


1-1-2011

# Characterization of splicing mechanisms by single-molecule fluorescence

Krishanthi Sanjeevani Karunatilaka  
*Wayne State University*

Follow this and additional works at: [http://digitalcommons.wayne.edu/oa\\_dissertations](http://digitalcommons.wayne.edu/oa_dissertations)

 Part of the [Biochemistry Commons](#), [Biophysics Commons](#), and the [Molecular Biology Commons](#)

---

## Recommended Citation

Karunatilaka, Krishanthi Sanjeevani, "Characterization of splicing mechanisms by single-molecule fluorescence" (2011). *Wayne State University Dissertations*. Paper 243.

**CHARACTERIZATION OF SPLICING MECHANISMS BY  
SINGLE-MOLECULE FLUORESCENCE**

by

**KRISHANTHI SANJEEWANI KARUNATILAKA**

**DISSERTATION**

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

**DOCTOR OF PHILOSOPHY**

2011

MAJOR: CHEMISTRY (Biochemistry)

Approved by:

---

Advisor

Date

---

---

---

## **DEDICATION**

This dissertation is dedicated to my loving parents and the Karunatilaka family who have supported and motivated me throughout my life.

## **ACKNOWLEDGEMENTS**

First and foremost, I would like to thank my advisor Professor David Rueda for his guidance and encouragement throughout my graduate studies. As a member of his first group of Ph.D. students, I really appreciate his immense support from the beginning to the end of my thesis work. I am also grateful for all his generous contributions of time, ideas and funding toward my research. I am thankful to my research committee members, Prof. Andrew Feig, Prof. Philip Cunningham and Prof. Tiffany Mathews for their feedback on my research as well as their valuable time spent reading the thesis and attending my research presentations. I am also thankful to our collaborators Prof. Anna Marie Pyle, Yale University, and Prof. Roland Sigel, University of Zurich, for their substantial time and continuous support that made my research career more productive.

I am thankful for all the present and past members of Rueda group, who have contributed in different ways at Wayne State University. I am truly grateful for Miriam Steiner, Chris Baker, Fardusee Ullah and Amanda Solem for their significant contributions toward my research. I am also grateful to Rui Zhao for his single-molecule microscopy support. Furthermore, I would like to especially thank Rui, Rajan, Amanda, Zhuojun, Elvin, Alfonso, Marcus, Gayan, Chandani, Hansini, Bishnu, Radek and Kyle for all their helpful discussions and for providing a productive environment to perform my studies.

I would like to thank Prof. Christine Chow for allowing me to use instrumentation in her lab and for providing valuable advice for my research. I am

also thankful to the RNA Club for the opportunities to discuss my research and to learn about other RNA research at Wayne State University. Additionally, I like to thank all the staff members in the department of Chemistry for their generous support during my studies.

I gratefully acknowledge all funding sources that made my research career more productive. I am thankful to the National Institutes of Health (NIH) and National Science Foundation (NSF) for funding this thesis work. I also thank Wayne State University for its financial support including the Graduate Dissertation Fellowship, which has allowed me to comfortably complete my dissertation.

Lastly, I would like to thank my family and friends for their immense support, love and encouragement. I owe my deepest gratitude to my loving parents for their continuous love, guidance and encouragement throughout my life. I am grateful for my brothers and sisters, who have supported me various ways to achieve my goals. I am also grateful for my fiancé, Kyle Vrtis, for his endless love and support he has given me to complete my thesis as well as for sharing all good and bad moments with me. Furthermore, I would like to express my gratitude to Kyle's family for all their support during last two years.

## TABLE OF CONTENTS

Dedication .....	ii
Acknowledgments .....	iii
List of Tables .....	ix
List of Figures .....	x
List of Abbreviations .....	xii
<b>CHAPTER 1: Introduction</b> .....	<b>1</b>
1.1 RNA and the origin of life .....	1
1.2 Structure and function of RNA .....	1
1.3 RNA folding .....	5
1.3.1 RNA folding pathways .....	6
1.3.2 Effect of metal ions on RNA folding .....	8
1.3.3 RNA folding and protein cofactors .....	9
1.4 Ribozymes .....	10
1.5 Group II Introns .....	12
1.5.1 Structure of group II Introns .....	13
1.5.2 Group II Intron splicing .....	17
1.6 Folding of group II introns .....	20
1.7 Protein-mediated group II intron splicing .....	24
1.7.1 Structure and function of Mss116 .....	25
1.7.2 Mss116-mediated group II intron splicing mechanism .....	28
1.8 Nuclear pre-mRNA splicing .....	29

1.9 The spliceosome .....	31
1.9.1 Assembly of the spliceosome .....	32
1.10 Spliceosomal U2-U6 snRNA complex.....	34
1.10.1 Structure of human and yeast U2-U6 snRNA complex .....	36
1.11 Fluorescence-based detection of biomolecules .....	38
1.11.1 Fluorescence theory .....	39
1.11.2 Fluorescence Resonance Energy Transfer (FRET) .....	39
1.11.3 Single-molecule fluorescence.....	42
1.11.4 Fluorescence anisotropy.....	46
1.12 Specific aims .....	48
<b>CHAPTER 2: Materials and Experimental Methods .....</b>	<b>50</b>
2.1 Materials.....	50
2.2 Single-molecule analysis of Mss116-mediated group II intron folding .....	51
2.2.1 Preparation of DNA, RNA and protein samples.....	51
2.2.2 Characterization of Mss116 activity .....	52
2.2.3 Single-molecule FRET experiments .....	53
2.3 Single-molecule studies of human spliceosomal U2-U6 snRNAs .....	55
2.3.1 RNA purification and labeling .....	55
2.3.2 Steady-state FRET (ssFRET).....	55
2.3.3 Single-molecule FRET (smFRET) .....	56
2.4 An RNA Aptamer Based Purification System for Fluorophore Labeled Proteins .....	57

2.4.1 RNA Synthesis and purification .....	57
2.3.2 Fluorophore labeling of protein .....	59
2.3.3 Fluorescence anisotropy measurements.....	59
2.3.4 Affinity column protein purification assay.....	61
<b>CHAPTER 3: Single-molecule Analysis of Mss116-mediated Group II Intron Folding .....</b>	<b>62</b>
3.1 Introduction.....	62
3.2 Experimental design.....	66
3.3 Results .....	70
3.3.1 Folding of group II introns alone requires high ionic strength .....	70
3.3.2 Mss116 facilitates group II intron folding in vitro.....	73
3.3.3 ATP hydrolysis by Mss116 is required to form the native state.....	81
3.3.4 Folding dynamics depend on Mss116 and ATP .....	85
3.4 Discussion .....	90
3.5 Conclusions.....	94
<b>CHAPTER 4: Single-molecule Studies of Human Spliceosomal U2-U6 snRNAs .....</b>	<b>97</b>
4.1 Introduction.....	97
4.2 Experimental design.....	102
4.3 Results .....	104
4.3.1 Mg <sup>2+</sup> -dependent structural dynamics by steady-state FRET .....	104
4.3.2 Mg <sup>2+</sup> -dependent structural dynamics by single-molecule FRET.....	106



4.3.3 Single-molecule detection of the human U2-U6 complex with modifications.....	110
4.3.4 Characterization of individual modifications in U2 stem I .....	112
4.4 Discussion .....	115
4.5 Conclusions.....	119
<b>CHAPTER 5: An RNA Aptamer Based Purification System for Fluorophore Labeled Proteins .....</b>	<b>122</b>
5.1 Introduction.....	122
5.2 Experimental design.....	126
5.3 Results and Discussion .....	128
5.3.1 SRB-2 binds TAMRA with high affinity and specificity.....	128
5.3.2 SRB-2 binding can be regulated by Mg <sup>2+</sup> ions and EDTA .....	130
5.3.3 RNA aptamer affinity purification of a TAMRA labeled protein .....	133
5.4 Conclusions.....	136
<b>CHAPTER 6: Conclusions and Future Directions .....</b>	<b>138</b>
References.....	145
Abstract.....	167
Autobiographical Statement.....	170

## LIST OF TABLES

### CHAPTER 3

Table 3.1: Peak centers, areas and peak heights.....77

Table 3.2: Single-molecule folding dynamics of D135-L14 ribozyme. ....87

Table 3.3: The percentage of transitions between conformational states.....89

### CHAPTER 4

Table 4.1: Difference between calculated free energy changes .....114

## LIST OF FIGURES

### CHAPTER 1

Figure 1.1: The central dogma of molecular biology.....	2
Figure 1.2: Four basic types of nucleotides in RNA.....	4
Figure 1.3: Secondary structure of ai5 $\gamma$ group II intron .....	15
Figure 1.4: Mechanisms of the group II intron splicing. ....	18
Figure 1.5: The proposed two-step folding pathway.....	22
Figure 1.6: Structure of Mss116 DEAD-box protein.....	26
Figure 1.7: Eukaryotic nuclear pre-mRNA splicing reaction .....	30
Figure 1.8: The spliceosome assembly pathway.....	33
Figure 1.9: Structure of human and yeast U2-U6 snRNA complexes.....	35
Figure 1.10: Jablonski energy diagram.....	41
Figure 1.11: Setup of TIRF single-molecule microscopy.....	45
Figure 1.12: The principle of fluorescence anisotropy. ....	47

### CHAPTER 3

Figure 3.1: The folding pathway of the D135-L14 ribozyme. ....	64
Figure 3.2: Secondary structure of the Sc. D135-L14 ribozyme.....	67
Figure 3.3: Single-molecule detection.....	69
Figure 3.4: Effect of ionic strength on the folding dynamics. ....	72
Figure 3.5: ATPase, unwinding and splicing activity of the Mss116. ....	74
Figure 3.6: Mss116 promotes folding of group II introns.....	76

Figure 3.7: The folding of group II introns at physiological conditions. ....	78
Figure 3.8: Effect of substrate on the protein-mediated folding. ....	80
Figure 3.9: Role of ATP in Mss116-mediated folding of group II introns. ....	82
Figure 3.10: Effects of mutant Mss116 on group II intron folding dynamics. ....	86
Figure 3.11: Percentage of dynamic and static molecules. ....	86
Figure 3.12: Transition density plots from HMM analysis. ....	88
Figure 3.13: Mss116-mediated group II intron folding pathway. ....	93
<b>CHAPTER 4</b>	
Figure 4.1: Structures of modified residues and U2 stem I. ....	100
Figure 4.2: Human U2-U6 snRNA complexes. ....	103
Figure 4.3: Steady-state fluorescence. ....	105
Figure 4.4: Mg <sup>2+</sup> -dependent structural dynamics. ....	107
Figure 4.5: The folding pathway of human U2-U6 complex. ....	109
Figure 4.6: Single-molecule analysis of the modified human U2-U6 complex. ....	111
Figure 4.7: Single-molecule FRET histograms of the modified constructs. ....	113
<b>CHAPTER 5</b>	
Figure 5.1: Selection of oligonucleotide aptamers by SELEX. ....	125
Figure 5.2: Structure of SRB-2 aptamer and fluorophores. ....	127
Figure 5.3: Fluorescence anisotropy. ....	129
Figure 5.4: Effect of Mg <sup>2+</sup> and EDTA concentrations on binding of SRB-2. ....	131
Figure 5.5: RNA aptamer affinity purification of a TAMRA labeled protein. ....	134

## LIST OF ABBREVIATIONS

AMPPNP	Adenylyl-imidodiphosphate
ATP	Adenosine-5'-triphosphate
D135	Domains 1, 3 and 5
DTT	Dithiothreitol
EBS	Exon binding sites
EDTA	Ethylenediamine-N,N,N',N'-tetraacetic acid
FRET	Fluorescence resonance energy transfer
IBS	Intron binding sites
IEP	Intron-encoded proteins
ISL	Intramolecular stem loop
MOPS	3-morpholinopropane-1-sulfonic acid
NMR	Nuclear magnetic resonance
NTP	Nucleoside 5'-triphosphate
PAGE	Polyacrylamide gel electrophoresis
Sc.	<i>Saccharomyces cerevisiae</i>
Sc.ai5 $\gamma$	Yeast mitochondrial ai5 $\gamma$ group II intron
snRNA	Small nuclear RNA
snRNP	Small nuclear ribonucleoproteins
TIRF	Total internal reflection fluorescence
TRIS	2-amino-2-(hydroxymethyl)propane-1,3-diol

## **CHAPTER 1**

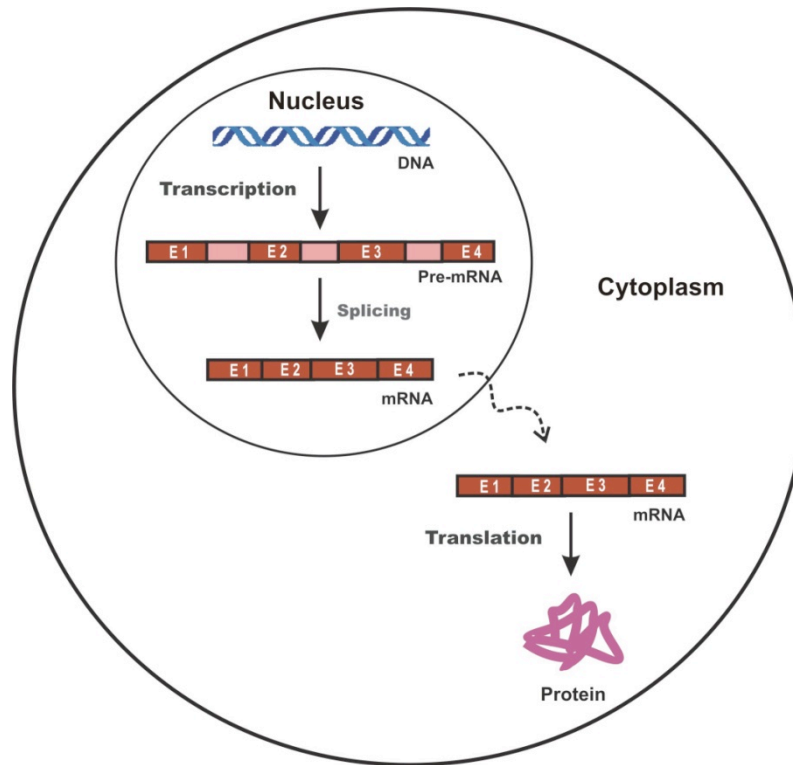
### **Introduction**

#### **1.1 RNA and the origin of life**

Gene expression is an essential biological process that converts genetic information to functional products in all domains of life. Regulation of gene expression is essential to maintain proper cellular activities in living organisms. As a general mechanism, gene expression mainly consists of two steps: transcription and translation (Fig. 1.1). Transcription is a highly regulated first step of gene expression that converts genetic information stored in deoxyribonucleic acid (DNA) to messenger ribonucleic acid (mRNA). Translation is a process that converts mRNA to functional protein with the help of macromolecular complexes known as the ribosomes. In contrast to protein coding genes, non-protein coding genes code for functional RNAs such as ribosomal RNA (rRNA) and transfer RNA (tRNA) as main gene products. In addition, some organisms use RNA to store their genetic information suggesting an important link between the RNA and the origin of life (1).

#### **1.2 Structure and function of RNA**

RNA is an important biological macromolecule that shows high functional versatility in living organisms. In addition to its well-known function as an information carrier, it can play an important role in catalysis and regulation of gene expression in living cells.



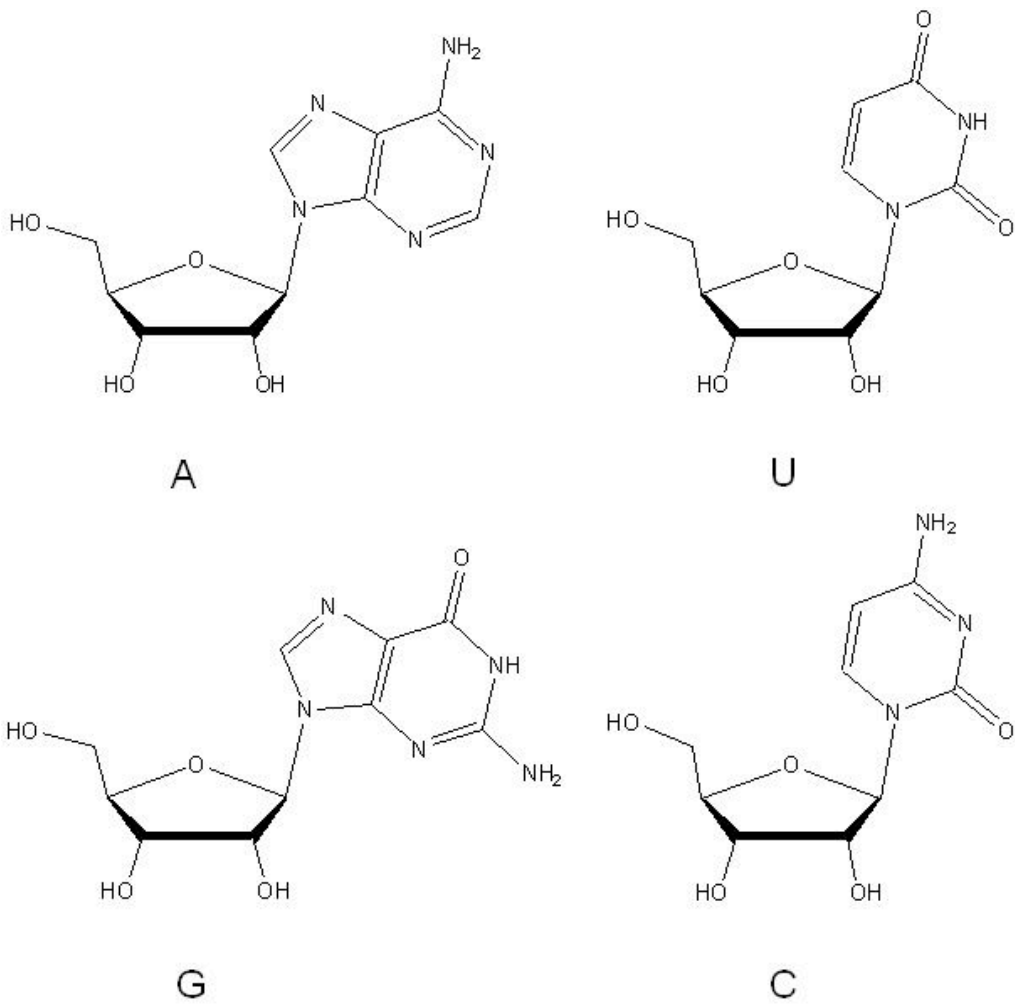
**Figure 1.1 The central dogma of molecular biology.** The central dogma describes the flow of genetic information from DNA to functional proteins through RNA intermediates. This process has two main steps: transcription and translation. Genetic information stored in DNA is converted into pre-mRNA by transcription in the nucleus. After RNA processing/splicing pre-mRNA forms the mature mRNA. Translation converts mRNA into functional proteins using ribosomes in the cytoplasm.

Although RNA molecules are only made of four types of basic nucleotides (Fig. 1.2), adenine (A), guanine (G), cytosine (C) and uracil (U), the ability of RNA molecules to fold into an infinite number of structures supports the functional diversity of these molecules. RNA consists of a long chain of nucleotides and each nucleotide contains a nitrogenous base, a ribose sugar, and a phosphate group. One of the major features of RNA that distinguishes it from DNA is the presence of a hydroxyl group at the 2' position of the ribose sugar. The presence of this functional group makes RNA conformationally flexible. As a result, RNA can form various secondary, tertiary and quaternary structures to perform different functions in cells.

RNA can perform an efficient regulatory role in living cells. Regulatory RNAs, including micro RNAs, small-interfering RNAs and small non-coding RNAs can regulate gene expression by various mechanisms either by direct translational suppression or by degradation of messenger RNA in cells (2-4). Riboswitches are one of the recently identified regulatory RNAs, which can regulate gene expression at both the transcriptional and the translational level providing more evidence for different roles of RNA at various stages in living cells (5,6).

In addition to a regulatory role, RNA can act as a catalyst in various biological processes. More than two decades ago, Cech and Altman first discovered the catalytic nature of RNA by demonstrating that RNA molecules have a unique ability to perform chemical reactions in the absence of proteins (7,8).





**Figure 1.2 Four basic types of nucleotides in RNA.** Adenine (A), Guanine (G), Cytosine (C) and Uracil (U). Each unit consists of a nitrogenous base and a ribose sugar. G forms Watson-Crick base pairs with C, and A base pairs with U.

The catalytic ability of RNA was further confirmed with the recognition of RNA as a primary participant in two vital biological processes: splicing and translation. The spliceosome, a multi-megadalton ribonucleoprotein complex consisting of five RNAs and over 150 proteins, catalyzes nuclear pre-mRNA splicing mainly through spliceosomal RNAs (9-13). The ribosome is another essential multi-megadalton ribonucleoprotein complex, which utilizes RNA components to catalyze mRNA translation in all living organisms (14-16).

In summary, RNA has the ability to store genetic information, and it can act as an enzyme to catalyze biological reactions supporting an interesting hypothesis called the “RNA world” (1). According to the RNA world hypothesis, the world was initially filled with life based on RNA and it evolved to become a modern DNA and protein world.

### **1.3 RNA folding**

In order to perform complex biological functions, an RNA strand must adopt a precise three-dimensional conformation in living organisms. Although RNA molecules are mostly comprised of only four basic nucleotides, folding of RNA can be as complicated as that of proteins, which are composed of 20 different amino acids. However, the mechanisms by which RNA can fold into active structures consist of the secondary and tertiary interactions are obscured in most biologically essential systems.

### ***1.3.1 RNA folding pathways***

Folding of a linear RNA strand into a compact active structure is a key requirement to ensure the proper function of mature RNA in cells. RNA folding is a hierarchical process and typically contains two main stages: formation of the secondary structure and formation of the highly compact tertiary structure. Formation of the secondary structure mostly depends on the Watson-Crick (A:U and G:C) base pairing between specific regions of the primary sequence and it can be stimulated by any group that has the ability to compensate the electrostatic penalty created from polyanionic RNA backbones in a close proximity. Therefore, monovalent cations, divalent cations, basic proteins and the other cationic molecules can facilitate the formation of RNA secondary structures by neutralizing the negative charges of RNA backbones (17-23). RNA mainly comprises of four basic secondary structural elements: helices, loops, bulges and junctions (24-26). These secondary structural elements, alone or combining with each other, promote the formation of specific structural elements such as stem-loops, tetraloops and pseudoknots that are directly involved in various biological functions (24-26).

In the final stage of RNA folding, the secondary structure collapses into a compact tertiary structure allowing duplex stacking and long-range intramolecular interactions between bases and backbone residues of RNA. In addition to the Watson-Crick base pairs, wobble and Hoogsteen base pairs can also present in the tertiary structure of functional RNAs. Unlike secondary structure, the formation

of tertiary structure usually requires divalent cations such as magnesium ions (19,23,27). The formation of RNA tertiary structure generally depends on several factors: 1) RNA sequence, 2) density of metal ions, 3) concentration of metal ions, 4) temperature, and 5) cofactors such as RNA binding proteins and polyamines. Therefore, individual RNA molecules can exhibit specific folding pathways under different reaction conditions providing evidence for a broad set of different RNA folding mechanisms.

RNA has a dynamic nature, which causes difficulties in defining a unique pathway for folding of large RNA molecules. Since RNA can be misfolded and trapped in stable inactive conformations, formation of the native conformations are highly challenging and leads to a common RNA folding problem (28). According to thermodynamic studies of the folding landscape (28,29), the native conformation contains the minimum free energy making it more stable than other inactive, misfolded conformations.

Because of the rugged nature of the folding potential energy landscape, RNA exhibits complex kinetic pathways with multiple transient intermediates before reaching the active conformation. Interactions with protein co-factors and metal ions can influence the stability of these intermediates by changing the RNA folding energy landscapes. Therefore, protein cofactors such as chaperones and metal ions can play an essential role in RNA folding.

### **1.3.2 Effect of metal ions on RNA folding**

Negatively charged phosphate groups in the backbone make RNA a charged molecule (polyanion) under physiological conditions and thus affect the formation of a compact structure. Positively charged groups are essential to screen these negative charges to promote RNA folding into complex three-dimensional structures.

Cations, monovalent and divalent, are major components required in RNA folding (17-19,23,30). Monovalent metal ions such as  $K^+$  and  $Na^+$  mainly contribute to the nonspecific electrostatic stabilization of the RNA backbone (18-20,30). Therefore, these ions play a major role in the formation of RNA secondary structures. The formation of highly compact tertiary structure requires specific electrostatic interactions of dehydrated higher valency metal ions such as  $Mg^{2+}$  for correct positioning of groups in space (19,23,27). In addition to higher valency ions, specific monovalent ions also rarely perform the similar behavior (18,31).

In addition to direct participation in RNA folding, metal ions are also required for RNA catalysis (17,32,33). During RNA-catalyzed reactions, metal ions can act as cofactors to promote efficient reaction catalysis. Cations can mediate the chemical reactions by activation of the nucleophile, stabilization of the transition state and protonation of the leaving group (34). As a result, most catalytic RNAs require divalent metal ions such as  $Mg^{2+}$  in the active site to perform catalytic reactions.

In accordance with the chemical structure, RNA contains a number of metal ion binding sites. In addition to unprotonated atoms in the bases, atoms in the phosphodiester linkage provide specific sites to interact with metal ions. Since the coordination affinity between metal ions and RNA is comparatively low, binding of metal ions with RNA is generally dynamic (34). Although, both monovalent and divalent metal ions are essential to regulate RNA structure and function, proteins and some positively charged molecules can replace metal ions from RNA and promote the formation of higher order structures.

### ***1.3.3 RNA folding and protein cofactors***

Proteins can interact with RNA specifically or non-specifically to assist folding of RNA into an active conformation in cells (22,35-37). Based on the general mechanism, these proteins can be divided into four main groups: 1) specific RNA-binding proteins, 2) RNA chaperones, 3) RNA annealers and 4) RNA helicases. Specific RNA-binding proteins are basic proteins that can mediate RNA folding by interacting with specific sequence or structure of RNA. Furthermore, these proteins promote RNA folding through stabilization of intermediate conformations or the functionally active tertiary structure (38).

In contrast, some proteins bind RNA transiently and non-specifically to promote the formation of the active conformation by resolving non-functional kinetic traps (28,35). These proteins are broadly defined as RNA chaperones and are involved in various biological processes including regulation of gene expression, RNA export and assembly of ribonucleoprotein (RNP) complexes.

RNA chaperones disrupt RNA-RNA interactions and unfold the RNA structure. However, this process does not depend on ATP binding and hydrolysis (22).

RNA annealers accelerate the annealing of unstructured RNAs by molecular crowding (39). Binding of these proteins with complementary RNAs increases the local concentration, thereby increasing the frequency of interactions between RNAs to promote the annealing process. Interestingly, some RNA annealers alter the RNA conformation after binding to enhance the inter-molecular interactions of RNA (40).

RNA helicases are members of the DEAD-box family of proteins that can mediate the rearrangement of structured RNAs using energy derived from ATP hydrolysis (41,42). Most of RNA helicases are non-processive proteins, thus they unwind short RNA duplexes (22). In contrast to RNA chaperones, RNA helicases promote remodeling of RNP complexes based on the RNA-dependent ATPase activity.

#### **1.4 Ribozymes**

RNA molecules that catalyze chemical reactions in the absence of proteins by forming active tertiary structures are known as catalytic RNAs or ribozymes (43,44). Functionally, ribozymes are very important and they can act as biosensors and therapeutic agents; furthermore they can play an important role in gene discovery (45-48). Because of this unique catalytic ability, ribozymes are ideal model systems to study the relationship between structure and function of RNA (36).

Catalytic RNA was first recognized independently by Thomas Cech and Sydney Altman in the 1980s with *Tetrahymena* group I introns and bacterial RNase P, respectively (7,8). Since then, a number of studies have been conducted to discover more naturally occurring catalytic RNAs ranging from small catalytic RNAs such as the hairpin ribozyme to large catalytic RNAs such as group I and group II introns (36,43,44). In addition to these ribozymes, spliceosomal and ribosomal RNAs associated in large ribonucleoprotein complexes also provide more evidence for catalytic ability of RNA.

Ribozymes are primarily classified into two groups based on their size and mechanisms: small ribozymes and large ribozymes. The hairpin ribozyme, the hammerhead ribozyme, the hepatitis delta virus (HDV) ribozyme and the Varkud Satellite (VS) ribozyme are some of the best-characterized naturally occurring small ribozymes. Most of these small ribozymes contain less than 200 nucleotides and catalyze the site-specific cleavage of the backbone in a single stranded region of the RNA molecule. Although small ribozymes use a common chemical reaction mechanism to cleave the phosphodiester bond by generating a 2'-3'-cyclic phosphate and a product with a 5'-OH terminus (44), the folding and cleavage pathways of these molecules differ based on the length of the RNA strand and the requirement of metal ions.

In contrast to small ribozymes, large ribozymes are structurally diverse and the length of RNA strand can vary from ~100 to ~3000 nucleotides. Nuclear ribonuclease (RNase) P, group I introns and group II introns are well-studied



examples of large ribozymes in nature. In addition to the size, large ribozymes greatly differ from small ribozymes in the catalytic activity. RNase P is an endoribonuclease, which site specifically cleaves precursor RNAs such as tRNA and 5S rRNA. RNase P associates with protein for its function and forms an essential RNP complex. Group I and group II introns belong to self-splicing introns and are necessary for the removal of non-coding sequences (introns) from coding sequences (exons) to form mature RNA for expression of specific proteins.

The cleavage reactions of large ribozymes are different from small ribozymes and they require external or distal nucleophiles. RNase P uses a hydroxyl group of a water molecule as an attacking nucleophile and generates products containing a 5'-phosphate and a 3'-OH after hydrolytic cleavage of the RNA(49). Although the cleavage reaction of both group I and II introns consists of two transesterification steps, group I introns use an external GTP and group II introns use a 2'-OH of an internal adenosine to attack the 5' splice site to initiate the first step of the splicing reaction. During the second step, the 3'-OH of the 5' exon attacks the 3' splice site and joins the two exons together forming a product with a 5'-phosphate and a 3'-OH (50,51).

### **1.5 Group II Introns**

Self-splicing introns are large RNA molecules that are able to catalyze their own excision from a pre-mRNA while covalently joining the flanking exon sequences. Group II introns are one of the largest natural self-splicing ribozymes (400-1000 nt) and they were initially identified in the organellar genomes of lower

eukaryotes and plants (52). However, bacterial genomic sequencing studies revealed that group II introns are widely distributed in bacteria suggesting group II introns may have evolved from the prokaryotic world (53). Furthermore, similarities between structures and catalytic mechanisms suggest that group II introns share a common ancestor with eukaryotic nuclear spliceosomal introns (54,55).

Group II introns have been found within protein coding mRNA as well as tRNA and rRNA genes. Although group II introns are important for expression of housekeeping genes in cells, bacterial group II introns can behave as mobile genetic elements and can be involved in genome diversification of bacteria (53,56). Interestingly, some group II introns contain an open reading frame (ORF) that encodes a protein. These intron-encoded proteins (IEP) can either facilitate intron splicing by stabilizing RNA tertiary structure or promoting site-specific insertion of group II introns into genomic DNA in bacteria (52,56).

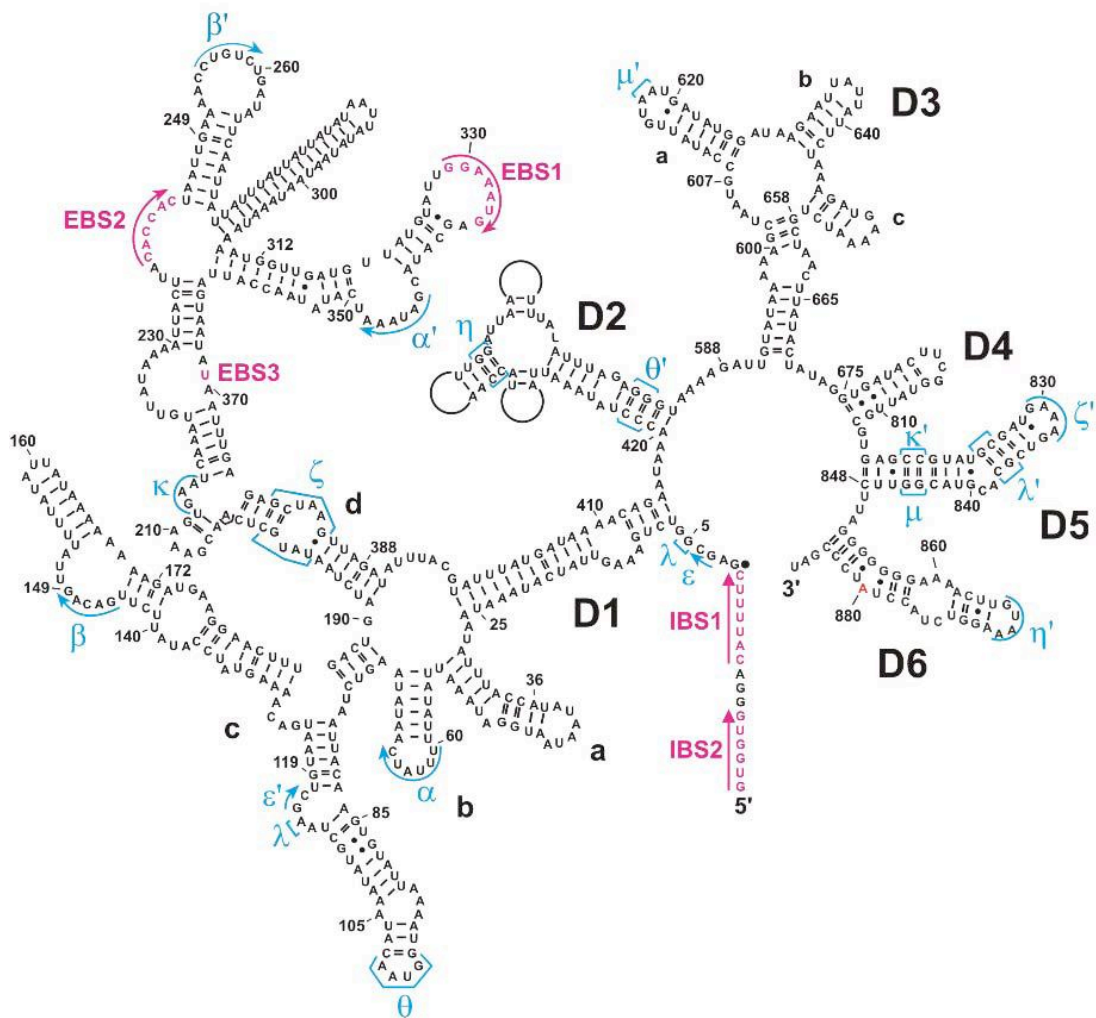
### ***1.5.1 Structure of group II Introns***

Based on the secondary structure, group II introns can be classified into three major subgroups: group IIA, group IIB and group IIC. Group IIA and IIB introns are different from each other based on the location of exon binding site 2 (EBS2), the length and composition of interdomain joiners and the tertiary interactions involved in recognition of the 3' splice site (57,58). Group IIC introns are significantly different from group IIA and IIB introns. Introns belonging to the group IIC family are small in size and exhibit simple secondary structures (58). Despite the diversity in primary sequences, group II introns share a highly

conserved secondary structure (Fig. 1.3) that consists of six domains (D1-D6) radiating from a central core (59).

Domain 1 is the largest intronic domain and serves as a scaffold for assembly of the other domains (60). D1 contains the exon-binding sites (EBS) and recognizes the 5'-exon (EBS1-IBS1 & EBS2-IBS2) and the 3'-exon (EBS3-IBS3) through base pairing interactions with intron-binding sites (IBS). According to structural studies conducted with the ai5 $\gamma$  group IIB intron (Fig. 1.3), derived from the *Saccharomyces cerevisiae* mitochondrial *cox1* gene, D1 forms intra-domain tertiary interactions ( $\alpha$ - $\alpha'$  &  $\beta$ - $\beta'$ ) as well as long range inter-domain tertiary interactions with D2 ( $\theta$ - $\theta'$ ) and D5 ( $\kappa$ - $\kappa'$ ,  $\zeta$ - $\zeta'$  and  $\lambda$ - $\lambda'$ ). The long range intra-domain interaction  $\alpha$ - $\alpha'$  facilitates the binding of 5'-exon (60) as well as playing a major role in intron folding (61). Interactions  $\kappa$ - $\kappa'$  and  $\zeta$ - $\zeta'$  between D1 and D5 are important for D5 docking (62,63) and the interaction  $\lambda$ - $\lambda'$  is directly involved in catalysis by facilitating the positioning of D5 close to 5' splice site (64). Conversely, the tetraloop-receptor interaction  $\theta$ - $\theta'$  between D1 and D2 is involved in the stabilization of the native structure rather than directly participating in catalysis (65).

D2 and D3 are not essential for group II intron function, but the presence of these domains enhances catalytic efficiency (66,67). Although D2 lacks phylogenetic conservation, it can interact with both D1 ( $\theta$ - $\theta'$ ) and D6 ( $\eta$ - $\eta'$ ) through inter-domain interactions and enhance the stability of the tertiary structure (65).



**Figure 1.3 Secondary structure of *Saccharomyces cerevisiae* mitochondrial *ai5 $\gamma$*  group II intron (*Sc. ai5 $\gamma$* ).** Six domains are numbered from D1-D6 and are arranged around a central core. Tertiary interactions are indicated by Greek letters in blue. Exon binding sites (EBS1, EBS2 and EBS3) and intron binding sites (IBS1 and IBS2) are shown in magenta. The branch point adenosine residue that is essential for branching pathway of intron splicing is shown in red. The black dot close to the 5' end represents the splice site of the group II intron (68).

