion (Hammann et al. 2001) or ligand (D'Souza and Summers 2004) has been measured by hybridization with complementary

However, measurement of individual lifetimes associated with interconversion between alternative RNA folds in equilibrium in the bulk phase is more difficult as a result of relatively short lifetimes as well as the reversible nature of the interconversion. Solution NMR offers an excellent approach for investigation of dynamics of such exchange reactions because the timescale (ms → sec) is within the NMR detection range. For example, the dynamics of a bi-stable RNA were probed by longitudinal exchange and Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion NMR spectroscopy monitoring a single $^{13}$C-methyl label (Kloiber et al. 2011), and recent work from Al-Hashimi and coworkers discovered a low abundance (~0.4%) transient RNA conformer with a long lifetime (~2 ms) by use of $^{13}$C and $^{15}$N R1$\rho$ relaxation dispersion experiment (Lee et al. 2014). However, because of the ease of incorporating a single $^{19}$F-labeled nucleotide by commercial chemical synthesis (as compared with the relative technical challenge associated with a single $^{13}$C label), and the demonstrated lack of structural or energetic perturbation associated with the $^{19}$F as compared with the methyl group (Kreutz et al. 2005; Puffer et al. 2009), as well as the greater sensitivity of the $^{19}$F label, monitoring of $^{19}$F has unique advantages.

In this chapter, I applied a method using $^{19}$F-$^{19}$F EXSY (EXchange SpectroscopY) (Jeener et al. 1979) experiments to investigate the dynamics of the conformational interconversion between RNA conformers. EXSY experiments utilize identical pulse sequence as that used in NOESY (Nuclear Overhauser Effect SpectroscopY) but with different purpose. A NOESY spectrum is used to detect through space correlations via Nuclear Overhauser cross relaxation between
nuclear spins during the mixing period. While EXSY spectra measure the chemical exchange between different conformers by observing the intensities of the exchange NOEs from a series of mixing times. An EXSY experiment is usually used for detecting the slower dynamics ranging from ten milliseconds to one second timescale (Bothe et al. 2011). A similar approach with $^{19}$F NMR has been used to analyze conformational exchange between inorganic beryllium fluorides on a millisecond timescale (Feeney et al. 1968), but it is a new approach to study equilibrium interconversion of RNA molecules.

To assess the feasibility of this approach, I first probed a model bi-stable RNA molecule that adopts two conformations in the equilibrium state, and used $^{19}$F-NMR to observe exchange between environments of a single substituted $^{19}$F-pyrimidine incorporated within the RNA oligomer. I followed this with application to the human U2-U6 snRNA complex to demonstrate and attempt to measure the equilibrium interconversion between conformers, a measurement that would likely be complicated with traditional $^1$H, $^{13}$C and $^{15}$N NMR because of the size of the sample.

### 3.2 Materials and methods

#### 3.2.1 Sample preparation

The RNA oligomer 5´-GAAGGCAACUUCGG($^{5}$FU)UG-3´, into which a single $^{5}$-19F-uridine (U) was incorporated, was purchased from Dharmacon and deprotected according to their protocols. The sequence was modified from that used by Kreutz et al. (Kloiber et al. 2011) to form shorter stems in each of the two major conformations and thus to facilitate more rapid interconversion. The gel purified bi-stable sample was heated to 90 °C, rapidly cooled on ice for 5 min, dried, and re-suspended in 95% H$_2$O/5% $^2$D$_2$O (Cambridge Isotope Laboratories). The
buffer used for NMR experiments contained 5 mM NaPi (pH 6.5), 50 mM NaCl, and 0.1 mM EDTA. Sample concentration was approximately 0.52 mM.

For the U2-U6 snRNA complex, the same sample was used as in Chapter II, although with an increase of RNA concentration to approximately 0.55 mM, in the same buffer condition.

### 3.2.2 Acquisition parameters

**1D $^{19}$F NMR:** One-dimensional $^{19}$F NMR experiments on the bi-stable model molecule were performed on a Varian INOVA 500 MHz (spectrometer frequency is 470.220 MHz for $^{19}$F) spectrometer equipped with a broadband probe with the same parameters as used in Chapter II except the relaxation delay was 7.5 s, the number of scans was 8000 and experiment time was ~19h.

**Measurements of Spin-lattice and Spin-spin relaxation time:** The Spin-lattice and Spin-spin relaxation time of the bi-stable model molecule were performed on the same Varian INOVA 500 MHz spectrometer as above. The Spin-lattice relaxation time was measured using Inversion recovery approach (Hahn 1949; Vold et al. 1968). The 90° pulse was 15µs. Relaxation delay was 20s. Seven recovery delays (0.0625s, 0.125s, 0.25s, 0.5s, 1s, 2s, and 4s) were used as the variable time to return to equilibrium. The total time of acquisition was approximate 120 hours.