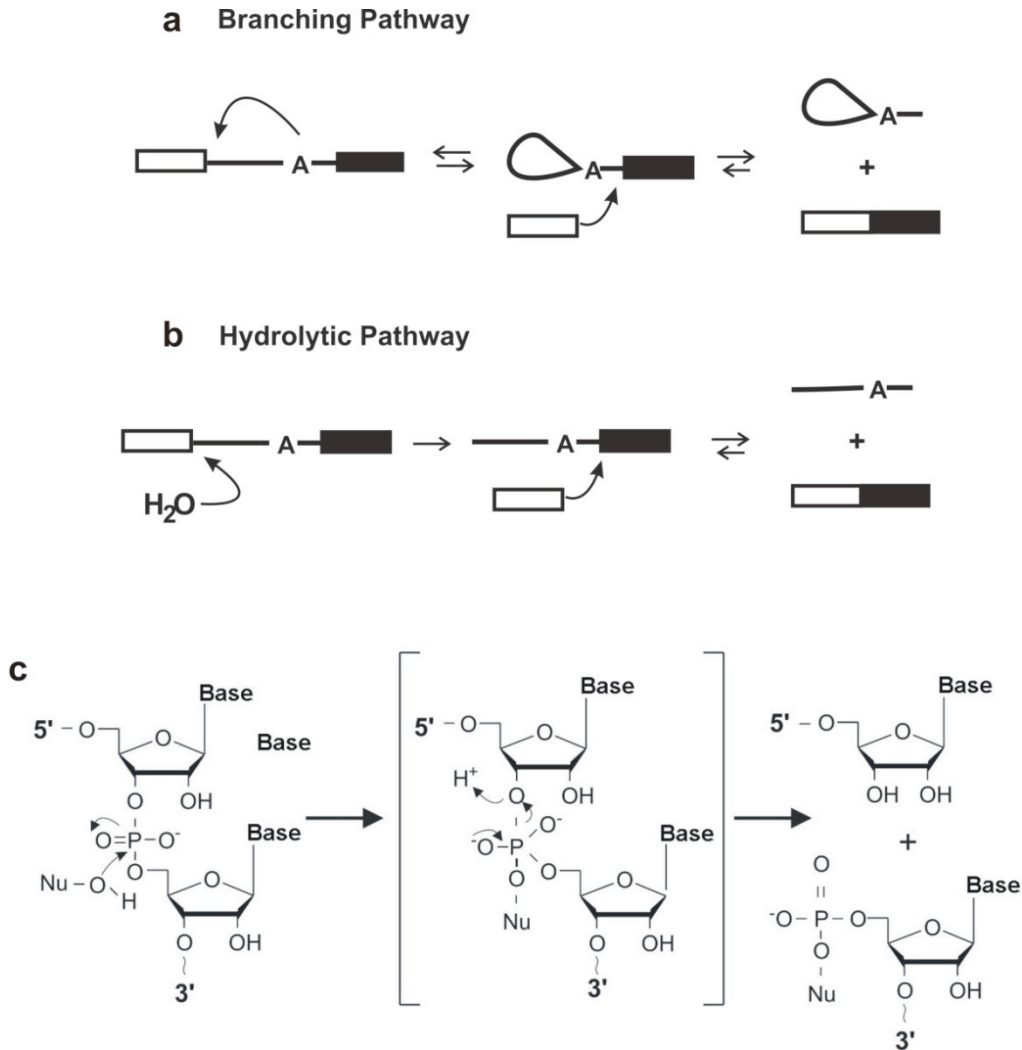
D3 forms tertiary contacts with D5 ( $\mu$ - $\mu'$ ) and acts as a catalytic effector during splicing reaction (60). The most variable domain in the intron is D4, which contains an open reading frame (ORF) for a multifunctional intron-encoded protein (IEP), maturase. IEPs of organellar introns are important for *in vivo* folding of the group II intron RNA into a catalytically active tertiary structure (54). However, IEPs in bacteria mainly facilitate the mobility of introns (56). According to a recent group II intron crystal structure, D4 points away from the intron active site to accommodate the comparatively large ORF without affecting the catalysis (59).

The most highly conserved region of the intron is D5, and it forms the catalytic center of the group II intron (54). Although D5 is a small stem loop structure (34 nt), it forms the network of tertiary contacts by interacting with the D1 and D3 (Fig. 1.3). Domain 5 docks onto the D1 scaffold through interactions between D1 and D5 ( $\kappa$ - $\kappa'$  and  $\zeta$ - $\zeta'$ ) as described (62,63). Therefore, D5 is highly shielded from solvent except the region close to D6. As the most essential part of the intron, D5 consists of several groups that are important for the catalytic reaction including the catalytic triad (5' -AGC), the tetra loop and the AC bulge. The catalytic triad, the most conserved region of D5, interacts with nucleotides in the linker between D2 and D3 (J2/3) and forms the triple helix structure (57). Therefore, this interaction merges two of the most essential parts together to promote the catalytic reaction of group II introns. Structural and mechanistic comparison between the spliceosome and group II introns indicates many similarities between D5 and the U6 snRNA in the spliceosome (55).

Domain 6 is important for the branching pathway of group II intron splicing (69). It contains the conserved branch point adenosine residue, which acts as a nucleophile during the first step of splicing reaction. D6 forms the tertiary contact with D2 ( $\eta$ - $\eta'$ ) and facilitates the branch point docking on the coordination loop in D1 (nucleotides 222-229 and 368-370) (70). According to biochemical studies of the ai5 $\gamma$  group II intron, D1 and D5 are the only domains strictly required for the minimal catalytic function in group II introns (52,71).

### ***1.5.2 Group II Intron splicing***

Group II introns can self-splice following two different reaction pathways: the branching pathway and the hydrolytic pathway (Fig. 1.4). Mechanisms of both group II intron splicing reactions consist of two consecutive transesterification steps (72,73). The branching pathway involves the branch point residue or the highly conserved adenosine located in a bulge of D6 as a nucleophile. In contrast, the hydrolytic pathway employs a solvent water molecule or a hydroxyl ion as a nucleophile to initiate the splicing reaction. The branching pathway is the most common two-step splicing mechanism in group II introns. The 2'-OH of the branch point adenosine initiates the first step of branching reaction by attacking the 5' splice site. During the second step, the 3'-OH of the 5' exon attacks the 3' splice site and ligates the two exons together while releasing the intron as a lariot intermediate with a 2'-5' phosphodiester bond (74).



**Figure 1.4 Mechanisms of group II intron splicing. a.** Schematic representation of the branching pathway **b.** Schematic representation of the water-mediated hydrolytic pathway. **c.** Group II intron cleavage mechanism. The attacking nucleophile can be an internal 2'-OH of adenine (Branching pathway) or a water molecule (Hydrolytic pathway). The cleavage reaction produces a 5'-phosphate and a 3'-hydroxyl terminus. Nu - nucleophile.

Both steps of the group II intron splicing are readily reversible (75,76). Interactions between intron-exon binding sites (IBS-EBS) are essential for substrate recognition as well as for proper positioning of different groups to mediate the splicing reaction. Alternatively, the hydrolytic pathway initiates the first step of splicing using an external hydroxyl ion or a water molecule as a nucleophile (73,77). The second step of splicing reaction ligates the 5' and 3' exons together and releases the linear intron providing an irreversible splicing reaction (Fig. 1.4).

As mobile genetic elements, group II introns can be inserted into specific sites of the genome by several mechanisms (53,56). Intron-encoded proteins or maturases are essential to mediate the mobility of introns. In order to perform this function, maturases contain several parts including an endonuclease (EN) domain and a reverse transcriptase (RT) domain (53). Reterohoming is the most efficient mobile pathway of group II introns, which occurs through a target DNA-primed reverse transcription (TPRT) mechanism (78). A ribonucleoprotein complex containing the excised intron RNA lariat and the reverse transcriptase (RT) encoded by the intron ORF is essential to mediate the homing process. According to the general retrohoming mechanism, first the intron lariat reverse splices directly into a target site of the DNA sense strand. Reverse transcriptase then cleaves the opposite DNA strand at a specific site and forms the cDNA of the inserted intron RNA through reverse transcription using the 3' end of the cleaved strand as a primer. Finally, the cDNA copy of the intron is integrated into the genome by



recombination or repair mechanisms (56,78). Retrotransposition is another mobile pathway of group II introns that occurs at low frequency in organisms (53).

Retrotransposition or ectopic transposition occurs by the insertion of intron RNA into nonallelic sites of the genome. There are two main proposed mechanisms for transposing intron RNA into ectopic sites. One mechanism involves reverse splicing of the intron RNA directly into a noncognate DNA target site. The second mechanism involves the insertion of intron RNA into nonallelic sites through an RNA intermediate (53). This mechanism involves reverse splicing of the intron into a site of the RNA followed by reverse transcription and cDNA integration through homologous recombination.

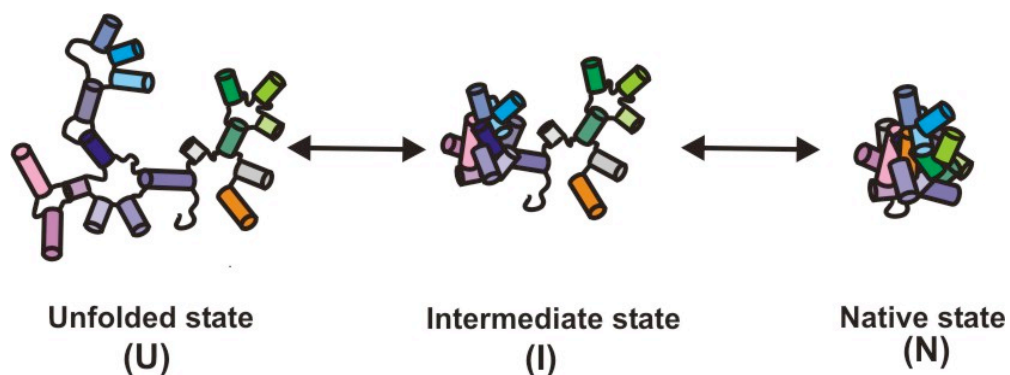
## **1.6 Folding of group II introns**

In order to perform catalysis, large multi-domain group II intron RNA must fold into the correct native conformation. Folding of large RNA molecules into functionally active conformations is a highly complex process. Some RNA molecules exhibit “smooth” folding pathways that consist of slow and direct folding of RNA into the native conformation (79,80). On the other hand, many large RNA molecules fold into the native conformation through “rugged” folding pathways with one or more stable misfolded intermediates or kinetic traps (81,82).

The folding pathway of group II introns has been extensively studied over the last decade using the *Sc. ai5 $\gamma$*  group II intron. The D135 ribozyme (618 nt), which is derived from the yeast *ai5 $\gamma$*  group II intron, is one of the best-characterized model systems for RNA catalysis. D135 lacks domain 6, and

domains 2 and 4 are shortened to hairpins, but it contains the original sequence of domains 1, 3 and 5 to retain catalytic activity (83). As a result, D135 can undergo multiple hydrolytic cleavage reactions with a short substrate (17/7 substrate RNA), which consists of the IBS1, IBS2 and the cleavage site of the 5'-exon (84). Since group II introns can self-splice efficiently under high salt conditions (500 mM KCl, 100 mM MgCl<sub>2</sub>) at 42°C, folding studies of the D135 RNA were primarily performed under these optimal *in vitro* splicing conditions (68,80).

Initial D135 folding studies were conducted using hydroxyl radical footprinting assays, which revealed the presence of solvent-inaccessible regions consistent with the tertiary contacts providing evidence for a compact tertiary structure (85). The global compaction of D135 was further characterized and two main collapsed states were initially identified using native gel analysis and equilibrium ultracentrifugation (80,83). A partially collapsed state with defined secondary structure that forms in the presence of K<sup>+</sup> ions is referred to as the unfolded state. The highly compact structure that forms after addition of Mg<sup>2+</sup> represents the folded state. In addition to that, ultracentrifugation studies revealed the Mg<sup>2+</sup> dependency of global compaction comparable to that of the hydroxyl radical footprinting studies ( $K_{Mg} \sim 20$  mM) (83,86). Urea denaturation studies have shown that the native state can unfold reversibly and both folding and unfolding reactions of D135 have similar thermodynamic parameters. Taken together, these structural studies proposed that the D135 RNA folds into the native conformation with an apparent two-step folding mechanism as shown in Figure 1.5 (80,86).



**Figure 1.5** The proposed two-step folding pathway of the *Sc. D135-ai5 $\gamma$*  group II intron. Domain 1 is shown in blue and purple, domain 3 is shown in green and domain 5 is shown in orange. The folding pathway of the group II intron consists of three structural conformations: U – unfolded state, I – intermediate state and N – native state. Adapted from reference 80.

Further characterization of ai5 $\gamma$ -derived RNA constructs suggested the folding of D1 is a rate limiting step of D135 folding and it requires comparatively high Mg<sup>2+</sup> concentration (80). Therefore, these D135 folding studies suggest the presence of a transient on-pathway folding intermediate consistent with the folding of other large RNA molecules. Although the D135 folding pathway consists of a slow folding rate ( $k_{obs} \sim 1 \text{ min}^{-1}$ ), there was no experimental evidence to support the presence of kinetic traps (83,85).

Specific atoms and functional groups that are critical for compaction of the D135 ribozyme were identified with nucleotide analog interference mapping (NAIM) (61). According to these mapping data, functional groups that are essential for D135 compaction are located in a small region of domain 1 that contains both D5 and D6 docking sites. Therefore, proper positioning of this element may trigger the folding of the entire intron into a compact structure.

In order to understand the function of group II introns *in vivo*, it is essential to characterize the intron RNA under native conditions. Folding studies of the ai5 $\gamma$  group II introns in near-physiological conditions (1 mM MgCl<sub>2</sub>, 100 mM KCl at 30°C) showed similar folding pathway as in optimal splicing conditions. However, the observed compaction of RNA is two orders of magnitude slower than that under optimal conditions (87). Furthermore, under low Mg<sup>2+</sup> concentrations, D1 of group II introns slowly collapses to form the near-native state, an on-pathway folding intermediate with a partially compact structure, which can be converted into the native state by increasing the concentration of magnesium ions (87). Since the

native state is not stable under near-physiological conditions, the group II intron folding pathway requires high ionic strength or protein cofactors for extra stabilization of RNA structures.

### **1.7 Protein-mediated group II intron splicing**

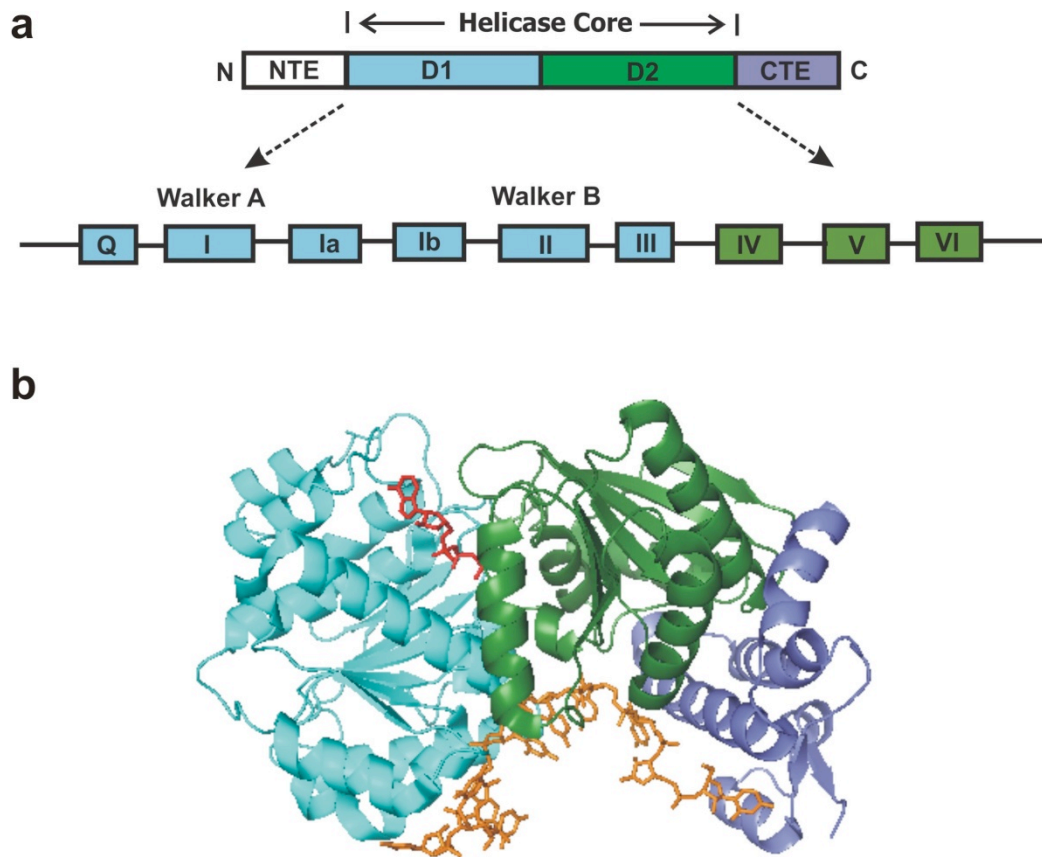
Group II introns can self-splice independently *in vitro* in the presence of high ionic strength and elevated temperatures. However, *in vivo* splicing of group II introns requires proteins to mediate RNA folding into a catalytically active structure. Intron-encoded proteins or maturases are the major proteins involved in the splicing of mobile group II introns. These proteins stabilize the active structure of group II introns after binding specifically with the intron RNA (53). Aside from intron-encoded proteins, several host-encoded proteins have been genetically identified as group II intron splicing factors. These protein factors can function individually or cooperate with other proteins to promote group II intron splicing (52).

Group II intron splicing factors promote intron splicing by stabilizing RNA folding intermediates and/or by resolving non-native structures known as “kinetic traps” in the RNA folding pathway by acting as RNA chaperones (88-90). In addition to mediating efficient splicing of introns, host-encoded splicing factors can also participate in other cellular activities. As a result, it is challenging to understand the exact mechanism of protein-mediated group II intron splicing by indirect biochemical approaches.

### **1.7.1 Structure and function of Mss116**

Mss116, mitochondrial splicing system 116, is a host-encoded protein required for efficient splicing of all mitochondrial (mt) introns in yeast (88-92). It was first identified in a genetic screen as a nuclear gene that affects the mitochondrial electron transport chain of yeast. As a member of the DEAD-box protein family, Mss116 exhibits an RNA-dependent helicase activity required for structural rearrangements of RNA or RNP complexes. Mss116 belongs to the DExH/D subfamily, which contains ATP-dependent nucleic acid remodeling motors required for several cellular processes including translation, ribosome assembly, RNA splicing and RNA degradation. Based on conserved signature motifs, Mss116 can be further classified as a member of the helicase superfamily 2 (SF2) that shows RNA binding, ATPase, RNA unwinding and remodeling activities.

The structure of Mss116 DEAD-box protein (Fig. 1.6) has several important parts including a helicase core, a short N-terminal extension (NTE), mitochondrial targeting sequence (Mt), an  $\alpha$ -helical C-terminal extension (CTE) and a basic tail (91). Comparable to other DEAD-box proteins, the helicase core of this protein contains two RecA-like domains (D1 and D2) joined by a flexible linker. Both C-terminal and N-terminal domains of the helicase core consist of several conserved motifs that are required for RNA and ATP binding as well as for interdomain contacts. The motifs I and II are highly conserved regions of Mss116 (are also known as the Walker A and the Walker B, respectively). Walker B consists of the conserved amino acids D-E-A-D that give the name to this protein family.



**Figure 1.6 Structure of Mss116 DEAD-box protein.** **a.** Schematic structure of Mss116 protein consisting of two RecA folds (cyan and green), an arginine rich C-terminal extension (purple) and an N-terminal extension (white). The helicase core of Mss116 consists of nine conserved motifs that are important for RNA and ATP binding. **b.** Crystal structure of Mss116 containing two domains in the helicase core and C-terminal extension in the presence of single stranded RNA (orange) and AMPPNP (red), a non-hydrolyzable ATP analog. Reproduced from reference (91), Copyright (2009), with permission from Elsevier.

According to recently published Mss116 crystal structures, RNA binds to a basic cleft perpendicular to the interface between two domains, and a non-hydrolysable ATP analog – AMPPNP (Adenylyl-imidodiphosphate) binds onto the opposite side (91). Binding of RNA and ATP changes the conformation of the helicase core from the open state to the closed state as indicated in the crystal structures. The  $\alpha$ -helical C-terminal extension is another essential part of the Mss116 that stabilizes the structure of the helicase core. It has been shown that point mutations within this region decrease RNA binding affinity and ATPase activity.

The crystal structure of Mss116 confirms these findings and indicates that CTE contributes to the RNA binding by extending the RNA-binding side of domain 2 (91). The C-terminal extension is followed by the basic tail, which is predicted to be an unstructured region of the Mss116. Deletion of the basic tail decreases the RNA binding affinity; therefore, it has been proposed that the basic tail of Mss116 may provide an additional RNA binding site for proper functioning of proteins.

Overall, these findings have suggested that the helicase core of Mss116 acts together with the basic C-terminal to perform its functions. Properties of Mss116 have been characterized by various biochemical studies over the last few years (88,89,93,94). As a DEAD-box protein, Mss116 can hydrolyze ATP upon binding of RNA and facilitates the unwinding of short RNA duplexes without apparent directionality (89). Similar to other members of the DEAD-box family, Mss116 is nonprocessive and exhibits comparable RNA-dependent ATPase



activity. In contrast to other RNA helicases (95), Mss116 unwinds short duplexes by local strand separation rather than by translocation along RNA duplexes (93). However, recent studies have shown that DEAD-box protein-mediated RNA unwinding only requires ATP binding, but not ATP hydrolysis (94). This suggests that binding of ATP with the protein changes the conformation of the helicase core and controls the affinity for the RNA. Therefore, it has been proposed that ATP hydrolysis is essential to recycle the enzyme by releasing the bound RNA (94).

### **1.7.2 Mss116-mediated group II intron splicing mechanism**

Previous studies have shown that Mss116 can promote the efficient splicing of more than 13 different mitochondrial group I and group II introns of *Saccharomyces cerevisiae in vivo* (90,92). Inhibition of group I and group II intron splicing in yeast due to defects in the Mss116 gene can be suppressed by the overexpression of CYT19 (90). The DEAD-box protein CYT19 of *Neurospora crassa* is structurally very similar to Mss116 and is primarily involved in proper folding of mt group I introns. In addition to *in vivo* data, recent studies have shown that both Mss116 and CYT19 can promote the splicing of mt group I and group II introns *in vitro* under near-physiological conditions (89,92).

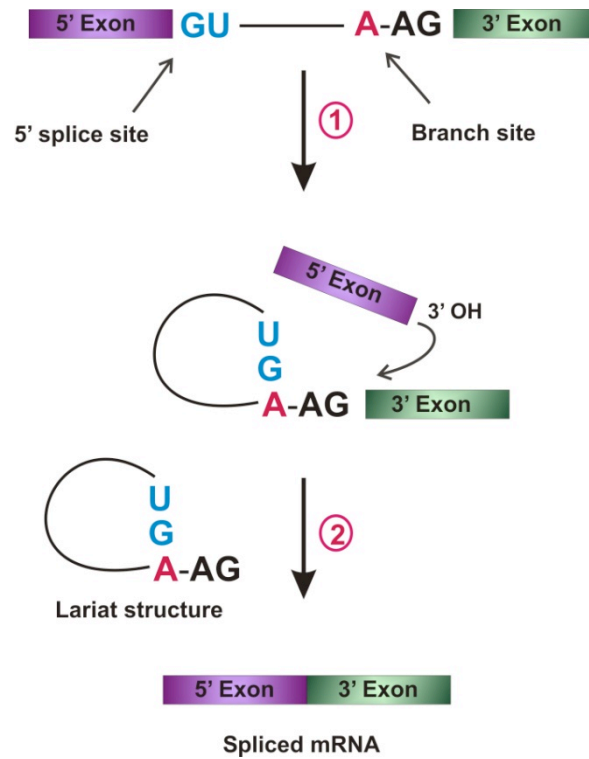
Biochemical studies have shown that CYT19 promotes the splicing of group I introns by resolving stable misfolded structures (kinetic traps) as an RNA chaperone (96). Since CYT19 can promote splicing of mt introns in Mss116 knockout yeast, it has been predicted that both proteins may use an ATP-dependent similar mechanism to facilitate splicing (90,92). However, experiments

conducted with Mss116 mutants that exhibit ATPase activity but are unable to unwind RNA duplexes indicated that unwinding is not strictly required for Mss116-mediated intron splicing as suggested by previous studies (89). Therefore, these studies hypothesized that Mss116 is more likely to promote the splicing of the group II introns by stabilizing on-pathway folding intermediates (89,97). Folding studies of the group II intron ai5 $\gamma$  further support this hypothesis as no evidence was provided for the presence of any off-pathway kinetic traps (68,80). Although there are some experimental data to support both hypotheses, the exact mechanism of the Mss116-mediated group II intron splicing has not been elucidated yet.

## **1.8 Nuclear pre-mRNA splicing**

Pre-mRNA splicing is an essential step of eukaryotic gene expression by which non-coding intron sequences are removed from primary transcripts to form the mature mRNA. Mechanistically the pre-mRNA splicing reaction resembles group II intron splicing and it consists of two transesterification steps. However, in the nucleus, this process is catalyzed by a large ribonucleoprotein complex called the spliceosome, which is a dynamic assembly of five small nuclear RNAs (snRNAs) and a large number of proteins (98).

Introns are often present within eukaryotic protein coding genes and they consist of the 5' splice site, the 3' splice site and the branch site adenosine residue that are essential for splicing reaction (Fig. 1.7).



**Figure 1.7 Eukaryotic nuclear pre-mRNA splicing reaction.** The mechanism of pre-mRNA splicing reaction consists of two consecutive transesterification steps. The first step consists of a nucleophilic attack by the 2'-OH of the conserved branch-point adenosine on the scissile phosphate at the 5'-splice site. In the second step, the 3'-OH of the 5' exon attacks the phosphate bond at the 3'-splice site and ligates the two exons to form the mature mRNA. Adapted from reference 99.

The 5' splice site contains an invariant GU sequence at the 5' end of the intron and the 3' splice site contains an invariant AG sequence at the 3' end of the highly conserved consensus region of the intron as shown in Figure 1.7 (99). The branch site adenosine residue, which is essential for the pre-mRNA splicing reaction, is present upstream of the invariant AG sequence and the polypyrimidine tract that contains several pyrimidines (C and U).

Similar to the branching pathway of group II introns, the spliceosome-catalyzed pre-mRNA splicing reaction (Fig. 1.7) consists of two sequential transesterification steps (99). In the first step, the 2' hydroxyl of a highly conserved adenosine (the branch point) attacks the 5' splice site resulting in a cleaved 5' exon and an intron-3'exon RNA loop known as the lariat.

During the second step, the 3' end of the 5' exon attacks the 3' splice site, cleaving the lariat intron and ligating the two exons together to form a mature mRNA that can be further processed for protein translation (99). Although the pre-mRNA splicing reaction consists of two simple steps, the regulation of this process is critical to maintain proper cellular activities. Defects in this splicing process can lead to numerous diseases including cancer, Alzheimer's and neurodegenerative disorders (100,101).

### **1.9 The spliceosome**

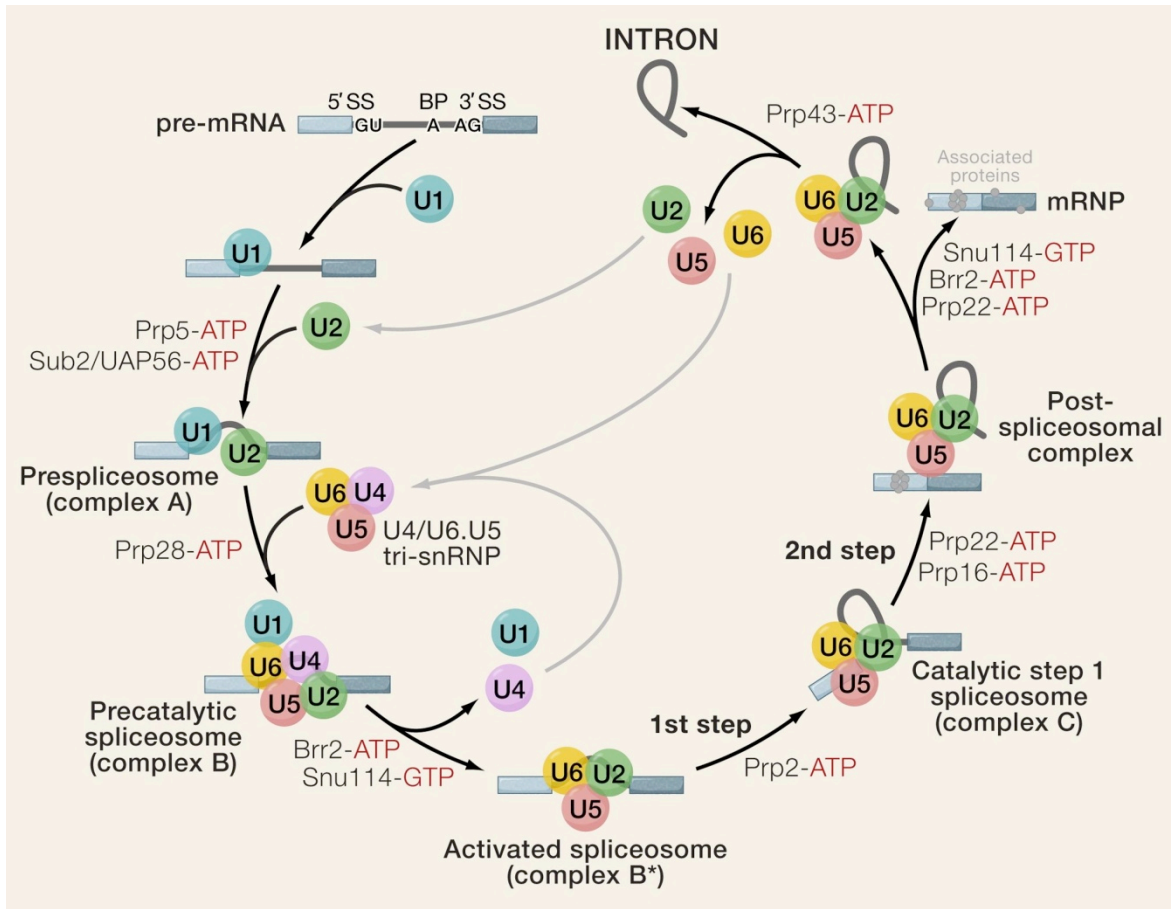
Five small nuclear RNAs – U1, U2, U4, U5, and U6 snRNAs assemble with more than 150 proteins to form small nuclear ribonucleoproteins (snRNPs) in the spliceosome (98,102). RNA-RNA and RNA-protein interactions between the

spliceosomal snRNPs as well as the pre-mRNA and the spliceosomal snRNPs are essential to regulate the splicing process (103).

### **1.9.1 Assembly of the spliceosome**

In order to perform eukaryotic pre-mRNA splicing, spliceosomal subunits must assemble with the pre-mRNA in a sequential manner (98). Assembly of the complex spliceosome has been extensively studied using several methods and it has been found that the pre-mRNA-directed spliceosome assembly pathway (Fig. 1.8) is an ATP-dependent process consisting of several complexes (103). Furthermore, it has been shown that this process is highly conserved from yeast to humans (104).

During the spliceosome assembly, U1 snRNP first recognizes the 5'-splice site and forms the early (E) complex ATP-independently. Association of U2 snRNP with pre-mRNA bulges out the branch point adenosine that is required for the first step of splicing reaction and forms the pre-spliceosome or the complex A. Binding of the U4/U6.U5 tri-snRNPs with the pre-spliceosome forms the pre-catalytic spliceosome or the spliceosomal complex B. The pre-catalytic spliceosome undergoes several structural rearrangements and releases both U1 and U4 snRNPs before it converts to the activated spliceosome or the complex B\* (Fig. 1.8). Both catalytic steps take place in the complex C that consists of U2, U6 and U5 snRNPs, which is known as the mature spliceosome. After the second catalytic step, the post-spliceosomal complex dissociates and releases the mature RNA for translation (103).

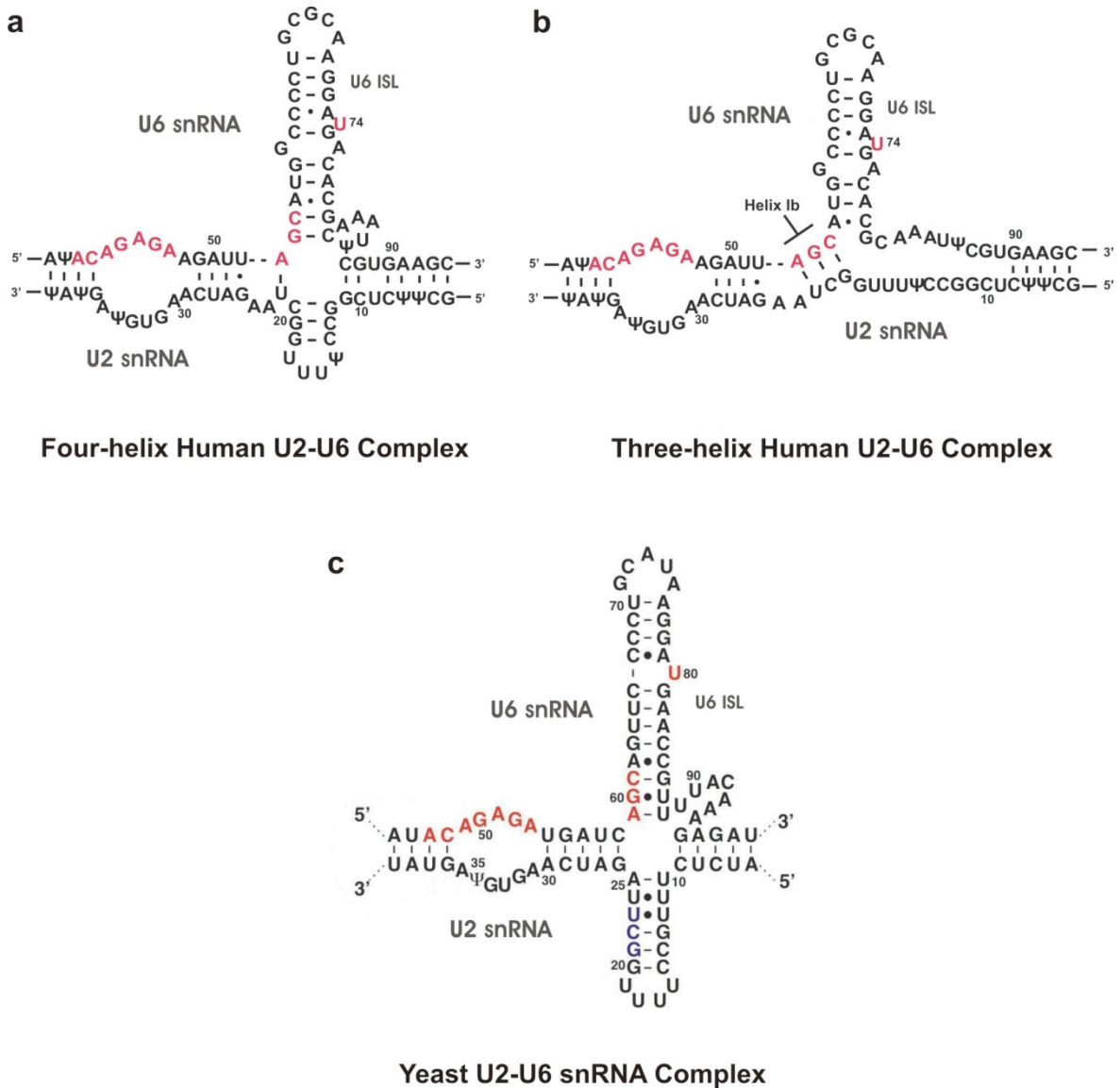


**Figure 1.8 The spliceosome assembly pathway.** Schematic representation of the spliceosome assembly pathway and the pre-mRNA splicing reaction. The nuclear pre-mRNA splicing reaction is catalyzed by the spliceosome. The assembly pathway of the spliceosome consists of stepwise integration of U1, U2, and U4/U6.U5 snRNPs. Reprinted from reference (103), Copyright (2009), with permission from Elsevier.

### 1.10 Spliceosomal U2-U6 snRNA complex

The active spliceosome, or the complex C, consists of U2, U5 and U6 snRNP complexes. However, splicing studies conducted *in vitro* and in mammalian cell extracts have shown that a highly conserved loop of U5 snRNA is dispensable for the catalytic steps of pre-mRNA splicing (105,106). Therefore, despite the abundance of proteins in the spliceosome, it has been suggested that U2 and U6 snRNAs are the critical components for nuclear pre-mRNA splicing (11,13,107).

In addition to mechanistic and structural parallels between the spliceosome and self-splicing group II introns (108-110), there are several pieces of evidence, which support the catalytic ability of spliceosomal RNAs. First, U2 and U6 form direct base pairing interactions with the 5' splice site and the branch point sequence during the first transesterification reaction. U2 binds the branch point directly (111), while the ACAGAG loop in U6 binds the 5' splice site (112). Second, extensive base pairing interactions between U2 and U6 snRNA form a helical junction structure in the active spliceosome, thus bringing the 5' splice site and the branch point into a close proximity. Third, Manley and Valadkhan have reported splicing related catalysis *in vitro* using a protein-free U2-U6 snRNA complex from humans, suggesting that proteins may not be directly involved in catalysis in the holo-spliceosome (11,13,107). Taken together, this evidence strongly suggests that U2 and U6 snRNAs (Fig. 1.9) form the catalytic core of the spliceosome and thus play an essential role during catalysis.



**Figure 1.9 Structure of human and yeast U2-U6 snRNA complexes.** **a.** Proposed structure of four-helix human U2-U6 snRNA complex. **b.** Proposed structure of three-helix human U2-U6 snRNA complex. **c.** Yeast U2-U6 snRNA complex. Top RNA strand is the U6 snRNA and bottom RNA strand is the U2 snRNA. Conserved regions of both human and yeast U2-U6 complex are shown in red (ACAGAGA sequence, AGC triad and  $Mg^{2+}$  binding residue U-74 in human and U-80 in yeast). U6 ISL – U6 intramolecular stem loop.



### **1.10.1 Structure of human and yeast U2-U6 snRNA complex**

Although it has been proposed that U2-U6 complex is essential for pre-mRNA splicing, the structure of U2-U6 complex is still a matter of debate (Fig. 1.9). Based on *in vivo* genetic studies, Guthrie and coworkers have proposed the three-helix structure for yeast U2-U6 complex (113). According to this structure, the highly conserved AGC triad in U6 base pairs with residues in U2 to form helix Ib. Mutational studies further indicated the formation of helix Ib as a requirement for both steps of splicing suggesting the previously proposed three-helix structure with helix Ib as a catalytically active conformation in yeast (114,115).

Similar to yeast U2-U6 complex, early mutational studies of human spliceosomal snRNAs suggests the three-helix structure for the U2-U6 complex in humans (116). However, later genetic studies conducted using a transient transfection assay with mutant U2-U6 snRNAs have indicated that residues involved in the formation of helix Ib in both U2 and U6 snRNAs participate in intramolecular rather than intermolecular base-pairing (117). Therefore, this compensatory mutational study suggests that the human U2-U6 complex forms the four-helix structure without the previously proposed helix Ib (Fig. 1.9).

Interestingly, NMR studies support the four-helix structure for yeast protein-free U2-U6 snRNA complex *in vitro*. According to NMR data, the AGC triad forms intramolecular base pairs within U6 snRNA and extends the U6 intramolecular stem-loop (U6 ISL) by forming the four-helix structure (108,117). Furthermore, these studies proposed that the four-helix structure brings the 5' splice site close

to the U6 ISL to assist the first step of splicing. Although, it has been proposed that the two structures correspond to different spliceosomal activation states, conformational changes between these structures are not well understood.

Despite the fact that spliceosomal snRNAs are highly conserved among different organisms, significant differences have been observed between the U2-U6 snRNA complex of humans and yeast (Fig. 1.9) (118). Human snRNAs contain numerous post-transcriptional modifications such as 2'-O-methyl groups and pseudouridines that are required for function (118,119). In addition, the stem I of human U2 snRNA contains C<sub>10</sub>-A<sub>23</sub> and G<sub>11</sub>-U<sub>22</sub> wobble base pairs instead of the tandem U-U wobble base pairs found in yeast (118). The observed structural variations between human and yeast U2-U6 complexes suggest that the function of these snRNAs in spliceosomal activation may have evolved somewhat differently in these species.

Recent NMR and UV melting studies by Butcher and coworkers have shown that the structures of the human and yeast stem-loop RNAs are nearly isosteric *in vitro*, but the post-transcriptional modifications within the human sequence increase its melting temperature by 18°C, and thus, significantly increase its stability (118). The formation of helix Ib requires the melting of base pairs from U2 stem I. Consequently, the increased stability of modified U2 stem I potentially hinders the formation of helix Ib. Since helix Ib has been implicated with the activation of the spliceosome for the second transesterification reaction, these dissimilarities suggest that different activation mechanisms may take place.

## 1.11 Fluorescence-based detection of biomolecules

Detection of biomolecules and their interactions is necessary for the proper understanding of the highly complex biological processes. There are two main strategies for the detection of biomolecules *in vivo* and *in vitro*: label-free and label-based techniques. Label-free methods allow detecting biomolecules based on their properties without modifying the structure. However, low sensitivity and specificity are main disadvantages of this technique. In contrast to the label-free methods, labeling of biomolecules with fluorescent dyes or radioisotopes increases the sensitivity of the detection method (120).

Fluorescent-based techniques are one of the common label-based detection methods that enable characterization of the structure and function of biomolecules with high sensitivity (121). Based on unique properties of the fluorescent labels, these techniques are widely used in both *in vivo* cellular imaging and *in vitro* assays for direct detection of biomolecular interactions and structural studies.

Development of fast and sensitive detection techniques together with efficient fluorophore-labeling methods has increased the applicability of fluorescence in RNA science. Over the last few years, fluorescent-based detection techniques have been extensively used to study the complex RNA folding pathways and interactions between RNA-RNA and RNA-proteins in various biological systems.

### **1.11.1 Fluorescence theory**

The emission of light from any molecule in the excited state is known as luminescence. Based on the properties of the excited state, luminescence can be divided into two types: fluorescence and phosphorescence. Phosphorescence occurs when a molecule emits the light from the triplet-excited state. In contrast, fluorescence occurs due to the emission of light from a molecule when it returns to the electronic ground state from the singlet-excited state.

Fluorescence is a common physical phenomenon that has fast emission rates ( $10^9 \text{ s}^{-1}$ ) and exhibits lifetimes of  $\sim 10 \text{ ns}$ . According to the Jablonski diagram (Fig. 1.10a), the absorption of light initiates the excitation of molecules from the ground state ( $S_0$ ) to the singlet-excited state ( $S_1$  or  $S_2$ ) containing a number of vibrational levels. Molecules in higher excited vibrational levels can be relaxed to the lowest vibrational levels through nonradiative processes such as vibrational relaxation (VR) and internal conversion (IC) within  $\sim 10^{-12} \text{ s}$ . Fluorescence occurs when molecules emit light and reach the ground state from the lowest excited vibrational state. However, the loss of energy by means of intersystem crossing (ISC) between the singlet state (S) and the triplet state (T) shifts emission light into a higher wavelength than that of absorption light.

### **1.11.2 Fluorescence Resonance Energy Transfer (FRET)**

Fluorescence resonance energy transfer (FRET) is a distance-dependent, non-radiative energy transfer process between a donor (D) and an acceptor (A) fluorophore attached to a biomolecule (122). Upon excitation of the donor, the

energy is transferred to the acceptor and increases its emission intensity while decreasing the emission intensity of the donor fluorophore in an anti-correlated manner (Fig. 1.10b). Energy transfer arises from a dipole-dipole interaction between fluorophores. The efficiency of this process ( $E_{FRET}$ ) is given by,

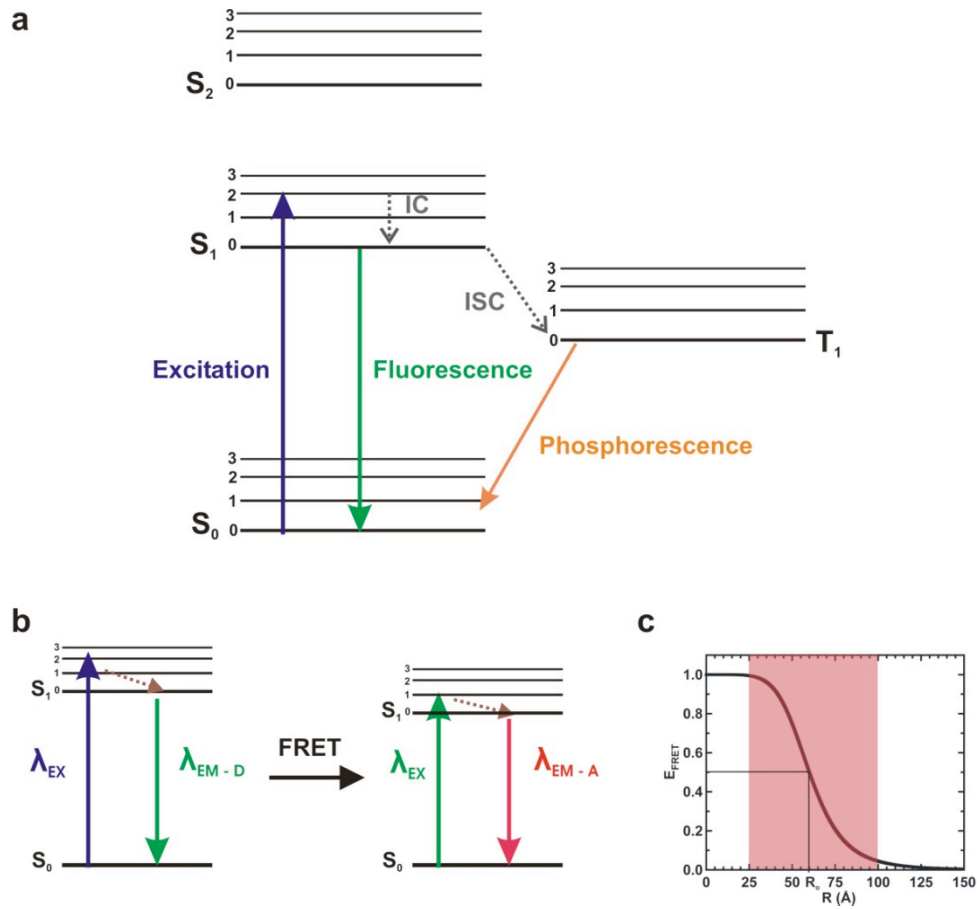
$$E_{FRET} = \frac{1}{[1 + (R/R_0)^6]} \quad (1)$$

$R$  is the distance between two fluorophores and  $R_0$  is the Förster distance, the distance at which 50% energy transferring efficiency is observed (Fig. 1.10c).  $R_0$  is specific for a given pair of fluorophores and typically range between 3-8 nm (36). Förster distance can be calculated using following equation:

$$R_0^6 = 8.8 \times 10^{23} \kappa^2 n^{-4} Q_D J \quad (2)$$

$\kappa^2$  is a factor describing the orientation of the dipoles,  $n$  is the refracting index of the medium,  $Q_D$  is the quantum yield of donor in the absence of acceptor fluorophore and  $J$  is the spectral overlap integral.

The efficiency of this energy transfer is related to the inverse sixth power of the ratio of the distances between fluorophores. Therefore, FRET can act as a powerful molecular ruler that measures the intra- and inter-molecular motion of biomolecules in the 2-10 nanometer distance range (122,123). In addition to the distance between donor and acceptor fluorophores, FRET efficiency depends on several factors including the orientation of the donor and acceptor dipole moments and the spectral overlap between the emission spectrum of the donor and the excitation spectrum of the acceptor.



**Figure 1.10 Jablonski energy diagram.** **a.** Jablonski energy diagram that explains the general theory of the fluorescence. **b.** Jablonski energy diagram for FRET. **c.** Förster distance ( $R_0$ ) for Cy3 and Cy5. Vibrational levels of the ground state ( $S_0$ ) and the excited states ( $S_1$  &  $S_2$  singlet states and  $T_1$  triplet state) are shown in black lines. Excitation, fluorescence and phosphorescence processes are indicated using arrows. IC - internal conversion; ISC - intersystem crossing;  $\lambda_{EX}$  – excitation wavelength and  $\lambda_{EM}$  – emission wavelength.

### **1.11.3 Single-molecule fluorescence**

Folding of RNA into an active conformation is a key requirement for proper functioning of RNA in cells. Due to the rugged nature of folding potential energy surfaces, RNA can follow complex kinetic pathways and adopt multiple intermediate conformations before reaching the native state. In order to understand the folding process, UV melting, Isothermal Titration Calorimetry (ITC), Circular Dichroism (CD) and chemical footprinting have been extensively used over the last few decades. However, all these traditional techniques deal with ensemble-averaged populations of a large number of molecules, which limits the amount of the information that can be obtained if multiple conformations are present in solution (36).

The emergence of single-molecule techniques has provided a huge advance in characterization of biomolecules, because it enables the direct observation of transient intermediates in isolated molecules, thereby resolving static and dynamic molecular heterogeneity and eliminating the requirement to synchronize all of the observed molecules to derive quantitative kinetic information. Therefore, single-molecule spectroscopy helps to deconvolute how complex biological systems function by revealing the kinetic steps otherwise hidden in ensemble-averaged bulk experiments.

Since the first single-molecule RNA folding study a decade ago, single-molecule fluorescence has become a widely used technique to study the mechanisms of folding and catalysis in various RNA-based systems. The power of

single-molecule detection combined with the sensitivity of FRET has made single-molecule FRET (smFRET) the most widely-used real-time detection technique to monitor structural dynamics, kinetics and thermodynamics of biopolymers. In order to achieve single-molecule sensitivity, it is critical to reduce all sources of background fluorescence (124). To this aim, two main techniques are often used: confocal and total internal reflection fluorescence (TIRF) microscopy. Both of these techniques reduce background fluorescence by limiting the detection volume.

In confocal microscopy, the out-of-focus light transmitted to the detector is reduced by a small aperture in the image plane of the microscope and detection is effectively limited to a small (femtoliter-size) volume, thus increasing the signal-to-noise ratio by suppressing the background signal. Confocal microscopy helps detect freely diffusing molecules in solution (125). However, the short dwell time of these molecules in the detection volume is a primary limitation of this technique. In order to overcome this limitation, two main strategies have been developed during the past years: immobilization of the sample in agarose or polyacrylamide gels (126) and the surface immobilization of molecules by specific interactions, such as a biotin-streptavidin linkage (127).

In contrast to confocal microscopy, the prism-based TIRF microscopy aids in the detection of molecules by exciting with an evanescent wave generated by total internal reflection of a laser beam (Fig. 1.11). In order to totally reflect the laser beam, the angle of the incidence ( $\theta$ ) has to be larger than the critical angle



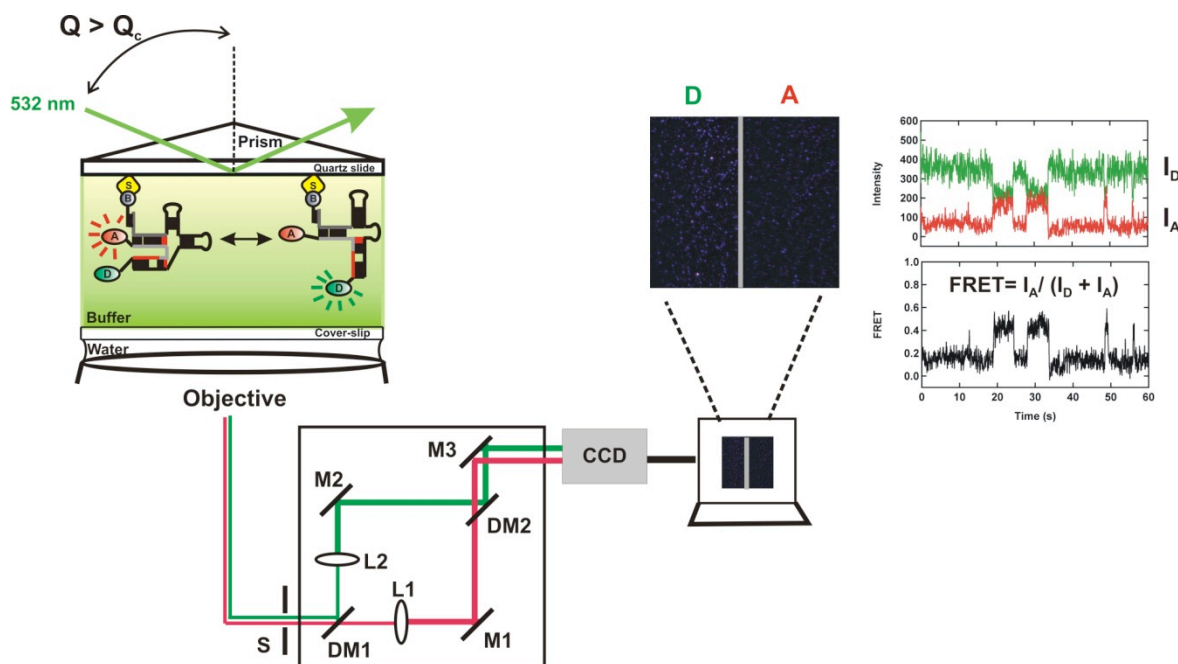
( $\theta_c$ ). The evanescent wave generated at the slide-water interface by the incident light can excite molecules up to ~100-150 nm away from the slide surface (128). The emission of the donor and acceptor fluorophores can be collected through an inverted microscopic objective and transferred into a light-sealed box through a slit.

The first dichroic mirror (DM1) physically separates the donor and acceptor intensities and allows them to pass through the lenses (L1 and L2) to amplify the image. The second dichroic mirror (DM2) obtains the separated donor and acceptor signals through the mirrors (M1 and M2) and recombines the separated emission signals side-by-side to redirect into a high quantum yield CCD camera. The CCD camera amplifies the signals and transfers the digitalized frames to a computer for data analysis (Fig. 1.11).

Based on emission fluorescence intensities, the efficiency of FRET ( $E_{FRET}$ ) can be obtained by the equation below,

$$E_{FRET} = \frac{I_A}{I_A + I_D} \quad (3)$$

where  $I_A$  and  $I_D$  are the acceptor and donor emission intensities, respectively. The reduction of the excitation volume to a thin sheet at the interface between the slide and the solution decreases the background signal and excessive photobleaching of molecules outside the detection volume. Furthermore, this technique facilitates the detection of a large number of molecules simultaneously in a wide observation field with the use of a CCD camera.



**Figure 1.11 Setup of prism-based total internal reflection fluorescence (TIRF) single-molecule microscopy.** Green line indicates the 532 nm excitation laser beam. The fluorescence emission intensity of the sample is collected by the objective and transferred to a light-sealed-box through a slit (s). Emission intensities of the donor (D) and acceptor (A) are separated using a dichroic mirror (DM) and directed to a CCD camera through the lenses (L) and a series of mirrors (M). Emission intensities are recorded by the CCD camera and the amplified signals are transferred to a computer to obtain corresponding FRET trajectories.  $Q$  – angle of incidence,  $Q_c$  – critical angle,  $I_A$  – acceptor emission intensity and  $I_D$  – donor emission intensity.

These features increase the applicability of this technique by alleviating some of the disadvantages of confocal microscopy. However, the time resolution that can be obtained by current CCD cameras is still lower than that of the avalanche photodiodes (APD) and photomultipliers in confocal fluorescence microscopy (36,129).

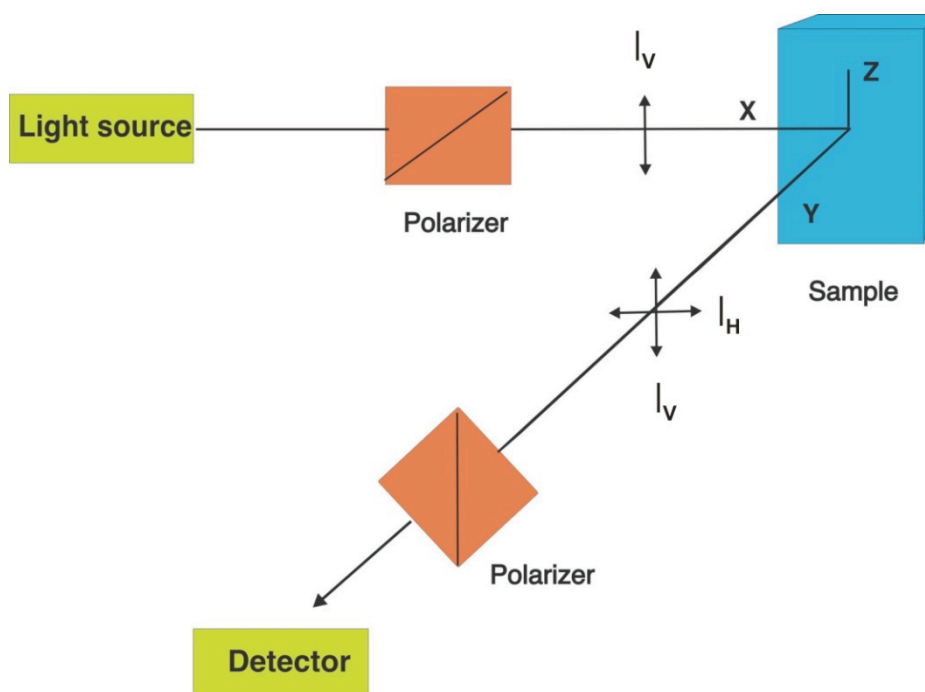
#### **1.11.4 Fluorescence anisotropy**

Fluorescence anisotropy is another widely used fluorescent-based technique that detects the rotational diffusion of a molecule based on the fluorescence polarization (130,131). Anisotropic measurements provide information about the size and the shape of molecules. Furthermore, this technique measures the binding interactions between two molecules; thereby permitting the determination of the binding constant or the dissociation constant for the specific interaction.

During fluorescence anisotropy assays, fluorophores are excited by vertical polarized light and emission intensities are measured in both vertical and horizontal directions through a polarizer (Fig. 1.12). Based on emission intensities, fluorescence anisotropy ( $r$ ) can be calculated by,

$$r = \frac{I_V - I_H}{I_V + 2I_H} \quad (4)$$

where  $I_V$  and  $I_H$  are the fluorescent intensities in both vertical and horizontal directions, respectively.



**Figure 1.12 Schematic diagram showing the principle of fluorescence anisotropy.** Fluorophore labeled molecules in the sample cuvette are excited with vertically polarized light and the emission intensity of the vertically polarized fluorescence ( $I_V$ ) and the horizontally polarized fluorescence ( $I_H$ ) are measured.

According to the theory of fluorescence anisotropy, excitation polarized light selectively excites the molecules that have an absorption transition dipole parallel to the electric vector of excitation light. Based on the rotational freedom of molecules, this selective excitation results in a partially oriented population of polarized fluorescence emission. Therefore, binding of a large molecule with fluorophore-labeled substance decreases the rotational movements or the rate of tumbling and increases the value of anisotropy due to highly polarized emission light. In contrast, small molecules show fast rate of tumbling and depolarize the emitted light relative to the excitation plane decreasing the value of anisotropy.

Fluorescence anisotropy has a lower limit of detection in the sub-nanomolar range and allows monitoring the real-time interactions between biomolecules without using any hazardous materials. Therefore, anisotropy based detection methods have numerous advantages over most of the traditional techniques that are used to study biological interactions between proteins and nucleic acids.

### **1.12 Specific aims**

Group II introns and nuclear pre-mRNA splicing have been studied for many years using various experimental methods. However, thorough characterization of these splicing mechanisms is not feasible with most of these ensemble methods, which are unable to identify transient intermediates. The main goal of my dissertation was to combine fluorescence-based detection techniques with single-molecule approaches to characterize RNA splicing mechanisms in eukaryotes. To achieve this goal, the following specific aims were studied:

**Specific aim 1:** *Characterization of group II intron folding pathway and the effect of Mss116 DEAD-box protein and ATP on the group II intron splicing mechanism.*

Group II introns can self-splice *in vitro* under high ionic strength conditions. *In vivo*, this process requires protein cofactors such as Mss116. The objective of this aim is to use single-molecule approaches to characterize group II intron splicing by monitoring protein-mediated RNA folding in real time.

**Specific aim 2:** *Characterization of structural dynamics of the U2-U6 snRNA complex that are essential for pre-mRNA splicing in humans.* Several studies have

suggested the primary contributors of spliceosome-catalyzed pre-mRNA splicing are the U2-U6 snRNAs. However, the structure and mechanism of the U2-U6 snRNA complex involved in splicing remains debatable. To address this issue, we have employed a novel single-molecule approach to investigate the structural dynamics of the human U2-U6 snRNA complex in real time.

**Specific aim 3:** *Development of an RNA aptamer based affinity chromatographic method for purification of fluorophore-labeled proteins under non-denaturing*

*conditions.* Although it has been proven that fluorescence-based detection techniques are capable of elucidating the structure and mechanisms of complex biological systems in detail, the application of these techniques can be limited due to poor labeling efficiencies of biomolecules. In order to address this issue, we sought to develop an RNA aptamer-based affinity chromatographic method for purification of fluorophore-labeled proteins from unlabeled molecules under native conditions.

## CHAPTER 2

### Materials and Experimental Methods

#### 2.1 Materials

Fluorophore-labeled (U6 and U6-1 RNA) and unlabeled RNA strands with or without modified bases (U2 and U6-2 RNA) as well as DNA oligonucleotides (Cy3-DNA, Cy5-DNA and Biotin-DNA) were purchased from The Howard Hughes Medical Institution Biopolymer/Keck Foundation Biotechnology Resource Laboratory (Yale University, New Haven, CT). D135-L14 group II intron RNA (638 nt) and SRB-2 RNA (54 nt) were synthesized by *in vitro* transcription using T7 RNA polymerase. SRB-2 template DNA was purchased from Sigma Genosys (Woodlands, Texas) and all NTPs (ATP, CTP, GTP & UTP) were purchased from Acros Organics (Morris Plains, New Jersey).

Tetramethylrhodamine-5-maleimide (TAMRAM) for protein labeling was purchased from AnaSpec, Inc., San Jose, CA and Cy5-NHS esters for RNA labeling was purchased from Amersham Biosciences. PEG-coated slides were made with 3-aminopropyltriethoxysilane (Vectabond, Vector Laboratories, Inc., Burlingame, CA), biotin polyethylene glycol succinimidyl carboxymethyl (BIO-PEG-SCM, 3400/5000 MW) and methoxy polyethylene glycol succinimidyl carboxymethyl (m-PEG-SCM, 5000 MW, Laysan Bio. Inc., Arab, AL). Mss116 was expressed and purified as explained in the methods section (37). <sup>32</sup>P-labeled nucleotide triphosphates were obtained from MP Biomedicals, LLC. (Solon, OH).

All chemicals used in these studies were purchased from either Sigma-Aldrich or Fisher Scientific.

## **2.2 Methods: Single-molecule analysis of Mss116-mediated group II intron folding**

### **2.2.1 Preparation of DNA, RNA and protein samples**

The D135-L14 RNA (638 nt) was obtained by *in vitro* transcription of the HindIII digested pT7D135-L14 plasmid DNA with T7 RNA polymerase under standard conditions (68). Biotinylated DNA strand (5'-Biotin-TGC ATG CCT GCA GGT CGA CTC TA-3'), Cy3-DNA strand (5'-Cy3-ACC AAG AGC GTT ATT AAT-5'), Cy5-DNA strand (5'-Cy5-ACG TAG TCC GAA ATA TAT-3'), wild type RNA substrate (17/7, 5'-CGU GGU GGG ACA UUU UCG AGC GGU-3') and slower cleaving RNA substrate (17/7-dC, 5'-CGU GGU GGG ACA UUU UdC GAG CGG U-3') were purchased from The Howard Hughes Medical Institution Biopolymer/Keck Foundation Biotechnology Resource Laboratory (Yale University, New Haven, CT). The 2'-OH positions were deprotected following the manufacturer's protocol. All samples were purified by denaturing 18% PAGE and subsequent C8 reversed-phase HPLC, as described (37).

Mss116 was overexpressed in Rosetta 2 cells (Novagen) and grown at 16°C overnight after induction. After purification of protein on a nickel column, it was cleaved with SUMO protease at 4°C overnight and purified using a gel filtration column. Untagged protein was concentrated by an Amicon ultra



concentrator and stored at  $-80^{\circ}\text{C}$ . The activity of Mss116 was confirmed by *in vitro* splicing, ATPase and unwinding assays.

### **2.2.2 Characterization of Mss116 activity**

Activity of purified Mss116 was characterized by conducting *in vitro* splicing, unwinding and ATPase assays as explained below.

**Splicing assay** - The body-labeled full-length ai5 $\gamma$  precursor RNA was incubated in 40 mM MOPS pH 7.5 at  $90^{\circ}\text{C}$  for 1 min followed by 3 min at  $30^{\circ}\text{C}$ . The RNA solution (1 nM final concentration) was added to a reaction mixture containing 20 nM Mss116, 100 mM KCl, 40 mM MOPS pH 7.5 and 8 mM  $\text{MgCl}_2$ , 5% glycerol, 2 mM DTT and 2U/ $\mu\text{L}$  RNase inhibitor. After incubation for 10 min at  $30^{\circ}\text{C}$ , the splicing reaction was initiated with the addition of ATP (1 mM final concentration). The sample aliquots were removed at 0 and 120 min time points and detected using 5% native polyacrylamide gel.

**Unwinding assay** - 250 nM Mss116 was pre-incubated with 0.1 nM end-labeled RNA duplex (12 bp) at  $30^{\circ}\text{C}$  for 10 min, and the unwinding reaction was initiated by the addition of ATP as described (89). The unwinding reaction was performed in 100 mM KCl, 40 mM MOPS pH 7.5, 4 mM  $\text{MgCl}_2$ , 0.01% IGEPAL, 2 mM DTT, 2U/ $\mu\text{L}$  RNase inhibitor and 4 mM ATP. Sample aliquots were removed at 0 and 30 min time points and detected using 10% native PAGE.

**ATPase assay** - 250 nM Mss116 protein was preincubated with or without 5  $\mu\text{M}$  single-stranded RNA at  $30^{\circ}\text{C}$  for 10 min, and all reactions were initiated by the addition of a mixture of P-32 labeled and unlabeled ATP. ATPase assays were

performed according to standard protocol (132) in 100 mM KCl, 40 mM MOPS, pH 7.5, 1 mM MgCl<sub>2</sub>, 2 mM DTT, 2U/μL RNase inhibitor and 1 mM ATP.

### **2.2.3 Single-molecule FRET experiments**

Protein-free single-molecule experiments were performed by heat-annealing (90°C for 45 seconds) Cy3-DNA, Cy5-DNA, Biotin-DNA (10 μM each) and D135-L14 RNA (1 μM) in 100 mM KCl, 40 mM MOPS pH 7.5 and 0.5% 2-mercaptoethanol. After the addition of 8 mM MgCl<sub>2</sub>, the reaction mixture was incubated at 30°C for 15-20 min. The fluorophore-labeled, biotinylated RNA/DNA complex was diluted to ~25 pM and bound to a streptavidin-coated quartz slide via a biotin-streptavidin linkage. Excess fluorophore-labeled and biotin-labeled DNA strands were removed from the slide by washing with reaction buffers. To minimize the non-specific binding of protein on the slide surface, all quartz slides were PEG-passivated using the BIO-PEG-SCM and the m-PEG-SCM.

In order to reduce photobleaching of fluorophores, an oxygen-scavenging system (OSS) consisting of 10% (w/v) glucose, 2% (v/v) 2-mercaptoethanol, 50 μg/ml glucose oxidase and 10 μg/ml catalase was used in all experiments. The smFRET experiments with Mss116 were performed by introducing 25 nM Mss116 with and without 1 mM ATP/AMPPNP in OSS solution as described (37). Similarly, smFRET experiments with the substrate (17/7 or 17/7-dC) were performed after addition of 25 nM substrate RNA into OSS solution containing 25 nM protein and 1 mM ATP. All TIRF-based smFRET experiments were performed with 33 ms time

resolution at room temperature under different experimental conditions, as described (68).

FRET histograms from different experimental conditions were constructed by collecting ~100 single-molecule trajectories showing the transitions between different structural conformations. Since we cannot distinguish molecules in the unfolded state (U, FRET ~ 0) from those containing only a donor fluorophore (FRET = 0, due to photobleaching, for example), we did not include them in the FRET histograms.

Folding rate constants ( $k_1$ ,  $k_{-1}$ ,  $k_2$  and  $k_{-2}$ ) of the D135 ribozyme under different experimental conditions were obtained by analyzing FRET trajectories with a Hidden Markov Model (HMM) (133). First, FRET trajectories were fit using HMM and then transition density plots (TDP) were constructed to obtain folding rate constants as described (37). Threshold FRET values 0.27 and 0.51 were used to separate transitions among the three different FRET states. The effect of Mss116 without ATP was further determined in the presence of glucose and hexokinase. Incubation of the reaction mixture with glucose and hexokinase removes the trace amount of ATP associated with sample preparation.

## **2.3 Methods: Single-molecule studies of human spliceosomal U2-U6 snRNAs**

### ***2.3.1 RNA purification and labeling***

5'-Cy3 labeled full length U6 RNA (5'-AUA CAG AGA AGA UUA GCA UGG CCC C**d**TG CGC AAG GAU GAC ACG CAA AUU CGU GAA GC-3'), U6-1 RNA (5'-Cy3-AUACAGAGAAGAUUAGCAUGGCCCCU-3'), U6-2 RNA (5'-AGG AUG ACA CGC AAA UUC GUG AAG C-3' Biotin) and U2 RNA (5' GCU UCU CGG CCU UUU GGC UAA GAU CAA GUG UAG UAU 3') with or without modified bases (pseudouridine and 2'-O-methyl guanosines) were purchased from The Howard Hughes Medical Institution Biopolymer/Keck Foundation Biotechnology Resource Laboratory (Yale University, New Haven, CT). The 2'-OH deprotection reaction was performed as the manufacture's protocol and the deprotected RNA was purified by denaturing 18% PAGE and subsequent C8 reversed-phase HPLC as described (68). The 3' end of Cy3-HU6-1 and the internal dT of full length HU6 RNA were labeled with Cy5 fluorophore and purified by reversed-phase HPLC.

### ***2.3.2 Steady-state FRET (ssFRET)***

The ratio of U6-1:U6-2:U2 was chosen as 1:2:4 for complete hybridization of fluorophore labeled U6-1 with the other two RNA strands in 130  $\mu$ l solution containing 100 mM NaCl, 50 mM TRIS-HCl pH 7.5 and 25 mM DTT. The samples were heat-annealed at 90°C for 1.5 min and incubated at room temperature for 15-20 min. The ssFRET assay was performed by adding different concentrations of

Mg<sup>2+</sup> (0-100 mM) at room temperature, and monitored using a Cary Eclipse fluorescence spectrometer. Cy3 was excited at 550 nm and emission of the donor and acceptor were recorded at 565 and 665 nm to obtain the FRET value using the equation,  $FRET = I_A / (I_A + I_D)$ . The subsequent FRET values were plotted against Mg<sup>2+</sup> concentration and fitted to a binding equation as explained (134).

$$FRET = FRET_0 + (FRET_0 - FRET_\infty) \frac{[Mg^{2+}]^n}{K_{Mg}^n + [Mg^{2+}]^n} \quad (5)$$

where  $FRET_0$  and  $FRET_\infty$  are the initial and final FRET efficiencies, respectively,  $K_{Mg}$  is the Mg<sup>2+</sup> dissociation constant, and n is the cooperativity coefficient.

### **2.3.3 Single-molecule FRET (smFRET)**

Fluorophore-labeled U6 (1 μM) and U2 (5-10 μM) were heat-annealed (90°C for 45 s) in 100 mM NaCl, 50 mM TRIS-HCl pH 7.5 with different Mg<sup>2+</sup> concentrations and incubated for 15-20 min at room temperature. In order to prevent U6 dimers, single-molecule experiments were performed in the presence of excess amount of U2 RNA (U6:U2 = 1:5 or 1:10) as suggested by previous studies (135) and the native polyacrylamide gel electrophoresis (native PAGE).

The fluorophore-labeled, biotinylated U2-U6 RNA complex was diluted to ~25 pM and bound to a streptavidin-coated quartz slide via a biotin-streptavidin linkage. Unbound RNA strands were removed from the slide by washing with reaction buffers. In order to reduce photobleaching of fluorophores, an oxygen-scavenging system consisting of 10% (w/v) glucose, 2% (v/v) 2-mercaptoethanol, 50 μg/ml glucose oxidase and 10 μg/ml catalase was used in all experiments.

FRET histograms at different  $[Mg^{2+}]$  were constructed by collecting more than 100 molecules showing transitions between different states. Dwell times were calculated and plotted in dwell time histograms to obtain the folding rate constants as explained (134). Differences between free energy changes associated with the formation of the low FRET state for unmodified (U) and modified (M) U2-U6 snRNA constructs were calculated as  $\Delta\Delta G = \Delta G_M - \Delta G_U$ , where  $\Delta G = -RT \ln(K)$  and  $K = A_{0.2}/A_{0.4}$  (A represents the area of a given FRET state in a FRET histogram). Error associated with  $\Delta\Delta G$  was analyzed using the bootstrap method as described (136). The bootstrap method generates an auxiliary histogram by resampling each original data point in a histogram. The standard deviation of this auxiliary histogram provides a better estimation of the error associated with original data.

## **2.4 Methods: An RNA Aptamer Based Purification System for Fluorophore Labeled Proteins**

### **2.4.1 RNA Synthesis and purification**

The SRB-2 aptamer RNA (5'-GGA ACC UCG CUU CGG CGA UGA UGG AGA GGC GCA AGG UUA ACC GCC UCA GGU UCC-3') was synthesized by *in vitro* transcription of the DNA template (5'-GGA ACC TGA GGC GGT TAA CCT TGC GCC TCT CCA TCA TCG CCG AAG CGA GGT TCC **TAT AGT GAG TCG TAT TAC TG**-3'). DNA was purchased from Sigma Genosys (Woodlands, Texas)

and all NTPs (ATP, CTP, GTP & UTP) were purchased from Acros Organics (Morris Plains, New Jersey).

*In vitro* transcription was carried out with 0.4  $\mu$ M of the DNA template and 0.4  $\mu$ M of the T7 promoter (5'-CAG TAA TAC GAC TCA CTA TA-3'), transcription buffer (80 mM HEPES-KOH pH 7.5, 40 mM DTT and 2 mM Spermidine), polyethylene glycol (80 mg/ml), NTPs (4 mM each),  $Mg^{2+}$  (24 mM) and T7 RNA Polymerase (0.15 mg/ml) in 1 ml. The reaction mixture was incubated at 37°C for 4 hrs and the transcribed products were purified by phenol-chloroform extraction, followed by ethanol precipitation. Transcripts were dissolved in 100  $\mu$ l of formamide gel loading buffer and purified using 20% denaturing polyacrylamide, 8 M urea gel electrophoresis followed by gel elution and ethanol precipitation. The resulting concentrations were determined by UV absorbance at 260 nm.

The transcription product was biotin labeled at the 3' terminus in order to immobilize the SRB-2 aptamer on streptavidin-coated beads. The biotinylation occurs as a two-step reaction; the first of which converts the 2' and 3' hydroxyl groups of the 3' terminal nucleotide to electrophilic aldehydes. This is followed by the nucleophilic addition of biotinhydrazine (137). In the first step, 20  $\mu$ L SRB-2 aptamer (700  $\mu$ M) were vacuum-dried and dissolved in 100  $\mu$ L oxidation buffer (10 mM Na-HEPES pH 7.5, 1 mM  $MgCl_2$ ). Then 14  $\mu$ L 50 mM  $KIO_4$  were added and the reaction was incubated for 20 minutes at 0°C. The RNA was then purified by ethanol precipitation for nucleophilic addition of biotinhydrazine. The dried RNA was dissolved in 100  $\mu$ L of 50 mM NaOAc, pH 5.0 and 1  $\mu$ L saturated

biotinhydrazine solution was added to the reaction mixture. Saturated biotin solution was prepared by dissolving 11 mg biotinhydrazine in 250  $\mu$ L distilled water. After incubating the reaction mixture overnight at 4°C, the biotinylated product was collected by ethanol precipitation and its concentration was measured by UV/Visible spectrophotometry. The biotinylation efficiency was estimated to be approximately 50% by HPLC.

### ***2.3.2 Fluorophore labeling of protein***

Modified maltose binding protein containing a single cysteine residue was labeled with tetramethylrhodamine-5-maleimide (TAMRAM) by a maleimide-cysteine reaction. The modified protein (2 nM) in 50 mM Tris-HCl pH 7.2 was incubated with 1  $\mu$ l of TCEP (132 nM/ $\mu$ l) for 1 hour at room temperature. Protein was labeled by incubating overnight at 4°C with a 5-fold molar excess of TAMRAM (AnaSpec, Inc., San Jose, CA). Excess fluorophore was removed from the labeled protein by dialysis with selection buffer (100 mM KCl, 10 mM Na-HEPES, pH 7.4 and 5 mM Mg<sup>2+</sup>) using a microcon YM-3 centrifugal filter as described by the manufacturer (Millipore Corporation). The concentrations of protein and tetramethylrhodamine (TAMRA) were calculated from their absorption at 280 and 550 nm, respectively, as described (138).

### ***2.3.3 Fluorescence anisotropy measurements***

Fluorescence anisotropy measurements were used to monitor the binding of TAMRA with SRB-2 aptamer in solution. TAMRA (AnaSpec, Inc., San Jose, CA)



was excited at 555 nm and all emissions were collected at 585 nm with 0.1 sec average time and 5 nm slit width (G factor = 2.1). Fluorescence anisotropy was calculated after measuring the emission intensities in vertical ( $I_V$ ) and horizontal ( $I_H$ ) directions. All anisotropic measurements were performed using a spectrofluorimeter (Cary Eclipse, Varian Inc.) in a solution containing 100 mM KCl, 10 mM Na-HEPES pH 7.4 and 250 nM TAMRA with varying concentration of SRB-2 aptamer,  $Mg^{2+}$  and EDTA. All measurements were taken in duplicate over a period of 3 minutes at 1 minute intervals. Initial anisotropy ( $r_0$ ) of a TAMRA-only solution was low compared to that of a mixture of TAMRA and RNA due to the free rotation of the fluorophore molecules. The binding of RNA molecules to TAMRA increases fluorescence anisotropy by restricting TAMRA's free rotation. The resulting SRB-2 aptamer,  $Mg^{2+}$  and EDTA titrations were fit to the following equations, respectively:

$$r = r_0 + (r_{\max} - r_0) \frac{(K_D + [R] + [SRB]) - \sqrt{(K_D + [R] + [SRB])^2 - (4[R][SRB])}}{2[R]} \quad (6)$$

$$r = r_0 + (r_{\max} - r_0) \frac{(K_{Mg} + [R.A] + [Mg]) - \sqrt{(K_{Mg} + [R.A] + [Mg])^2 - (4[R.A][Mg])}}{2[R.A]} \quad (7)$$

$$r = r_{\max} - (r_{\max} - r_0) \frac{[EDTA]^n}{(K_{EDTA}^n + [EDTA]^n)} \quad (8)$$

where  $r_0$  and  $r_{\max}$  are the fluorescence anisotropies of free and SRB-2 bound TAMRA at saturation and  $[R]$  and  $[R.A]$  are the initial TAMRA and TAMRA- aptamer concentrations, respectively.

### **2.3.4 Affinity column protein purification assay**

Streptavidin coated 6% crosslinked agarose beads (Pierce Biotechnology, Rockford, IL) solution (1 mL) was loaded into a 2 mL column and allowed to settle. The resulting column bed was then washed by passing 5 column volumes (5 x 500 $\mu$ L) of binding buffer (100 mM KCl, 10 mM Na-HEPES, 5 mM MgCl<sub>2</sub>, pH 7.4) through the column. Then, 6.5 nmol of biotinylated SRB-2 aptamer were added to the column bed followed by 500  $\mu$ L binding buffer. The biotinylated SRB-2 was allowed to bind with streptavidin-coated beads by incubating at room temperature for 30 minutes, before washing the column with 5 column volumes of binding buffer. Next, 0.4 nmol of the TAMRA-labeled maltose binding protein mixture was added to the column in one 200  $\mu$ L aliquot and allowed to bind 30 minutes at room temperature. Excess protein, unlabeled protein and impurities were washed from the column using 5-8 column volumes of binding buffer. Each column volume was then analyzed by both fluorescence and UV-Visible spectrophotometric methods.