The Spin-spin relaxation time was measured using CPMG pulse sequence \(90 - [\tau_{cp} - 180 - \tau_{cp}2n]\) (Carr and Purcell 1954; Meiboom and Gill 1958). The 90° pulse was 15µs and relaxation delay was 2.5s. Spin-echo cycle time \(\tau_{cp}\) was set to 1ms. Seven different CPMG relaxation times (4ms, 8ms, 16ms, 32ms, 64ms, 128ms and 256ms) were used for each experiment. Data were processed and corresponding relaxation times were calculated using Varian VNMRJ software. The values are reported as the mean of the repeated experiments.
**2D $^{19}$F-$^{19}$F EXSY experiments**: Two-dimensional $^{19}$F-$^{19}$F EXSY experiments of bi-stable sample were performed on the same Varian INOVA 500 MHz spectrometer equipped with a broadband probe as used for 1D $^{19}$F NMR experiments. EXSY spectra with different mixing times ranging from 2 to 760 ms were collected at 25°C. $^{19}$F excitation pulse length was 15 μs, spectral width was 10527.7 Hz, acquisition time was 0.195 s, and relaxation delay was 2.5 s. The number of points was 4096 x 200. The number of scans was 200 and experiment time was ~24h.

Two-dimensional $^{19}$F-$^{19}$F EXSY experiments of human U2-U6 snRNA complex spectra were acquired on a Bruker 600-MHz spectrometer equipped with dedicated fluorine CyroProbe (spectrometer frequency of $^{19}$F is 564.603 MHz) at 25 °C. $^{19}$F excitation pulse length was 10.42 μs, spectral width was 14,115.1 Hz, acquisition time was 0.036 s, and relaxation delay was 1.8 s. The number of points was 1024x64, and the number of scans was 256.

**Fitting of data**: The intensity of the auto (diagonal) peaks for the two states and the exchange peaks between them in 2D $^{19}$F-$^{19}$F EXSY experiment can be expressed by the following equations (Farrow et al. 1994; Latham et al. 2009):

\[
I_{SS}(T) = I_5(0)\left\{-\left(\lambda_2 - a_{11}\right)e^{-\lambda_1 T} + \left(\lambda_1 - a_{11}\right)e^{-\lambda_2 T}\right\} / (\lambda_1 - \lambda_2) 
\]

\[
I_{DD}(T) = I_5(0)\left\{-\left(\lambda_2 - a_{22}\right)e^{-\lambda_1 T} + \left(\lambda_1 - a_{22}\right)e^{-\lambda_2 T}\right\} / (\lambda_1 - \lambda_2) 
\]

\[
I_{SD}(T) = I_5(0)\left\{a_{21}e^{-\lambda_1 T} - a_{21}e^{-\lambda_2 T}\right\} / (\lambda_1 - \lambda_2)
\]

\[
I_{DS}(T) = I_5(0)\left\{a_{12}e^{-\lambda_1 T} - a_{12}e^{-\lambda_2 T}\right\} / (\lambda_1 - \lambda_2)
\]

where $a_{11} = R_s + k_{SD}, a_{21} = -k_{DS}, a_{22} = -k_{SD}, a_{22} = R_D + k_{DS},$ and $\lambda_{1,2} = 1/2 \left\{ (a_{11} + a_{22}) \pm \sqrt{(a_{11} - a_{22})^2 + 4k_{SD}k_{DS}} \right\}^{1/2}$. $R_s$ and $R_D$ represent the relaxation rate of single-stranded and double-stranded conformations, respectively. $k_{SD}$ and $k_{DS}$ are kinetics of the exchange.
A simultaneous non-linear least squares fit to Equations (1)-(4) was applied to extract the exchange rates. Fitting of the curve was achieved in collaboration with Dr. Ranajeet Ghose using Prism 6 suite (GraphPad Software Inc.) with in-house written scripts.

### 3.3 Results

The RNA sequence 5’-GAAGGCAACUUCGG(5FU)UG-3’ adopts two conformations characterized by GCAA or UUCG tetraloops, in which the 5'-19F labeled uridine (U) is located in a single- or double-stranded environment, respectively (Figure 3.1A and B). Observation of exchangeable imino 1H peaks corresponding to the G•A and G•U pairs in the proposed tetraloops confirms with formation of the two conformations (Figure 3.1C) (Zhao et al. 2014b).

The two environments result in 19F peaks of chemical shifts at -165.2 ppm and -169.0 ppm, respectively, assigned according to published values (Puffer et al. 2009) with a distribution of 0.57: 1 at 25°C (Figure 3.2) (Zhao et al. 2014b). Observation of sharp peaks at the predicted locations implies that interconversion between conformers is in the slow exchange realm of the NMR timescale. To characterize the dynamics of the bi-stable sample and provide necessary information for performance of the EXSY experiments, measurements were made of the spin-lattice relaxation time (T1) and spin-spin relaxation time (T2) for each of the conformers. T1 is useful for setting up delay between scans (usually 5*T1) to allow the system to equilibrate; T2 relates to the linewidth at half-height (Δν1/2) of the NMR signal. The data for the peak corresponding to the labelled nucleotide in the single-stranded region, T1 = 0.82 ± 0.02s, and for the peak corresponding to the labelled nucleotide in the double-stranded region, T1 = 0.81 ± 0.01s, are consistent with the size of the sample. These data suggest that the maximum mixing time of the EXSY experiments (which should be less than T1) is approximate 800 ms before the
signal relaxes. Measurements showed that $T_2 = 0.020 \pm 0.002\text{s}$ for the single-stranded peak, and $T_2 = 0.025 \pm 0.001\text{s}$ for double-stranded peak; these values are much less than values of $T_1$ for the corresponding peaks, indicating relaxation from chemical exchange between the conformers.

In order to characterize the kinetics of the interconversion, time-dependent $^{19}\text{F}-^{19}\text{F}$ exchange was measured in a series of 2D EXSY spectra (Jeener et al. 1979; Farrow et al. 1994; Latham et al. 2009) with mixing times ranging from 2 ms to 760 ms (Figure 3.3) (Zhao et al. 2014b). These spectra exhibit a build-up and decay of cross peaks (Figure 3.4). We observed no evidence of any intermediate between the two conformers in either $^1\text{H}$ or $^{19}\text{F}$ spectra, which would have been detected by additional peaks or relatively broader peaks with respect of the molecular weight of this molecule. I therefore applied a two-state exchange model to analyze the rates of exchange between each of the major conformations, from which we obtained the values of $k_{SD} = 24.2 \pm 1.8 \text{s}^{-1}$, $k_{DS} = 13.1 \pm 1.0 \text{s}^{-1}$, where $k_{SD}$ is the rate constant of interconversion from single-stranded 5-$^{19}\text{F}$-U to double-stranded 5-$^{19}\text{F}$-U and $k_{DS}$ is the rate constant of the reverse reaction. The equilibrium constant $K_{eq} = k_{DS}/k_{SD} = 0.54 : 1$, and the exchange rate $k_{ex} = k_{SD} + k_{DS} = 37.3 \pm 2.8 \text{s}^{-1}$ (Zhao et al. 2014b).

The calculated $K_{eq}$ value is in close agreement with the value measured from a fully relaxed condition in a 1D $^{19}\text{F}$ spectrum ($I_S/I_D = 0.57 : 1$). Also, the rates extracted imply that the interconversion between the two states was cooperative. This latter finding is consistent with the result by $^{13}\text{C}$ NMR studies on a related sequence (Kloiber et al. 2011), which (as in this case) did not display any evidence of a fully open single-stranded state or other intermediates. These findings indicate that the $^{19}\text{F}-^{19}\text{F}$ EXSY experiment is a valid approach to measure the kinetics of dynamic exchange between RNA conformations in equilibrium.
Figure 3.1 (A & B) Proposed secondary structures of the bi-stable sequence. The 5'-F labeled uridine (U, in red) was in a single- or double-stranded environment, depending upon the fold. (C) One-dimensional $^1$H spectrum of the bi-stable sample at 10°C. The imino peaks correspond to the G•A and G•U base pairs in each conformation were labeled in the figure.
Figure 3.2 One-dimensional $^{19}$F spectrum of the bi-stable sample at 25°C. The downfield peak at -165.2 ppm corresponds to the single-stranded environment of the $^{19}$F-U in the structure in Figure 3.1A and the upfield peak at -169.0 ppm corresponds to the double-stranded environment of the $^{19}$F-U in the structure in Figure 3.1B (Zhao et al. 2014b). Resonances were referenced with external neat trifluoroacetic acid (TFA) to -78.5 ppm and assignments were obtained according to Puffer et al. (2009).
Figure 3.3 Two-dimensional $^{19}$F-$^{19}$F EXSY spectra of the bi-stable sample at mixing times of 2 ms, 10 ms, 100 ms and 700 ms, as marked. Auto- and cross peaks, corresponding to the steady state of each conformation and the interconversion between them are labeled in the figure. The intensity of auto peaks decreases with the increase of mixing time because of relaxation of the magnetization. In contrast, the intensity of the cross peaks undergo a build-up and decay (see Figure 3.4) as a result of the chemical exchange (Zhao et al. 2014b).
**Figure 3.4** Plot of the intensities of exchange peaks vs. mixing times (2–760 ms) from which exchange values were calculated (Zhao et al. 2014b). $I_{DD}$ and $I_{SS}$ represent intensities of the auto peaks for double-stranded 5-$^{19}$F-U and single-stranded 5-$^{19}$F-U, respectively; $I_{SD}$ and $I_{DS}$ represent the intensities of exchange peaks from “S” to “D” and vice versa. Fitting of the curve was achieved using Prism 6 with in-house written scripts. From these data, one can track the decay of each of the “auto” peaks on the diagonal from exchange and return to the equilibrium state with time (green and red points and traces), as well as the build-up and decay of each of the exchange peaks (blue and black points and traces). It is observed that the behavior of the exchange peaks $I_{SD}$ and $I_{DS}$ is essentially identical.
Table 3.1. Extracted parameters from 2D $^{19}$F-$^{19}$F EXSY experiments of the bi-stable sample:

<table>
<thead>
<tr>
<th>$I_S(0) \times 10^6$</th>
<th>$I_D(0) \times 10^6$</th>
<th>$R_S$ (s$^{-1}$)</th>
<th>$R_D$ (s$^{-1}$)</th>
<th>$k_{SD}$ (s$^{-1}$)</th>
<th>$k_{DS}$ (s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2 ± 0.7</td>
<td>4.1 ± 0.6</td>
<td>3.0 ± 1.4</td>
<td>0.4 ± 0.7</td>
<td>24.2 ± 1.8</td>
<td>13.1 ± 1.0</td>
</tr>
</tbody>
</table>
Measurement of exchange in the U2-U6 snRNA complex of the protein-free human spliceosome

I then applied this method to analyze exchange processes in a larger, biologically important U2-U6 snRNA complex. In the previous chapter, I used 1D $^{19}\text{F}$ spectra of the protein-free U2-U6 snRNA bearing a single 5-F-cytidine (C) label in a region near the central junction to show that the complex adopts at least two major conformations. The spectrum exhibits a large sharp peak (with a shoulder) at a chemical shift corresponding to 5-$^{19}\text{F}$-C in a double-stranded region and is attributed to formation of a four-helix model; there is also a lesser very broad peak totaling approx. 15% the area of the sharp peak, with a range of chemical shifts corresponding to 5-$^{19}\text{F}$-C in a single-stranded region. The very broad nature of this peak (or collection of peaks) suggests an ensemble of local variants of a three-helix junction conformer (schematically shown in Figure 2.1) (Zhao et al. 2013). The fact that the single-stranded peak is distinct from the double-stranded peak, and that the location of each overlaps with the control resonance chemical shifts, suggest that any exchange between the conformations is in slow NMR chemical exchange time scale and thus the broad single-stranded peak was not likely the result of “breathing”, i.e. the open up of the double-stranded peak. Additional measurements on mutated complexes verified the presence of the three-helix structure which is very dynamic, thus suggesting that the two ensembles of conformations observed in the one-dimensional spectrum are associated with alternate base pairing patterns in the region of the central junction. However, what was not clear was whether the two major conformers are in an equilibrium exchange with each other on a timescale that could be detected by solution NMR.

In order to investigate this question, the same approach of $^{19}\text{F}$-$^{19}\text{F}$ EXSY experiments was applied to human U2-U6 snRNA complex. Due to the low distribution and the broadness of the
single-stranded peak, a more sensitive probe is required as compared with the broadband probe used for the bi-stable sample in order to obtain acceptable spectra. Therefore, EXSY experiments on U2-U6 snRNA complex were acquired by Dr. Clemens Anklin from Bruker BioSpin at 600 MHz spectrometer equipped with dedicated fluorine CyroProbe.

A series of 2D $^{19}$F-$^{19}$F EXSY spectra with mixing times ranging between 50 and 700 ms were acquired on human U2-U6 snRNA complex. Similar as in the case of the model bi-stable stem loop, I observed a rapid buildup of cross peaks at 100 ms and a long decay between 200 and 500 ms. However, a notable difference between buildup behavior of cross peaks in measurements of the of bi-stable stem loop and that of the U2-U6 snRNA is that the latter system displayed multiple cross peaks corresponding to different regions of the very broad peak in a single-stranded environment and the sharp peak (and shoulder) of the double-stranded environment. Fluorine is the only frequency labeled in the system, so multiple EXSY cross peaks (or exchange NOEs) observed in 2D spectra suggest a number of exchange events involving alternative conformations or environments involving the segment in which the 5-$^{19}$F-C resides (representative spectra shown in Figure 3.5) (Zhao et al. 2014a). Interestingly, different cross peaks exhibition maximal intensity at different times in the millisecond-to-second timescale. For example, the cross peak between the double-stranded peak and the more upfield single-stranded peak was building up from 100 ms to 200 ms while the cross peak between the double-stranded peak and the more downfield single-stranded peak has started to decay. This finding suggests that these exchange processes are independent events, i.e there are several pairs of conformers interconverting with each other at different timescales. Because of the limited resolution afforded by these spectra and significant spectral overlap of the multiple cross peaks, these EXSY data alone are insufficient for a comprehensive analysis of the exchange behavior in the human U2-
Figure 3.5 Two-dimensional $^{19}$F-$^{19}$F EXSY spectra of human U2-U6 snRNA complex acquired at 600 MHz with mixing times of 100 ms, 200 ms, 400 ms and 500 ms, as marked (Zhao et al. 2014a). The appearance of multiple exchange NOEs between conformers associated with each of the two major models is consistent with interconversion (identified in the figure) between the two models shown in Figure 2.1 and 2.2, likely including multiple alternative or intermediate folds, on a sub-second time scale.
U6 snRNA complex. However, our approach provides qualitative evidence to support spontaneous exchange involving multiple folds of this complex in which the junction region adopts alternative conformations. The U2-U6 snRNA complex operates in situ in conjunction with numerous U2 snRNP proteins that are likely to stabilize and/or favor particular conformational features. These data represent the first demonstration of conformational exchange within the junction region that may be associated with spliceosomal function.