UV-Visible absorbance was measured to monitor the presence of MBP ( $A_{280}$ ) and TAMRA ( $A_{550}$ ) in buffer washings, and fluorescence spectra were recorded at  $\lambda_{\text{ex}} = 550$  nm to determine the presence of excess TAMRA ( $\lambda_{\text{em}} = 580$  nm) labeled proteins. Labeled proteins were then eluted by adding one column volume elution buffer (5 mM EDTA, 100 mM KCl, 10 mM Na-HEPES pH 7.4). The chelation of Mg<sup>2+</sup> was carried out for 10-15 minutes at room temperature. Complete elution of labeled proteins was achieved by passing 5 additional column volumes of elution buffer. Each column volume was then analyzed as described above.

## CHAPTER 3

# Single-molecule Analysis of Mss116-mediated Group II Intron Folding

(Adapted from Karunatilaka, K. S. *et. al. Nature*. 2010)

### 3.1 Introduction

Group II introns are large multi-domain RNA molecules that can self-splice *in vitro* providing more evidence for the catalytic ability of RNA. The structure, splicing mechanism and folding pathway of this intron have been introduced in detail in the chapter 1.

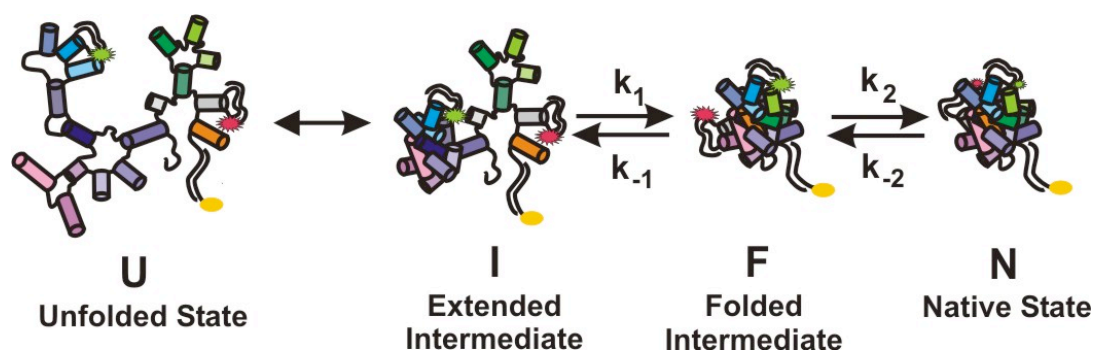
The D135 ribozyme, the minimal active form derived from the group II intron ai5 $\gamma$  of *Saccharomyces cerevisiae*, consists of the necessary components for catalysis and is used as a model system to study group II intron folding and catalysis (Fig. 1.3). Previous studies have described the two-step folding pathway (U  $\leftrightarrow$  I  $\leftrightarrow$  N) for the D135 ribozyme (Fig. 1.5) which includes one obligatory intermediate under optimal *in vitro* splicing conditions (86). The unique folding pathway of this group II intron consists of an initial slow compaction step of domain I (D1), followed by rapid folding of the remaining domains onto the D1 scaffold to form the native state (N).

The unfolded state (U) is an extended form of the group II intron that forms in the presence of monovalent ions, thus it consists of secondary interactions, but lacks tertiary interactions. In the presence of both monovalent and divalent ions,

the U state folds slowly and forms the transient intermediate (I) with a collapsed D1 structure. This slow folding step ( $U \leftrightarrow I$ ) is recognized as the rate limiting step of the group II intron folding pathway (80). Furthermore, it has been found that the initial collapse of domain 1 is mainly controlled by the region between  $\kappa - \zeta$  in the center of D1 (61). This region is also involved in the binding of metal ions and the docking of the highly conserved D5 that forms the active site of the group II intron (61,139). The second step of folding ( $I \leftrightarrow N$ ) is comparatively fast and has not been well characterized by traditional detection methods.

Previous studies have shown that the addition of monovalent ions decreases the hydrodynamic radius ( $R_H$ ) of the molecule from 146 to 81 Å due to the formation of an RNA conformation containing secondary interactions (86). Upon addition of divalent ions, the  $R_H$  further decreases to 61 Å indicating the formation of tertiary interactions (86). Single-molecule FRET (smFRET) has a unique ability to study the folding of molecules within 100 Å in real-time by detecting the transient intermediates that can be hidden in ensemble-averaged experiments. Therefore, the smFRET technique has been used to study the folding pathway of group II introns using D135 RNA as a model system (36).

Characterization of the fast folding step by TIRF-based smFRET experiments revealed three distinct structural conformations with a previously unidentified on-pathway folding intermediate (Fig. 3.1): I - extended intermediate state, F - folded intermediate state, and N - native state (68).



**Figure 3.1 The folding pathway of the D135-L14 ribozyme.** The folding pathway consists of four different structural conformations: U-unfolded state, I-extended intermediate state, F-folded intermediate state, and N-native state.  $k_1$ ,  $k_{-1}$ ,  $k_2$  and  $k_{-2}$  are the rate constants for the structural transitions between different states. Adapted from reference (68) with permission from PNAS.

Furthermore, the observed fast folding rates between different states suggests these conformations are connected by small activation barriers. FRET time trajectories indicated that the native conformation occurs only transiently at 20 mM or higher  $Mg^{2+}$  concentrations. However, the binding of the substrate with the intron stabilizes the high FRET conformation; thus, this structure likely acts as the active conformation of the ribozyme *in vitro*. Interestingly, it was found that increasing  $[Mg^{2+}]$  leads to an increase in structural dynamics. The fraction of dynamic molecules versus  $[Mg^{2+}]$  yields a dissociation constant that coincides with the values for bulk folding and cleavage (68). This result relates the D135 structural dynamics with its catalytic activity and implies that domain motion may contribute to efficient catalytic activity of introns.

The splicing of group II introns *in vivo* is assisted by protein cofactors such as Mss116 (90). Mss116 is a DEAD-box protein belonging to the helicase superfamily 2 that has RNA-dependent ATPase and helicase activities. In addition to *in vivo* activity of Mss116, recent studies have shown its ability to facilitate the splicing of group II introns under near-physiological conditions (100 mM KCl, 40 mM MOPS pH 7.5 and 8 mM  $MgCl_2$  at 30°C) *in vitro* (88,89).

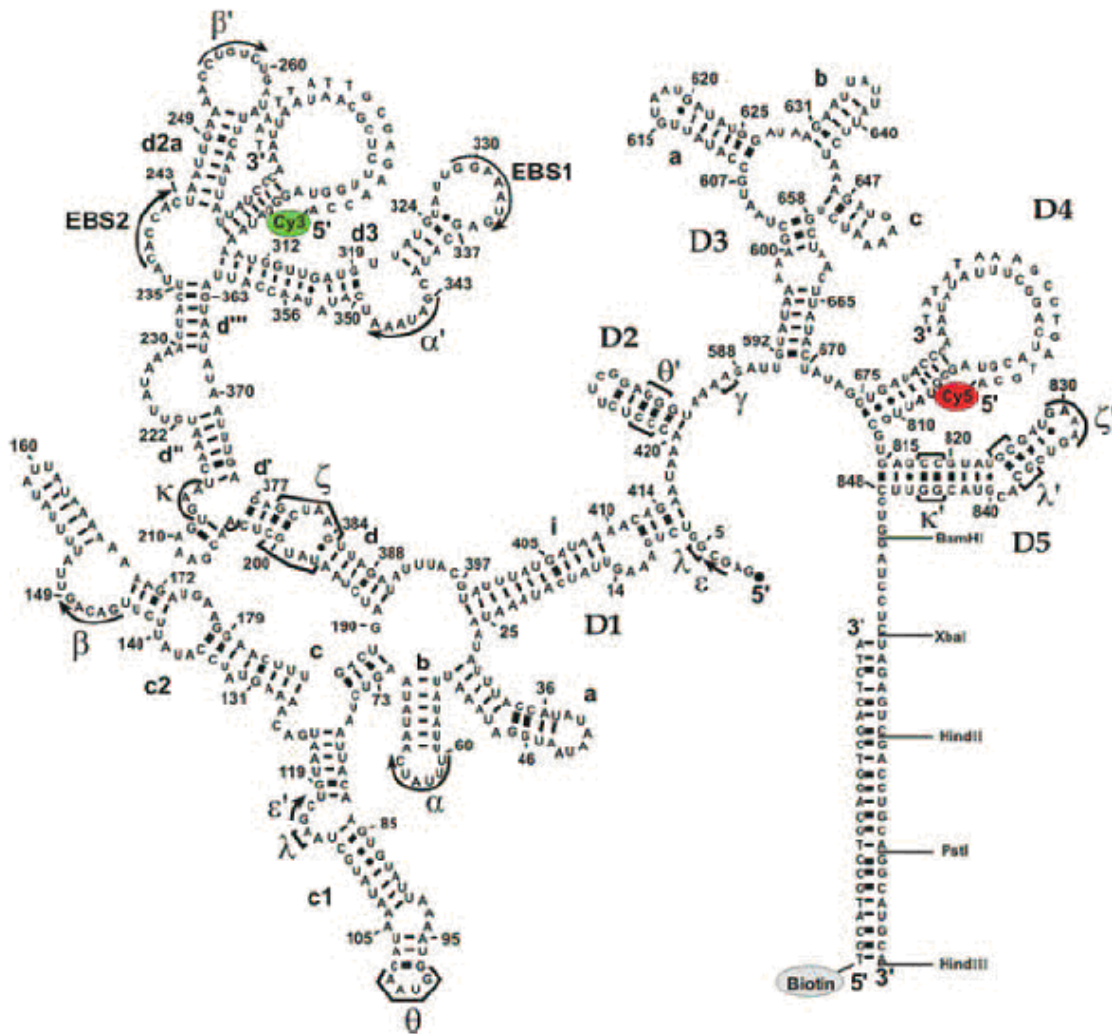
Based on the properties of Mss116 DEAD-box protein, it has been proposed that Mss116 can promote group II intron splicing either by stabilizing on-pathway intermediates or by disrupting misfolded structures (88,89,97). However, the exact mechanism of Mss116 activity on the group II intron splicing has not been elucidated yet. Studies conducted using other systems have shown that both

mechanisms are possible. For example, the LtrA protein encoded by a group IIA intron promotes self-splicing by binding the intron RNA (140). RNA chaperones, such as StpA, bind nonspecifically and unfold RNA to allow refolding into the active conformation (141,142).

Previous work has used the catalytic activity of group II introns to report on Mss116 function. However, to further understand how a DEAD-box protein facilitates RNA-folding, it is necessary to observe RNA folding directly. Therefore, we have characterized the effect of Mss116 on the folding dynamics of a fluorophore-labeled D135 ribozyme (D135-L14) using single-molecule fluorescence resonance energy transfer (68). Thereby we will be able to understand the mechanism of protein-mediated splicing *in vivo*. Furthermore, direct detection of interactions between Mss116 and its natural partner, ai5 $\gamma$  group II intron, will expand the existing knowledge of DEAD-box proteins in the RNA field.

### **3.2 Experimental design**

In order to perform single-molecule fluorescence experiments, it is essential to label D135 RNA with a fluorophore pair (donor and acceptor) that can satisfy all the requirements of FRET. Small RNA strands (<100 nts) can be easily fluorophore-labeled by attaching the fluorescent tags at terminal positions or internally during the solid-phase chemical RNA synthesis.



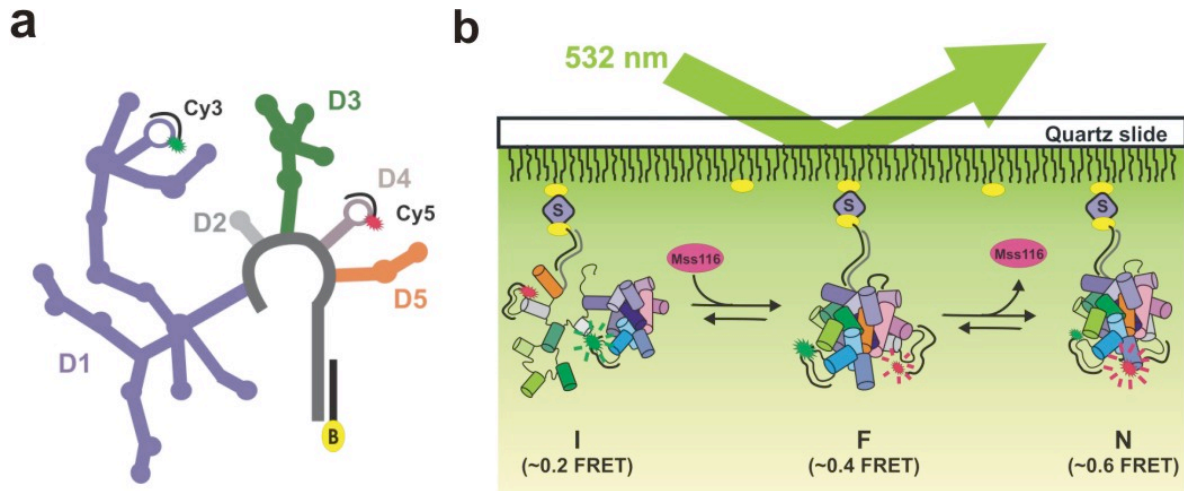
**Figure 3.2 Secondary structure of the *Sc.* D135-L14 ribozyme.** D135-L14 ribozyme is (638 nt) derived from the group II intron *Sc.ai5 $\gamma$*  in the *cox I* gene of *Saccharomyces cerevisiae* (*Sc.*) D1 and D4 domains hybridize with DNA oligonucleotides labeled with Cy3 (green) and Cy5 (red) fluorophores, respectively. The 3' end of the intron is elongated to bind with 5'-biotinylated complementary DNA strand for surface immobilization in smFRET experiments. The tertiary interactions are labeled by Greek letters. Adapted from reference (68) with permission from PNAS.



In contrast to small RNA constructs, the labeling of large RNA molecules (>100 nts) with fluorescent dyes is highly challenging, because long RNA molecules are difficult to chemically synthesize with a significant yield. Recent studies have found several methods for fluorophore labeling of large RNA molecules. The ligation-based labeling technique is one successful approach that can be applied for fluorophore labeling of large RNA molecules (143). However, the efficiency of this method is highly dependent on the length of the RNA construct.

In addition to RNA ligation, hybridization of short fluorophore-labeled DNA or RNA oligonucleotides with single-stranded loop regions permits efficient labeling of large RNA molecules (144). This method promotes labeling of RNA molecules without forming covalent interactions. Because of its large size, D135 RNA was fluorophore-labeled using the hybridization technique as previously described (68).

In order to label D135 RNA with donor and acceptor fluorophores, two single-stranded loop regions have been introduced into two different locations (D1 and D4) that are not absolutely essential for catalytic activity of group II introns (Fig. 3.2). Then, modified D135-L14 intron RNA was hybridized with DNA probes (Cy3-DNA and Cy5-DNA) as described (68). Furthermore, the 3' end of D135 RNA was extended and hybridized with another DNA strand labeled with biotin (Biotin-DNA) to immobilize the whole fluorophore-labeled, biotinylated complex on a quartz slide (Fig. 3.3).



**Figure 3.3 Single-molecule detection by Total Internal Reflection (TIR) fluorescence microscopy.** **a.** The secondary structure of fluorophore-labeled and biotinylated D135-L14 RNA. Fluorophore-labeled DNA oligonucleotides (Cy3 Donor - green, Cy5 Acceptor - red) and biotin (yellow) labeled DNA oligonucleotide are shown in black. **b.** single-molecule setup showing the D135 RNA complex immobilized on PEG-passivated quartz slide via a biotin-streptavidin linkage. The sample was excited by a laser beam at 532 nm and the emission was recorded by the CCD camera through an objective.

Folding dynamics of fluorophore-labeled group II introns in the presence and absence of Mss116 and ATP were detected by the TIRF-based single-molecule fluorescence. To minimize non-specific binding of protein, the fluorophore-labeled and biotinylated D135-L14 ribozyme construct was immobilized on polymer (Polyethylene Glycol - PEG) passivated quartz slide via a biotin-streptavidin linkage (Fig. 3.3).

### **3.3 Results**

#### ***3.3.1 Folding of group II introns alone requires high ionic strength***

The folding pathway of the group II intron D135-L14 has been studied using single-molecule FRET in variable ionic conditions (68). Consistent with previous results, in high ionic strength conditions (8 mM  $MgCl_2$ , 500 mM KCl and 80 mM MOPS pH 7.5 at 22°C), the smFRET time trajectories of D135-L14 display repeated stochastic transitions among three distinct conformational states (Fig. 3.4a): an extended intermediate (I, FRET  $\sim$  0.1), a folded intermediate (F, FRET  $\sim$  0.4) and the native state (N, FRET  $\sim$  0.6). Among all transitions, fewer than 2% of molecules show direct transitions between the extended intermediate and the native states, indicating that the folded intermediate state is an obligatory folding intermediate, as previously described (68).

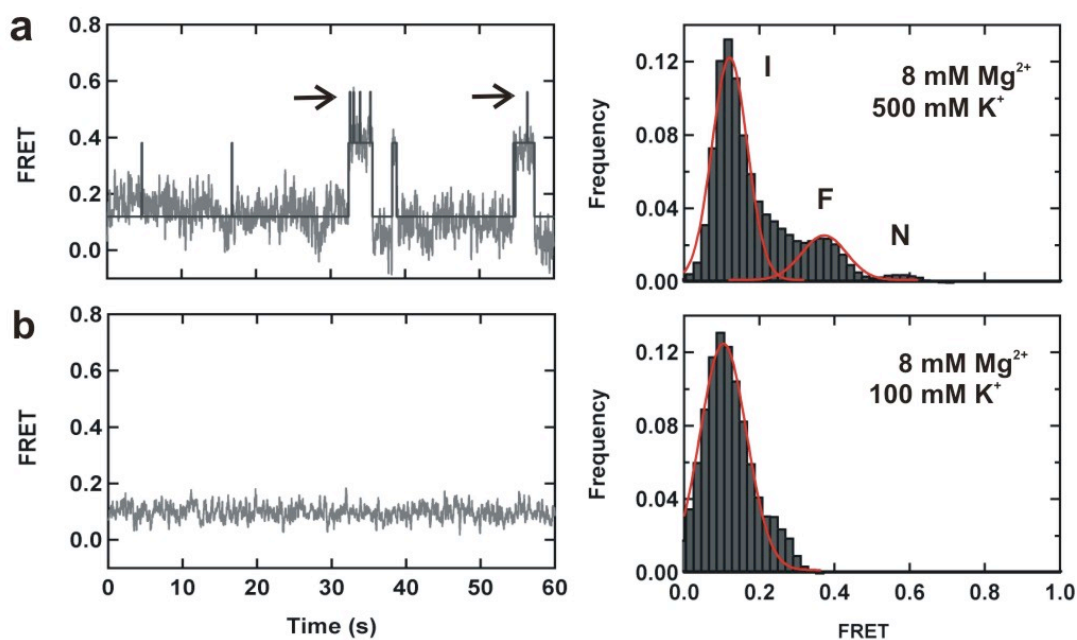
In orthogonal studies employing gel-shift methodologies, the I, F and N states of D135 co-migrate as a single 'compact' species, and the functional native state can only be differentiated by the presence of catalytic activity (61,87).

Nonetheless, these two approaches are complementary because the gel shift assay can easily report on the transition between the unfolded state (U, FRET  $\sim$  0) and the intermediate state (61,87,97), but the smFRET assay cannot because the U state is indistinguishable from species that are only labeled with the donor fluorophore (68,145).

In this study, the fluorophore-labeled D135-L14 was surface-immobilized via a biotin-streptavidin linkage on poly(ethylene glycol)-coated (PEG) quartz slides to reduce non-specific binding of RNA and proteins (146,147). However, the observed FRET histogram and traces are similar to those previously reported indicating that the PEG coating does not affect the behavior of D135-L14 alone (68).

In order to understand the behavior of group II introns *in vivo*, the folding dynamics of D135 RNA were monitored *in vitro* at the single-molecule level under near-physiological conditions (8 mM MgCl<sub>2</sub>, 100 mM KCl and 40 mM MOPS pH 7.5 at 22°C). In contrast to the optimal splicing conditions described earlier, the D135-L14 ribozyme displays only one FRET distribution (FRET  $\sim$ 0.1) under near-physiological conditions (Fig. 3.4b). According to the observed FRET trajectories, the vast majority of molecules reside in the extended intermediate state under these low ionic strength conditions (Fig. 3.4b).

Compared to RNA molecules in high ionic strength conditions, most of the observed molecules in low ionic strength were static and did not exhibit structural transitions between different conformations (compare Fig. 3.4a and b).



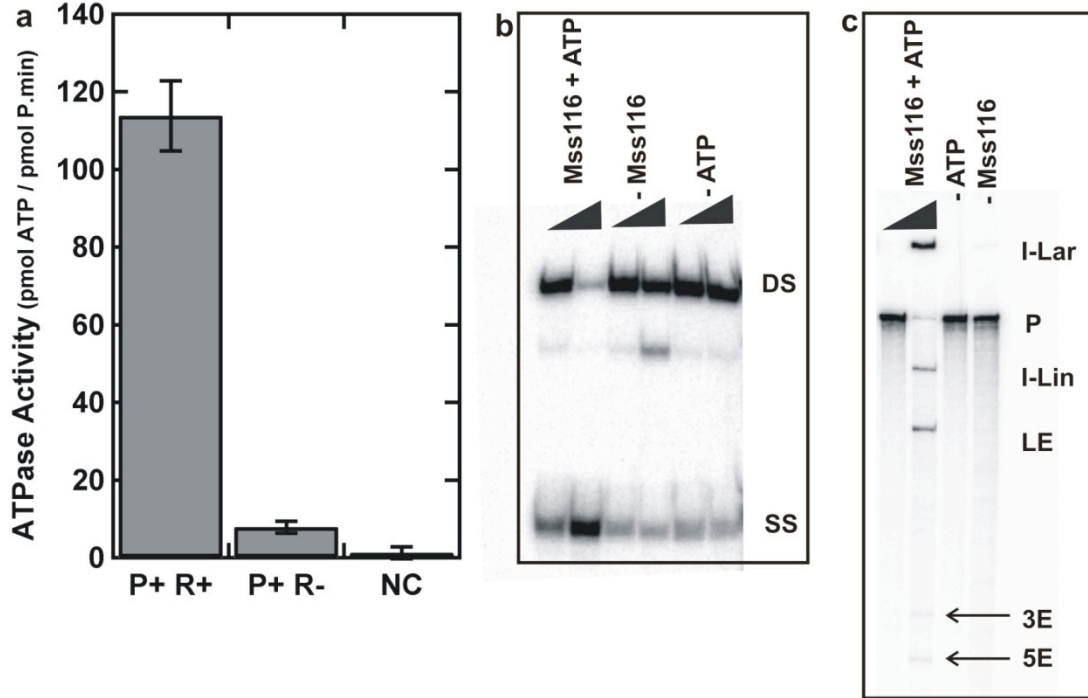
**Figure 3.4 Effect of ionic strength on the folding dynamics of D135-L14 ribozyme.** **a.** Typical smFRET trajectory (left) showing the dynamics of a single D135-L14 ribozyme under high ionic strength conditions, and FRET histogram representing the distribution of FRET states of ~100 trajectories (right). The FRET trace indicates the presence of three different conformations: **I** (FRET~0.1), **F** (FRET~0.4) and **N** (FRET~0.6). Black line represents idealized HMM and arrows indicate rapid transient excursions to the N state. **b.** Typical FRET trajectory (left) showing the dynamics of a single D135-L14 ribozyme under near-physiological conditions, and histogram representing the distribution of FRET states from ~100 trajectories (right).

The differences in the FRET distributions of the D135-L14 ribozyme in the two different ionic strengths confirm that under low ionic strength conditions the RNA folds primarily to the intermediate conformation and lacks structural dynamics, as previously observed (68,87).

### **3.3.2 *Mss116 facilitates group II intron folding in vitro***

Since the splicing of ai5 $\gamma$  group II introns *in vivo* requires Mss116 as a protein cofactor, we characterized the effect of Mss116 on the folding of D135-L14 under near-physiological conditions. The activity of our recombinant Mss116 was first confirmed by ATPase, unwinding and *in vitro* splicing assays (Fig. 3.5).

Mss116 is a DEAD-box protein, which uses RNA-dependent ATP hydrolysis activity to perform its function. Therefore, the initial experiments were conducted in the presence of 1 mM ATP. The observed smFRET trajectories (Fig. 3.6a, b and c) and distributions of D135-L14 in the presence of Mss116 exhibit transitions between all three conformational states (I, F and N), in contrast to the smFRET distribution in the absence of protein. The appearance of the higher FRET states in the presence of the protein cofactor indicates that Mss116 facilitates folding of the RNA. In the presence of Mss116, ~30% of molecules reach the native state transiently or stably including a minor population of molecules (~9%) that remain static in the N state. To examine this effect more closely, we monitored the FRET ratio of a single D135-L14 ribozyme before and after addition of Mss116 and ATP (Fig. 3.6d).



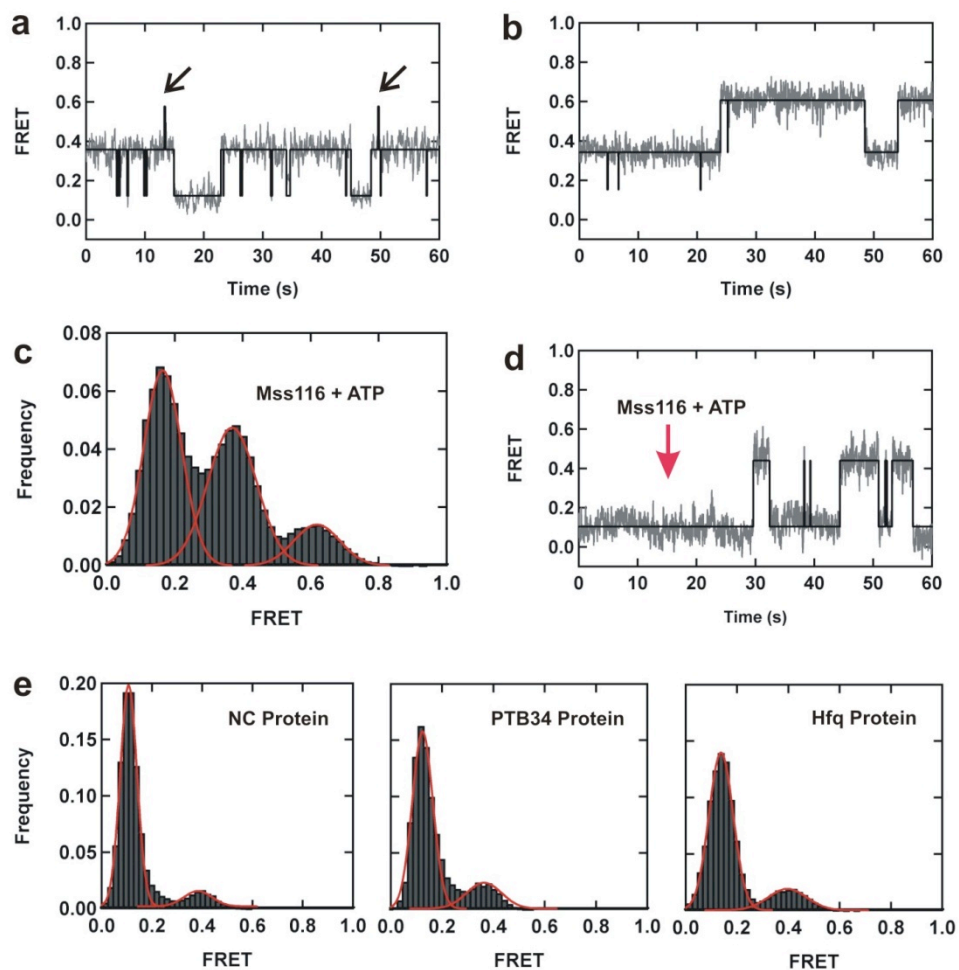
**Figure 3.5 ATPase, unwinding and splicing activity of the Mss116.** **a.** ATPase activity of Mss116 in the presence and absence of RNA. NC represents the negative control in the absence of both protein and RNA. **b.** Unwinding of a 12 bp duplex with a 3'-overhang in the presence of Mss116, RNA and KCl. The figure shows the results obtained in the presence of Mss116 and ATP, without Mss116 and without ATP at 0 and 30 min time points. DS - double-stranded and SS - single-stranded RNA. **c.** Splicing of ai5y in the presence of Mss116 and ATP at 0 and 120 min time points. Splicing was not observed in the absence of ATP or Mss116 even after 120 min. I-Lar – intron lariat, P – precursor, I-lin – linear and/or broken lariat intron, LE – ligated exons, 3E – 3' exon and 5E – 5' exon.

Before introduction of Mss116 and ATP, D135-L14 appears to be static in the extended intermediate state (FRET  $\sim 0.1$ ). After addition of Mss116 and ATP ( $\sim 15$  s), transitions from the low FRET conformation to higher FRET conformations were observed, providing direct evidence of Mss116-mediated RNA folding.

To evaluate the protein specificity of Mss116-induced folding, we examined the folding dynamics of D135-L14 in the presence of three different basic RNA-binding proteins (Fig. 3.6e): HIV nucleocapsid - NC (148), Polypyrimidine tract binding - PTB (149) and Hfq proteins (150). The corresponding smFRET histograms show all three proteins can populate the folded intermediate to a lesser extent than Mss116, and none promotes folding into the native state. Among the molecules observed in the presence of each non-specific protein, more than 90% appear static in either the I or F state.

Lowering the  $[\text{Mg}^{2+}]$  to 1 mM yielded similar results, showing that Mss116 can also help D135-L14 form F and N states under physiological conditions (Fig. 3.7). However, we also observe a higher number of molecules in the 0 FRET state, suggesting that the number of molecules in the unfolded state is higher at 1 mM  $[\text{Mg}^{2+}]$ . Interestingly, at both 8 and 1 mM  $\text{Mg}^{2+}$ , we find that the average FRET ratio of the extended intermediate increases to  $\sim 0.18$  in the presence of Mss116 (Table. 3.1), indicating that the protein induces a slight overall compaction of this conformer.



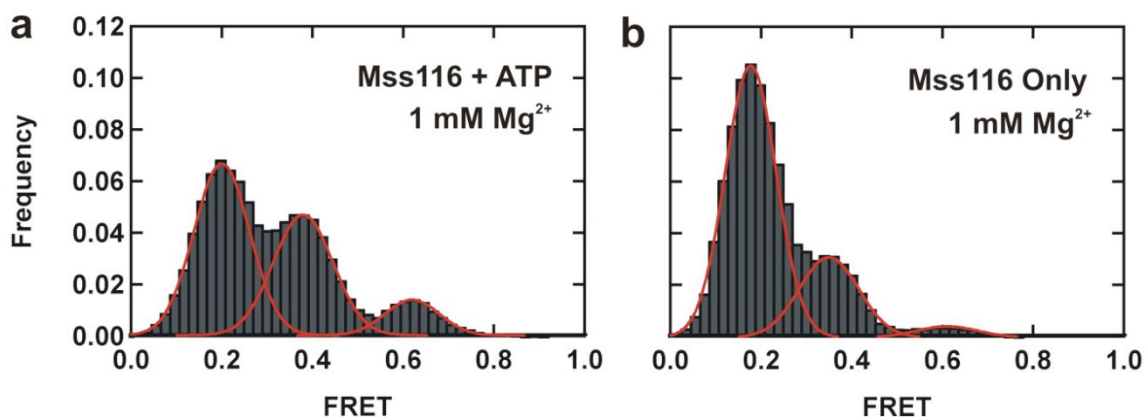


**Figure 3.6 Mss116 promotes folding of group II introns at near-physiological conditions.** **a** and **b**. FRET trajectories of a single D135-L14 ribozyme in the presence of Mss116 and ATP. Black lines represent idealized HMM and arrows indicate transient excursions to the native state. **c**. Histogram of the distribution of three structural conformations. **d**. FRET trajectory of the dynamic behavior of a single ribozyme upon addition of Mss116 and ATP at 15-20 seconds. **e**. FRET state distribution in the presence of the indicated non-specific RNA binding proteins. NC - nucleocapsid protein, PTB34 - Polypyrimidine tract binding protein, Hfq – host factor for RNA phage Q beta replication protein.

Unlike Mss116, none of these non-specific proteins affects the average FRET ratio of the extended intermediate, indicating that Mss116 may recognize this conformation more specifically (Table. 3.1). Taken together, these data show that the formation of the folded intermediate (F), which is the slow, rate-determining step along the folding pathway (151), can be facilitated by a diverse set of basic RNA binding proteins. However, the native state is only formed in the presence of Mss116 and ATP. Thus, Mss116 plays a specific role during the final, rapid stages of ai5 $\gamma$  folding.

Experimental Conditions	I – State			F – State			N- State		
	Peak Center	Area	Peak Height	Peak Center	Area	Peak Height	Peak Center	Area	Peak Height
500 mM KCl	0.12	0.70	0.121	0.37	0.15	0.024	0.61	0.03	0.003
100 mM KCl	0.11	0.85	0.124	ND	ND	ND	ND	ND	ND
P Only	0.16	0.60	0.086	0.37	0.30	0.042	0.63	0.05	0.006
P + ATP	0.16	0.40	0.067	0.37	0.40	0.048	0.62	0.13	0.015
P + AMPPNP	0.17	0.60	0.078	0.37	0.30	0.034	0.62	0.05	0.006
P + ATP+17/7	0.18	0.45	0.069	0.38	0.30	0.043	0.60	0.12	0.019
P+ATP+17/7dC	0.18	0.50	0.073	0.38	0.20	0.041	0.62	0.16	0.021

**Table 3.1 Peak centers, areas and peak heights obtained by fitting the FRET histograms to Gaussians.** All experiments in the presence of Mss116 (P) with or without ATP, AMPPNP and substrate RNA (17/7 or 17/7dC) were performed under near physiological conditions (8 mM MgCl<sub>2</sub>, 100 mM KCl and 40 mM MOPS pH 7.5). ND – Not Detected.

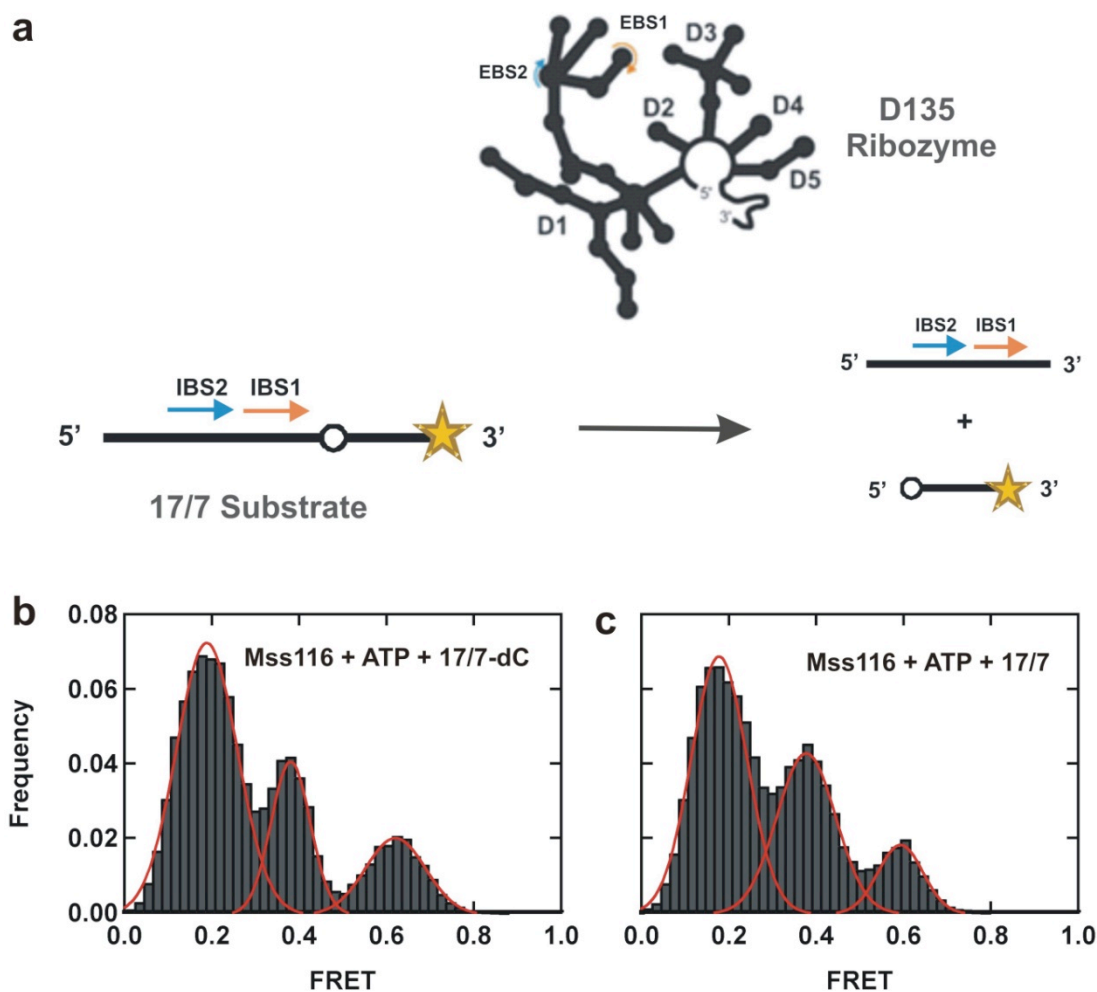


**Figure 3.7 Mss116 can promote the folding of group II introns at physiological conditions.** **a.** FRET histogram showing the distribution of three structural conformations of D135-L14 in the presence of Mss116 and ATP in physiological conditions (1 mM MgCl<sub>2</sub>, 100 mM KCl and 40 mM MOPS pH 7.5). **b.** FRET histogram showing the distribution of three structural conformations of D135-L14 in the presence of Mss116 without ATP in physiological conditions.

In the cell, the group II intron also interacts with its substrate (Fig. 1.3), the exons, which are important components of the active site. The D135 ribozyme has an ability to actively cleave an exonic substrate under single- and multiple-turnover conditions. In a trans cleavage assay, D135 ribozymes can cleave a 17/7 exonic RNA substrate (24 nt) containing the last 17 nts of the 5' exon and the first 7 nts of the intron, comprising the 5' splice site using a hydrolytic reaction as shown in Figure 3.8a (68).

Our previous single-molecule experiments have shown that the presence of an exon substrate (17/7 or 17/7-dC) stabilizes the native state of the group II introns (68). Indeed, under near-physiological conditions and in the presence of Mss116, ATP and the RNA substrate (17/7), all three folded states are populated (Fig. 3.8b). Additionally there is ~30% higher population in the native state (FRET ~0.6) than that without RNA substrate (compare Fig 3.6 and 3.8).

Similar behavior was also observed with a slow-cleaving RNA substrate (17/7-dC, Fig. 3.8c), as expected (68). However, without Mss116 and ATP, the substrate RNA alone does not promote group II intron folding into the native state under near-physiological conditions. Overall, these substrate-based single-molecule FRET data provide more evidence for the native state as a catalytically active conformation of group II introns.