### 3.4 Discussion

The rates of interconversion extracted from the exchange data imply that the interconversion between the two states of the bi-stable sample was cooperative without complete open of the helix as an intermediate, which would be much slower than the observed value $k_{ex} = 37.3 \pm 2.8 \text{ s}^{-1}$. This finding is consistent with the result by $^{13}$C NMR studies on a related sequence (Klöiber et al. 2011), which (as in this case) did not display any evidence of a fully open single-stranded state or other intermediates. However, the calculated exchange rate $k_{ex}$ is much faster (approximate 70 times faster) in this research as compared to that from a similar bi-stable sequence (Klöiber et al. 2011), which can likely be attributed to the shortened helix from 6 base pairs to 4 in one of the conformers. These findings support our conclusions that the $^{19}$F-$^{19}$F EXSY experiment is a valid approach to measure the kinetics of dynamic exchange between RNA conformations in equilibrium.

The bi-stable RNA stem loop examined here represents a single model RNA sequence; however, I note that this same approach can be applied to a wide range of RNA sequences in order to gain further insight into the dynamics and thermodynamics associated with RNA folding.
and function. Such examples may include, but not be limited to, studies of small nuclear RNAs, ribozymes, and riboswitches.

For human U2-U6 snRNA complex, my data provided the qualitative evidence for the dynamic interconversion at sub-second timescale between four-helix conformation and three-helix conformation. The multiple exchange NOEs suggest that the junction region is quite flexible and thus is capable of interconversion at different environments, corresponding to different complex in the process of splicing reaction. Due to the size and the dynamic nature of the molecule, it would be difficult to assign the peaks in the junction in the traditional $^1$H NMR. Thus $^{19}$F NMR provides an important alternative approach in studies of RNA structures.

Biochemical studies using chiral phosphorothioate in the substrates suggest that the active site undergoes conformational change between steps (Moore and Sharp 1993). The results were reinforced by functional studies of yeast U2-U6 snRNA complex, which suggest different conformations at different steps of splicing (Madhani and Guthrie 1992; Mefford and Staley 2009; McPheeters and Abelson 1992). Although metal ions and many snRNP proteins are involved in the splicing reaction in vivo, the ability of RNA alone to undergo conformational rearrangement to form the active site is important for understanding of the role of RNA in catalytic activity. My data present the first direct evidence of conformational interconversion within human U2-U6 snRNA complex, thus providing a new insight for relating structure features to functions.

In contrast with the conformational change model (Moore and Sharp 1993), some researchers proposed a single conformation between the two steps of splicing (Steitz and Steitz 1993; Anokhina et al. 2013). However, it is still unclear how the active spliceosome assembles and rearranges in vivo. We note that the spontaneous interconversion between folds of the
protein-free human U2-U6 snRNA complex suggested in this research provides the kinetic and thermodynamic basis for analysis of spliceosomal conformational rearrangements associated with the steps of splicing.

In conclusion, we have demonstrated the application of $^{19}$F NMR spectroscopy to probe the dynamics of RNA conformational exchange. The novelty, experimental advantage, and impact of this approach is the ability to determine kinetic rates of RNA conformational exchange from data derived from $^{19}$F spectra, using oligomers with single $^{19}$F label made by commercial nucleic acid synthesis. The application of this approach on human U2-U6 snRNA complex supports the interconversion between different conformers.
Chapter IV

Conclusions and Future directions

4.1 Conclusions

4.1.1 Fold of human U2-U6 snRNA complex

The research described in this dissertation focuses on the fold of protein free human U2-U6 snRNA complex. The functional studies of the yeast sequence suggest the importance of the formation of Helix Ib, and thus the three-helix model (Madhani and Guthrie 1992) in at least one (Hilliker et al. 2007) and possibly both (Mefford and Staley 2009) cleavage steps of pre-mRNA splicing. In contrast, genetic experiments on the U2-U6 snRNA complex in human cells support the importance of the four-helix model from the appearance of U2 Stem I during both steps of splicing (Sun and Manley 1995) and found no evidence for Helix Ib. The difference between these two models relates to the relative positions of metal ion binding sites and the catalytically important AGC triad as a result of different base pairing patterns: formation of additional intramolecular base pairs in U6 snRNA in the four-helix model extend the U6 (with the AGC triad at its base) or intermolecular base pairs in the three-helix model to include the AGC triad in Helix Ib. The difference in the relative positioning of catalytically essential metal ion binding sites within the AGC triad and a bulged region within the U6 ISL, which have been proposed to approach each other to form a two metal-ion center in the spliceosome’s active site similar to that formed by analogous regions in a crystallographic model of a spliced Group II intron (Toor et al. 2008; Keating et al. 2010). Thus these alternative models have a large impact on understanding the assembly and function of the active spliceosome.
Many structural studies have been performed on the protein-free complex *in vitro*. The fact that catalytic activity has been observed by the protein-free human U2-U6 snRNA complex (albeit not the identical two-step reaction as *in situ* and at a low rate and yield) suggests that conformations formed by the protein-free complex resemble those within the intact spliceosome (Valadkhan and Manley 2001; Valadkhan et al. 2007; Valadkhan et al. 2009). NMR studies of the truncated yeast U2-U6 complex revealed the formation of two consecutive U•U base pairs predicted to occur in U2 Stem I, thus suggesting formation of the four-helix model *in vitro* (Sashital et al. 2004). However, a later model derived by the same group from a combination of solution NMR, small angle x-ray scattering, and computer modeling of a sequence identical with the previously studied complex with the exceptional of an additional four base pairs near the termini of Helix II, indicated formation of a three-helix junction structure (Burke et al. 2012).

Although the human and yeast spliceosomes perform identical reactions, it is possible that the corresponding two complexes have different conformational distribution because of the different sequence in the junction region. Chemical structural probing of human U2-U6 snRNA complex at different stages of the splicing reaction suggests the prevalence of the three helix junction in the active spliceosome (Anokhina et al. 2013), but this method focuses more on the overall fold with less detailed local conformation specifically in the junction. Also, a careful observation of their published structure probing results reveals ambiguities, shown by appearance of bands corresponding to accessible, *i.e.* unpaired, nucleotides that would only be present in the four-helix fold as well as those present in the three-helix fold. Enzymatic structure probing studies performed by Ravichandra Bachu in our laboratory showing similar ambiguities, consistent with contributions of both four- and three-helix folds in the protein-free human U2-U6 snRNA complex (Zhao et al. 2013) prompted further analysis of the possible heterogeneity by
$^{19}$F NMR presented here. While computational simulation suggests both conformations for the complex (Cao and Chen 2006), the results in this dissertation provide the first direct structural evidence and distribution of the conformational heterogeneity of human U2-U6 snRNA complex under protein free conditions. The four-helix conformation was supported by the formation of G-$^{19}$C base pair in U2 Stem I from the double-stranded chemical shift of the fluorine resonance.

Analysis of the $^{19}$F NMR data of the broad peak at single-stranded chemical shift, as well as the results from mutations stabilizing each of the two models respectively, suggests the presence of a small fraction of conformations similar to the proposed three-helix model. Furthermore, the $^{19}$F-$^{19}$F EXSY experiments imply spontaneous interconversion between multiple conformers of the three-helix junction fold with a more homogeneous four-helix fold (one major conformer with a lesser contribution of a second environment displayed as a “shoulder” in the NMR spectra), with exchange rates in the $2–10$ sec$^{-1}$ range. These interconversions support the possibility or rearrangement of the active site that has been proposed to accompany the two steps of splicing (Moore and Sharp 1993), although convincing evidence is still lacking for such a mechanism (Steitz and Steitz 1993). These findings provide useful information of the ground state conformation of human U2-U6 snRNA complex as well as its dynamic interconversion, and thus serve as the basis for further investigation relating structures and functions of the RNA components throughout the process of pre-mRNA splicing.