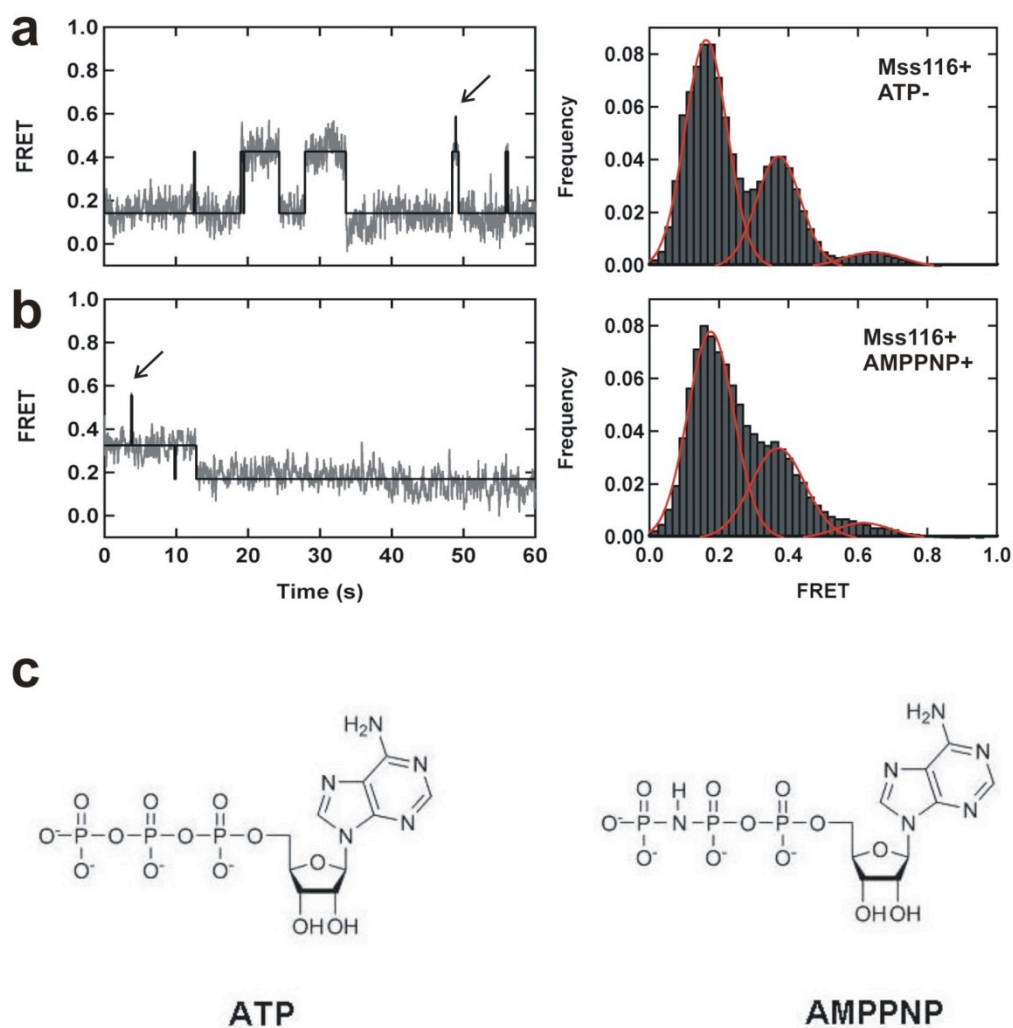


**Figure 3.8 Effect of substrate binding on the protein-mediated folding dynamics and native-state stabilization of group II introns. a.** D135 ribozyme can cleave the 17/7 RNA substrate containing the last 17 nts of the 5' exon & the first 7 nts of the intron, comprising the 5' splice site. **b.** Histogram showing the FRET distribution of the fluorophore-labeled D135-L14 ribozyme in the presence of the slower cleaving substrate 17/7-dC. **c.** Histogram showing the FRET distribution of the D135-L14 ribozyme in the presence of the wild type substrate 17/7. Both experiments were performed under near-physiological conditions (8 mM MgCl<sub>2</sub>, 100 mM KCl and 40 mM MOPS pH 7.5) with 25 nM Mss116 and 1 mM ATP.

### **3.3.3 ATP hydrolysis by Mss116 is required to form the native state**

In order to investigate the role of ATP on Mss116-mediated folding of D135, we monitored the folding dynamics under various conditions (Fig. 3.9). In the absence of ATP, the smFRET trajectories show transitions among all three conformations of the D135 ribozyme, as observed in the presence of ATP (Fig. 3.9a). However, the population of the native state was ~2.5-fold lower in the absence of ATP, showing that ATP enhances formation of the native state. This small native state population may be due to traces of ATP that co-purify with Mss116 or D135-L14. However, control experiments in the presence of glucose and hexokinase, which remove traces of ATP from the protein solution, still show low levels of the native state population. These findings indicate that, while Mss116 facilitates folding to the native state in an ATP-dependent manner, the native state can also be inefficiently populated in the absence of ATP.

To examine the ATP dependence more carefully, we examined the folding dynamics of D135 with the non-hydrolysable ATP analog AMPPNP (Fig. 3.9b). In the presence of AMPPNP, we also observe the three structural conformations with a low population in the native state, as in the absence of ATP. Nonetheless, the population of molecules in the native state is ~2.5-fold lower than in the presence of ATP. In the absence of Mss116, however, ATP or AMPPNP alone are not sufficient to form the folded intermediate state or the native state. These data show that ATP binding and hydrolysis by Mss116 are important for the efficient formation of native D135 molecules, but hydrolysis is not strictly required.



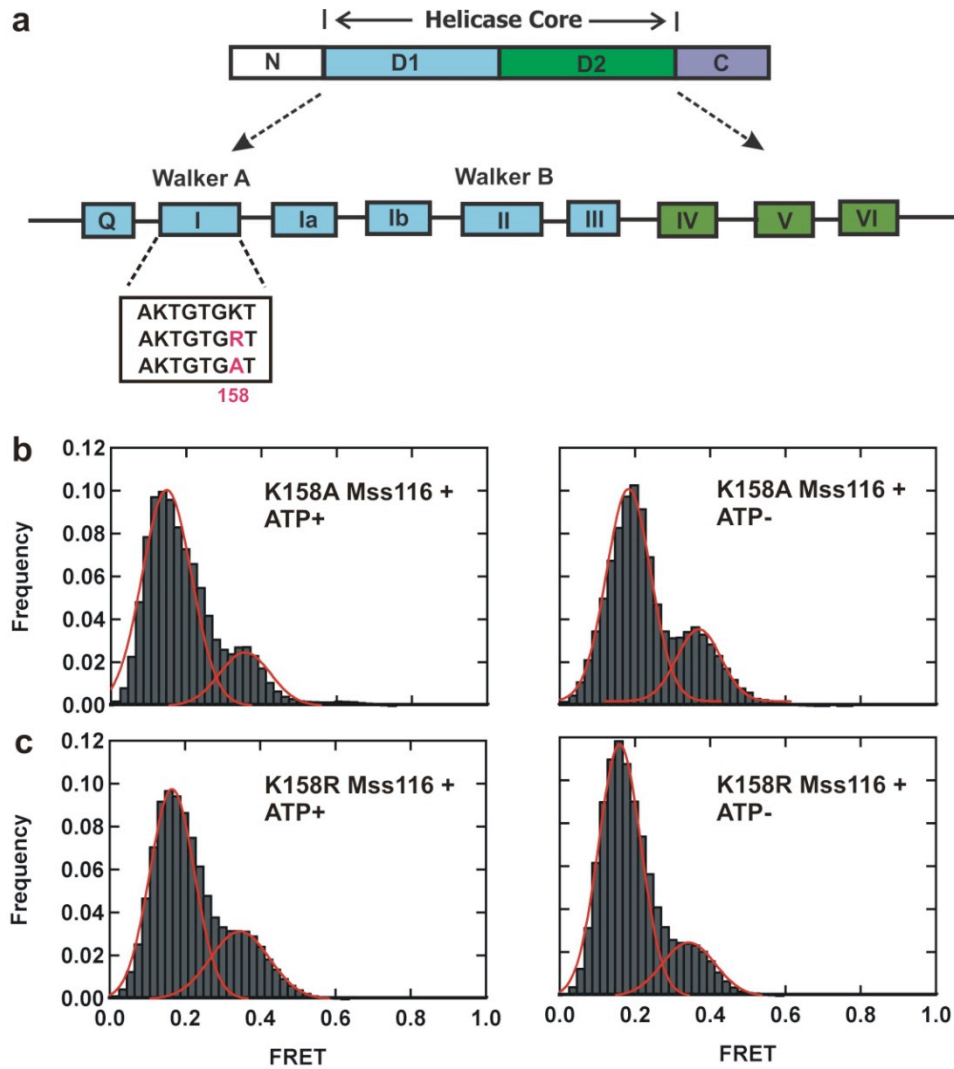
**Figure 3.9 Role of ATP in Mss116-mediated folding of group II introns.** **a.** A typical FRET trace (left) and FRET histogram (right) showing the distribution of different conformations of the Mss116-mediated folding pathway in the presence of wild type Mss116 without ATP. **b.** A typical FRET trace (left) and FRET histogram (right) showing the distribution of different conformations in the presence of wild type Mss116 and the non-hydrolysable ATP analog, AMPPNP. Black lines represent idealized HMM and arrows indicate rapid transient excursions to the native state. **c.** Chemical structure of ATP and non-hydrolysable ATP analog - AMPPNP.

To further clarify the role of ATP binding and hydrolysis on group II intron folding, we studied the effect of two previously characterized mutations (K158A and K158R) in the Walker A box (Fig. 3.10), the most highly conserved region of the protein that is required for helicase activity of DExH/D proteins. According to structural characterization of these mutant protein constructs, the K158A mutant significantly binds with ATP but has greatly reduced ATP hydrolysis activity. In contrast, the K158R mutant is unable to bind and hydrolyze ATP under similar reaction conditions.

The corresponding smFRET histograms of both Mss116 mutant constructs clearly show the presence of the extended and the folded intermediates (FRET  $\sim 0.2$  and  $\sim 0.4$ , respectively, Fig. 3.10) with or without ATP. Further supporting our previously proposed hypothesis, these mutational data suggest that not only ATP binding, but also the hydrolysis of ATP is necessary to form the native conformation of group II introns. Rare transient excursions between the folded intermediate and the native state were observed with the K158A mutant in the presence of ATP, consistent with the previously reported low ATPase activity of this mutant.

In agreement with smFRET data obtained in the presence and absence of ATP and AMPPNP, these mutational studies show that efficient formation of the native state requires ATP hydrolysis but the folded intermediate state can be formed independent from ATP.



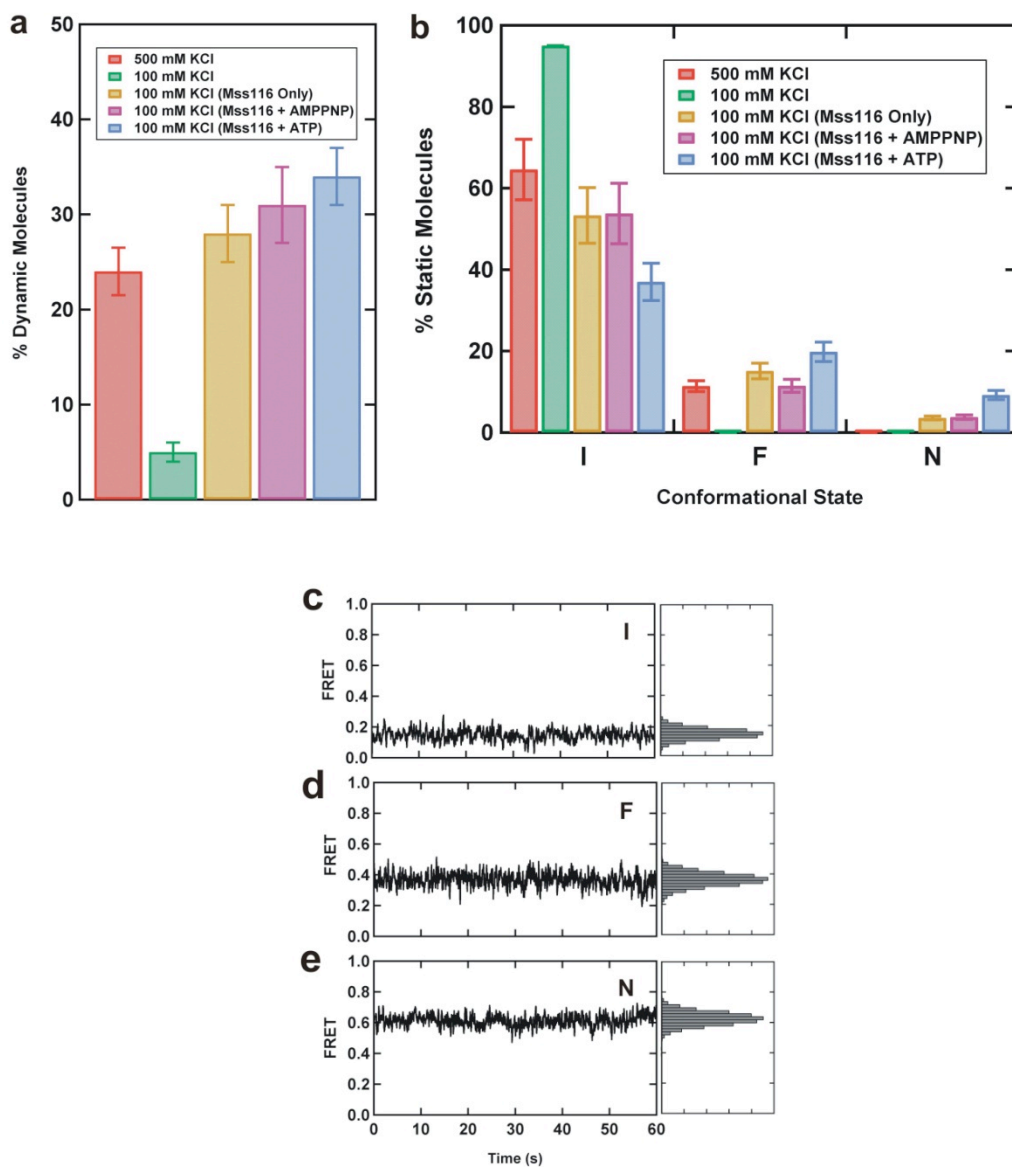


**Figure 3.10 Effects of mutant Mss116 on group II intron folding dynamics. a.** Schematic structure of Mss116 showing mutations (K158A and K158R) in the Walker A region. **b.** FRET histogram showing the distribution of different conformations in the presence of mutant K158A Mss116 with and without ATP. **c.** FRET histogram showing the distribution of different conformations in the presence of mutant K158R Mss116 with and without ATP.

### **3.3.4 Folding dynamics depend on Mss116 and ATP**

To define the folding rate constants ( $k_1$ ,  $k_{-1}$ ,  $k_2$  and  $k_{-2}$ ), we analyzed the dynamic FRET trajectories (Fig. 3.11a) using a Hidden Markov Model – HMM (Fig. 3.12, (133)). Without Mss116 and under high ionic strength conditions, a large subpopulation (>75%) appears static in either the I or F states during the two-minute timescale of our observation window (Fig. 3.11). Among the subpopulation of dynamic molecules (~24%), the observed folding rate constants between the extended and folded intermediates ( $k_1 = 0.5 \pm 0.3 \text{ s}^{-1}$  and  $k_{-1} = 0.2 \pm 0.1 \text{ s}^{-1}$ ) agree closely with previous values (68).

Under near-physiological conditions without Mss116, the dynamic subpopulation decreases to  $\leq 5\%$  (Fig. 3.11a). Therefore, folding rate constants could not be determined under these conditions. However, upon addition of Mss116 and ATP, the fraction of dynamic molecules increases to ~34%, and transitions to both F and N were observed (Fig. 3.11). The resulting folding rate constants (Table. 3.2) show that  $k_1$  and  $k_{-1}$  are comparable to the high ionic strength conditions ( $p > 0.02$ ), suggesting that Mss116 and  $\text{Mg}^{2+}$  stabilize these states similarly. However, in the presence of Mss116,  $k_2$  is ~3-fold higher and  $k_{-2}$  is ~2-fold lower than in 100 mM  $\text{Mg}^{2+}$  (68), indicating that the protein stabilizes the native state more than  $\text{Mg}^{2+}$  ions (Table. 3.2).

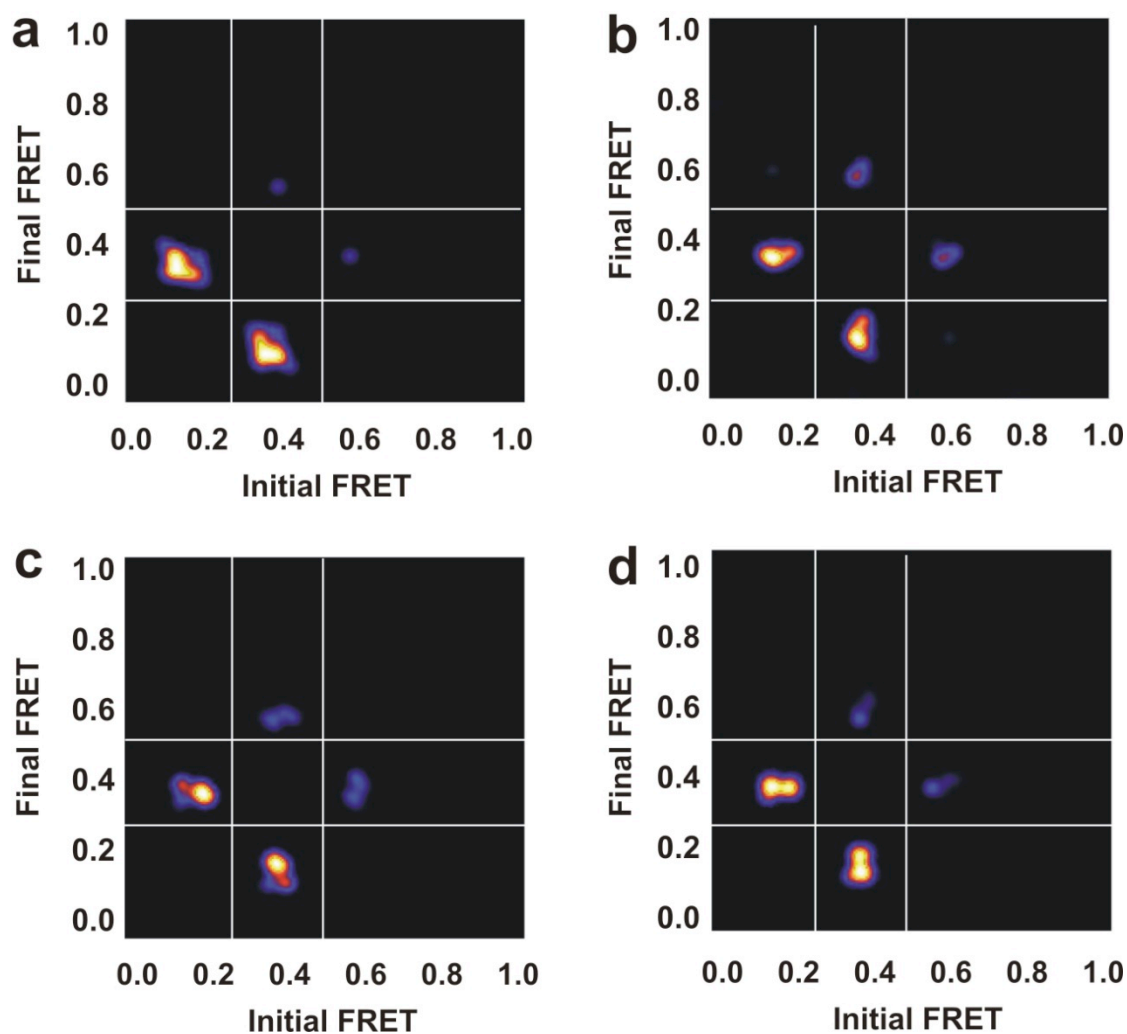


**Figure 3.11 Percentage of dynamic and static molecules and FRET traces showing the static behavior.** **a.** Percentage of dynamic molecules under different reaction conditions (see figure legend). **b.** Percentage of static molecules in each conformational state and under different reaction conditions. **c, d, and e.** Characteristic FRET trajectory of a static molecule in the extended intermediate state ( $I \sim 0.1$  FRET), in the folded intermediate state ( $F \sim 0.4$  FRET) and in the native state ( $N \sim 0.6$ ), respectively. Experiments were performed in either high ionic strength (8 mM  $MgCl_2$ , 500 mM KCl and 80 mM MOPS pH 7.5) or near-physiological conditions (8 mM  $MgCl_2$ , 100 mM KCl and 40 mM MOPS pH 7.5) as mentioned.

In the presence of Mss116 without ATP or with AMPPNP, ~30% of molecules exhibited dynamics. The resulting folding rate constants  $k_1$  and  $k_{-1}$  are comparable to those with ATP present ( $p > 0.02$ ). However,  $k_2$  is 3 to 6-fold higher ( $p < 0.02$ ) than in the presence of ATP (Table. 3.2). This indicates that ATP hydrolysis, and not just binding, contributes to efficient formation of the native state. In the presence of Mss116 without ATP or with AMPPNP, the majority of molecules that dwell in F show transitions to I (~75%) rather than N (~25%). However, with both Mss116 and ATP, the percentage of F to I transitions decreases (58%) while the percentage of F to N transitions increases (42%). Therefore, ATP hydrolysis also increases the probability that F molecules sample N instead of I (Table 3.3).

Experimental Conditions		$k_1$ (s <sup>-1</sup> )	$k_{-1}$ (s <sup>-1</sup> )	$k_2$ (s <sup>-1</sup> )	$k_{-2}$ (s <sup>-1</sup> )
Mss116	ATP-	0.5 ± 0.1	0.6 ± 0.2	0.2 ± 0.1	5.1 ± 1.3
	AMPPNP+	0.6 ± 0.1	0.2 ± 0.1	0.3 ± 0.1	2.8 ± 0.7
	ATP+	0.3 ± 0.1	0.2 ± 0.1	0.2 ± 0.1	0.9 ± 0.2

**Table 3.2 Single-molecule folding dynamics of D135-L14 ribozyme.**  $k_1$ ,  $k_{-1}$ ,  $k_2$  and  $k_{-2}$  represent the observed folding rate constants among three different conformations (I, F and N) as shown in figure 3.12. Folding rate constants under different experimental conditions were obtained by analyzing FRET traces with a Hidden Markov Model. All experiments were performed in 8 mM MgCl<sub>2</sub>, 100 mM KCl and 40 mM MOPS pH 7.5. The experiments with Mss116 were conducted in the presence or absence of ATP or AMPPNP as mentioned. Folding rate constants of D135 RNA in the presence of 100 mM Mg<sup>2+</sup> under high salt conditions (500 mM KCl and 80 mM MOPS pH 7.5) without Mss116 and ATP are  $k_2 = 0.07$  s<sup>-1</sup> and  $k_{-2} = 1.7$  s<sup>-1</sup>.



**Figure 3.12 Transition density plots (TDP) from HMM analysis.** **a.** TDP showing transitions between different FRET states under high ionic strength conditions (8 mM MgCl<sub>2</sub>, 500 mM KCl and 80 mM MOPS pH 7.5) without Mss116 and ATP. **b.** TDP showing transitions in the presence of Mss116 and ATP. **c.** TDP showing transitions in the presence of Mss116 without ATP. **d.** TDP showing transitions in the presence of Mss116 and AMPPNP. All experiments with Mss116 with or without ATP and AMPPNP were performed in near-physiological conditions (8 mM MgCl<sub>2</sub>, 100 mM KCl and 40 mM MOPS pH 7.5).

Detailed analysis of static molecules in each state provides additional evidence for ATP-dependent Mss116 activity. In the presence of Mss116 without ATP, the fraction of static molecules in I decreases from 95% to 53%, and they begin to appear in both F and N (15% and 4%, respectively, Fig 3.11). In the presence of Mss116 with ATP, the fraction of static molecules in N increases significantly (2.5-fold,  $p = 0.02$ ), while that in F does not change significantly ( $p = 0.16$ ). These data support an important role for ATP hydrolysis for efficient Mss116-induced folding of the intron ai5 $\gamma$ , as previously proposed (89,92).

Experimental Conditions	Percentage of Transitions	
	F $\rightarrow$ I	F $\rightarrow$ N
Mss116 Only	73	27
Mss116 + AMPPNP	74	26
Mss116 + ATP	58	42

**Table 3.3 The percentage of transitions between conformational states.** The percentage of transitions observed from the folded intermediate state (F) to either the extended intermediate state (I) or the native state (N) in the presence of Mss116 with or without ATP and AMPPNP.

### 3.4 Discussion

Based on previous biochemical and folding studies (80,87-89,92), it has been proposed that Mss116 can promote group II intron splicing by stabilizing on-pathway intermediates and/or by disrupting off-pathway misfolded structures. Previous work has shown that ai5 $\gamma$  folds through a slow but smooth pathway devoid of kinetic traps (80,87), and an Mss116 mutant with a significant helicase defect still retains the ability to promote splicing (89,152). Other work supports a mechanism where Mss116 unwinds kinetic traps to promote splicing, even at low levels of helicase activity (88). Both models are primarily based on indirect studies of ai5 $\gamma$  splicing, thus, any putative role for Mss116 in RNA folding is purely hypothetical. To specifically examine the effect of Mss116 on ai5 $\gamma$  folding, we have employed a well-characterized smFRET assay that enables us to directly monitor the role of Mss116 and ATP on the intron folding.

Previous experiments showed that under near-physiological conditions, group II introns alone do not stably form the native state (87). Our data show they cannot even transiently sample the native state. The large subpopulation of static molecules in I (~95%) indicates a high activation barrier between I and F (Fig. 3.12). The appearance of all three states in the presence of Mss116 shows that Mss116 can promote RNA folding by lowering the activation barriers between folded states, consistent with recent data (97,153).

We observe two distinct effects on folding. First, F state folding is promoted by Mss116 and other RNA-binding proteins even without ATP (Figs. 3.6 and 3.9).

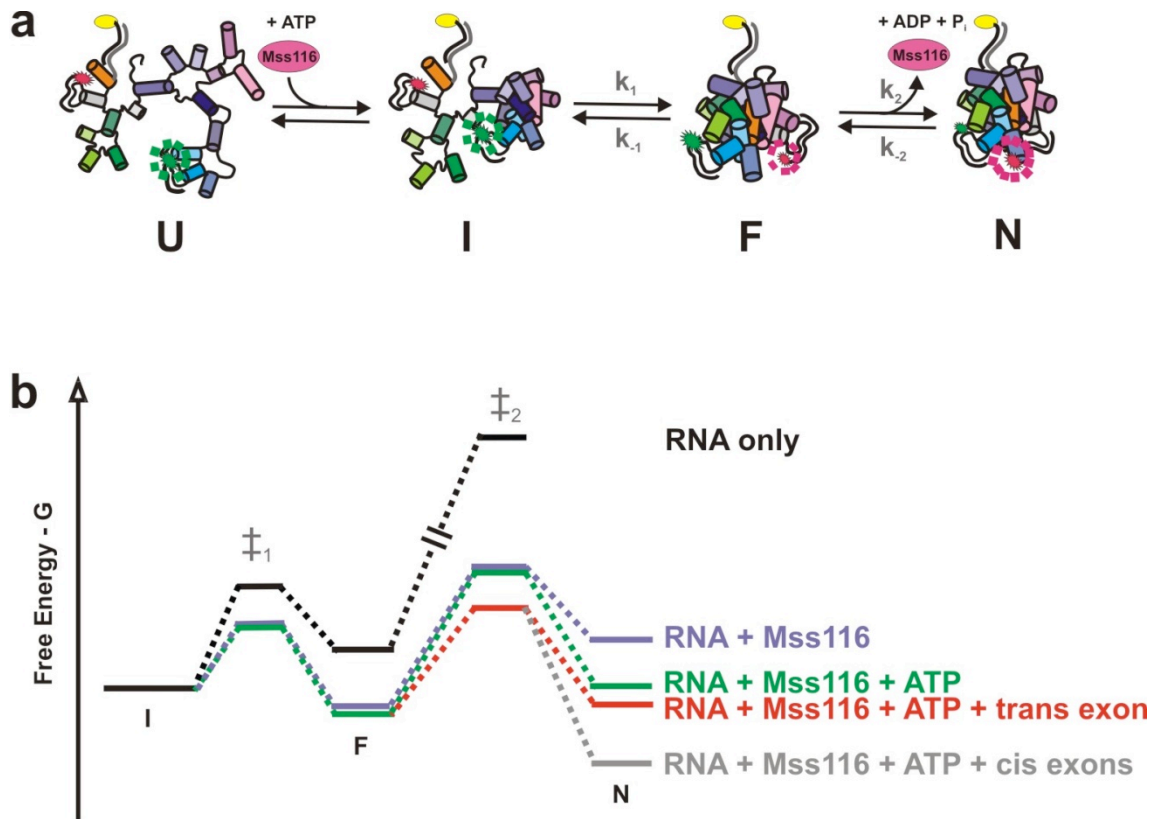
The fact that diverse, basic RNA-binding proteins promote this stage of folding suggests that formation of the folded intermediate state is contingent on electrostatic stabilization or annealing. The lack of ATP dependence also shows that mechanical events such as translocation or duplex unwinding are not involved in the obligate early steps of ai5 $\gamma$  folding. Although diverse proteins stimulate the I $\rightarrow$ F transition, only Mss116 increases the FRET ratio of I (~0.12 to ~0.18), indicating a specific interaction with the extended intermediate (Table. 3.1). The higher FRET value suggests a slightly more collapsed I state in the presence of Mss116, and that this protein is ideally suited to folding of ai5 $\gamma$ . Second, Mss116 promotes folding to the native state through a mechanism involving ATP hydrolysis. Without ATP or with AMPPNP, only a small fraction of molecules reach the native state, indicating that ATP hydrolysis contributes to the function of Mss116 specifically during native state formation.

Several possible models explain the function of Mss116 during native state formation. In a stabilization model, Mss116 may bind and stabilize the  $\kappa$ - $\zeta$  region, a substructure of domain 1 that must form before the intron can productively fold (61). This same substructure contains a binding site for the catalytic domain 5, which docks late in the assembly pathway (87). If Mss116 stimulates folding by stabilizing the  $\kappa$ - $\zeta$  element, it will then be required to dissociate before D5 can stably dock within the core. Thus, ATP hydrolysis may stimulate a conformational change in Mss116 that allows the protein to be released during the last stage of the folding process, consistent with previously published findings indicating that



ATP hydrolysis is required for Mss116 turnover and dissociation from RNA (94,152). The formation of a small fraction of the native population in the absence of ATP is consistent with this model, as Mss116 is likely to have a finite off-rate. *In vivo*, efficient recycling of Mss116 may be necessary to maintain a pool of active Mss116 enzymes, and may allow the cell to overcome non-functional binding events while preventing inhibition of properly folded RNAs (152).

However, other models, such as unwinding during the final transition from the folded intermediate to the native state, also remain possible. According to this model, the protein would first facilitate the slow, early stages of folding (including D1 collapse) through electrostatic interactions, but formation of the final native state may be impeded by one or more small misfolded structures. To form the correctly folded structure, the protein would resolve the small misfolded structures through strand exchange or unwinding by local strand separation. This ATP-dependant remodeling event would provide an opportunity for the native contacts to form. Although we do not have direct evidence for the presence of misfolded structures, we have previously shown that I, F and N are likely heterogeneous (68). Thus, it is possible that misfolded intermediates with FRET ratios similar to on-pathway intermediates are present but not readily distinguishable. However, recent *in vivo* experiments show that the ability of Mss116 to hydrolyze ATP correlates with function even with different degrees of helicase activity, in agreement with a recycling function for ATPase activity (152). Helicase activity may play additional roles in the presence of long exons (154).



**Figure 3.13 Mss116-mediated group II intron folding pathway.** **a.** D135-L14 minimal folding pathway: the unfolded state (U), the extended-intermediate state (I), the folded-intermediate state (F) and the native state (N). Folding rate constants of I, F and N are  $k_1$ ,  $k_{-1}$ ,  $k_2$  and  $k_{-2}$  respectively. **b.** Hypothetical free energy diagram for folding of the D135-L14 based on pre-steady state data and previous work (68,97). In the absence of Mss116 and ATP (black), extended intermediate state is the most stable conformation of the ribozyme. Mss116 alone (purple), results in formation of the folded intermediate state. Addition of both Mss116 and ATP with and without trans exon-17/7 RNA (red and green, respectively) enhances native state formation (68). Cis exons (grey) further stabilize the native state of group II introns (97).

Taken together, we observe multiple roles for Mss116 in the folding pathway of ai5 $\gamma$  and we propose that Mss116 mediates group II intron folding by stabilizing on-pathway intermediates and transition states from the unfolded state to the folded states. We also observe that efficient transition from the folded intermediate to the native state requires ATP hydrolysis, possibly for Mss116 release and recycling (Fig. 3.13). The substrate further stabilizes the native state, raising the interesting possibility that *in vivo* and in the presence of all the group II intron domains and Mss116, the native state becomes the most stable conformation.

### 3.5 Conclusions

DEAD-box proteins are highly conserved across all domains of life, and they are involved in nearly all aspects of RNA metabolism; however, the mechanism of many of these proteins remains unclear. The discovery that Mss116 facilitates splicing of all group I and II introns in yeast mitochondria (90) has provided a rare opportunity to study how DEAD-box proteins might promote folding of their native RNA substrates. Therefore, studies of the Sc.ai5 $\gamma$  group II intron derivative and its natural cofactor, Mss116, provide key insights into both protein-mediated RNA folding and catalysis.

Previous biochemical and folding studies (80,87-89,92) have proposed that Mss116 can promote group II intron splicing by stabilizing on-pathway intermediates and/or by disrupting misfolded structures. However, the exact mechanism of protein-mediated splicing is still under debate. Group II intron ai5 $\gamma$

has been shown to fold through a slow but smooth pathway devoid of kinetic traps (80,87). Furthermore, an Mss116 mutant with a significant helicase defect still retains the ability to promote group II intron splicing (88,89,155). Based on these observations, it has been suggested that Mss116 uses a stabilization mechanism. In contrast, some evidence supports a mechanism by which Mss116 unwinds kinetic traps to promote splicing, even at low levels of helicase activity (88,89). Both mechanisms, however, are primarily based on indirect studies of ai5 $\gamma$  splicing, and thus any putative role for Mss116 in RNA folding is purely hypothetical. As a result, the effect of Mss116 on ai5 $\gamma$  folding has been characterized using a single-molecule fluorescence assay that enables the direct observation of the influence of Mss116 and ATP on the group II intron folding.

According to the smFRET data, group II introns cannot reach to the catalytically active native state alone under near-physiological conditions due to large activation barriers between states. However, the DEAD-box protein Mss116 or non-specific basic RNA binding proteins can lower the activation barriers and promote the folding of intron RNA to the higher FRET conformations. Although the transitions from the extended intermediate state to the folded intermediate state are independent on ATP hydrolysis, the binding and hydrolysis of ATP is required for efficient formation of the native state under these conditions. Based on previous studies that have shown DEAD-box proteins need ATP hydrolysis to release RNA from proteins (94) and our experimental observation that early rapid folding steps are independent from ATP hydrolysis, we proposed that ATP

hydrolysis occurs in the final steps and facilitates the formation of the folded native state by recycling the Mss116. Furthermore, in the presence of Mss116 and ATP, it has been shown that the RNA substrate containing exons can enhance the formation of the native state providing additional evidence for the native state as the catalytically active conformation.

In summary, we observe that Mss116 has multiple roles in the folding pathway of ai5 $\gamma$  and we propose that Mss116 mediates group II intron folding by stabilizing on-pathway intermediates upon transition from the unfolded state to the folded state. However, we also observe that the transition from the folded intermediate state to the native state requires ATP hydrolysis as illustrated by our free-energy diagram (Fig. 3.13). According to this scheme, unfolded group II intron ribozymes (U) slowly collapse to the extended intermediate state (I) under near physiological conditions, as previously observed (87). This transition is so slow that we do not expect to observe it in our experiments. The extended intermediate can interact with Mss116 to further compact this state. Next, Mss116 (or other proteins) induce the formation of the folded intermediate (F). Finally, the native state is stably formed once ATP is hydrolyzed, possibly upon Mss116 release. Our data also show that the presence of substrate further stabilizes the native state, raising the interesting possibility that *in vivo* and in the presence of all the group II intron domains and Mss116, the native state becomes the most stable conformation.

## CHAPTER 4

### Single-molecule Studies of Human Spliceosomal U2-U6 snRNAs

#### 4.1 Introduction

Spliceosome-catalyzed nuclear pre-mRNA splicing is an essential process in eukaryotic gene expression (Fig. 1.1). The assembly pathway of the spliceosome and the mechanism of pre-mRNA splicing have been introduced in previous chapters (98,99). Although the spliceosome consists of five snRNAs and more than 150 proteins (98,102), it has been shown that the active spliceosome consists of only U2, U5 and U6 snRNPs. Additionally, several studies have proposed the U2-U6 snRNA complex is the catalytic core of the spliceosome (11,13,107).

Spliceosomal U2 and U6 snRNAs play an important role in pre-mRNA splicing in all eukaryotes. The structure and the catalytic ability of this U2-U6 snRNA complex have been discussed in detail in Chapter 1 (108,113-118). In summary, two main structures have been proposed for both human and yeast U2-U6 snRNA complexes (Fig. 1.9). *In vivo* genetic studies support the three-helix structure containing helix Ib for yeast U2-U6 snRNA complex (113,114). It has also been postulated that the three-helix structure is required for both steps of pre-mRNA splicing in yeast. In contrast, NMR studies support the four-helix structure for the yeast protein-free U2-U6 complex *in vitro* (108). Although it has been proposed that the U2-U6 complex can adopt two conformations that reflect

different spliceosomal activation states, there was no direct experimental evidence to support that hypothesis.

In order to understand the structural dynamics of U2-U6 spliceosomal RNAs, protein-free yeast U2-U6 snRNA complexes were monitored using single-molecule fluorescence (134). Structural dynamics of the wild type and mutant yeast U2-U6 complex, which have mutations in highly conserved regions, show the presence of at least three distinct conformations in dynamic equilibrium: the low FRET state ( $\sim 0.2$ ), the mid FRET state ( $\sim 0.4$ ) and the high FRET state ( $\sim 0.6$ ). As a result, the single-molecule data led to the hypothesis of a minimal two-step folding pathway for yeast U2-U6 complex with an obligatory folding intermediate (134). Furthermore, single-molecule data suggested  $Mg^{2+}$  can facilitate the conformational changes; notably, the first step strongly depended on the concentration of magnesium ions.

Based on the single-molecule studies and previous experimental data, the high FRET state was assigned the four-helix structure proposed by the NMR data, the low FRET state was assigned the three-helix structure proposed by the genetic data and the mid FRET state was assigned the intermediate conformation. Mutant U2-U6 snRNA complexes with mutations in highly conserved residues corresponding to the formation of helix Ib further confirmed that the low FRET state was the three-helix structure with helix Ib (134). Taken together, single-molecule studies of yeast U2-U6 spliceosomal RNAs indicate that the U2-U6 complex can adopt both previously proposed structures *in vitro*.

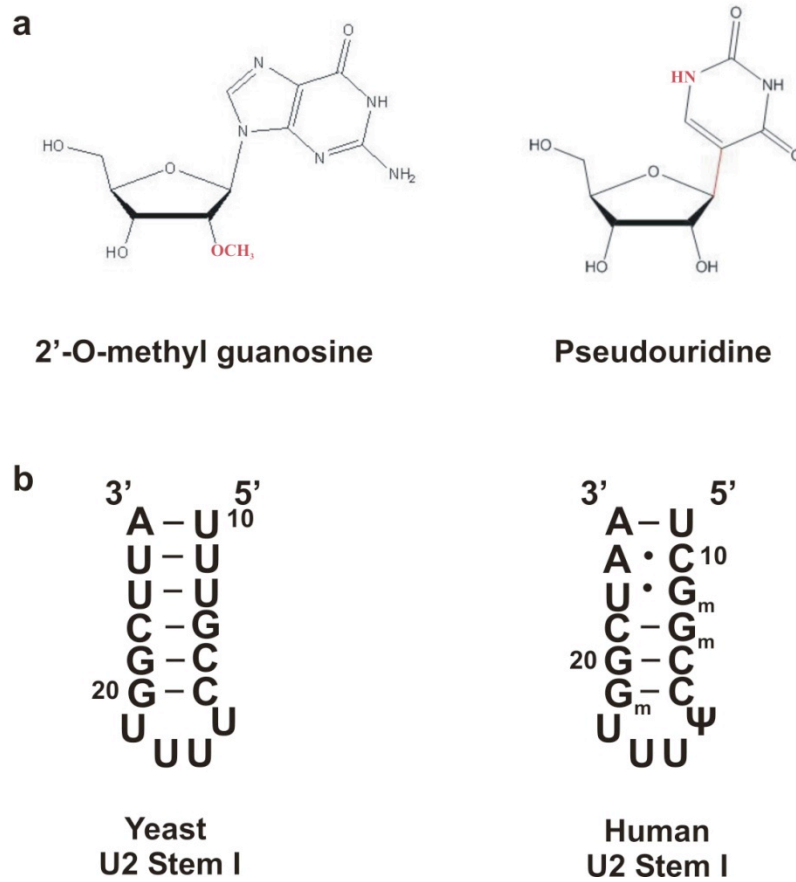
Similar to the previously proposed yeast U2-U6 three-helix structure, early *in vivo* genetic studies suggest that human U2-U6 spliceosomal RNAs also form the three-helix structure with helix Ib (116). However, later studies conducted with mutant U2-U6 constructs have proposed that helix Ib residues in both U2 and U6 snRNAs participate in intramolecular rather than intermolecular base pairing (117). As a result, the catalytic structure of the human U2-U6 snRNA complex is still uncertain.

The sequences of spliceosomal snRNAs are highly conserved among different organisms. However, significant differences exist between human and yeast U2-U6 snRNAs suggesting these two organisms may have different activation mechanisms (Figs. 1.9 and 4.1). There are two major differences between human and yeast U2-U6 snRNA complexes (118).

- 1) Human snRNAs contain numerous functional post-transcriptional modifications such as 2'-O-methyl groups and pseudouridines (118).
- 2) Stem I of human U2 snRNA contains C<sub>10</sub>-A<sub>23</sub> and G<sub>11</sub>-U<sub>22</sub> wobble base pairs instead of the tandem U-U wobble base pairs found in yeast (118).

The majority of these conserved post-transcriptional modifications are clustered in functionally important regions of snRNAs and perform either RNA-dependent or protein-dependent functions. Out of five snRNAs, U2 snRNA is highly post-transcriptionally modified and contains ten 2'-O-methylated residues and thirteen pseudouridines.





**Figure 4.1 Structures of modified residues and U2 stem I of yeast and humans.** **a.** Chemical structures of the modified residues in human spliceosomal snRNAs. **b.** Structural differences between U2 stem I of yeast and human spliceosomal snRNAs.

Interestingly, recent NMR and UV melting studies by Butcher and coworkers have shown that the structures of the human and yeast stem-loop RNAs are nearly isosteric *in vitro*, but the post-transcriptional modifications (Fig. 4.1) within the human sequence increase its melting temperature by 18°C, and thus, significantly increase its stability (118). Since stem I plays an important role in the formation of helix Ib, the higher stability of human U2 stem I may affect the formation of helix Ib in humans. Previous studies have proposed that the formation of helix Ib is essential for the activation of the spliceosome for the second transesterification reaction in yeast. Therefore, dissimilarities between these two species suggest that different spliceosomal activation mechanisms are possible during the splicing reaction.

In addition to structural differences between human and yeast spliceosomal RNAs, it has been shown that protein-free human U2-U6 snRNA complexes can catalyze a splicing related reaction *in vitro* (11,13,107). Furthermore, consistent with the structural complexity, previous studies have found that human spliceosomal snRNAs associate with more proteins than yeast snRNAs.

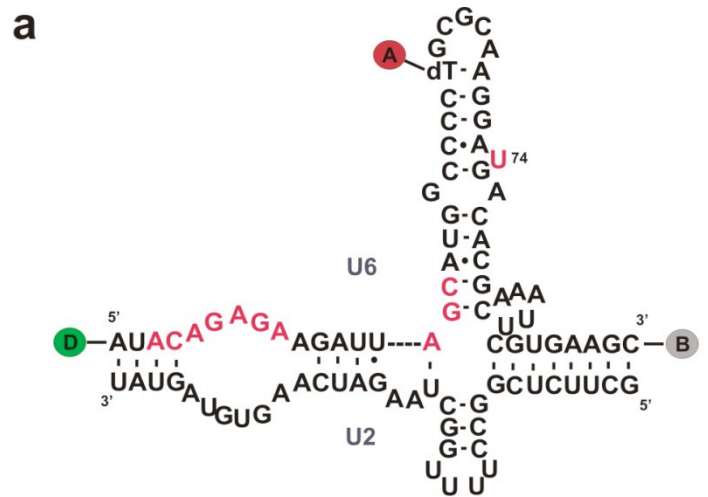
The primary objective of this study is to understand and compare the structural dynamics of human and yeast U2-U6 spliceosomal snRNAs. In order to achieve that goal, Mg<sup>2+</sup>-dependent structural dynamics were monitored at the single-molecule level. Furthermore, to investigate the roles of highly stable U2 stem I modifications on structural dynamics, single-molecule fluorescence was used to directly observe and quantify the effects of these modifications (2'-O-

methyl groups and pseudouridines) on the stability and dynamics of the human U2-U6 complex.

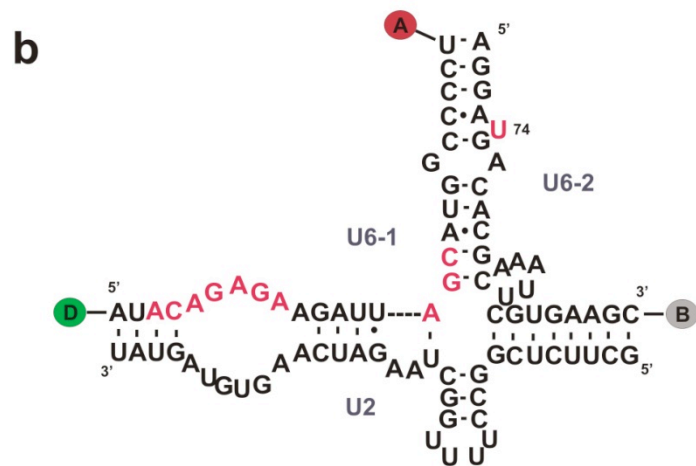
## 4.2 Experimental design

To reveal structural dynamics of human spliceosomal snRNAs, a fluorophore-labeled U2-U6 snRNA complex was studied using single-molecule fluorescence. Based on the proposed structure of the U2-U6 complex, the FRET donor (Cy3) was incorporated at the 5' end of U6 and the U6 ISL was internally labeled with the FRET acceptor (Cy5). The 3' end of the U6 strand was labeled with biotin to immobilize the fluorophore-labeled U2-U6 construct on a slide (Fig. 4.2). The fluorophore-labeled and biotinylated U2-U6 complex was immobilized on a quartz slide via a biotin-streptavidin linkage and the folding dynamics of the U2-U6 snRNA complex were directly monitored using TIRF-based single-molecule microscopy.

Since the labeling efficiency of the full length U6 snRNA in the two-strand U2-U6 complex was low, a three-strand U2-U6 complex was used to characterize the folding dynamics using steady-state fluorescence. To enhance the labeling efficiency, U6 snRNA has divided into two strands (U6-1 and U6-2). Similar to the yeast U2-U6 complex, the FRET donor (Cy3) was incorporated at the 5' end of U6-1, and the FRET acceptor (Cy5) was attached to the 3' end of U6-1. Furthermore, the 3' end of U6-2 was biotinylated as previously described (Fig. 4.2).



**Two-strand human U2-U6 complex**



**Three-strand human U2-U6 complex**

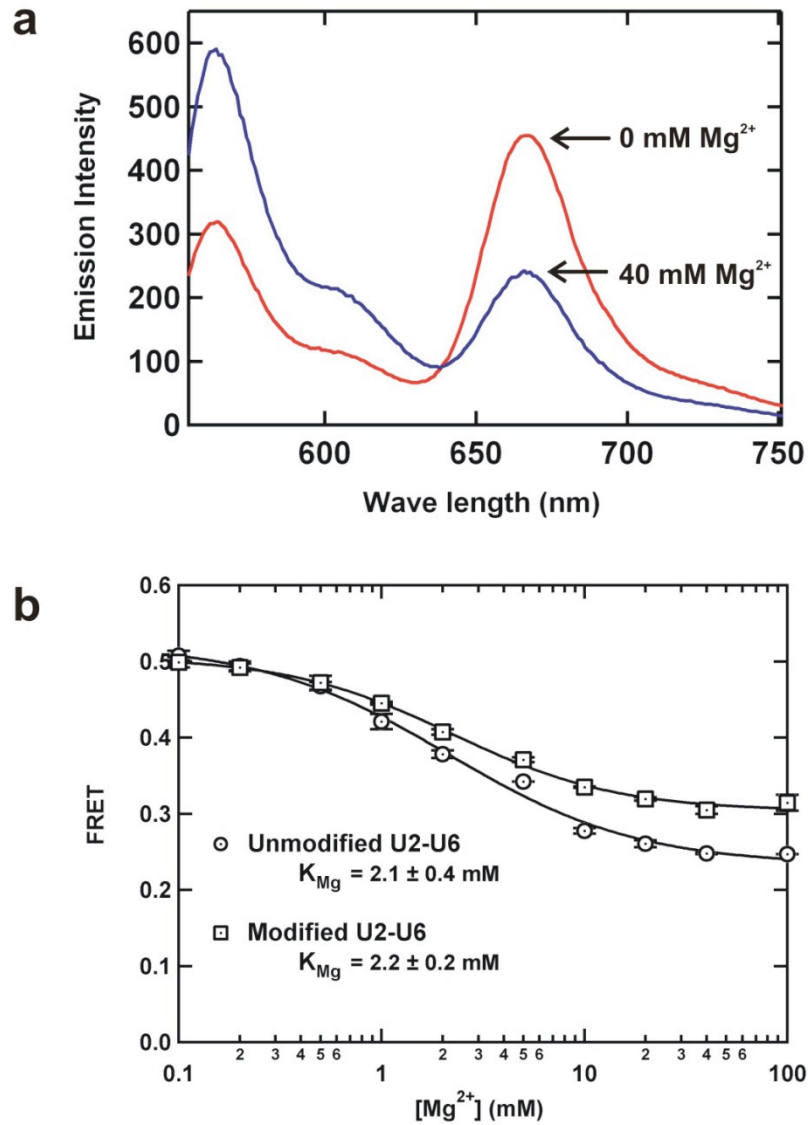
**Figure 4.2 Human U2-U6 snRNA complexes.** **a.** Two-strand (U6 and U2) human U2-U6 snRNA complex. The 5' end of the U6 snRNA was labeled with the FRET donor (Cy3) and the U6 snRNA was internally labeled with the FRET acceptor (Cy5). Biotin was attached to the 3' end of the U6 snRNA for surface immobilization. **b.** Three-strand (U6-1, U6-2 and U2) human U2-U6 snRNA complex. The 5' end of the U6-1 was labeled with the FRET donor (Cy3) and the 3' end of the U6-1 was labeled with the FRET acceptor (Cy5). Biotin was attached to the 3' end of the U6-2 for surface immobilization. Highly conserved regions (the ACAGAGA sequence, the AGC triad and a magnesium binding residue U-74) are shown in red.

## 4.3 Results

### 4.3.1 $Mg^{2+}$ -dependent structural dynamics by steady-state FRET

Magnesium ions play an essential role during the pre-mRNA splicing reaction by facilitating RNA folding and catalysis.  $Mg^{2+}$ -dependent folding dynamics of the three-strand U2-U6 complex were first characterized using the steady-state FRET (ssFRET, Fig. 4.3) in standard buffer conditions (50 mM TRIS-HCl pH 7.5 and 100 mM NaCl). Steady-state FRET is a very sensitive technique to monitor global structural changes of small RNAs. In the absence of  $Mg^{2+}$ , the emission intensity of donor was lower than that of the acceptor providing the high FRET state (FRET  $\sim$ 0.5). As suggested for the yeast U2-U6 complex, the high FRET state was assigned to the four-helix structure containing the highly conserved ACAGAGA loop and the U6 internal stem-loop (ISL) in close proximity.

Upon addition of 40 mM  $Mg^{2+}$ , the emission intensity of the donor increased substantially while decreasing the acceptor intensity providing evidence for  $Mg^{2+}$  induced conformational change (Fig. 4.3). The observed low FRET state (FRET  $\sim$  0.2) was assigned to the three-helix structure with helix 1b, where the ACAGAGA loop was separated from the U6 ISL. The  $Mg^{2+}$  dependence of the U2-U6 conformational change was determined by plotting the FRET efficiency as a function of  $Mg^{2+}$  concentration (Fig. 4.3). The resulting isothermal titration was fitted to a  $Mg^{2+}$  binding equation (Eq. 5, page 56).



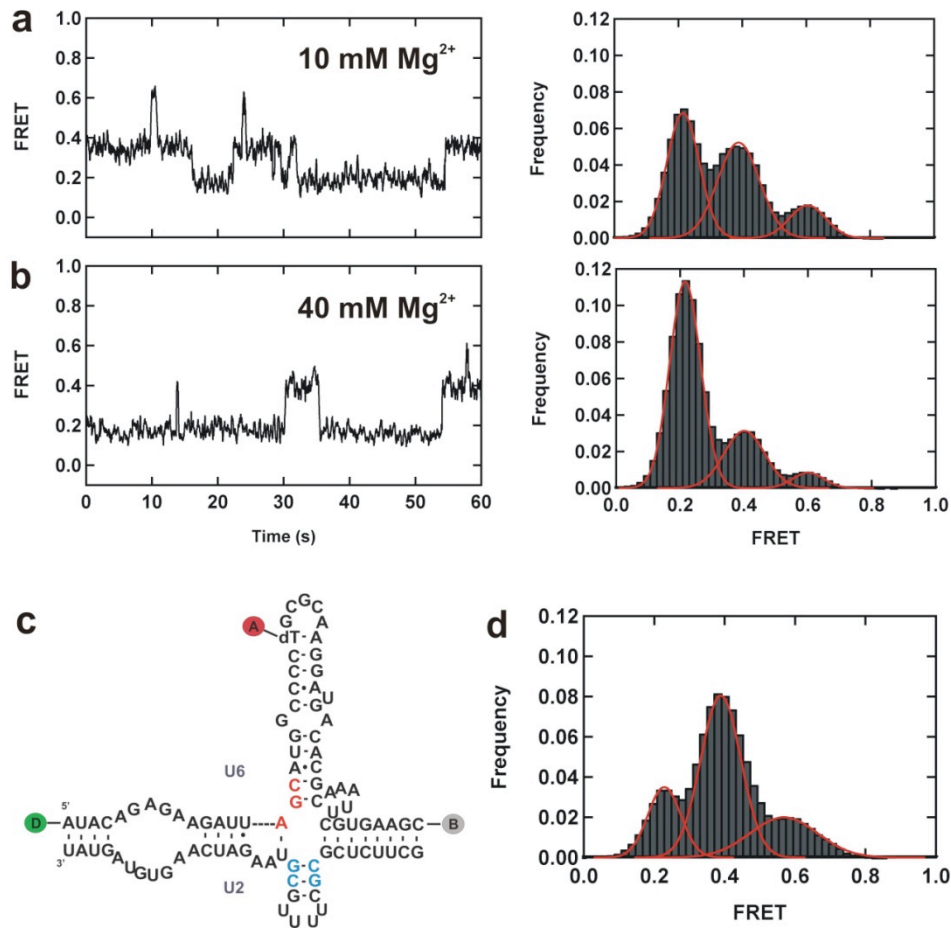
**Figure 4.3 Steady-state fluorescence.** **a.** The emission spectra of fluorophore-labeled U2-U6 snRNA complex at 0 mM (red) and 40 mM (blue)  $Mg^{2+}$ . **b.**  $Mg^{2+}$  dependence of the FRET efficiency of fluorophore-labeled modified and unmodified human U2-U6 complexes. Circles represent the unmodified U2-U6 complex data and squares represent the modified U2-U6 complex data. The black lines are fittings of steady-state FRET data to the binding equation (Eq. 5, page 56).

The observed  $Mg^{2+}$  dissociation constant of human U2-U6 complex,  $K_{Mg} = 2.1 \pm 0.4$  mM, is consistent with yeast spliceosomal snRNAs ( $K_{Mg} = 3.3 \pm 0.7$  mM). Those data suggest the metal ions that have been found to be important factors in the activation of the spliceosome may contribute to the conformational changes of the U2-U6 complex.

#### **4.3.2 $Mg^{2+}$ -dependent structural dynamics by single-molecule FRET**

Single-molecule fluorescence has an ability to dissect RNA folding pathways by revealing transient folding intermediates that can be hidden in bulk experiments. Therefore, single-molecule fluorescence was applied to the two-strand human U2-U6 complex to elucidate the  $Mg^{2+}$ -dependent folding dynamics in detail.

The effect of  $Mg^{2+}$  concentration on the structural dynamics of individual U2-U6 snRNA complexes was measured at several  $Mg^{2+}$  concentrations ranging from 10 to 40 mM (Fig. 4.4). The resulting smFRET trajectories at different  $Mg^{2+}$  concentrations clearly identified the existence of distinct FRET efficiencies corresponding to three different conformational states in dynamic equilibrium as previously observed in the yeast U2-U6 snRNA complex (134): the low FRET state (FRET  $\sim 0.2$ ), the mid FRET state (FRET  $\sim 0.4$ ) and the high FRET state (FRET  $\sim 0.6$ ). At high  $Mg^{2+}$  concentration (40 mM), the majority of molecules populated the 0.2 FRET state. However, when we decreased the  $Mg^{2+}$  concentration to 10 mM, the population of the low FRET state conformation decreased while the population of higher FRET states increased (Fig. 4.4a and b).

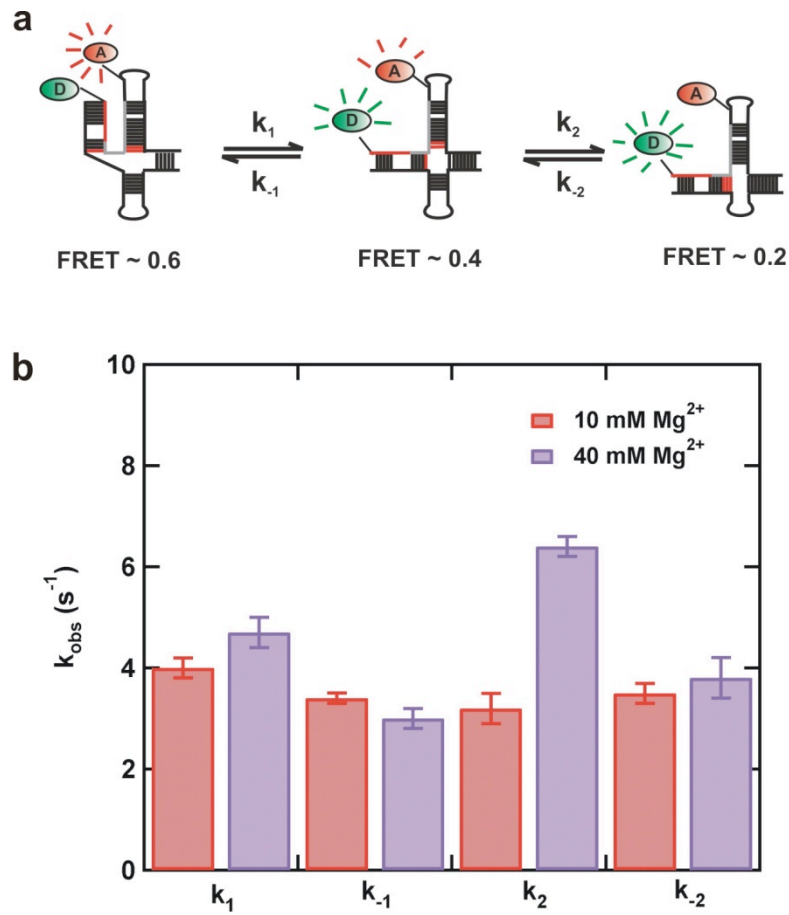


**Figure 4.4**  $Mg^{2+}$ -dependent structural dynamics of the human U2-U6 snRNA complex. **a.** A typical FRET trajectory and FRET histogram of human U2-U6 complex in 10 mM  $Mg^{2+}$ . **b.** A typical FRET trajectory and FRET histogram of human U2-U6 complex in 40 mM  $Mg^{2+}$ . **c.** Structure of the human U2-U6 snRNA complex containing flipped bases (blue) in the U2 stem I (4-fold mutant). The FRET donor – Cy3 (green), the FRET acceptor – Cy5 (red) and the biotin (grey) are shown using circles. **d.** FRET histogram of the human 4-fold mutant U2-U6 complex in 10 mM  $Mg^{2+}$ .



Based on smFRET data and sequence similarities between human and yeast U2-U6 snRNAs, we propose a two-step folding pathway for the human U2-U6 snRNAs similar to yeast (Fig. 4.5a). We assigned the high FRET state ( $\sim 0.6$ ) to the four-helix structure, the low FRET state ( $\sim 0.2$ ) to the three-helix structure and the mid FRET state ( $\sim 0.4$ ) as an obligatory folding intermediate. In order to confirm the low FRET state corresponds to the three-helix structure with helix Ib, folding dynamics were monitored using a four-fold mutant that prevents helix Ib formation. Four bases in the U2 stem I were flipped to prevent base pairing of residues in U2 stem I with the catalytic triad of U6 snRNA (Fig. 4.4c). As a result, the four-fold mutant prevents the formation of helix Ib while maintaining the junction structure of the U2-U6 complex. According to the smFRET data, the population of the low FRET state significantly decreased with the mutant construct providing additional evidence for the low FRET state as the three-helix structure with the helix Ib (Fig. 4.4d).

To further characterize the structural dynamics of the human U2-U6 snRNA complex, the folding rate constants ( $k_1$ ,  $k_{-1}$ ,  $k_2$  and  $k_{-2}$ ) were obtained by dwell time analysis (Fig. 4.5). Folding rate constants  $k_1$ ,  $k_{-1}$ , and  $k_{-2}$  did not show significant changes with varying  $Mg^{2+}$  concentrations ( $p > 0.02$ ). However, the folding rate constant  $k_2$  increased by  $\sim 2$ -fold ( $p < 0.02$ ) when  $Mg^{2+}$  concentration increased from 10 mM to 40 mM (Fig. 4.5b).



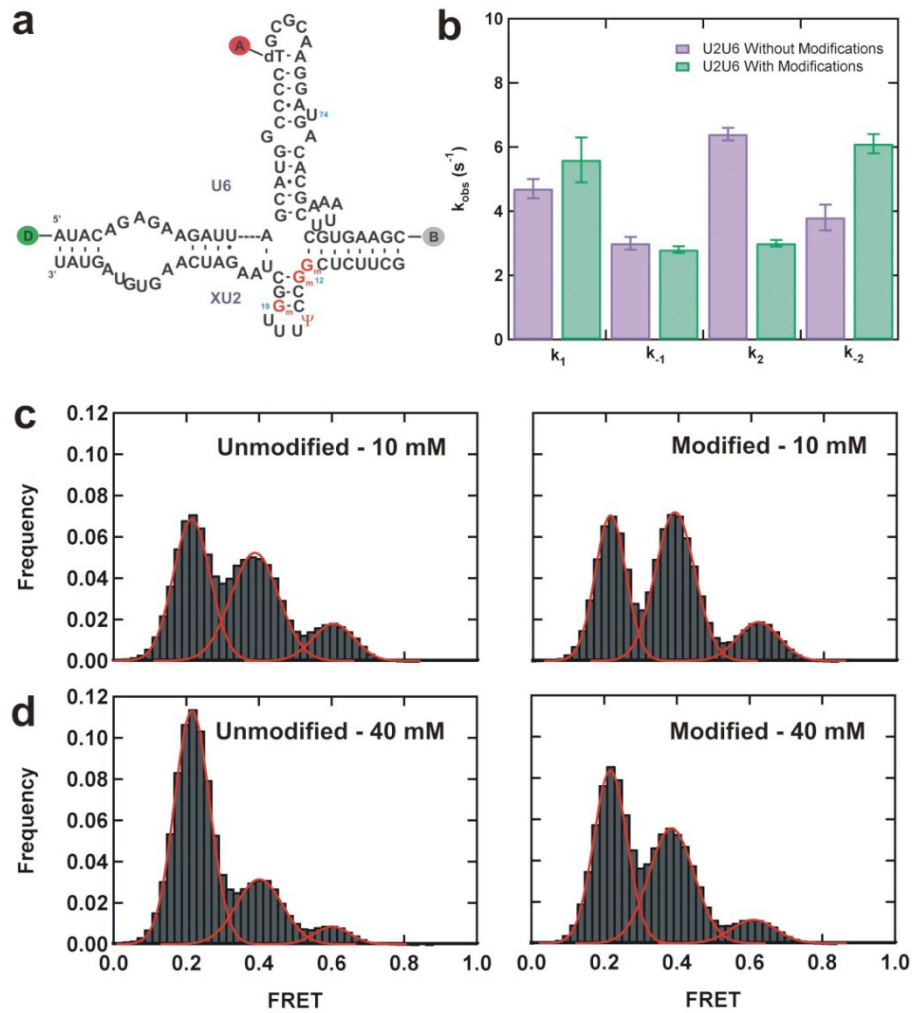
**Figure 4.5 The folding pathway and dynamics of human U2-U6 complex. a.** Proposed two-step minimal folding pathway of the human U2-U6 snRNA complex consisting of 3 states: the low FRET state (FRET ~ 0.2), the mid FRET state (FRET ~ 0.4) and the high FRET state (FRET ~ 0.6). **b.** Folding dynamics of the human U2-U6 snRNA complex in 10 mM and 40 mM Mg<sup>2+</sup>. Folding rate constants of the U2-U6 complex represented by k<sub>1</sub>, k<sub>-1</sub>, k<sub>2</sub> and k<sub>-2</sub>.

Taken together, these data indicate that in the absence of post-transcriptionally modified bases, high concentration of divalent ions facilitates the formation of the three-helix structure of the human U2-U6 complex similar to that of yeast snRNAs (134).

#### ***4.3.3 Single-molecule detection of the human U2-U6 complex with modifications***

Human U2-U6 snRNA complex contains a large number of posttranscriptional modifications in both strands (Fig. 1.9). To understand the effect of these modifications on the structural dynamics of the human U2-U6 complex, we incorporated 2'-O-methylguanosines - G<sub>m</sub> at position 11, 12, 19 and a pseudouridine at position 15 into U2 stem I (Fig. 4.6a). The human U2-U6 complex containing all four modified residues exhibited a similar Mg<sup>2+</sup> titration curve and Mg<sup>2+</sup> dissociation constant ( $K_{Mg} = 2.2 \pm 0.2$  mM) comparable to that of the unmodified spliceosomal RNA construct in ssFRET analysis (Fig. 4.3b). However, the final FRET value of the modified construct is slightly higher than the unmodified construct in 40 mM Mg<sup>2+</sup> suggesting that modifications may destabilize the low FRET conformation.

To reveal the effect of these modified residues in detail, folding dynamics of modified RNA constructs were monitored at the single-molecule level (Fig. 4.6). The smFRET trajectories of modified human U2-U6 complexes also showed Mg<sup>2+</sup>-dependent conformational changes with three FRET states as previously observed for the unmodified construct (Fig. 4.6).



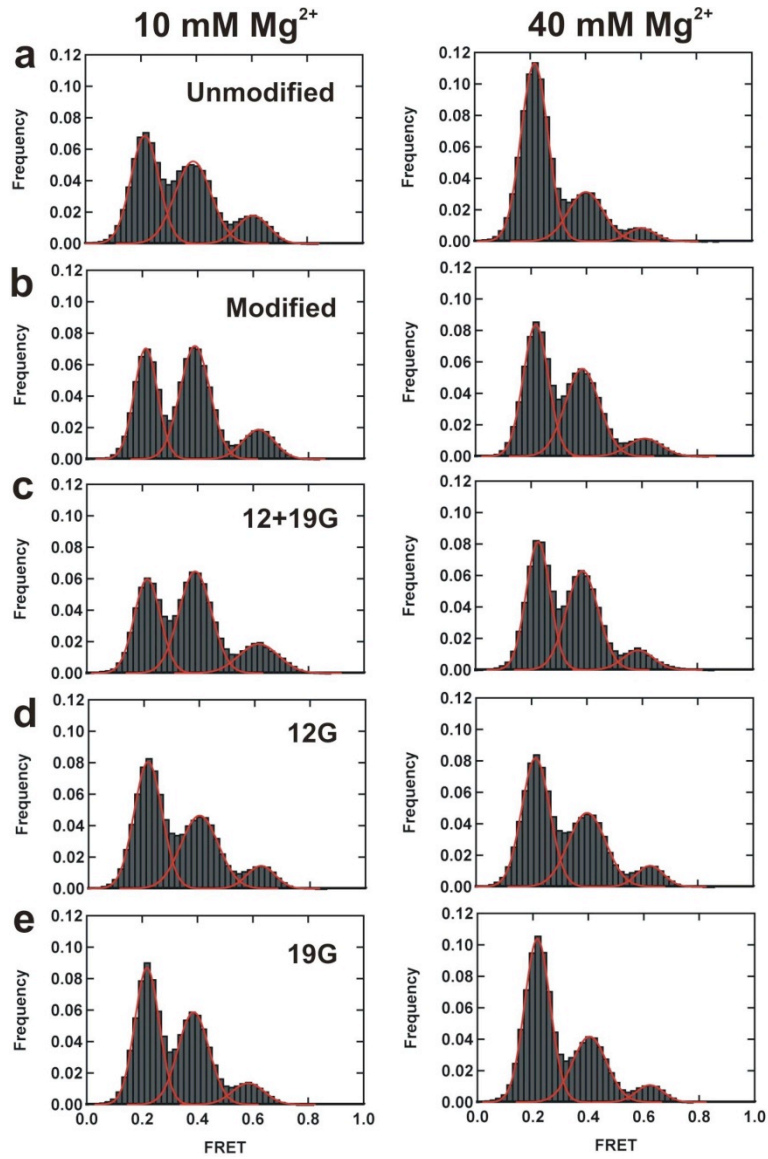
**Figure 4.6 Single-molecule analysis of the modified human U2-U6 complex.** **a.** Structure of the post-transcriptionally modified human U2-U6 snRNA complex. **b.** Folding rate constants of the human U2-U6 snRNA complex in 40 mM Mg<sup>2+</sup>. **c** and **d.** FRET histograms of the human U2-U6 complex with and without modifications in 10 mM and 40 mM Mg<sup>2+</sup>.

Analysis of the FRET histograms shows that the incorporation of modified bases in U2 stem I increases the population of the mid FRET state while decreasing the population of the low FRET state at both  $Mg^{2+}$  concentrations. However, this effect was more prominent in 40 mM  $Mg^{2+}$  than that of 10 mM  $Mg^{2+}$  (Fig. 4.6). Overall this data indicate that the above mentioned modifications may stabilize the four-helix structure while decreasing the stability of the three-helix structure of the U2-U6 complex.

Comparison of the folding rate constants of the U2-U6 complex with and without modifications shows that the folding rate constants  $k_1$  and  $k_{-1}$  did not vary significantly with modifications in both magnesium ion concentrations ( $p > 0.02$ ), but modified bases directly affect the folding rate constant  $k_2$  and  $k_{-2}$  in 40 mM  $Mg^{2+}$ . Under high magnesium ion conditions, the human U2-U6 snRNA construct with modifications decreases  $k_2$  by ~2-fold ( $p < 0.02$ ) and increases  $k_{-2}$  by ~1.5-fold ( $p < 0.02$ ). Therefore, modified bases may destabilize the formation of the low FRET structure corresponding to the three-helix structure of the U2-U6 complex (Fig. 4.6b).

#### **4.3.4 Characterization of individual modifications in U2 stem I**

To understand the effect of individual bases on the structural dynamics of the U2-U6 complex, single-molecule experiments were performed with U2-U6 snRNA constructs containing 2'-O-methylguanosine ( $G_m$ ) at positions 12 ( $12G_m$ ), 19 ( $19G_m$ ), and both 12 and 19 ( $12G_m + 19G_m$ ) on the U2 snRNA strand (Fig. 4.7).



**Figure 4.7 Single-molecule FRET histograms of human U2-U6 snRNA complexes.** **a.** FRET histogram of the unmodified construct. **b.** FRET histogram of the modified construct containing 2'-O-methylguanosines at positions 11, 12, and 19, and pseudouridine at position 15. **c.** FRET histogram of the modified construct containing 12G<sub>m</sub> and 19G<sub>m</sub>. **d.** FRET histogram of the modified construct containing 12G<sub>m</sub>. **e.** FRET histogram of the modified construct containing 19G<sub>m</sub>.

U2U6 snRNA Complex	$\Delta\Delta G$ (kcal/mol)
Unmodified	0
11 $G_m$ + 12 $G_m$ + 19 $G_m$ + 15 $\psi$	$0.53 \pm 0.17$
12 $G_m$ + 19 $G_m$	$0.51 \pm 0.18$
12 $G_m$	$0.30 \pm 0.16$
19 $G_m$	$0.19 \pm 0.13$

**Table 4.1** Difference between calculated free energy changes ( $\Delta\Delta G$ ) of the unmodified and modified human U2-U6 complexes.  $\Delta\Delta G > 0$  destabilizes the low FRET state. Error associated with  $\Delta\Delta G$  was analyzed using the bootstrap method.  $\psi$  - pseudouridine and  $G_m$  - 2'-O-methylguanosine.

The influence of post-transcriptional modifications on the structural dynamics of RNA was determined by the differences between free energy changes ( $\Delta\Delta G$ ) of the unmodified and modified U2-U6 complexes (Fig. 4.7 and Table 4.1). According to the smFRET histograms and free energy calculations, the U2-U6 construct containing both 12G<sub>m</sub> and 19G<sub>m</sub> on the U2 strand showed similar behavior to the U2-U6 construct containing all four modified bases described earlier. As a result, the pseudouridine in position 15 and the 2'-O-methylguanosine at position 11 of the U2 strand did not show an overall effect on the U2-U6 structural dynamics under these conditions. Furthermore, the spliceosomal RNA construct containing individual modifications slightly destabilized the low FRET state of the U2-U6 complex compared to the construct without modified bases. Therefore, these smFRET data suggest a collective effect of modified bases during the destabilization of the three-helix structure of the U2-U6 snRNA complex.

#### **4.4 Discussion**

The U2-U6 snRNA complex is the main component of the catalytic core of the eukaryotic spliceosome. The human U2-U6 snRNA complex can catalyze a splicing related reaction in the absence of proteins, providing evidence for the catalytic ability of these spliceosomal RNAs (11,13,107). However, the specific conformation of the human U2-U6 snRNA complex that triggers the splicing reaction is not well understood. Therefore, the human U2-U6 complex was



characterized using single-molecule fluorescence to investigate the structural dynamics in real time.

Consistent with previous studies, our smFRET data suggest that the human U2-U6 snRNA complex can act as a  $Mg^{2+}$ -dependent conformational switch. Similar to single-molecule folding dynamics of yeast U2-U6 snRNAs (134), human U2-U6 snRNA complexes without post-transcriptionally modified bases in the U2 stem I show the presence of at least three distinct FRET states in dynamic equilibrium corresponding to different structural conformations (Fig. 4.5). Based on sequence similarities between human and yeast spliceosomal RNAs and mutational studies, the observed FRET states have been assigned as the three-helix structure (FRET  $\sim 0.2$ ), the four-helix structure (FRET  $\sim 0.6$ ), and the previously unidentified intermediate structure (FRET  $\sim 0.4$ ).

Considering the positions of the FRET donor and acceptor fluorophores, we propose that the highly conserved ACAGAGA sequence and the magnesium binding site U-74 in U6 ISL are in close proximity in the high FRET state. Therefore, the high FRET state resembles the four-helix structure with the extended U6 ISL containing the AGC triad proposed by NMR studies (Fig. 4.5). Furthermore, this structure supports the important properties of the active conformation exhibiting close association of the highly conserved regions that have been shown to participate in the splicing reaction.

In high magnesium concentrations, the U2-U6 complex prefers the low FRET state, which has two fluorophores far away from each other. As we

previously observed in yeast, the human U2-U6 complex can also adopt the three-helix structure with helix Ib by rearranging the junction structure (Fig. 4.5). The existence of the three-helix structure with helix Ib was further confirmed using a four-fold mutant that prevents the formation of helix Ib (Fig. 4.4). Although it can maintain the junction structure, this mutant construct containing the flipped bases in the U2 stem I greatly decreases the low FRET state population providing a direct evidence for this conformation as the previously proposed three-helix structure.

In addition to the conformations discussed above, smFRET data of U2-U6 snRNAs exhibited the existence of a previously unidentified folding intermediate with FRET  $\sim 0.4$ . Since previous studies have shown that the stability of this conformation is independent of the mutations at U-80 in yeast (comparable to U-74 in humans) U6 snRNA, we propose this mid FRET state has the extended U6 ISL structure, but lacks the tertiary interactions between the ACAGAGA sequence and the U6 ISL (Fig. 4.5). However, the role of this conformation in pre-mRNA splicing is not well understood in both humans and yeast.

Consistent with yeast smFRET data (134), the population of low FRET state (FRET  $\sim 0.2$ ) increases with  $Mg^{2+}$  concentration indicating a magnesium-mediated stabilization of the low FRET state. As a result, under high  $Mg^{2+}$  concentrations, the human U2-U6 complex without post-transcriptionally modified bases showed a higher population of molecules in the low FRET state than the higher FRET states.

In contrast, the human U2-U6 snRNAs with modified bases in U2 stem I (11G<sub>m</sub>, 12G<sub>m</sub>, 19G<sub>m</sub> and pseudouridine at position 15) show a clear stabilization effect of the mid FRET state while lowering the population of the low FRET state corresponding to the three-helix structure (Fig. 4.6).

Interestingly, smFRET data obtained from the U2-U6 construct with 2'-O-methylguanosines at positions 12 and 19 (12G<sub>m</sub> + 19G<sub>m</sub>) exhibited similar folding dynamics to the fully modified construct (Fig 4.7 and Table 4.1). Therefore, these results suggest that 11G<sub>m</sub> and pseudouridine at position 15 of U2 stem I do not exhibit an overall effect on the folding dynamics of the U2-U6 complex. This unique behavior can be due to either cancellation of opposite effects or due to lack of direct stabilization/destabilization effect from the two individual modifications. Furthermore, the human U2-U6 complex with individual modifications (12G<sub>m</sub> or 19G<sub>m</sub>) showed only a slight destabilization effect on the low FRET state compared to the U2-U6 complex without modified bases (Table 4.1).

Taken together, the smFRET data suggest that modified bases collectively affect the structural dynamics of the U2-U6 complex by stabilizing U2 stem I. Thus, these modifications may prevent the formation of the three-helix structure consisting of helix Ib. Although the modified bases destabilize the low FRET state, the effect of modifications on the stability of U2 stem I is less significant in the U2-U6 complex than in the U2 stem I alone as shown in UV melting studies (118). Therefore, we propose that post-transcriptionally modified residues may be involved in protein recognition and interactions rather than the direct stabilization

of U2-U6 RNA structures *in vivo*. Additionally, these modified bases may participate in structural stabilization during the early assembly of the spliceosome as suggested previously (156).

#### **4.5 Conclusions**

Nuclear pre-mRNA splicing is an essential process in eukaryotic gene expression that is catalyzed by a large RNA-protein complex known as the spliceosome. Although the spliceosome consists of five snRNAs and more than 150 proteins (98), previous studies have shown that the spliceosomal U2 and U6 snRNAs interact with the pre-mRNA and are the primary contributors to the formation of the spliceosomal catalytic core. Therefore, the U2-U6 snRNA complex plays a critical role in pre-mRNA splicing in all eukaryotes (11,13,107).

To elucidate the conformations of the U2-U6 snRNA complex involved in the splicing reaction, the spliceosomal U2-U6 RNAs have been characterized using various genetic and biochemical methods. Initial genetic studies have suggested the human U2-U6 complex also adopts the three-helix structure to mediate the splicing reaction as observed in yeast (116). However, genetic complementation studies have shown residues involved in the formation of the helix Ib prefer to form intramolecular base pairing rather than intermolecular base pairing between human U2 and U6 snRNAs (117). Thus, they proposed human U2-U6 snRNAs form the four-helix structure similar to the yeast U2-U6 complex proposed by NMR studies (108,117). As a result of these two contradicting views,

the structure of the U2-U6 complex participating in pre-mRNA splicing is not well understood.

Although it has been hypothesized, the ability of the U2-U6 complex to adopt at least two different conformations corresponding to different activation states has not been well characterized experimentally. In order to address this issue, we used single-molecule prism-based TIRF microscopy to characterize the dynamics of human spliceosomal RNAs. Our single-molecule fluorescence studies of human spliceosomal RNAs show that both previously proposed structures are not only present, but also in a dynamic equilibrium in the presence of magnesium ions similar to yeast U2-U6 complex (134). In addition, single-molecule data indicate the presence of at least one obligatory intermediate conformation in the U2-U6 folding pathway. Furthermore, we suggest the three-helix structure of the human U2-U6 complex contains helix Ib, which forms between the catalytic triad of U6 snRNA and U2 stem I residues.

One of the major differences between human and yeast spliceosomal RNAs is the presence of several post-transcriptional modifications in human snRNAs compared to yeast snRNAs (118). To understand the effect of these modifications on structural dynamics, single-molecule experiments were performed in the presence of modifications in U2 stem I. Although the U2-U6 complex without these modifications shows the similar behavior to yeast, the spliceosomal RNAs containing modified bases destabilize the low FRET state while increasing the population of the mid FRET state. However, the comparison of free energy

calculations based on the smFRET data and UV melting data (118) suggests modified bases may involve in different functions such as protein recognition and/or early assembly of the spliceosome *in vivo* (156), but not in direct stabilization of RNA structures.

In summary, the human U2-U6 complex follows the minimal two-step folding pathway consisting of the previously proposed two structures and at least one obligatory intermediate conformation. Furthermore, RNA constructs without post-transcriptionally modified bases exhibited similar dynamics as yeast spliceosomal RNAs (134). Although modified bases can affect the structural dynamics of human U2-U6 snRNAs, the structural effects of the U2-U6 complex were less than that of U2 stem I alone (118) suggesting a different role rather than direct stabilization of RNA structures in cells.

## CHAPTER 5

# An RNA Aptamer Based Purification System for Fluorophore Labeled Proteins

### 5.1 Introduction

RNA-protein interactions play an essential role in living organisms. These interactions are involved in the regulation of various processes including mRNA splicing, translation and RNA maturation. In order to properly understand the regulating mechanisms, it is important to directly study the RNA-protein interactions by genetic, biochemical and biophysical approaches (157). Considering fluorescence-based techniques have proven very useful in studying the structure and function of biomolecules (36), researchers have recently utilized fluorescently-labeled proteins in the investigation of complex biological processes both *in vitro* and *in vivo* (158-160).

Although a variety of methods for fluorophore labeling proteins of interest are available (158), these methods are rarely 100% efficient and often result in mixtures of labeled and unlabeled proteins. Unlabeled proteins in solution pose a problem in fluorescence assays because they act as competitors to their labeled counterparts. This competition challenges the ability to conduct kinetic and thermodynamic analyses by fluorescence. Further purifying fluorophore-labeled proteins can be particularly challenging if the protein denatures easily. In order to overcome all of these challenges, we have developed an efficient method to

separate fluorophore labeled proteins from unlabeled proteins under non-denaturing conditions.

Many analytical methods are based on molecular recognition. RNA aptamers exhibit interesting molecular recognition properties by binding a desired ligand with high affinity and specificity (161). Aptamers can bind a wide range of target molecules ranging from large protein domains, such as the HIV-1 nucleocapsid protein (162), to small organic molecules, such as the fluorophore Malachite Green (163), with dissociation constants that often range from micromolar to low picomolar (164,165). These qualities make RNA aptamers ideal molecular recognition tools in analytical, diagnostic and therapeutic applications.

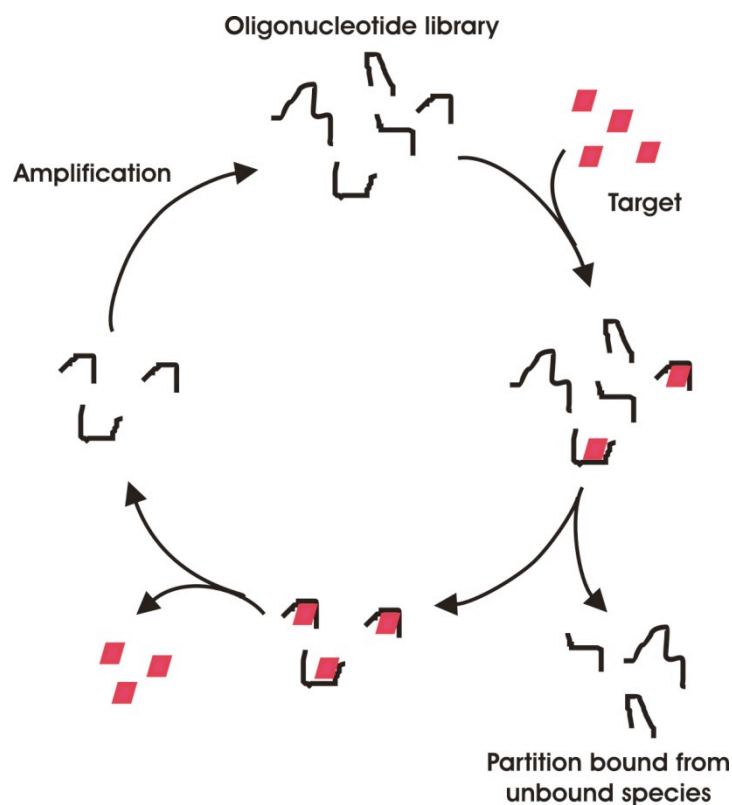
With the development of SELEX (Systematic Evolution of Ligands by Exponential Enrichment), an *in vitro* selection tool, it is possible to quickly isolate and amplify oligonucleotide sequences that specifically bind a desired ligand (161,166). SELEX involves repeating cycles of a selective binding followed by an amplification (Fig. 5.1). The process begins with the synthesis of a single-stranded library of oligonucleotides containing a central region of random sequence flanked by a 5' and 3' region of defined sequences. Since all members of the library contain a unique sequence, the complexity of the synthetic library depends on the number of randomized nucleotides in the central region. During the selection process, the pool of random oligonucleotides is first incubated with the desired target ligand and then the high-affinity binding sequences are separated from non-specific or low-affinity binding sequences by methods such as affinity



chromatography (161) or, more recently, high efficiency electrophoretic methods (167). The oligonucleotide sequences exhibiting high affinity for the target molecules are amplified by RT-PCR (for RNA) or PCR (for DNA). Since PCR methods yield double stranded oligonucleotide products, an additional step is required to produce a single-stranded oligonucleotide library enriched with high-affinity sequences for the next selection cycle.

RNA aptamers display remarkably high affinity and selectivity for their target ligands compared to those of some monoclonal antibodies (168). Thus aptamers can be considered a valid alternative to antibodies or other bio-mimetic receptors. Moreover, chemically synthesizable aptamers possess several advantages over antibodies making them a promising alternative in the field of affinity chromatography (136,169). Roming and co-workers have developed an aptamer specific to L-selectin to purify human L-selectin receptor globulin (LS-Rg) fusion protein (170). Deng *et al.* used an aptameric stationary phase to monitor the level of adenosine in the brain of live rats (171).

The above-mentioned experiments demonstrate that aptamer-chromatographic methods provide a valid alternative to other methods such as HPLC. Consequently, we have taken advantage of the specific properties of aptamers to develop an aptamer-based affinity purification method to purify fluorophore-labeled proteins.



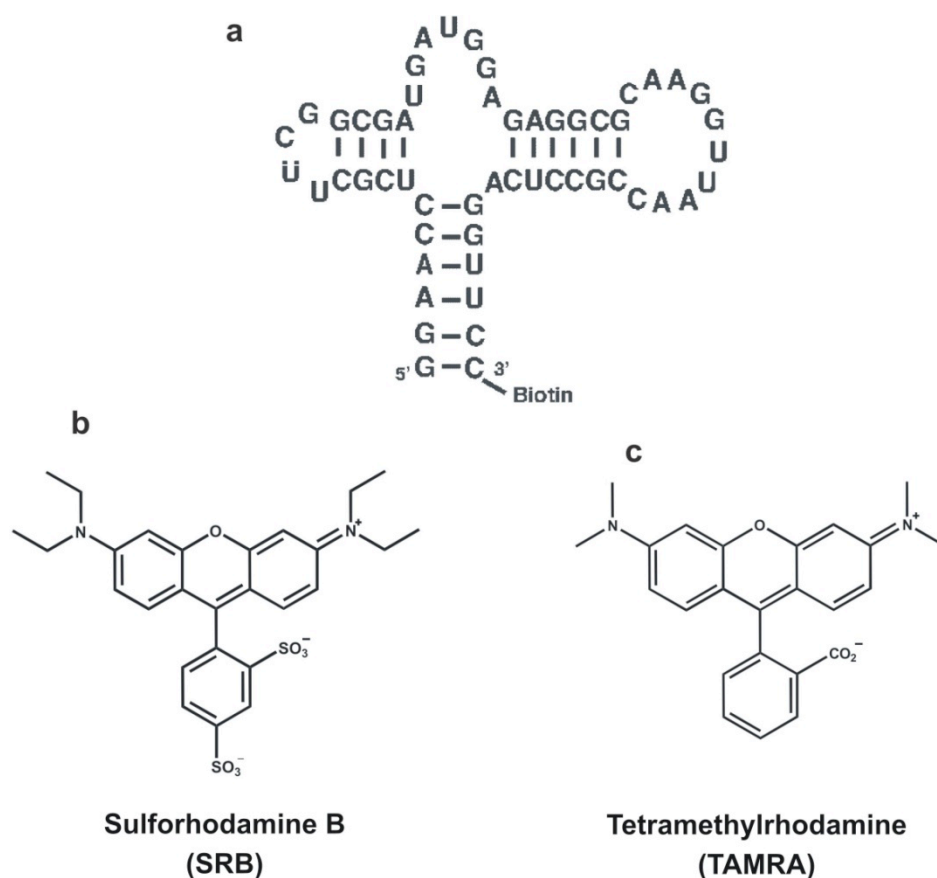
**Figure 5.1** Scheme of *in vitro* selection of oligonucleotide aptamers by **SELEX**. The first step involves the generation of a large pool of random oligonucleotide sequences. After incubating these sequences with target molecules, high affinity binding sequences can be separated from non-specific species. Then these specific sequences can be amplified by Polymerase Chain Reaction (PCR) to make a new enriched library for the next cycle.

Tetramethylrhodamine (TAMRA) is a common fluorophore in numerous fluorescence-based assays (172,173). Therefore, we have chosen this fluorophore as a model to show that RNA aptamers can be used to purify fluorophore-labeled proteins. Previous work has identified an RNA aptamer, SRB-2, which binds sulforhodamine B tightly and selectively ( $K_D = 310$  nM). SRB-2 is a 54 nucleotide RNA aptamer that binds the fluorophore by recognizing both its planar aromatic ring system and its negatively-charged sulfonate group at the ortho position (174). Strong similarities in the structure of sulforhodamine and the more commonly employed tetramethylrhodamine species suggest that this aptamer would be suitable for binding TAMRA (Fig. 5.2). Discussed here is a method by which the affinity of SRB-2 for TAMRA is applied to the purification of TAMRA-labeled maltose binding protein (MBP).

## 5.2 Experimental design

In order to develop an efficient chromatographic method for purification of fluorophore-labeled proteins from the unlabeled fraction under mild reaction conditions, we wanted to construct an affinity purification column using an aptamer RNA specific for the given fluorophore.

Fluorescence anisotropy is a widely used fluorescent-based technique that detects the binding of molecules with fluorophores based on the rotational diffusion of a molecule. Therefore, the specificity and binding properties between the aptamer and the fluorophore of interest were characterized using fluorescence anisotropy under specific buffer conditions as explained in the methods section.



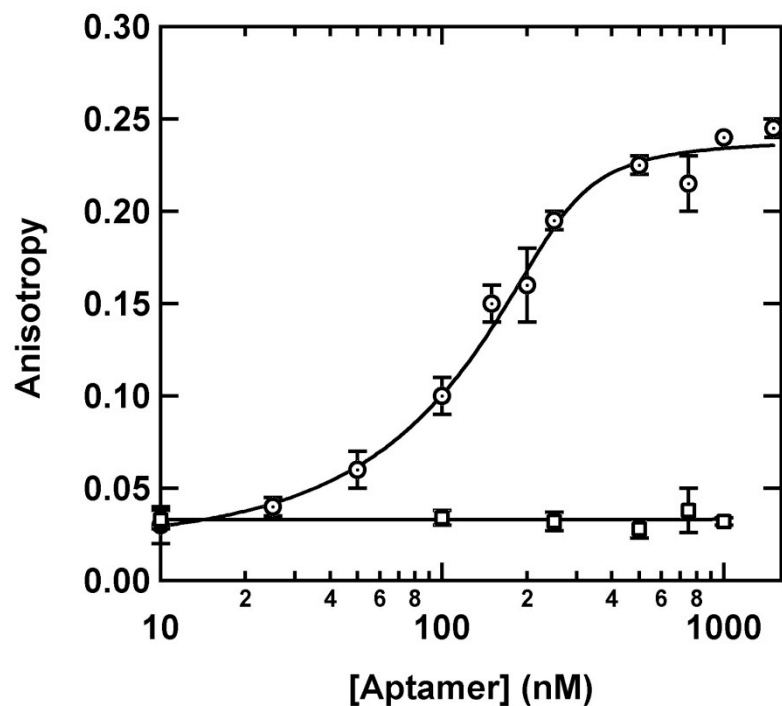
**Figure 5.2 Structure of SRB-2 aptamer and fluorophores** **a.** Secondary structure of the SRB-2 aptamer. The aptamer consists of a three-way helical junction with two large unpaired regions. The 3'-end is biotinylated for column immobilization. **b.** Structure of sulforhodamine B (SRB) **c.** The subject ligand, tetramethylrhodamine (TAMRA), contains a 3-ring aromatic system and a meta-positioned carboxylate group that have previously been shown to be recognized by the SRB-2 aptamer.

After optimizing the binding and elution conditions for the SRB-2 aptamer and tetramethylrhodamine, an affinity chromatographic column was constructed using biotin-labeled SRB-2 aptamer and streptavidin coated 6% crosslinked agarose beads (Fig. 5.2). To characterize the specific activity of this assay, tetramethylrhodamine-labeled maltose binding protein was purified using an SRB-2 aptamer based affinity chromatographic column under optimized reaction conditions.

## 5.3 Results and Discussion

### 5.3.1 SRB-2 binds TAMRA with high affinity and specificity

We first tested the ability of the SRB-2 aptamer to bind TAMRA by fluorescence anisotropy. Fluorescence anisotropy reports on the rotational freedom of the fluorophore, and therefore, we expect to observe a significant increase in TAMRA fluorescence anisotropy upon binding the RNA aptamer. Figure 5.3 shows the fluorescence anisotropy of 250 nM TAMRA as a function of SRB-2 concentration. In the absence of SRB-2, the observed anisotropy of the TAMRA was very low ( $0.04 \pm 0.01$ ), indicating that TAMRA can rotate freely in solution. In the presence of 500 nM aptamer, the fluorescence anisotropy increases to  $0.24 \pm 0.01$ . This high anisotropy value suggests the aptamer has bound the fluorophore and restricts its free rotation. A fit to an SRB-2 titration (Eq. 6, page 60) provides a dissociation constant  $K_D = 15 \pm 6$  nM, indicating that SRB-2 binds TAMRA tighter than sulforhodamine B for which it was initially selected ( $K_D \approx 310$  nM) (174).

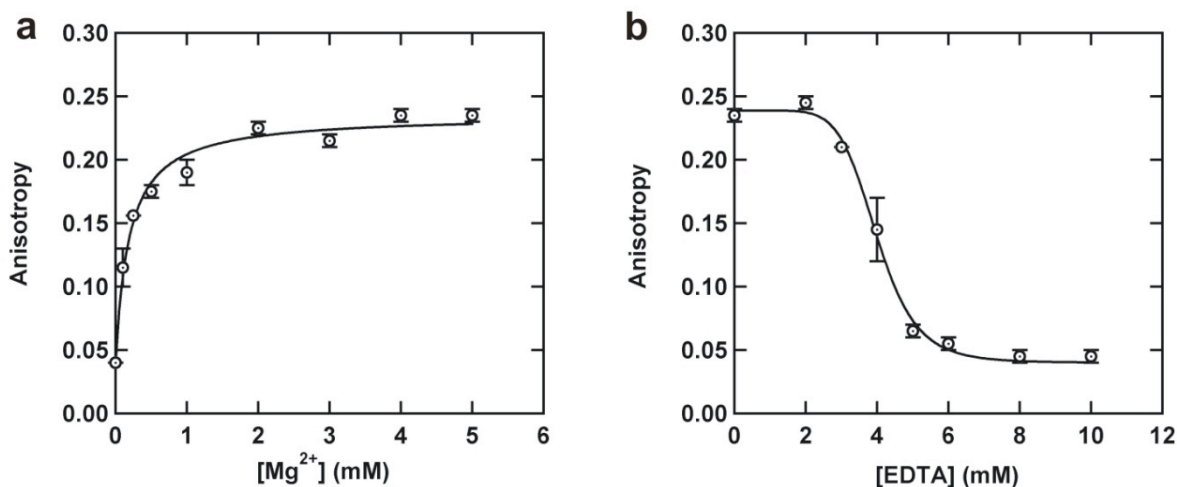


**Figure 5.3 Fluorescence anisotropy value as a function of SRB-2 concentration.** Circles represent the binding of the SRB-2 aptamer and squares represent the binding of control RNA with TAMRA indicating specificity for the TAMRA ligand by SRB-2. Solid line represents the binding curve, fitting to Equation 6 (page 60) results in a dissociation constant  $K_D = 15 \pm 6$  nM. All anisotropy measurements were done in buffer conditions: 100 mM KCl, 5 mM  $MgCl_2$ , 10 mM Na-HEPES, pH 7.4 with 250 nM rhodamine.

The binding specificity of SRB-2 for TAMRA was further tested by anisotropic titration of a negative control RNA sequence (36 nucleotides) under identical buffer conditions (Fig. 5.3). No appreciable change in TAMRA anisotropy was observed with increasing RNA concentrations. We also measured the kinetics of the SRB-2/TAMRA complex formation by monitoring the fluorescence anisotropy increase as a function of time. Our results show that the fluorescence anisotropy increases faster than our time resolution ( $<10$  s, data not shown), indicating that the complex forms rapidly. These results show that SRB-2 binds TAMRA with high affinity and specificity, and thus this aptamer is suitable to develop an affinity purification assay for TAMRA-labeled proteins.

### ***5.3.2 SRB-2 binding can be regulated by $Mg^{2+}$ ions and EDTA***

The SRB-2 aptamer requires the presence of magnesium ions to bind its ligand (174). To control binding and dissociation from the aptamer, we sought to use  $Mg^{2+}$  ions and EDTA, respectively. First, we determined the  $Mg^{2+}$  ion requirement for SRB-2 binding TAMRA by measuring the TAMRA fluorescence anisotropy as a function of  $Mg^{2+}$  concentration. Figure 5.4a shows the fluorescence anisotropy of 250 nM TAMRA in the presence of saturating SRB-2 (500 nM) in selection buffer and as a function of  $[Mg^{2+}]$ . In the absence of  $Mg^{2+}$ , the initial anisotropy is  $0.04 \pm 0.01$ , comparable to that of free TAMRA, indicating that in the absence of  $Mg^{2+}$  ions, the aptamer does not fold into its functional structure and thus, it cannot bind TAMRA.



**Figure 5.4 Effect of  $Mg^{2+}$  and EDTA concentrations on binding of SRB-2 aptamer with TAMRA.** **a.** TAMRA fluorescence anisotropy as a function of  $Mg^{2+}$  concentration. Anisotropy was measured in solutions containing 100 mM KCl, 10 mM Na-HEPES, pH 7.4, 250 nM TAMRA and 500 nM SRB-2 at different concentrations of  $Mg^{2+}$ . The data indicates a significant  $Mg^{2+}$ -binding dependence, a property easily exploited for elution of bound ligands. Curve fitting to Equation 7 (page 60) results in a dissociation constant  $K_{Mg} = 0.19 \pm 0.03$  mM. **b.** TAMRA fluorescence anisotropy as a function of EDTA concentration. The effect of EDTA on binding of SRB-2 with TAMRA was determined by measuring the anisotropy in solutions containing 100 mM KCl, 5 mM  $MgCl_2$ , 10 mM Na-HEPES, pH 7.4, 250 nM TAMRA and 500 nM SRB-2 aptamer. Curve fitting to Equation 8 (page 60) results in  $K_{EDTA} = 4.0 \pm 0.1$  mM and co-operativity coefficient,  $n = 7 \pm 1$ .



Addition of 5 mM  $\text{Mg}^{2+}$  results in an anisotropy increase to  $0.24 \pm 0.01$ , comparable to that of fully bound TAMRA, indicating that in the presence of  $\text{Mg}^{2+}$  ions SRB-2 is capable of binding TAMRA. A fit to a  $\text{Mg}^{2+}$  titration (Eq. 7, page 60) yields a  $K_{\text{Mg}} = 190 \pm 30 \mu\text{M}$  indicating the presence of a strong magnesium ion binding site that is important for TAMRA binding.

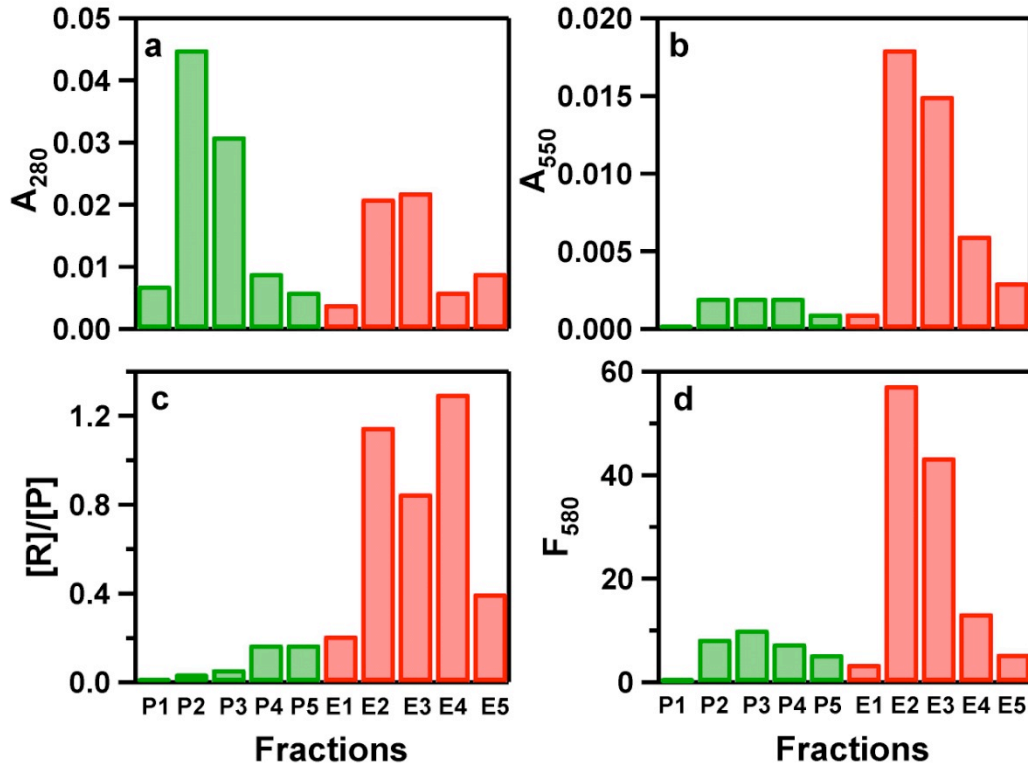
Next, we sought to characterize the amount of EDTA required to release TAMRA from the aptamer by measuring its fluorescence anisotropy in the presence of SRB-2 and 5 mM  $\text{Mg}^{2+}$ , and as a function of EDTA concentration (Fig. 5.4b). In the absence of EDTA, the initial anisotropy is  $r_{\text{max}} = 0.24 \pm 0.01$ , indicating that TAMRA is fully bound. Addition of 5 mM EDTA, results in an anisotropy decrease to  $r_0 = 0.05 \pm 0.01$ , indicating that EDTA has chelated the  $\text{Mg}^{2+}$  ions in solution, thus unfolding the aptamer, which in turn releases TAMRA in solution.

A fit to the resulting EDTA titration (Fig. 5.4b, Eq. 8, page 60) results in a  $K_{\text{EDTA}} = 4.0 \pm 0.1 \text{ mM}$  and a cooperativity coefficient,  $n = 7 \pm 1$ . This coefficient indicates that chelation of RNA-bound  $\text{Mg}^{2+}$  by EDTA is a highly cooperative process and suggests that the aptamer unfolds as  $\text{Mg}^{2+}$  ions become more loosely bound to the RNA. These results show that TAMRA binding to the SRB-2 aptamer and release can be readily regulated under mild, non-denaturing conditions by simply regulating the  $[\text{Mg}^{2+}]$  in solution with EDTA.

### **5.3.3 RNA aptamer affinity purification of a TAMRA labeled protein**

After optimizing the binding conditions, we fluorophore-labeled maltose binding proteins (MBP) with TAMRA and sought to purify the labeled protein by SRB-2 aptamer-based affinity chromatography. In order to perform the purification assay, fluorophore-labeled protein was added to a column containing SRB-2 aptamer bound agarose beads. Fluorophore-labeled MBP was purified and eluted using both washing and elution buffers as described in materials & methods. The column washings and elutions were analyzed by UV-Visible absorbance as well as by fluorescence emission spectroscopy (Fig. 5.5).

Figure 5.5a shows the trend of UV absorbance at 280 nm as a function of column volume, indicating the presence of the maltose binding protein in solution. Fractions labeled P1 through P5 (Fig. 5.5, green bars) of those column volumes eluted in aptamer binding conditions (100 mM KCl, 5 mM MgCl<sub>2</sub>, 10 mM Na-HEPES, pH 7.4) intended to remove unlabeled proteins from the column. Those labeled E1 through E5 (Fig. 5.5, red bars) represent the volumes taken in EDTA elution conditions (5 mM EDTA, 100 mM KCl, 10 mM Na-HEPES, pH 7.4) intended to remove bound TAMRA-labeled proteins. A significant increase in UV absorbance at 280 nm is observed at volume P2, which diminishes in the following three column volumes. An increase in UV absorbance is seen at volumes E2 and E3. Clearly two populations of protein are removed from the column: one under aptamer binding conditions and another under elution conditions.



**Figure 5.5 RNA aptamer affinity purification of a TAMRA labeled protein. a.** UV absorbance at 280 nm as a function of column volume. Volumes labeled P1 through P5 are the result of column washings in binding conditions intended to remove unlabeled proteins, whereas those labeled E1 through E5 are the result of EDTA elutions intended to remove TAMRA-labeled proteins from the aptamer. The significant increases in UV absorbance in P2, P3 and E2, E3 indicate two distinct populations of protein being eluted from the column. **b.** UV-Vis absorbance at 550 nm as a function of column volume. The significant increase in UV-Vis absorbance at E2 indicates the elution of the TAMRA fluorophore with EDTA elution buffer. **c.** Ratio of fluorophore concentration to protein concentration as a function of column volume. The concentration ratios, determined by UV analysis, are approximately 1:1 in elutions E2 through E4, indicating the elution of singly labeled proteins. **d.** Fluorescence emission intensity as a function of column volume. The increase in fluorescence emissions at E2 verifies the elution of the TAMRA fluorophore with EDTA elution buffer.

Figure 5.5b shows the UV-Vis absorbance at 550 nm of the same column volumes as above, indicating the presence of the TAMRA fluorophore species. A small baseline value for volumes P1 through P5 may indicate a small percentage of fluorophore-labeled proteins binding on part of the immobilized SRB-2. A significant increase of UV absorbance is observed at volumes E2 and E3, and congruent with the previous figure, indicates the elution of TAMRA-labeled protein in these volumes.

Figure 5.5c shows the ratio of concentrations of the TAMRA fluorophore to those of the MBP protein in successive column volumes calculated from the absorbance data presented in Figures 5.5a and b. Values smaller than 1 in volumes P1 through P5 show an abundance of MBP with respect to TAMRA, indicating predominantly unlabeled proteins present in solution, whereas values approximately equal to 1 seen in elution E2-E4 indicate the fraction of the singly TAMRA-labeled MBP proteins.

The data presented in Figure 5.5d, fluorescence emission intensity ( $\lambda_{\text{ex}} = 555 \text{ nm}$ ,  $\lambda_{\text{em}} = 580 \text{ nm}$ ) as a function of elution fraction, compliments the data presented in Figure 5.5b. An increase in fluorescence emission intensity at E2 and E3 indicates that the rise in UV absorbance seen in Figure 5.5b is in fact brought about by the elution of the TAMRA fluorophore species in the respective column volumes.

## 5.4 Conclusions

With the objective of isolating fluorophore labeled-proteins from mixtures of labeled and unlabeled proteins such as those resulting from inefficient labeling reactions, we have developed a purification method based on an RNA aptamer specific to the fluorophore species. The modified maltose binding protein (MBP) labeled with the tetramethylrhodamine (TAMRA) fluorophore species by a cysteine-maleimide reaction has been effectively isolated from unlabeled MBP by passing the mixture through an affinity chromatographic column containing sulforhodamine binding aptamer (SRB-2) immobilized by a biotin-streptavidin interaction (Fig. 5.2).

The binding of SRB-2 to the TAMRA fluorophore species was first characterized, and then binding and elution conditions optimized by a fluorescence anisotropic assay (Figs. 5.3 and 5.4). Following such characterizations, the aptamer was 3'-biotinylated and immobilized in an LC column bed of streptavidin-coated agarose beads. The column elutions were analyzed by UV-Visible and fluorescence spectrophotometry for their protein and fluorophore contents (Fig. 5.5). The result is a well-resolved separation of fluorophore-labeled proteins from identical proteins, which were not effectively labeled in the labeling reaction. The majority of labeled proteins eluted over only two column volumes.

The method described herein poses several unique advantages for the purification of fluorophore labeled proteins. Certainly the primary advantage of this approach is the ability to purify protein samples in native buffer and pH conditions.

The benefit of this method is that it ensures the purified protein will not have been denatured, thus preserving the structure and function of the subject protein. Another advantage of this method is its inherent versatility. The growing field of aptamer chemistry continues to identify new aptamer-ligand systems. Therefore, this method may be easily expanded to purify proteins modified *in vitro* by other labeling reactions. Furthermore, this method may be useful in the isolation of proteins modified *in vivo* by employing aptamers selected against substrates of such modifications. Finally, adaptations of this method to employ DNA aptamers may yield highly stable affinity purification systems with extended shelf lives.

## CHAPTER 6

### Conclusions and Future Directions

Single-molecule fluorescence has a unique ability to characterize RNA folding pathways by identifying transient intermediates in real times that can be otherwise hidden in ensemble-averaged experiments (36). Therefore, the folding pathway of group II introns and the effect of Mss116 DEAD box proteins on group II intron folding as well as the dynamics of the human U2-U6 spliceosomal RNA complex were characterized using the TIRF-based single-molecule microscopy (37).

Group II introns can self-splice *in vitro* under high ionic strength conditions. Previous single-molecule studies have been able to successfully characterize the three-step folding pathway of group II introns under these optimum *in vitro* splicing conditions (68). However, *in vivo*, the splicing of group II introns is primarily assisted by protein cofactors such as the Mss116 DEAD-box protein. Biochemical and folding studies (80,87-89,92) have proposed that Mss116 can promote group II intron splicing by stabilizing on-pathway intermediates and/or by disrupting misfolded structures. However, the exact mechanism of protein-mediated splicing is still debatable.

Previous studies have revealed that the yeast group II intron ai5 $\gamma$  can form a catalytically active native structure through a slow but smooth folding pathway devoid of kinetic traps (80,87). Furthermore, it has been shown that a helicase-deficient Mss116 mutant still retains the ability to promote group II intron splicing

(88,89,155). Based on these observations, it has been suggested that Mss116 catalyzes splicing by stabilizing the RNA structures. In contrast, some evidence supports a mechanism by which Mss116 unwinds kinetic traps to assist splicing, even at low levels of helicase activity (88,89). Both proposed mechanisms, however, are primarily based on indirect studies of ai5 $\gamma$  group II intron splicing. To address this issue, a single-molecule fluorescence assay has been used to characterize the effect of Mss116 on ai5 $\gamma$  folding by directly monitoring the influence of Mss116 and ATP on group II intron folding (37).

Group II introns alone cannot reach the catalytically active native state under near-physiological conditions. According to single-molecule studies, large energy barriers between states prevent the formation of the native state under these conditions in the absence of protein cofactors. However, the DEAD-box protein Mss116 or non-specific basic RNA binding proteins can lower the activation barriers and promote the folding of intron RNA to the active native state. Although transitions from the extended intermediate state to the folded intermediate state are independent from the ATP hydrolysis, the binding and hydrolysis of ATP is required for efficient formation of the native state under these conditions as illustrated in the free-energy diagram (Fig. 3.12) (37).

In summary, Mss116 has multiple roles in the folding pathway of ai5 $\gamma$ . Based on smFRET data, we propose that Mss116 mediates group II intron folding by stabilizing on-pathway intermediates upon transition from the unfolded state to the folded state. However, we also observe that the transition from the



folded intermediate state to the native state requires ATP hydrolysis. Taken together, we propose a three-step folding pathway for ai5 $\gamma$  in the presence of Mss116 and ATP. First, unfolded group II intron ribozymes (U) slowly collapse to the extended intermediate state (I) under near physiological conditions, as previously observed (87). As expected, this transition is so slow that we did not observe it in our experiments. The extended intermediate can interact with Mss116 to further compact this state. Next, Mss116 (or other proteins) induces formation of the folded intermediate (F). Finally, the native state is stably formed once ATP is hydrolyzed, possibly upon Mss116 release. Our data also show that the presence of substrate further stabilizes the native state, raising the interesting possibility that *in vivo* and in the presence of all the group II intron domains and Mss116, the native state becomes the most stable conformation (37).

In order to confirm the proposed Mss116-mediated folding pathway, we are planning to monitor the folding of group II introns using the fluorophore-labeled Mss116. Thereby, we can simultaneously detect the RNA folding and the interactions between the RNA and the protein *in vitro*. Furthermore, the single-molecule assay can be used to understand the effect of the other DEAD-box proteins such as CYT19 and Deb1 on the folding of group II introns. Finally, this assay will allow us to characterize the essential structural interactions of group II introns using mutant ai5 $\gamma$  constructs.

The structure and catalytic mechanism of group II introns are analogous to the spliceosome-catalyzed nuclear pre-mRNA splicing in eukaryotes (57).

Although the spliceosome consists of five snRNAs and more than 150 proteins, previous studies have shown that the spliceosomal U2 and U6 snRNAs interact with the pre-mRNA and are the primary contributors to the formation of the spliceosomal catalytic core. Therefore, the U2-U6 snRNA complex plays an important role in pre-mRNA splicing in all eukaryotes (11,13,107).

In order to understand the structural role of the U2-U6 snRNA complex in the splicing reaction, the spliceosomal U2-U6 RNAs have been characterized using various genetic and biochemical methods. Initial genetic studies have suggested the human U2-U6 complex can adopt a three-helix structure to mediate the splicing reaction as observed in yeast (116). However, genetic complementation studies have shown residues involved in the formation of helix Ib prefer to form intramolecular base pairing rather than intermolecular base pairing between human U2 and U6 snRNAs (117). Thus, they proposed human U2-U6 snRNAs form the four-helix structure similar to the yeast U2-U6 complex proposed by NMR studies (108,117). As a result of these conflicting views, the catalytically active structure of the human U2-U6 complex participating in pre-mRNA splicing is not well understood.

It has been hypothesized that the U2-U6 complex can adopt different conformations corresponding to different activation states; however, these dynamics have not been well characterized experimentally. In order to address this issue, we used single-molecule prism-based TIRF microscopy to characterize the dynamics of human spliceosomal RNAs. Our single-molecule fluorescence

studies of human spliceosomal RNAs show that both previously proposed structures are not only present, but also in a dynamic equilibrium in the presence of magnesium ions similar to yeast U2-U6 complex (134). In addition, single-molecule data indicate the presence of at least one obligatory intermediate conformation in the U2-U6 folding pathway. Furthermore, we suggest the three-helix structure of the human U2-U6 complex contains helix Ib, which forms between the catalytic triad of U6 snRNA and U2 stem I residues.

One of the major differences between human and yeast spliceosomal RNAs is the presence of several post-transcriptional modifications in human snRNAs compared to yeast snRNAs (118). To understand the effect of these modifications on structural dynamics, single-molecule experiments were performed in the presence of a few important modifications in U2 stem I (2'-O-methylguanosines - G<sub>m</sub> at positions 11, 12 and 19, and pseudouridine at position 15). Although the U2-U6 complex without these modifications shows similar behavior to yeast, the spliceosomal RNAs containing modified bases destabilize the low FRET state while increasing the population of the mid FRET state. However, comparison of free energy calculations from smFRET data and UV melting data (118) suggests modified bases may be involved in different functions such as protein recognition and early assembly of the spliceosome (156), but not in direct stabilization of RNA structures.

In summary, the human U2-U6 complex follows the minimal two-step folding pathway consisting of the previously proposed two structures and at least

one obligatory intermediate conformation. Furthermore, RNA constructs without post-transcriptionally modified bases exhibited similar dynamics as yeast spliceosomal RNAs (134). Although modified bases can affect the structural dynamics of human U2-U6 snRNAs, the structural effects of the modified bases on the U2-U6 complex were less than the effects on the U2 stem I alone (118). This suggests the modified bases may have an alternative role besides direct stabilization of RNA structures in cells. Therefore, we propose modified bases predominantly participate in protein recognition and/or early assembly of the spliceosome (156) rather than direct stabilization of the U2-U6 snRNA complex *in vivo*.

In order to further understand the structural dynamics of spliceosomal U2-U6 snRNAs that are essential for pre-mRNA splicing and to clarify the specific role of post-transcriptionally modified bases of snRNAs in splicing, we can monitor the dynamics of the fluorophore-labeled U2-U6 snRNA complex in the presence of the cell extract containing all necessary spliceosomal components. Furthermore, these experiments will provide an opportunity to monitor the pre-mRNA splicing in real time revealing fundamental structural rearrangements between RNA and protein components in the spliceosome.

Overall, we were able to successfully combine FRET with single-molecule detection to dissect RNA splicing mechanisms. In the process, we revealed the role of Mss116 DEAD-box protein on the group II intron folding pathway and monitored structural dynamics of human U2-U6 snRNAs that are essential for pre-

mRNA splicing *in vitro*. Characterization of these fundamental splicing mechanisms in eukaryotes will not only expand the existing knowledge in the splicing field, but also pave the way for future research.

**REFERENCES**

1. Cech, T. R. (2009) Evolution of biological catalysis: ribozyme to RNP enzyme. *Cold Spring Harb Symp Quant Biol.* **74**, 11-16
2. Bartel, D. P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell.* **116**, 281-297
3. Carthew, R. W., and Sontheimer, E. J. (2009) Origins and Mechanisms of miRNAs and siRNAs. *Cell.* **136**, 642-655
4. Valencia-Sanchez, M. A., Liu, J., Hannon, G. J., and Parker, R. (2006) Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes Dev.* **20**, 515-524
5. Tucker, B. J., and Breaker, R. R. (2005) Riboswitches as versatile gene control elements. *Curr Opin Struct Biol.* **15**, 342-348
6. Winkler, W. C., and Breaker, R. R. (2005) Regulation of bacterial gene expression by riboswitches. *Annu Rev Microbiol.* **59**, 487-517
7. Cech, T. R., Tanner, N. K., Tinoco, I., Jr., Weir, B. R., Zuker, M., and Perlman, P. S. (1983) Secondary structure of the Tetrahymena ribosomal RNA intervening sequence: structural homology with fungal mitochondrial intervening sequences. *Proc Natl Acad Sci U S A.* **80**, 3903-3907
8. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., and Altman, S. (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell.* **35**, 849-857

9. Abelson, J. (2008) Is the spliceosome a ribonucleoprotein enzyme? *Nat Struct Mol Biol.* **15**, 1235-1237
10. Valadkhan, S. (2007) The spliceosome: a ribozyme at heart? *Biol Chem.* **388**, 693-697
11. Valadkhan, S., and Manley, J. L. (2001) Splicing-related catalysis by protein-free snRNAs. *Nature.* **413**, 701-707
12. Valadkhan, S., Mohammadi, A., Wachtel, C., and Manley, J. L. (2007) Protein-free spliceosomal snRNAs catalyze a reaction that resembles the first step of splicing. *RNA.* **13**, 2300-2311
13. Valadkhan, S. (2005) snRNAs as the catalysts of pre-mRNA splicing. *Curr Opin Chem Biol.* **9**, 603-608
14. Ban, N., Nissen, P., Hansen, J., Moore, P. B., and Steitz, T. A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science.* **289**, 905-920
15. Nissen, P., Hansen, J., Ban, N., Moore, P. B., and Steitz, T. A. (2000) The structural basis of ribosome activity in peptide bond synthesis. *Science.* **289**, 920-930
16. Noller, H. F., Hoffarth, V., and Zimniak, L. (1992) Unusual resistance of peptidyl transferase to protein extraction procedures. *Science.* **256**, 1416-1419
17. Pyle, A. M. (2002) Metal ions in the structure and function of RNA. *J Biol Inorg Chem.* **7**, 679-690

18. Shiman, R., and Draper, D. E. (2000) Stabilization of RNA tertiary structure by monovalent cations. *J Mol Biol.* **302**, 79-91
19. Woodson, S. A. (2005) Metal ions and RNA folding: a highly charged topic with a dynamic future. *Curr Opin Chem Biol.* **9**, 104-109
20. Das, R., Kwok, L. W., Millett, I. S., Bai, Y., Mills, T. T., Jacob, J., Maskel, G. S., Seifert, S., Mochrie, S. G., Thiyagarajan, P., Doniach, S., Pollack, L., and Herschlag, D. (2003) The fastest global events in RNA folding: electrostatic relaxation and tertiary collapse of the Tetrahymena ribozyme. *J Mol Biol.* **332**, 311-319
21. Schroeder, R., Grossberger, R., Pichler, A., and Waldsich, C. (2002) RNA folding in vivo. *Curr Opin Struct Biol.* **12**, 296-300
22. Rajkowitsch, L., Chen, D., Stampfl, S., Semrad, K., Waldsich, C., Mayer, O., Jantsch, M. F., Konrat, R., Blasi, U., and Schroeder, R. (2007) RNA chaperones, RNA annealers and RNA helicases. *RNA Biol.* **4**, 118-130
23. Draper, D. E. (2008) RNA folding: thermodynamic and molecular descriptions of the roles of ions. *Biophys J.* **95**, 5489-5495
24. Hendrix, D. K., Brenner, S. E., and Holbrook, S. R. (2005) RNA structural motifs: building blocks of a modular biomolecule. *Q Rev Biophys.* **38**, 221-243
25. Wyatt, J. R., Puglisi, J. D., and Tinoco, I., Jr. (1989) RNA folding: pseudoknots, loops and bulges. *Bioessays.* **11**, 100-106



26. Leontis, N. B., Lescoute, A., and Westhof, E. (2006) The building blocks and motifs of RNA architecture. *Curr Opin Struct Biol.* **16**, 279-287
27. Batey, R. T., Rambo, R. P., and Doudna, J. A. (1999) Tertiary Motifs in RNA Structure and Folding. *Angew Chem Int Ed Engl.* **38**, 2326-2343
28. Herschlag, D. (1995) RNA chaperones and the RNA folding problem. *J Biol Chem.* **270**, 20871-20874
29. Solomatin, S. V., Greenfeld, M., Chu, S., and Herschlag, D. (2010) Multiple native states reveal persistent ruggedness of an RNA folding landscape. *Nature.* **463**, 681-684
30. DeRose, V. J. (2003) Metal ion binding to catalytic RNA molecules. *Curr Opin Struct Biol.* **13**, 317-324
31. Shcherbakova, I., Gupta, S., Chance, M. R., and Brenowitz, M. (2004) Monovalent ion-mediated folding of the *Tetrahymena thermophila* ribozyme. *J Mol Biol.* **342**, 1431-1442
32. Piccirilli, J. A., Vyle, J. S., Caruthers, M. H., and Cech, T. R. (1993) Metal ion catalysis in the *Tetrahymena* ribozyme reaction. *Nature.* **361**, 85-88
33. Strobel, S. A., and Cochrane, J. C. (2007) RNA catalysis: ribozymes, ribosomes, and riboswitches. *Curr Opin Chem Biol.* **11**, 636-643
34. Sigel, R. K., and Pyle, A. M. (2007) Alternative roles for metal ions in enzyme catalysis and the implications for ribozyme chemistry. *Chem Rev.* **107**, 97-113

35. Schroeder, R., Barta, A., and Semrad, K. (2004) Strategies for RNA folding and assembly. *Nat Rev Mol Cell Biol.* **5**, 908-919
36. Karunatilaka, K. S., and Rueda, D. (2009) Single-Molecule Fluorescence Studies of RNA: A Decade's Progress. *Chem Phys Lett.* **476**, 1-10
37. Karunatilaka, K. S., Solem, A., Pyle, A. M., and Rueda, D. (2010) Single-molecule analysis of Mss116-mediated group II intron folding. *Nature.* **467**, 935-939
38. Mohr, G., Zhang, A., Ganelos, J. A., Belfort, M., and Lambowitz, A. M. (1992) The neurospora CYT-18 protein suppresses defects in the phage T4 td intron by stabilizing the catalytically active structure of the intron core. *Cell.* **69**, 483-494
39. Yang, Q., and Jankowsky, E. (2005) ATP- and ADP-dependent modulation of RNA unwinding and strand annealing activities by the DEAD-box protein DED1. *Biochemistry.* **44**, 13591-13601
40. Muller, U. F., and Goring, H. U. (2002) Mechanism of the gBP21-mediated RNA/RNA annealing reaction: matchmaking and charge reduction. *Nucleic Acids Res.* **30**, 447-455
41. Jankowsky, E., and Fairman, M. E. (2007) RNA helicases--one fold for many functions. *Curr Opin Struct Biol.* **17**, 316-324
42. Cordin, O., Banroques, J., Tanner, N. K., and Linder, P. (2006) The DEAD-box protein family of RNA helicases. *Gene.* **367**, 17-37

43. Lilley, D. M. (2005) Structure, folding and mechanisms of ribozymes. *Curr Opin Struct Biol.* **15**, 313-323
44. Wu, Q., Huang, L., and Zhang, Y. (2009) The structure and function of catalytic RNAs. *Sci China C Life Sci.* **52**, 232-244
45. Bonetta, L. (2009) RNA-based therapeutics: ready for delivery? *Cell.* **136**, 581-584
46. Sullenger, B. A., and Gilboa, E. (2002) Emerging clinical applications of RNA. *Nature.* **418**, 252-258
47. Famulok, M., and Verma, S. (2002) In vivo-applied functional RNAs as tools in proteomics and genomics research. *Trends Biotechnol.* **20**, 462-466
48. Soukup, G. A., and Breaker, R. R. (1999) Nucleic acid molecular switches. *Trends Biotechnol.* **17**, 469-476
49. Kazantsev, A. V., and Pace, N. R. (2006) Bacterial RNase P: a new view of an ancient enzyme. *Nat Rev Microbiol.* **4**, 729-740
50. Cech, T. R. (1990) Self-splicing of group I introns. *Annu Rev Biochem.* **59**, 543-568
51. Jacquier, A. (1996) Group II introns: elaborate ribozymes. *Biochimie.* **78**, 474-487
52. Fedorova, O., and Zingler, N. (2007) Group II introns: structure, folding and splicing mechanism. *Biol Chem.* **388**, 665-678
53. Toro, N., Jimenez-Zurdo, J. I., and Garcia-Rodriguez, F. M. (2007) Bacterial group II introns: not just splicing. *FEMS Microbiol Rev.* **31**, 342-358

54. Michel, F., and Ferat, J. L. (1995) Structure and activities of group II introns. *Annu Rev Biochem.* **64**, 435-461
55. Keating, K. S., Toor, N., Perlman, P. S., and Pyle, A. M. (2010) A structural analysis of the group II intron active site and implications for the spliceosome. *RNA.* **16**, 1-9
56. Lambowitz, A. M., and Zimmerly, S. (2004) Mobile group II introns. *Annu Rev Genet.* **38**, 1-35
57. Toor, N., Keating, K. S., and Pyle, A. M. (2009) Structural insights into RNA splicing. *Curr Opin Struct Biol.* **19**, 260-266
58. Toor, N., Hausner, G., and Zimmerly, S. (2001) Coevolution of group II intron RNA structures with their intron-encoded reverse transcriptases. *RNA.* **7**, 1142-1152
59. Toor, N., Keating, K. S., Taylor, S. D., and Pyle, A. M. (2008) Crystal structure of a self-spliced group II intron. *Science.* **320**, 77-82
60. Qin, P. Z., and Pyle, A. M. (1998) The architectural organization and mechanistic function of group II intron structural elements. *Curr Opin Struct Biol.* **8**, 301-308
61. Waldsich, C., and Pyle, A. M. (2007) A folding control element for tertiary collapse of a group II intron ribozyme. *Nat Struct Mol Biol.* **14**, 37-44
62. Costa, M., and Michel, F. (1995) Frequent use of the same tertiary motif by self-folding RNAs. *EMBO J.* **14**, 1276-1285

63. Boudvillain, M., and Pyle, A. M. (1998) Defining functional groups, core structural features and inter-domain tertiary contacts essential for group II intron self-splicing: a NAIM analysis. *EMBO J.* **17**, 7091-7104
64. Boudvillain, M., de Lencastre, A., and Pyle, A. M. (2000) A tertiary interaction that links active-site domains to the 5' splice site of a group II intron. *Nature.* **406**, 315-318
65. Costa, M., Deme, E., Jacquier, A., and Michel, F. (1997) Multiple tertiary interactions involving domain II of group II self-splicing introns. *J Mol Biol.* **267**, 520-536
66. Griffin, E. A., Jr., Qin, Z., Michels, W. J., Jr., and Pyle, A. M. (1995) Group II intron ribozymes that cleave DNA and RNA linkages with similar efficiency, and lack contacts with substrate 2'-hydroxyl groups. *Chem Biol.* **2**, 761-770
67. Fedorova, O., Mitros, T., and Pyle, A. M. (2003) Domains 2 and 3 interact to form critical elements of the group II intron active site. *J Mol Biol.* **330**, 197-209
68. Steiner, M., Karunatilaka, K. S., Sigel, R. K., and Rueda, D. (2008) Single-molecule studies of group II intron ribozymes. *Proc Natl Acad Sci U S A.* **105**, 13853-13858
69. Lehmann, K., and Schmidt, U. (2003) Group II introns: structure and catalytic versatility of large natural ribozymes. *Crit Rev Biochem Mol Biol.* **38**, 249-303

70. Hamill, S., and Pyle, A. M. (2006) The receptor for branch-site docking within a group II intron active site. *Mol Cell*. **23**, 831-840
71. Koch, J. L., Boulanger, S. C., Dib-Hajj, S. D., Hebbar, S. K., and Perlman, P. S. (1992) Group II introns deleted for multiple substructures retain self-splicing activity. *Mol Cell Biol*. **12**, 1950-1958
72. Peebles, C. L., Benatan, E. J., Jarrell, K. A., and Perlman, P. S. (1987) Group II intron self-splicing: development of alternative reaction conditions and identification of a predicted intermediate. *Cold Spring Harb Symp Quant Biol*. **52**, 223-232
73. Jarrell, K. A., Peebles, C. L., Dietrich, R. C., Romiti, S. L., and Perlman, P. S. (1988) Group II intron self-splicing. Alternative reaction conditions yield novel products. *J Biol Chem*. **263**, 3432-3439
74. Padgett, R. A., Podar, M., Boulanger, S. C., and Perlman, P. S. (1994) The stereochemical course of group II intron self-splicing. *Science*. **266**, 1685-1688
75. Augustin, S., Muller, M. W., and Schweyen, R. J. (1990) Reverse self-splicing of group II intron RNAs in vitro. *Nature*. **343**, 383-386
76. Morl, M., and Schmelzer, C. (1990) Group II intron RNA-catalyzed recombination of RNA in vitro. *Nucleic Acids Res*. **18**, 6545-6551
77. van der Veen, R., Kwakman, J. H., and Grivell, L. A. (1987) Mutations at the lariat acceptor site allow self-splicing of a group II intron without lariat formation. *EMBO J*. **6**, 3827-3831

78. Dickson, L., Huang, H. R., Liu, L., Matsuura, M., Lambowitz, A. M., and Perlman, P. S. (2001) Retrotransposition of a yeast group II intron occurs by reverse splicing directly into ectopic DNA sites. *Proc Natl Acad Sci U S A*. **98**, 13207-13212
79. Chauhan, S., Behrouzi, R., Rangan, P., and Woodson, S. A. (2009) Structural rearrangements linked to global folding pathways of the *Azoarcus* group I ribozyme. *J Mol Biol*. **386**, 1167-1178
80. Su, L. J., Waldsich, C., and Pyle, A. M. (2005) An obligate intermediate along the slow folding pathway of a group II intron ribozyme. *Nucleic Acids Res*. **33**, 6674-6687
81. Treiber, D. K., Rook, M. S., Zarrinkar, P. P., and Williamson, J. R. (1998) Kinetic intermediates trapped by native interactions in RNA folding. *Science*. **279**, 1943-1946
82. Pan, J., and Woodson, S. A. (1998) Folding intermediates of a self-splicing RNA: mispairing of the catalytic core. *J Mol Biol*. **280**, 597-609
83. Swisher, J. F., Su, L. J., Brenowitz, M., Anderson, V. E., and Pyle, A. M. (2002) Productive folding to the native state by a group II intron ribozyme. *J Mol Biol*. **315**, 297-310
84. Michels, W. J., Jr., and Pyle, A. M. (1995) Conversion of a group II intron into a new multiple-turnover ribozyme that selectively cleaves oligonucleotides: elucidation of reaction mechanism and structure/function relationships. *Biochemistry*. **34**, 2965-2977

85. Swisher, J., Duarte, C. M., Su, L. J., and Pyle, A. M. (2001) Visualizing the solvent-inaccessible core of a group II intron ribozyme. *EMBO J.* **20**, 2051-2061
86. Su, L. J., Brenowitz, M., and Pyle, A. M. (2003) An alternative route for the folding of large RNAs: apparent two-state folding by a group II intron ribozyme. *J Mol Biol.* **334**, 639-652
87. Fedorova, O., Waldsich, C., and Pyle, A. M. (2007) Group II intron folding under near-physiological conditions: collapsing to the near-native state. *J Mol Biol.* **366**, 1099-1114
88. Del Campo, M., Tijerina, P., Bhaskaran, H., Mohr, S., Yang, Q., Jankowsky, E., Russell, R., and Lambowitz, A. M. (2007) Do DEAD-box proteins promote group II intron splicing without unwinding RNA? *Mol Cell.* **28**, 159-166
89. Solem, A., Zingler, N., and Pyle, A. M. (2006) A DEAD protein that activates intron self-splicing without unwinding RNA. *Mol Cell.* **24**, 611-617
90. Huang, H. R., Rowe, C. E., Mohr, S., Jiang, Y., Lambowitz, A. M., and Perlman, P. S. (2005) The splicing of yeast mitochondrial group I and group II introns requires a DEAD-box protein with RNA chaperone function. *Proc Natl Acad Sci U S A.* **102**, 163-168
91. Del Campo, M., and Lambowitz, A. M. (2009) Structure of the Yeast DEAD box protein Mss116p reveals two wedges that crimp RNA. *Mol Cell.* **35**, 598-609



92. Halls, C., Mohr, S., Del Campo, M., Yang, Q., Jankowsky, E., and Lambowitz, A. M. (2007) Involvement of DEAD-box proteins in group I and group II intron splicing. Biochemical characterization of Mss116p, ATP hydrolysis-dependent and -independent mechanisms, and general RNA chaperone activity. *J Mol Biol.* **365**, 835-855
93. Del Campo, M., Mohr, S., Jiang, Y., Jia, H., Jankowsky, E., and Lambowitz, A. M. (2009) Unwinding by local strand separation is critical for the function of DEAD-box proteins as RNA chaperones. *J Mol Biol.* **389**, 674-693
94. Liu, F., Putnam, A., and Jankowsky, E. (2008) ATP hydrolysis is required for DEAD-box protein recycling but not for duplex unwinding. *Proc Natl Acad Sci U S A.* **105**, 20209-20214
95. Jankowsky, E. (2011) RNA helicases at work: binding and rearranging. *Trends Biochem Sci.* **36**, 19-29
96. Mohr, S., Matsuura, M., Perlman, P. S., and Lambowitz, A. M. (2006) A DEAD-box protein alone promotes group II intron splicing and reverse splicing by acting as an RNA chaperone. *Proc Natl Acad Sci U S A.* **103**, 3569-3574
97. Fedorova, O., Solem, A., and Pyle, A. M. (2010) Protein-facilitated folding of group II intron ribozymes. *J Mol Biol.* **397**, 799-813
98. Stark, H., and Luhrmann, R. (2006) Cryo-electron microscopy of spliceosomal components. *Annu Rev Biophys Biomol Struct.* **35**, 435-457
99. Newman, A. (1998) RNA splicing. *Curr Biol.* **8**, R903-905

100. Licatalosi, D. D., and Darnell, R. B. (2006) Splicing regulation in neurologic disease. *Neuron*. **52**, 93-101
101. Kalnina, Z., Zayakin, P., Silina, K., and Line, A. (2005) Alterations of pre-mRNA splicing in cancer. *Genes Chromosomes Cancer*. **42**, 342-357
102. Ritchie, D. B., Schellenberg, M. J., and MacMillan, A. M. (2009) Spliceosome structure: piece by piece. *Biochim Biophys Acta*. **1789**, 624-633
103. Wahl, M. C., Will, C. L., and Luhrmann, R. (2009) The spliceosome: design principles of a dynamic RNP machine. *Cell*. **136**, 701-718
104. Brow, D. A. (2002) Allosteric cascade of spliceosome activation. *Annu Rev Genet*. **36**, 333-360
105. O'Keefe, R. T., Norman, C., and Newman, A. J. (1996) The invariant U5 snRNA loop 1 sequence is dispensable for the first catalytic step of pre-mRNA splicing in yeast. *Cell*. **86**, 679-689
106. Segault, V., Will, C. L., Polycarpou-Schwarz, M., Mattaj, I. W., Branlant, C., and Luhrmann, R. (1999) Conserved loop I of U5 small nuclear RNA is dispensable for both catalytic steps of pre-mRNA splicing in HeLa nuclear extracts. *Mol Cell Biol*. **19**, 2782-2790
107. Valadkhan, S., and Manley, J. L. (2003) Characterization of the catalytic activity of U2 and U6 snRNAs. *RNA*. **9**, 892-904

108. Sashital, D. G., Cornilescu, G., McManus, C. J., Brow, D. A., and Butcher, S. E. (2004) U2-U6 RNA folding reveals a group II intron-like domain and a four-helix junction. *Nat Struct Mol Biol.* **11**, 1237-1242
109. Gordon, P. M., Sontheimer, E. J., and Piccirilli, J. A. (2000) Metal ion catalysis during the exon-ligation step of nuclear pre-mRNA splicing: extending the parallels between the spliceosome and group II introns. *RNA.* **6**, 199-205
110. Sontheimer, E. J., Gordon, P. M., and Piccirilli, J. A. (1999) Metal ion catalysis during group II intron self-splicing: parallels with the spliceosome. *Genes Dev.* **13**, 1729-1741
111. Parker, R., Siliciano, P. G., and Guthrie, C. (1987) Recognition of the TACTAAC box during mRNA splicing in yeast involves base pairing to the U2-like snRNA. *Cell.* **49**, 229-239
112. Lesser, C. F., and Guthrie, C. (1993) Mutations in U6 snRNA that alter splice site specificity: implications for the active site. *Science.* **262**, 1982-1988
113. Madhani, H. D., and Guthrie, C. (1992) A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. *Cell.* **71**, 803-817
114. Hilliker, A. K., and Staley, J. P. (2004) Multiple functions for the invariant AGC triad of U6 snRNA. *RNA.* **10**, 921-928

115. Mefford, M. A., and Staley, J. P. (2009) Evidence that U2/U6 helix I promotes both catalytic steps of pre-mRNA splicing and rearranges in between these steps. *RNA*. **15**, 1386-1397
116. Wolff, T., and Bindereif, A. (1993) Conformational changes of U6 RNA during the spliceosome cycle: an intramolecular helix is essential both for initiating the U4-U6 interaction and for the first step of slicing. *Genes Dev.* **7**, 1377-1389
117. Sun, J. S., and Manley, J. L. (1995) A novel U2-U6 snRNA structure is necessary for mammalian mRNA splicing. *Genes Dev.* **9**, 843-854
118. Sashital, D. G., Venditti, V., Angers, C. G., Cornilescu, G., and Butcher, S. E. (2007) Structure and thermodynamics of a conserved U2 snRNA domain from yeast and human. *RNA*. **13**, 328-338
119. Yu, Y. T., Shu, M. D., and Steitz, J. A. (1998) Modifications of U2 snRNA are required for snRNP assembly and pre-mRNA splicing. *EMBO J.* **17**, 5783-5795
120. Ray, S., Mehta, G., and Srivastava, S. (2010) Label-free detection techniques for protein microarrays: prospects, merits and challenges. *Proteomics*. **10**, 731-748
121. Kessler, C. (1994) Non-radioactive analysis of biomolecules. *J Biotechnol.* **35**, 165-189
122. Stryer, L. (1978) Fluorescence energy transfer as a spectroscopic ruler. *Annu Rev Biochem.* **47**, 819-846

123. Selvin, P. R. (1995) Fluorescence resonance energy transfer. *Methods Enzymol.* **246**, 300-334
124. Rueda, D., and Walter, N. G. (2005) Single molecule fluorescence control for nanotechnology. *J Nanosci Nanotechnol.* **5**, 1990-2000
125. Deniz, A. A., Dahan, M., Grunwell, J. R., Ha, T., Faulhaber, A. E., Chemla, D. S., Weiss, S., and Schultz, P. G. (1999) Single-pair fluorescence resonance energy transfer on freely diffusing molecules: observation of Forster distance dependence and subpopulations. *Proc Natl Acad Sci U S A.* **96**, 3670-3675
126. Lu, H. P., Xun, L., and Xie, X. S. (1998) Single-molecule enzymatic dynamics. *Science.* **282**, 1877-1882
127. Ha, T., Ting, A. Y., Liang, J., Caldwell, W. B., Deniz, A. A., Chemla, D. S., Schultz, P. G., and Weiss, S. (1999) Single-molecule fluorescence spectroscopy of enzyme conformational dynamics and cleavage mechanism. *Proc Natl Acad Sci U S A.* **96**, 893-898
128. Zhao, R., and Rueda, D. (2009) RNA folding dynamics by single-molecule fluorescence resonance energy transfer. *Methods.* **49**, 112-117
129. Zhuang, X. (2005) Single-molecule RNA science. *Annu Rev Biophys Biomol Struct.* **34**, 399-414
130. Gradinaru, C. C., Marushchak, D. O., Samim, M., and Krull, U. J. (2010) Fluorescence anisotropy: from single molecules to live cells. *Analyst.* **135**, 452-459

131. Lundblad, J. R., Laurance, M., and Goodman, R. H. (1996) Fluorescence polarization analysis of protein-DNA and protein-protein interactions. *Mol Endocrinol.* **10**, 607-612
132. Shuman, S., Spencer, E., Furneaux, H., and Hurwitz, J. (1980) The role of ATP in in vitro vaccinia virus RNA synthesis effects of AMP-PNP and ATP gamma S. *J Biol Chem.* **255**, 5396-5403
133. McKinney, S. A., Joo, C., and Ha, T. (2006) Analysis of single-molecule FRET trajectories using hidden Markov modeling. *Biophys J.* **91**, 1941-1951
134. Guo, Z., Karunatilaka, K. S., and Rueda, D. (2009) Single-molecule analysis of protein-free U2-U6 snRNAs. *Nat Struct Mol Biol.* **16**, 1154-1159
135. Valadkhan, S., and Manley, J. L. (2000) A tertiary interaction detected in a human U2-U6 snRNA complex assembled in vitro resembles a genetically proven interaction in yeast. *RNA.* **6**, 206-219
136. Efron, B., Gong, G. (1983) A Leisurely Look at the Bootstrap, the Jackknife, and Cross-Validation. *The American Statistician.* **37**, 36-48
137. Newby Lambert, M., Vocker, E., Blumberg, S., Redemann, S., Gajraj, A., Meiners, J. C., and Walter, N. G. (2006) Mg<sup>2+</sup>-induced compaction of single RNA molecules monitored by tethered particle microscopy. *Biophys J.* **90**, 3672-3685
138. Rueda, D., Hsieh, J., Day-Storms, J. J., Fierke, C. A., and Walter, N. G. (2005) The 5' leader of precursor tRNA<sup>Asp</sup> bound to the *Bacillus subtilis*

- RNase P holoenzyme has an extended conformation. *Biochemistry*. **44**, 16130-16139
139. Waldsich, C., and Pyle, A. M. (2008) A kinetic intermediate that regulates proper folding of a group II intron RNA. *J Mol Biol*. **375**, 572-580
140. Noah, J. W., and Lambowitz, A. M. (2003) Effects of maturase binding and Mg<sup>2+</sup> concentration on group II intron RNA folding investigated by UV cross-linking. *Biochemistry*. **42**, 12466-12480
141. Zhang, A., Derbyshire, V., Salvo, J. L., and Belfort, M. (1995) Escherichia coli protein StpA stimulates self-splicing by promoting RNA assembly in vitro. *RNA*. **1**, 783-793
142. Clodi, E., Semrad, K., and Schroeder, R. (1999) Assaying RNA chaperone activity in vivo using a novel RNA folding trap. *EMBO J*. **18**, 3776-3782
143. Moore, M. J., and Sharp, P. A. (1992) Site-specific modification of pre-mRNA: the 2'-hydroxyl groups at the splice sites. *Science*. **256**, 992-997
144. Smith, G. J., Sosnick, T. R., Scherer, N. F., and Pan, T. (2005) Efficient fluorescence labeling of a large RNA through oligonucleotide hybridization. *RNA*. **11**, 234-239
145. Steiner, M., Rueda, D., and Sigel, R. K. (2009) Impact of Ca<sup>2+</sup> on Single Molecule Folding of a Group II Intron Ribozyme. *Angewandte Chem. Int. Ed*. **48**, 9739-9742

146. Ha, T., Rasnik, I., Cheng, W., Babcock, H. P., Gauss, G. H., Lohman, T. M., and Chu, S. (2002) Initiation and re-initiation of DNA unwinding by the *Escherichia coli* Rep helicase. *Nature*. **419**, 638-641
147. Christian, T. D., Romano, L. J., and Rueda, D. (2009) Single Molecule Measurements of Synthesis by DNA Polymerase with Base-Pair Resolution. *Proc Natl Acad Sci U S A*. **106**, 21109-21114
148. Levin, J. G., Guo, J., Rouzina, I., and Musier-Forsyth, K. (2005) Nucleic acid chaperone activity of HIV-1 nucleocapsid protein: critical role in reverse transcription and molecular mechanism. *Prog Nucleic Acid Res Mol Biol*. **80**, 217-286
149. Lamichhane, R., Daubner, G. M., Thomas-Crusells, J., Auweter, S. D., Manatchal, C., Austin, K. S., Valniuk, O., Allain, F., and Rueda, D. (2010) RNA looping by PTB: evidence using FRET and NMR spectroscopy and for a role in splicing repression. *Proc Natl Acad Sci U S A*. **107**, 4105-4110
150. Hopkins, J. F., Panja, S., McNeil, S. A., and Woodson, S. A. (2009) Effect of salt and RNA structure on annealing and strand displacement by Hfq. *Nucleic Acids Res*. **37**, 6205-6213
151. Pyle, A. M., Fedorova, O., and Waldsich, C. (2007) Folding of group II introns: a model system for large, multidomain RNAs? *Trends Biochem Sci*. **32**, 138-145



152. Bifano, A. L., Turk, E. M., and Caprara, M. G. (2010) Structure-guided mutational analysis of a yeast DEAD-box protein involved in mitochondrial RNA splicing. *J Mol Biol.* **398**, 429-443
153. Liebeg, A., Mayer, O., and Waldsich, C. (2010) DEAD-box protein facilitated RNA folding in vivo. *RNA Biol.* **7**, 103-111
154. Zingler, N., Solem, A., and Pyle, A. M. (2010) Dual roles for the Mss116 cofactor during splicing of the ai5 $\gamma$  group II intron. *Nucleic Acids Res.* **38**, 6602-6609
155. Bifano, A. L., Turk, E. M., and Caprara, M. G. Structure-guided mutational analysis of a yeast DEAD-box protein involved in mitochondrial RNA splicing. *J Mol Biol.* **398**, 429-443
156. Donmez, G., Hartmuth, K., and Luhrmann, R. (2004) Modified nucleotides at the 5' end of human U2 snRNA are required for spliceosomal E-complex formation. *RNA.* **10**, 1925-1933
157. Gopinath, S. C. (2009) Mapping of RNA-protein interactions. *Anal Chim Acta.* **636**, 117-128
158. Marks, K. M., and Nolan, G. P. (2006) Chemical labeling strategies for cell biology. *Nat Methods.* **3**, 591-596
159. Christian, T. D., Romano, L. J., and Rueda, D. (2009) Single-molecule measurements of synthesis by DNA polymerase with base-pair resolution. *Proc Natl Acad Sci U S A.* **106**, 21109-21114

160. Yan, Y., and Marriott, G. (2003) Analysis of protein interactions using fluorescence technologies. *Curr Opin Chem Biol.* **7**, 635-640
161. Ellington, A. D., and Szostak, J. W. (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature.* **346**, 818-822
162. Held, D. M., Kissel, J. D., Patterson, J. T., Nickens, D. G., and Burke, D. H. (2006) HIV-1 inactivation by nucleic acid aptamers. *Front Biosci.* **11**, 89-112
163. Baugh, C., Grate, D., and Wilson, C. (2000) 2.8 Å crystal structure of the malachite green aptamer. *J Mol Biol.* **301**, 117-128
164. Mairal, T., Ozalp, V. C., Lozano Sanchez, P., Mir, M., Katakis, I., and O'Sullivan, C. K. (2008) Aptamers: molecular tools for analytical applications. *Anal Bioanal Chem.* **390**, 989-1007
165. Tombelli, S., Minunni, M., and Mascini, M. (2005) Analytical applications of aptamers. *Biosensors & bioelectronics.* **20**, 2424-2434
166. Tuerk, C., and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science.* **249**, 505-510
167. Mendonsa, S. D., and Bowser, M. T. (2004) In vitro selection of high-affinity DNA ligands for human IgE using capillary electrophoresis. *Anal Chem.* **76**, 5387-5392
168. Jenison, R. D., Gill, S. C., Pardi, A., and Polisky, B. (1994) High-resolution molecular discrimination by RNA. *Science.* **263**, 1425-1429

169. Clark, S. L., and Remcho, V. T. (2002) Aptamers as analytical reagents. *Electrophoresis*. **23**, 1335-1340
170. Romig, T. S., Bell, C., and Drolet, D. W. (1999) Aptamer affinity chromatography: combinatorial chemistry applied to protein purification. *Journal of chromatography*. **731**, 275-284
171. Deng, Q., Watson, C. J., and Kennedy, R. T. (2003) Aptamer affinity chromatography for rapid assay of adenosine in microdialysis samples collected in vivo. *Journal of chromatography*. **1005**, 123-130
172. Kunzelmann, S., and Webb, M. R. (2010) A fluorescent, reagentless biosensor for ADP based on tetramethylrhodamine-labeled ParM. *ACS Chem Biol*. **5**, 415-425
173. Wayment, J. R., and Harris, J. M. (2009) Biotin-avidin binding kinetics measured by single-molecule imaging. *Anal Chem*. **81**, 336-342
174. Holeman, L. A., Robinson, S. L., Szostak, J. W., and Wilson, C. (1998) Isolation and characterization of fluorophore-binding RNA aptamers. *Fold Des*. **3**, 423-431

**ABSTRACT****CHARACTERIZATION OF SPLICING MECHANISMS BY SINGLE-MOLECULE  
FLUORESCENCE**

by

KRISHANTHI SANJEEWANI KARUNATILAKA

May 2011

**Advisor:** Prof. David Rueda**Major:** Chemistry (Biochemistry)**Degree:** Doctor of Philosophy

Group II introns rank amongst the largest self-splicing ribozymes found in bacteria and organellar genomes of various eukaryotes. Interestingly, these ribozymes can self-splice via two transesterification reactions that resemble nuclear pre-mRNA splicing in eukaryotes (54,55). Despite the diversity in primary sequences, group II introns possess highly conserved secondary structures consisting of six domains (D1-D6). To perform its function, the large multidomain group II intron RNA must adopt the correctly folded structure (52). As a result, *in vitro* splicing of these introns requires high ionic strength and elevated temperatures. *In vivo*, this process is mainly assisted by protein cofactors (37). However, the exact mechanism of protein-mediated splicing of group II intron RNA is still not known.

In order to elucidate the mechanism of protein-mediated splicing of group II introns, we have studied the folding dynamics of the D135 ribozyme, a minimal active form of the yeast *ai5γ* group II intron, in the presence of its natural cofactor, the DEAD-box protein Mss116, using single-molecule fluorescence. Consistent with folding studies at very high magnesium concentrations, our single-molecule data show that Mss116 can promote the folding of group II introns under near physiological conditions *in vitro*. Furthermore, smFRET data indicate that the Mss116-mediated group II intron folding pathway is a multi-step process that consists of both ATP-independent and ATP-dependent steps (37).

Structurally and mechanistically group II introns are similar to spliceosome-catalyzed pre-mRNA splicing. Out of five snRNAs, only the highly conserved U2 and U6 snRNAs are required in both steps of RNA splicing. The U2-U6 snRNA complex forms the active site of the spliceosome and has been shown to undergo splicing-related catalysis in the absence of proteins (11,13,107). Single-molecule studies of yeast U2-U6 snRNAs show a  $Mg^{2+}$  induced conformational change, which may be involved in spliceosomal activation *in vivo* (134). In contrast to yeast, human U2 and U6 snRNAs contain a large number of post-transcriptional modifications. Recent studies have shown these modifications make human snRNAs more stable than that of yeast indicating a possibility of having different spliceosomal activation states.

In order to understand and compare the catalytic mechanisms, we used single-molecule fluorescence to characterize the conformational changes of human

U2-U6 complex in the presence and absence of modifications using  $Mg^{2+}$  as a divalent metal ion. Our FRET data clearly show a  $Mg^{2+}$  induced conformational change with three FRET states. Based on smFRET data, we propose a minimal two-step folding pathway for human snRNAs similar to yeast. Although unmodified snRNAs exhibit similar folding dynamics as yeast (134), modified bases destabilize the low FRET state of the U2-U6 complex. However, comparison of FRET and UV melting data (118) suggests modified bases may be involved in protein recognition and/or early assembly of the spliceosome rather than direct stabilization of RNA structures *in vivo*.

## AUTOBIOGRAPHICAL STATEMENT

**Name:** Krishanthi Sanjeevani Karunatilaka

**Education: B.Sc.** Biochemistry and Molecular Biology,  
University of Colombo, Colombo, Sri Lanka

**Ph.D.** Chemistry - Major in Biochemistry  
Wayne State University, Detroit, Michigan, USA

### Publications:

1. Karunatilaka, K. S., Solem, A., Pyle, A. M., Rueda, D. Single-molecule analysis of Mss116-mediated group II intron folding. *Nature* **467**, 935-939 (2010).
2. Guo, Z., Karunatilaka, K. S., Rueda, D. Single-molecule analysis of protein-free U2-U6 snRNAs. *Nature Struct. & Mol. Biol.* **16**, 1154-1159 (2009).
3. Karunatilaka, K. S., Rueda, D. Single-molecule fluorescence studies of RNA: A decade's progress. *Chem. Phy. Let.* **476**, 1-10 (2009).
4. Steiner, M., Karunatilaka, K. S., Sigel, R. K. O., Rueda, D. Single-molecule studies of group II intron ribozymes. *PNAS* **105**, 13853-13858 (2008).
5. Karunatilaka, K. S., Baker, C., Rueda, D. An RNA aptamer-based purification system for fluorophore-labeled proteins (in preparation).
6. Karunatilaka, K. S., Rueda, D. Single-molecule analysis of human U2-U6 snRNAs (in preparation).

### Presentations:

#### Oral

1. *Single-molecule Analysis of Mss116-Mediated Group II Intron Folding*. The Annual Meeting of the RNA Society, Seattle, Washington (June, 2010).
2. *Single-molecule Analysis of Mss116-Mediated Group II Intron Folding*. Michigan RNA Meeting, East Lansing, Michigan (March, 2010).

#### Poster

1. *Single-molecule Analysis of Mss116-Mediated Group II Intron Folding*. Biophysical Society Meeting, San Francisco, CA (February, 2010).
2. *Single-molecule Studies of Human Spliceosomal snRNAs*. The Annual Meeting of the RNA Society, Madison, Wisconsin (May, 2009).
3. *Development of an RNA Aptamer Based Purification Method for Fluorophore Labeled Proteins*. The Annual Meeting of the RNA Society, Madison, Wisconsin (May, 2007).