4.1.2 Effect of magnesium ion

The dependence of the catalytic activity of the complex upon several specifically bound Mg$^{2+}$ ions (Sontheimer et al. 1997; Yean et al. 2000; Huppler et al. 2002; Valadkhan and Manley 2002; Yuan et al. 2007) underscores the importance of metal ion in the splicing reaction.
Comparisons of the human U2-U6 snRNA complex with the mechanistically and structurally parallel Group II intron (Toor et al. 2008), suggest that the metal ion binding sites in the U2-U6 snRNA complex are brought together to form a two metal ion center (Steitz and Steitz 1993) in the activate complex. This prediction suggests the role of magnesium ion in structurally assisting the formation of activate sites and/or in catalysis. Single molecule FRET studies on both human and yeast U2-U6 snRNA complex also suggest conformational change induced by binding of magnesium ion (Guo et al. 2009; Karunatilaka and Rueda 2014). However, it is still not clear how much of the magnesium ion dependence relates to an alteration within the junction to form the different structural models, ion-dependent changes in the approach between the fluorescent Cy5 donor probe on the unpaired region of the 5’ region of U6 snRNA and the Cy3 acceptor on the truncated ISL, or a combination of the two. To probe the effect of magnesium ion on the secondary structure of the junction, I acquired one-dimensional $^{19}$F NMR spectra for the sample in the absence and presence of 5mM Mg$^{2+}$. The major difference between our samples and those used in the FRET studies is that we included sufficiently long stem corresponding to Helix III to limit any structural changes to the junction region, whereas the sample used by Guo et al. (2009) had only a few base pairs in the proposed Helix III range, and the authors suggested that at least part of the Mg$^{2+}$-dependent changes observed in their studies may have been related to an alternative position of the unpaired sequence (Guo et al. 2009). Comparison of $^{19}$F NMR the spectrum with 5 mM Mg$^{2+}$ to that in its absence suggests a small shift in the equilibrium distribution towards the three-helix conformation (from ~14% to 17%), consistent with the findings from enzymatic probing by Ravichandra Bachu from our laboratory on the same samples. Therefore we speculate that the Mg$^{2+}$ has very little effect on the base pairing pattern of the human U2-U6 snRNA complex in the region of the junction and it alone is
not likely to be sufficient to facilitate the rearrangement required suggested by the single molecule FRET data. Based on the observations of the decrease in radius of gyration from small angle x-ray scattering (SAXS) data (Burke et al. 2012), as well as the FRET data from our laboratory (Chu and Greenbaum, unpublished data), we speculate that the role of Mg$^{2+}$ may alter the tertiary fold of the complex and makes it more compact.

4.1.3 Dynamics of human U2-U6 snRNA complex

The data of one-dimensional $^{19}$F NMR of human U2-U6 snRNA complex and its mutations suggest that any exchange between the conformations is within the slow NMR chemical exchange time scale (100-500 ms). In contrast, the “breathing” of base pairs is anticipated to reside within the rapid chemical exchange time scale (sub millisecond range) and would be demonstrated by a single broad peak instead of two distinct peaks that each overlapped with that of the corresponding control oligomer; therefore, we expect that the presence of the difference peaks does correspond with the conformational rearrangements. Our $^{19}$F-$^{19}$F EXSY data provide the evidence for the interconversion between different conformers in a ~100-500 millisecond timescale, consistent with the low energy barrier calculated from the equilibrium distribution. It has previously been shown that the two steps of splicing implies different stereochemistry (Moore and Sharp 1993) and each of two chemical reactions is reversible (Tseng and Cheng 2008), thus we can speculate that the interconversion we observed here may relate to altered active sites for the two steps of splicing. The finding of interconversion in such a complicated system helps us understand how the flexible nature of RNA molecules contributes to their functions.
The application of $^{19}$F-$^{19}$F EXSY experiments on a bi-stable model molecule confirmed the feasibility of this method on RNA dynamics. By comparison, in the case of human U2-U6 snRNA complex, the multiple conformations with different dynamics make it too complicated to present a comprehensive analysis of the intensity of each peak thus extract the exchange parameters. The presence of the exchange NOEs suggests the dynamics within the complex, which may be difficult for other approaches because of the size of the complex. We speculate that the pronounced heterogeneity within the three-helix fold reflects flexibility associated with the protein-free state and may not occur in the assembled spliceosome. It is also possible that association with spliceosomal proteins shifts the distribution of conformations further. However, data shown here provide evidence that this technique is a valuable and unambiguous approach to study rearrangement of functional RNA associated with biochemical function, and can readily be applied to biologically significant RNA systems.

4.2 Future directions

4.2.1 Characterize the relative orientation between Helix I and U6 ISL of human U2-U6 snRNA complex

Compared with spectra of $^{19}$F-$^{19}$F EXSY experiments on a bi-stable model molecule at different mixing times, which provided build-up and decay of cross peaks between resonances corresponding to unique single- and double-stranded environments, comparable experiments on the a protein-free sample of the human U2-U6 snRNA complex at different mixing times illustrated multiple cross peaks with different patterns of build-up and decay. These findings are consistent with a more complex pattern of exchange involving multiple junction arrangements.
The time scale and comparison with other data (Guo et al. 2009) suggests that these alternative conformations are likely to represent exchange between three- and four-helix folds. However, there were multiple cross peaks involving resonances in the single-stranded region of the spectrum, implying multiple flexible conformations in exchange in the junction which is lack of careful characterization. There is too much overlap in the $^1$H spectra of human U2-U6 snRNA complex to obtain unambiguous assignments in the junction area, but such assignments would assist in characterizing the features of multiple conformation. Measurements of the relative life time of each conformation would help understand how rearrangement may be accomplished between steps of splicing. Because this junction region may well interact with proteins in vivo (Stark and Lührmann 2006), the structural features in protein-free complex present a starting point for investigation of spliceosome structure and dynamics.

In order to minimize the relaxation and line broadening associated with the dynamic rearrangement of the complex, several mutants stabilizing three-helix or four-helix conformation can be used as comparisons. With the help of the spectroscopic fingerprint of an AUU vs. AAU duplex (Lu et al. 2011) (in this case, requiring mutation of U2 G$_{25}$→A), selective deuteration (Lu et al. 2010), and selective (only label one type of nucleotides in the complex) or full $^{13}$C$^{15}$N isotopic labeled samples, we may be able to obtain the unambiguous assignments of less dynamic region in the complex. Once we have the assignments, residual dipolar coupling (RDC) (Al-Hashimi et al. 2002) can be used to characterize the relative orientation between different modules of the complex with and without metal ions, such as Helix I and U6ISL etc. This information will help with understanding of the tertiary fold of the complex as well as the conformational change induced by metal ions. It may also be valid to examine the behavior of the complex with and without the intron substrates to see if the intron induces any change.
4.2.2 Structural features of the junction in the human U2-U6 snRNA complex

With the resonances assignments of the junction in human U2-U6 snRNA complex, we can characterize the structural features of the junction including the stacking of Helix I, the metal ion binding site in the conserved AGC triad, the stacking of the AAAUU loop in U6 snRNA (A82-84, U84, 85), the structural features of the G:A base pair in U6ISL etc. Also, 2-amino purine can be used to substitute A23, A24 in U2 snRNA and thus use fluorescence spectroscopy to determine the stacking of these two A’s in Helix I. Local structural features in the junction will explain the difference between three-helix and four-helix conformation and help with understanding the dynamic nature of the complex. Also, the comparison of the structure features of the activate sites with those from the crystal structure of Group II introns (Toor et al. 2008) will help to arrive at a more general model of the catalytic core of the pre-mRNA splicing apparatus, and further our understanding of the splicing mechanism across different platforms.

4.2.3 Develop a three-dimensional model of human U2-U6 snRNA complex

Due to the dynamic nature of human spliceosome, little is known about the tertiary structure of the U2-U6 snRNA complex. X-ray structure of the self-splicing Group II intron has suggested that the distinct metal ion binding sites in primary sequence are brought into close proximity to each other in tertiary structure (Toor et al. 2008). Based on the mechanistic parallels and sequence and structural similarities between functionally RNA sequences in Group II intron and spliceosome, one can anticipate similar tertiary fold of U2-U6 snRNA complex upon binding of divalent metal ions. Additionally, it is not certain if the conformational interconversion between different secondary structures also involves tertiary fold change. Thus, it is very important to characterize the three-dimensional fold of U2-U6 snRNA complex under different conditions.
Small angle x-ray scattering (SAXS) is a useful approach for detecting tertiary structural information of macromolecules and relative orientation between different fragments in solution (Zhang et al. 2014). We can collect SAXS data on the mutated constructs that stabilize either the three-helix or four-helix conformation to obtain the shape of the complex and the relative orientation between Helix III and U6 ISL, Helix II and U6 ISL and Helix II and Helix III. With the information from the constructs that are primarily homogeneous in conformation, we can deconvolute the SAXS data for human U2-U6 snRNA complex. And with the relative orientation obtained from NMR RDC measurements and SAXS, as well as the local structural features of the junction area, we can calculate the three-dimensional structural of U2-U6 snRNA complex by computational simulation, which is essential for understanding structural and functional relationship in human spliceosome.

4.2.4 Comparison of U2-U6 snRNA complex with U12-U6atac complex from minor spliceosome

The natural occurrence of a functional analogue of U2-U6 snRNA complex, the U12-U6atac complex of the low-abundance “minor” spliceosome found in eukaryotes (Levine and Durbin 2001), provides valuable information for investigating the mechanism involved in the splicing reaction catalyzed by spliceosome. In spite of the identical chemistry catalyzed by both the major and minor spliceosomes (Tarn and Steitz 1996; Collins and Guthrie 2000), these two complexes share limited (~60%) sequence similarities, which suggests the catalytic function relates more to the structure of the complex than the sequence content (Tarn and Steitz 1996; Incorvaia and Padgett 1998; Otake et al. 2002). Interestingly, the Helix II formed in U2-U6 snRNA complex is missing from the U12-U6atac complex, and there is no evidence for a structural analogue of Stem I (found in the four-helix model of the U2-U6 snRNA complex); thus there is no central
junction present in the U12-U6atac complex (Shukla and Padgett 2001; Shukla and Padgett 2002). Alternatively, we propose that any required flexibility in the central region required for folding of the complex and for positioning of the AGC triad may be facilitated by flexibility of the backbone within a short single-stranded region of U6atac snRNA opposing the opening formed by the 5' terminus of the U12 snRNA and the 3' end of the U6atac snRNA. The comparisons of structure features within these two complexes may provide insight into how the metal ion binding sites within the AGC triad and U6 ISL bulge are positioned, which will help us understand how the flexibility in the junction of U2-U6 snRNA complex assists in assembly of the active spliceosome.
References:


