WATCHING DYNAMICS AND ASSEMBLY OF SPLICEOSOMAL COMPLEXES AT SINGLE MOLECULE RESOLUTION

A thesis submitted for the degree of

DOCTOR OF PHILOSOPHY

2015

CHANDANI MANOJA WARNASOORIYA

Section of Virology, Department of Medicine

Imperial College London

Copyright Declaration

Declaration of Origin

I hereby declare that this project was entirely my own work (except the gel images in figures 4.2-4.5 in the chapter 4) and that any additional sources of information have been properly cited.

I hereby declare that any internet sources, published and unpublished works from which I have quoted or drawn reference have been referenced fully in the text and in the reference list. I understand that failure to do so will result in failure of this project due to plagiarism.

Chapter 3 was in collaboration with Prof. Samuel Butcher and Prof. David Brow at University of Wisconsin-Madison, Wisconsin, USA. The Prp24 full length protein and the truncated protein (234C) were provided by them.

Chapter 4 was in collaboration with Prof. Kiyoshi Nagai and his lab members Dr. John Hardin and Yasushi Kondo at LMB Cambridge University, Cambridge, UK. All the Electrophoretic mobility shift assays for the full length U4/U6 complex and the calculations of binding affinities were done by Dr. John Hardin and Yasushi Kondo at LMB Cambridge University (Figures 4.2-4.5 in the chapter 4) and these figures were obtained from Dr. John Hardin and Yasushi Kondo with permission.

DEDICATION

To My Loving Amma because you sacrificed all your wellbeing
To give me a better life

To my loving Thattha because you sacrificed all your happiness
To make me happy

To my loving husband because you sacrificed all your dreams
To bring me all this long

To my loving daughter because you sacrificed all your cuddling times
To let me fulfil my task

ACKNOWLEDGEMENTS

First of all I want to convey my sincere gratitude to my advisor, Dr David Rueda, for giving me the opportunity to work in his lab, which opens the path to learn an emerging technique and his guidance, encouragement and support throughout my research work. I'm fortunate to be a part of his lab and able to do promising research. Thank you very much Dr. Rueda for all the help and opportunities you have given me, and for the diligent effort you made to train me as a good scientist. Also I would like to thank him for giving me the opportunity to join and obtain my PhD from Imperial College London.

I would also like to thank my thesis committee in Department of Chemistry, Wayne State University, USA, Prof. Louis Romano and Prof. Arthur Suits, for their valuable suggestions and feedback on my research during the first three years of my PhD. My special thanks go to Prof. Christine Chow for her cooperation, guidance and support through my difficult times. I express my gratitude to my research progress panel from section of Virology, Imperial College London, Prof. Wendy Barclay and Dr. Goedele Maertens for their valuable suggestions and scientific discussion during my early stage and late stage assessments. I would also like to thank Prof. Ramon Vilar and Dr. Ray O'Keefe for agreeing to be in my PhD viva committee.

I am also thankful to our collaborators Prof. Samuel Butcher, Prof. David Brow and Ashley Richie from University of Wisconsin-Madison for sending me the Prp24 protein and important discussions and suggestions. Also Prof. Kiyoshi Nagai, Dr. John Hardin and Yasushi Kondo from LMB, Cambridge, UK for providing U4/U6 proteins, carried out the initial experiments and valuable discussions and suggestions.

Then, I would like to acknowledge all my lab members. My biggest thank goes to Zhuojun Guo for teaching me single molecule experiments and everything not only about research but also about life. Thank you so much Zhuojun for everything you taught me and for the fun times we had in the lab. I also want to give a big thanks to Dr. Amanda Solem for her continuous support and valuable suggestions and Dr. Elvin Aleman and Dr. Alfonso Brenlla for the technical support in single molecule setup. I also want to thank my former lab mates Rajan, Krishanthi, Rui, Sharla, May Eric, Pramodha, Gayan and Marcus for their immense guidance and support. I would also want to thank my fellow lab mates from Wayne State University, Bishnu, Hansini and Imali, for the valuable discussions we had and making the stressful graduate school life an endurable one. I will cherish all those lovely times that we spent together. I also like to thanks Chow lab members from Wayne State University for their help for my research, especially Gayani and Daya for helping me with MALDI and Xun for encouraging me and cheering me up in my difficult times. I would also like to thank my current lab members in Imperial College Sheila, Hailey, Kathy, Amit, Adam, Kotryna, Maria and Martin to all your help with my research work as well as with my personal life. Thank you very much Sheila, for being a wonderful friend to me and thank you for being there for me whenever I needed someone to talk in hard times. A special thank goes to Sheila, Hailey, Kathy and Adam for spending lot of time on reading and correcting my thesis, I really appreciate your help on that. Thank you all for making an enjoyable working environment.

I want to convey my sincere appreciation to Wayne State University Chemistry department administrative staff specially Melissa Nestor, Debbie, Diane and all the

others. I would also like to thanks people in Imperial College who helped me in many ways. I would like to convey my sincere appreciation to everyone in the 10th floor, Commonwealth building; Guido lab members, Paul Coote, Cecilia George for all the help you gave. Also I would like to thank Hayley Kendall and Brett Onslow for helping me with the administrative work.

Then my heartiest gratitude goes to the Wayne state - Sri Lankan student Association, Michigan Sri Lankan community and all my Sri Lankan friends, who help me to make a second home in USA. I'm thanking to all my best friends in Sri Lanka who were there for me to cheer me up and help me. Also I thank all my Sri-Lankan friends in London for making life in London much endurable one. Thank you being like my own family and helping me during my pregnancy and taking care of my baby during past year. Without you all it will not be easy for me to continue with my PhD while having baby.

Finally, I would like to convey my heartfelt appreciation to my Parents; Upali Warnasooriya and Chandra Warnasooriya for bringing me up as a better human and for everything you did throughout my life. Also I thank to my brother Vindana and sister-in-law Pradeepa for encouraging me and also for taking care of our parents while I'm far away from them. Your love means a lot to me, especially to bear this stressful life in a foreign country. Then I want to express my utmost gratitude to guardian of my life, my loving husband Chinthaka Sanjeewa for your enormous support, encouragement and love which means so much for me. Since the day we met, you were there by my side for last 10 years, helping me in every way to fulfil my dreams and come this far in my career. Without your love, protection and guidance, I can't imagine how I would have

come this far. Thank you very much for being with me in every step in my life and I'm sorry for all the hardship you have to face while you are helping me to accomplish my dreams. Last but not least I would like to thank my darling daughter, Chanudi who brought all the happiness to our lives. I'm really sorry that I couldn't spend more time with you during your first year, the period you really need to be with your mother. Thank you for enduring the times that I'm away from you, working on the thesis all night without sleeping with you. I'm really blessed to have a lovely daughter as you, who never gave me a hard time and behave so nicely like you understand my situation. I love you so much my loving daughter.

All these six years, I wouldn't be able to focus on and successfully complete my PhD without all the help and support I had from all of you. Thank you all loving people !!!

TABLE OF CONTENT

TABLE OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

DNA	Deoxyribonucleic acid
RNA	Ribonucleic acid
Pre-mRNA	Precursor messenger RNA
D1-D6	Domain 1-6
EBS	Exon-binding sites
IBS	Intron-binding sites
ORF	Open reading frame
SS	Splice site
BS	Branch site
snRNP	Small nuclear ribonucleoproteins
snRNA	Small nuclear RNA
U snRNA	Uridine-rich snRNA
SL	Stem-loop
SF	Splicing factor
Prp	Pre-mRNA processing proteins
RRM	RNA recognition motif
TALS	Taybi-Linder syndrome
ISL	Intramolecular stem loop
T_m	Melting temperature
smFRET	Single-molecule fluorescence resonance energy transfer
EMSA	Electrophoresis mobility shift assay

Cryo-EM	Cryo-electron microscopy
AFM	Atomic force microscopy
CoSMoS	Colocalization single-molecule spectroscopy
Sm-FLIM	Single-molecule fluorescence life time imaging
TCSPC	Time-correlated single photon counting
FRAP	Fluorescence recovery after photobleaching
FIONA	Fluorescence imaging with one-nanometer accuracy
TIRF	Total internal reflection fluorescence microscopy
CCD	Charge-coupled device
NHS	N-hydroxysuccinimide
PEG	Polyethylene glycol
HPLC	High performance liquid chromatography
OSS	Oxygen scavenging system
MALDI-MS	Matrix assisted laser desorption ionization-mass spectroscopy

ABSTRACT

Splicing plays a major role in eukaryotic gene expression by processing pre-mRNA to form mature mRNA. Splicing is catalysed by the spliceosome; a ribonucleoprotein complex consisting of five small nuclear ribonucleoproteins (snRNPs) and over 100 proteins. Proper assembly of spliceosomal components is critical for its function, and thus assembly defects can be lethal. Several diseases have been associated with splicing defects, such as cancer (breast cancer, leukaemia), cystic fibrosis, duchenne muscular dystrophy (DMD), retinitis pigmentosa (RP) and taybi-linder syndrome (TALS). Studying the structural dynamics and distinct functions of snRNA complexes and the factors that affect the stability of those complexes provides an overall idea regarding the structure and function of the spliceosome, which can guide us to discover novel therapeutics for splicing related diseases. Hence, the aim of this study is employing smFRET technique to monitor the structural dynamics and assembly of snRNA complexes and the effect of protein factors on those dynamics with single molecule resolution. Three specific aims have been addressed in this thesis work to achieve the main goal.

The first part of this study is focused on understanding how spliceosomal components are recycled. This work shows that binding of Prp24; U6 snRNP specific protein unwinds U2 from U2/U6 complex and stabilizes U6 at a low FRET conformation, suggesting a novel role for Prp24 as a recycling factor. The second part of this study is focused on understanding the assembly of sub-spliceosomal complexes and their global structure. This work shows that although the binding of individual proteins slightly

changes the conformation of U4/U6 duplex; overall it maintains a rigid structure. This suggests that the U4/U6 adopts a preformed conformation and act as a scaffold for protein binding, while preventing U6 from premature activation. The third part of this study is focused on understanding structural and functional similarities between minor and major spliceosomal complexes. This work shows that minor spliceosomal U12/U6atac complex adopts a conformation similar to the three-helix junction structure of major spliceosomal U2/U6.

CHAPTER 1: Introduction

1.1: Gene expression

According to the central dogma of molecular biology, in living cells, genetic information stored in DNA first transfers to messenger RNA (mRNA) by transcription and then translates into functional proteins (Figure 1.1)[1]. However, in eukaryotes, before transport out of the nucleus and protein synthesis, the transcript needs to undergo at least three major processing steps, adding a 5' cap structure, adding a poly A tail to the 3'-end and splicing (Figure 1.1)[1,2].

1.2: Pre-mRNA splicing

The majority of the eukaryotic protein coding genes (exons) are disrupted by some sequences named introns[3,4]. During transcription, both introns and exons are transcribed into a single transcript, known as precursor mRNA (pre-mRNA)[4,5]. Before translation, the transcript must be processed to remove these introns and join the protein coding sequences together (Figure1.2a)[4,6]. The removal of introns and ligation of exons is known as splicing and the processed transcript is called a mature mRNA[7,8]. This processed mature mRNA will then be transported to the cytoplasm and be recognized by the ribosome complex in order to translate the mRNA into an amino acid sequence which eventually folds into a functional protein[1,8,9].

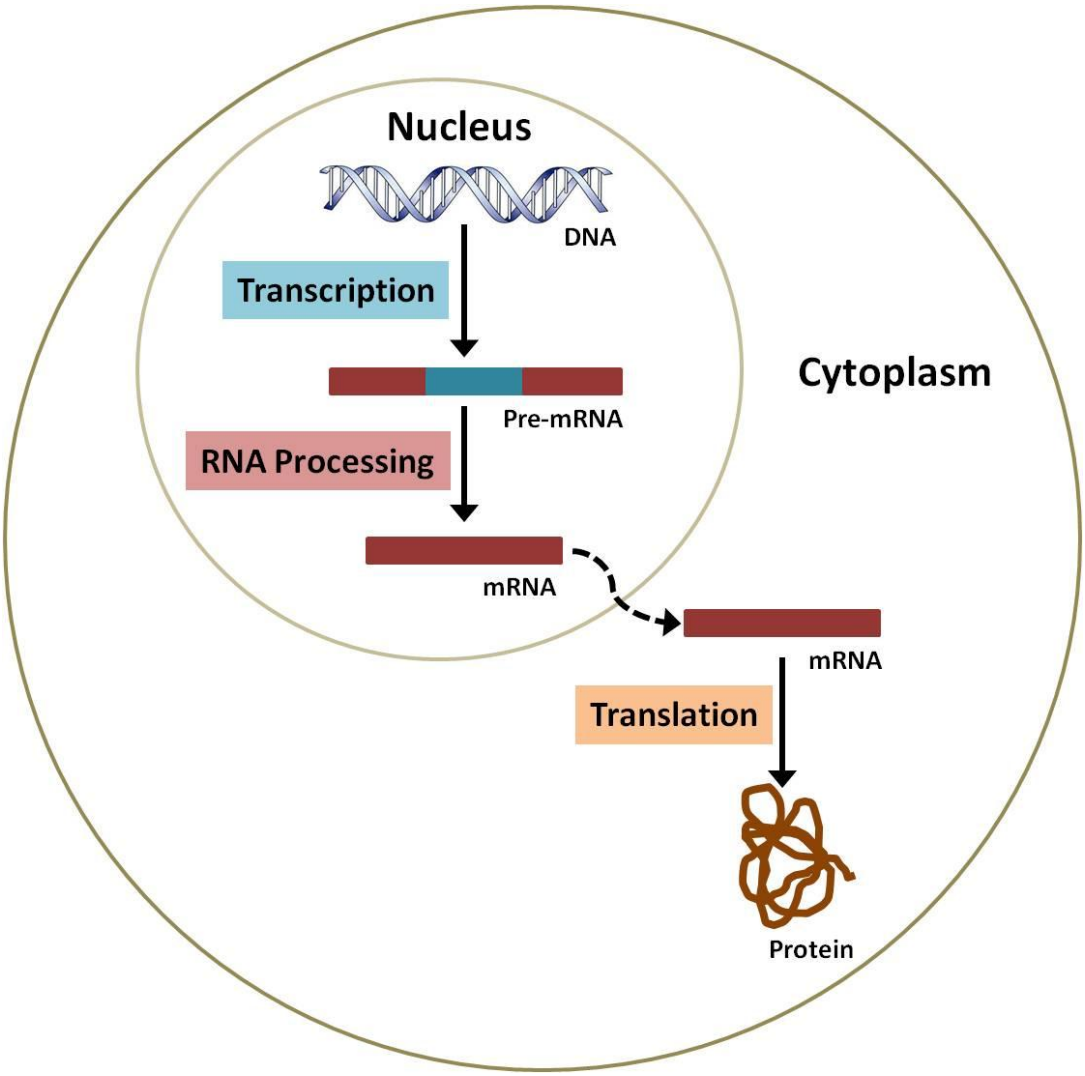

Figure 1.1: The central dogma of molecular biology. Flow of genetic materials from DNA to protein is explained by the central dogma of molecular biology. First the genetic information stored in DNA is transferred into pre-mRNA by transcription followed by the RNA processing which transform pre-mRNA into mature mRNA. These two processes occur in the nucleus. Processed mRNA is then transported into the cytoplasm where mRNA coded for the formation of functional proteins by a process named as translation.

1.2.1: Alternative splicing

Splicing plays a major role in eukaryotic gene expression[1,5,10], and it is an essential checkpoint for mRNA maintenance, which helps cells to regulate gene expression[11]. Most exons undergo constitutive splicing; in which all the exons are always included in mature mRNA[1]. In contrast, some eukaryotic genes carryout alternative splicing, which produces multiple mature mRNAs from a single transcript by shuffling different combinations of introns and exons (Figure 1.2b)[1,8]. In some cases exons can be skipped and sometimes those can be lengthened or shortened by altering the position of the splice sites[12]. In this way, some transcripts can contain multiple to thousands splicing patterns, which results in various protein isoforms from a single transcript[7,8,12]. It has been estimated that transcripts from >95% of human transcripts undergo alternative splicing[6,13,14]. Hence, the alternative splicing play an important role in protein diversity and gene regulation in cells[12,13]. This is an evolutionary advancement in higher eukaryotes to overcome the complexity of an organism that requires vast number of genes to produce the necessary proteins[15].

1.3: Discovery and evolution of introns

One of the greatest findings about eukaryotic cells is that the eukaryotic genes consist of introns, which are derived from the term “intragenic regions”. This was discovered independently by Phillip A. Sharp and Richard J. Roberts in 1977, which lead them to receive the Nobel Prize in physiology or medicine in 1993. The number of introns present in organisms is highly variable; ranging from few per genome in some species to hundreds of thousands per genome in vertebrates and plants[15,16]. As an example the

dystrophin gene forms a >2 million nucleotides long pre-mRNA, which is then processed in to a 14000 nucleotides long mature mRNA, by removing 78 introns[17]. When and why introns have evolved is not clear yet but there are two main hypotheses to explain this; 'Introns Early (IE)' and 'Introns Late (IL)' which are still on debate[15,18,19]. According to the 'IE' concept, introns are very old and present in all early organisms. With evolution, some organisms have lost their introns resulting in fewer or no introns[3,15,18]. In contrast the 'IL' concept suggests that introns are inserted into eukaryotic genes at a later time point during the evolution[3,15,18]. At present no direct evidence has been found to favour one hypothesis over the other and therefore it can only be speculated that some introns have been present from a very early stage of life, where as some of them added to the genes after the divergence of eukaryotes[20]. The enormous variation of intron number and the position of introns among eukaryotic genes and also within different species, suggest that introns have either been deleted or gained or both throughout the evolution[15,19].

1.4: Different types of Introns

There are four major classes of introns that have been discovered; group I, group II, tRNA and nuclear introns (or spliceosomal introns)[19-22]. Among these four, tRNA introns are short introns which are spliced through a different mechanism than the others[22]. Therefore these introns are more accurately categorized as a part of tRNA processing as they are less related to other groups. Group I and II introns exhibit self-splicing, where they catalyze their own excision[23]. Moreover, group II and nuclear introns share a similar reaction mechanism[23].

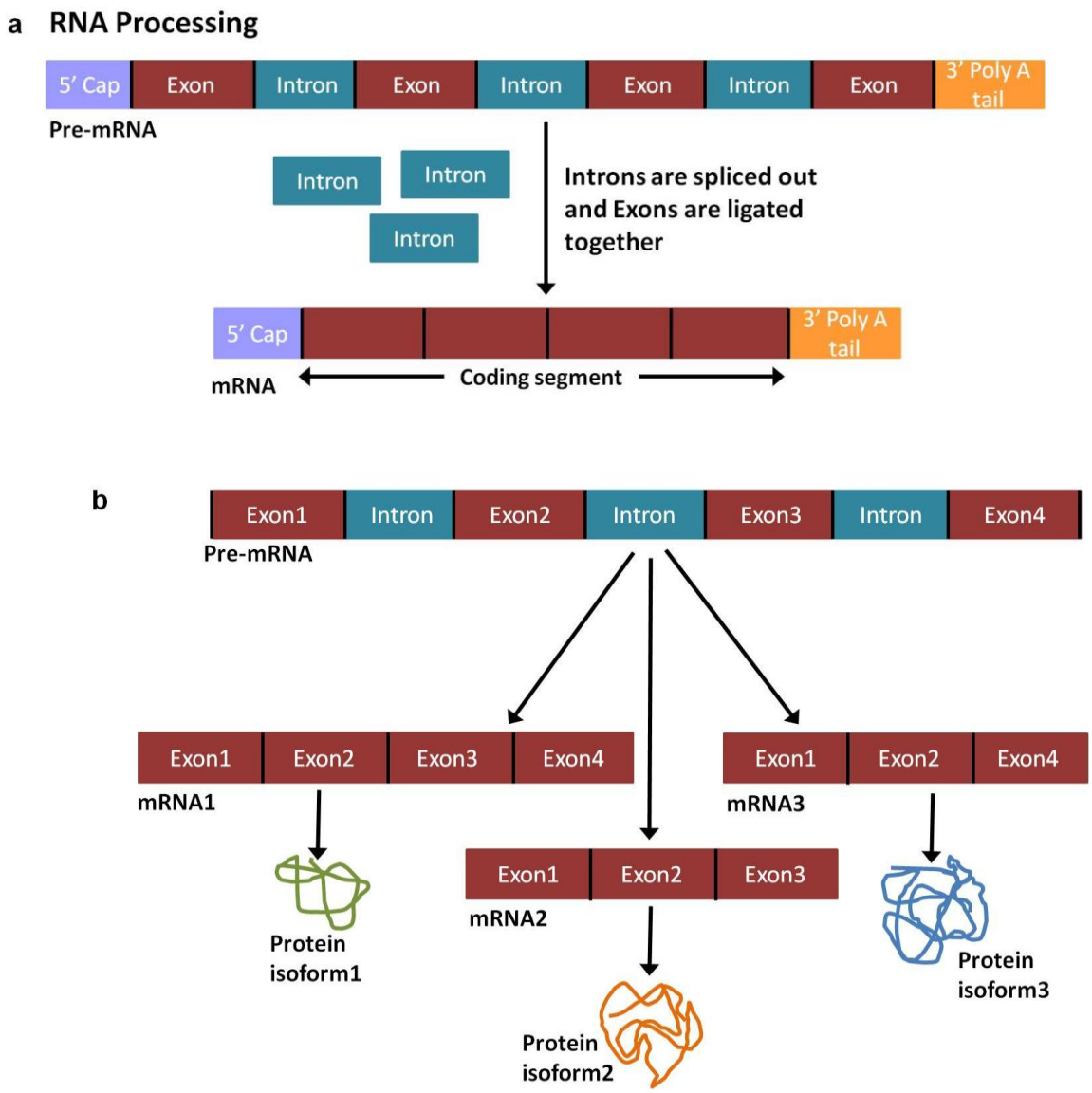

Figure 1.2: RNA processing and splicing. Pre-mRNA consists of protein coding sequences known as exons and non protein coding sequences known as introns. Before translation, the Pre-mRNA is needed to be processed. (a) During RNA processing, pre-mRNA undergoes splicing where introns are excised and exons are ligated together to form the mature-mRNA, a 5' cap structure and a poly A tail are added to the 5' end and 3' end of the mRNA respectively. (b) Alternative splicing increases the diversity of proteins in higher eukaryotes. In this process, the same mRNA is spliced differently by selecting different set of introns/exons, resulting in different mature-mRNA leading to the formation of multiple isoforms of proteins from a single gene.

1.4.1: Group I introns

Group I introns can be found in various organisms, genes and genomes including fungal and plant mitochondrial DNAs, chloroplasts and bacteriophages[24]. They can also be found in nuclear rRNA genes of *Tetrahymena* and other lower eukaryotes[25]. Their presence in a wide variety of organisms and early primitive species suggest that group I intron could be the ancestral element of modern introns. Group I introns adopt highly conserved secondary and tertiary structures which are important for their role in self splicing. Folding of group I introns results in the formation of an active site by bringing the key residues together, which are otherwise located distantly[26]. In general, group I introns share a similar core structure at their active site. Despite their sequence differences, all group I introns have a series of short, conserved sequence residues, which are base pairing in the conserved structure (Figure1.3a)[26-28]. The conserved secondary structure of group I consists of paired regions (P1-P10) separated by single stranded joining regions (J) or capped by loops (Figure1.3a and b)[22,24,26,28]. The catalytic core consists of two helical domains with paired (P) regions; P4-P5-P6 and P3-P7-P9[22,24,27,28]. Among these paired regions, P3-P9 domain has shown to be important, since it has the guanosine binding site[22,24,26-28]. Many recent crystal structures of group I introns provide more information about this general architecture and also revealed important interactions within the domains which are important for the folding and catalysis of group I introns[28-30].

1.4.2: Group II introns

Group II introns are present in bacterial and organellar genomes in many eukaryotes including yeast and plants but not in animal genomes[20,23,31]. This type of intron has very divergent primary sequences, but share a highly conserved secondary structure[23]. Group II intron secondary structure is composed of six domains branching out from a central region (Figure 1.3c and d)[20,22,23,27,32]. The largest among all six domains is domain 1 (D1), which forms inter- and intra-domain tertiary interactions and thus acts as a scaffold for the assembly of other domains[23]. As an example, interactions between D1 and D5 are important for D5 docking and for the positioning of D5 close to the 5' SS, which is essential for the catalysis. Also this contains exon-binding sites (EBS) and hence recognises 5' and 3' SS through base pairing interactions with intron-binding sites (IBS)[23,27,32]. The next two domains, domain 2 and 3 (D2 and D3) have not shown any essential role in group II function[23]. However the presence of D2 and D3 increase the catalytic efficiency. Domain 4 contains the open reading frame (ORF) and domain 6 (D6) is important since it contains the branch point adenosine residue[27,32]. Among the six domains, domain 5 (D5) is the most conserved and vital domain, which forms the catalytic centre of the group II intron (Figure 1.3c)[23,32]. Despite its short sequence it contains several regions that are highly conserved as well as essential for the catalysis, including the AGC triad, asymmetric bulge and the formation of tertiary interactions (Figure 1.3c)[27,32].

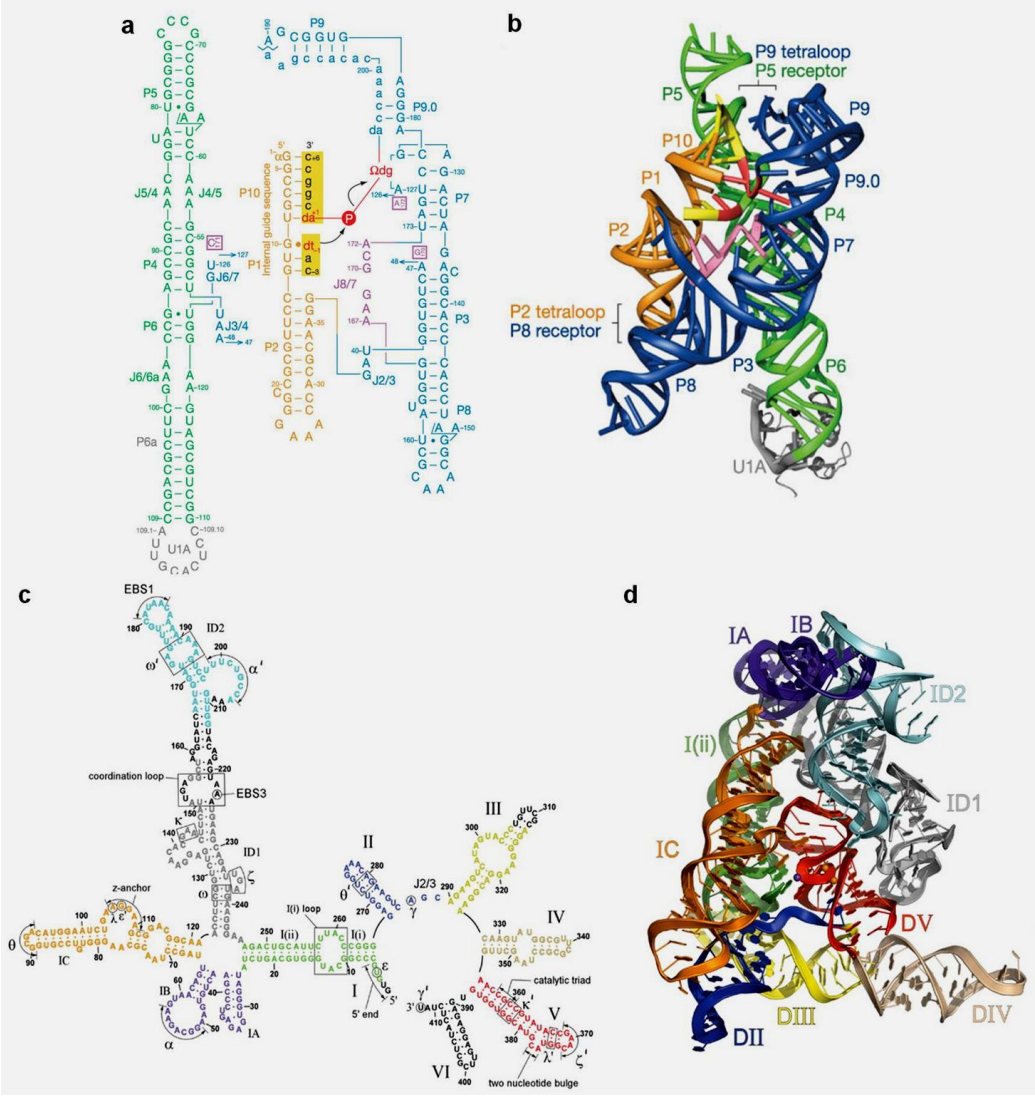

Figure 1.3: Secondary structure and crystal structures of Group I and Group II introns. (a) Secondary structure and (b) crystal structure (in ribbon and cylinder representation) of the group I intron of the purple bacterium Azoarcus sp. BH7228. The intron, exon sequences and the structural elements (P and J) are colour coded in both structures in a similar manner. The RNA transcript in capital letters, residues from two chimaeric oligonucleotides (for intron/3' exon segment and 5' exon segment) are in lower-case letters. Reprinted by permission from Macmillan Publishers Ltd: Nature (28), copyright (2004). (c) Secondary structure and (d) the crystal structure (in ribbon representation) for group II intron of Oceanobacillus iheyensis32. Six domains are showing as Domain I-VI and colour coded in both structures in a similar manner. Reprinted from (32) with permission from AAAS.

1.4.2: Nuclear Introns (or spliceosomal introns)

Nuclear introns are found in almost all eukaryotic nuclear genomes. This type of introns are absent in prokaryotes and the number of introns among eukaryotes change tremendously[3,19,20]. Nuclear introns are spliced by an RNA-protein complex called the spliceosome, which will be discussed in detailed later in this chapter.

1.5: Highly conserved regions in pre-mRNA

An erroneous splicing of an intron, even by one nucleotide can entirely change the end product resulting in the formation of non-functional proteins or mutant proteins that can eventually cause cancer or other cellular defects[33,34]. Hence, during the splicing process, precise identification of intron-exon junctions is crucial. In favour of this, introns contain some short, highly conserved sequences to mark the boundaries of introns that help in precise excision of introns[1]. These sequences are named as the 5' splice site (SS), the 3' splice site and the branch site (BS, Figure 1.4a)[5,6].

The 5' SS, is a highly conserved region at the junction between the 5' exon and downstream intron. In higher eukaryotes this 5' SS composes of AG/GURAGU sequence (where '/' denotes the exon-intron junction and 'R' is a purine)[2,5,6]. In majority of the organisms (~95-99%) the dinucleotide 'GU', marks the 5' intron terminus, followed by a less conserved sequence. The 3' SS is a highly conserved region that can be found at the junction of an intron and the adjacent 3' exon and contains a YAG/G sequence (where Y is a pyrimidine)[2,5,6]. Similar to the 5' SS, AG is the highly conserved dinucleotide that marks the 3' end of the intron. The branch site (BS) is a highly conserved adenosine located within the intron, ~18-40 nucleotides upstream to the 3'

splice site[5]. This adenosine is surrounded by a consensus sequence of CURACU (where A is the branch site adenosine) and followed by a polypyrimidine tract in higher eukaryotes[2,5,6]. These sequences in yeast (*Saccharomyces cerevisiae*) are highly conserved when compared to those in metazoans (Figure 1.4a)[5]. These three regions play a major mechanistic role in splicing.

1.6: Chemistry of splicing mechanisms

Group I introns splice via two trans-esterification reactions using an external Guanosine (exoG) as a cofactor (Figure 1.4b)[22,25,27]. Before the first step, the exoG needs to bind to a pocket in the catalytic core of the group I intron, which is known as the G-binding site[22,25,27]. During the first step, the 3' hydroxyl group (OH) of exoG carries out a nucleophilic attack on the 5' SS resulting in covalent binding of exoG to the 5' end of the intron and release of the upstream exon[22,25,27]. The free OH group at the 3' end of the exon then attacks the 3' SS, followed by the release of exoG by replacing it with the last nucleotide of the intron, which is always a G and known as ωG[27]. This results in the removal of intronic RNA and ligation of two adjacent exons.

Both group II introns and nuclear introns show a similar mechanism for splicing, whereby intron removal occurs via two trans-esterification reactions using 2' OH of the highly conserved adenosine in the BS as the nucleophile (Figure 1.4c)[21,23,27,35]. In the first reaction, the 2' OH adenosine in the BS carries out a nucleophilic attack on the 5' splice site[27,31]. This reaction results in a free 3' hydroxyl on the 5' exon and a circular lariat intermediate with a 2'-5' linkage at the branch site adenosine[27,31]. Secondly, the free 3' hydroxyl group on the 5' exon carries out a nucleophilic attack on the 3'-splice

site resulting in a removal of the lariat structure, followed by religation of the two exons to form the mature mRNA which will eventually be transported into the cytoplasm[5,31,36]. The resulting lariat intron can be degraded or further processed into other encoded RNAs such as small nucleolar RNAs, microRNAs, and long non-coding RNAs[37]

Other than the branching pathway described above, group II introns also exhibit an alternative splicing mechanism known as the hydrolytic pathway, where a water molecule acts as the nucleophile for the first step (Figure 1.4d)[23,31]. The second step in this pathway is identical to that in the branching pathway.

1.7: Splicing of nuclear introns catalyzed by the spliceosome

In higher eukaryotes the splicing of nuclear introns is catalyzed by the spliceosome, a multi-mega-Dalton ribonucleoprotein complex[1,5,6]. The spliceosome consists of five small nuclear ribonucleoproteins (snRNPs) known as U1, U2, U4, U5 and U6 and numerous non-snRNP protein splicing factors.[5,38] Studies have shown that the yeast spliceosome contains ~80 proteins, whereas human spliceosome contains ~170 different proteins, and it has homologous counterparts for almost all yeast spliceosomal proteins[1,39,40].

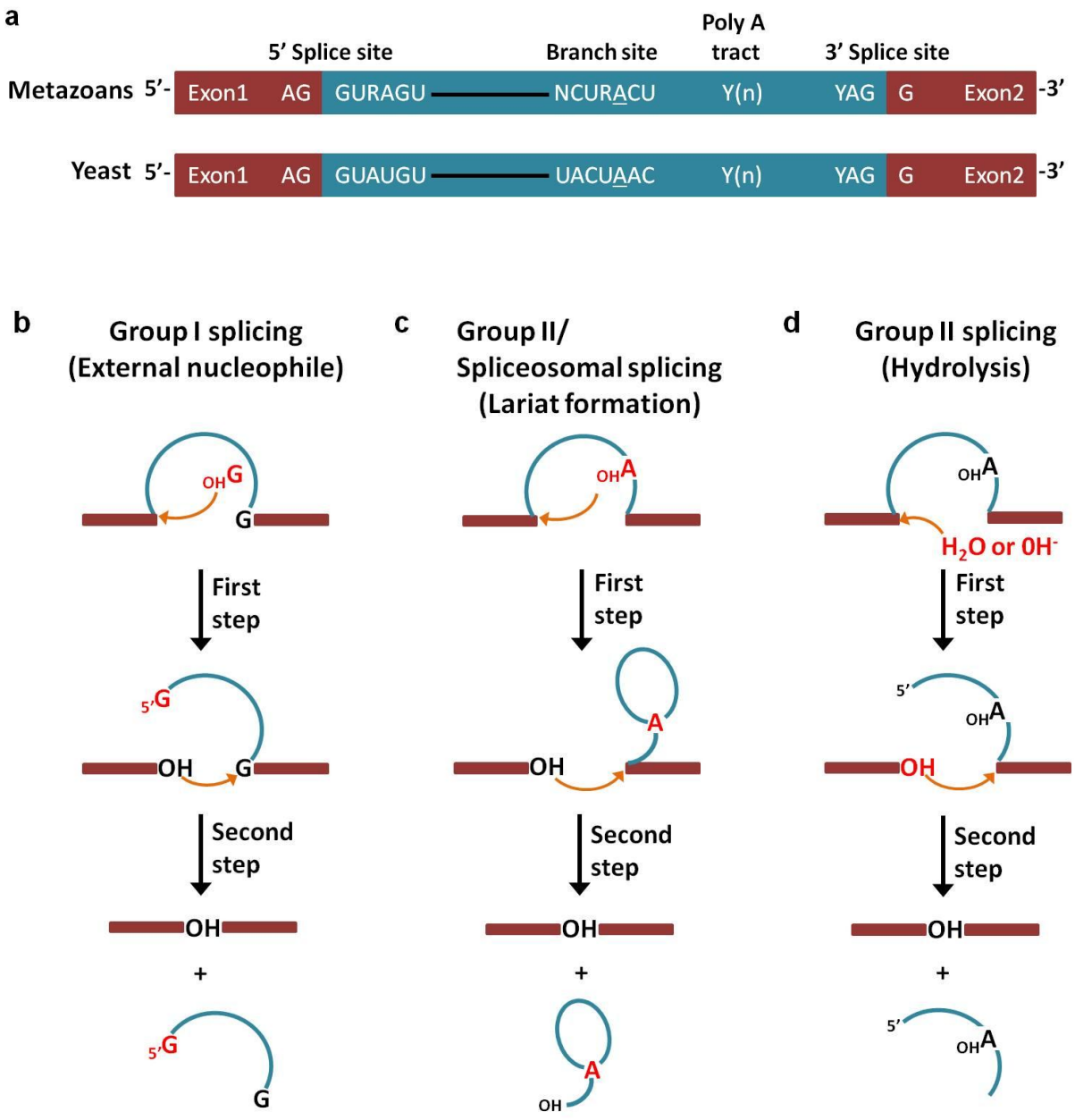

Figure 1.4: Splicing mechanism. (a) Highly conserved regions between metazoan (upper) and yeast (bottom) pre-mRNA that are essential for splicing. (b) Self splicing in group I, using an external guanosine. (c) Splicing mechanism in group II and spliceosomal introns via lariat formation. (d) Hydrolytic pathway for splicing by group II introns. Nucleophile in each mechanism is shown in red.

1.7.1 Assembly of the spliceosome

SnRNAs and related protein factors repeatedly undergo a highly ordered and stepwise pathway during the spliceosomal assembly and catalysis (Figure 1.5)[1,5,41]. First, the U1 snRNP recognizes the 5' splice site in the pre-mRNA and U2 binds to the branch point, subsequently forming the pre-spliceosome (complex A). The U4/U6 di-snRNP associates with U5 to form the U4/U6•U5 tri-snRNP complex. This preformed tri-snRNP binds to the pre-mRNA to form the pre-catalytic spliceosome (complex B). This binding causes major structural rearrangements within RNA-RNA and RNA-protein interactions which results in the release of U1 and U4 from the complex, forming the activated spliceosome (complex B^{act}). The B^{act} complex is further activated by ATP-dependent helicases to form the B* complex, which catalyzes the first step of splicing. This gives rise to complex C, which catalyzes the second catalytic step. After both steps of splicing have occurred and the mature mRNA is formed, the spliceosomal components dissociate and start another cycle of splicing[1,5,10,36]. During the splicing reaction, only U2, U5 and U6 remain in the spliceosome, and thus it has been suggested that they form the catalytic core of the splicing machinery[42,43]. Assembly of the spliceosomal components and their recycling after both splicing reactions are highly conserved[5].

1.7.2: Structure and assembly of U snRNPs

Each snRNP consists of a Uridine-rich small nuclear RNA (U snRNA), several snRNP-specific proteins, and a common set of proteins known as the core domain which is made up with seven Sm or Like-Sm (LSm) proteins[5,36,44].

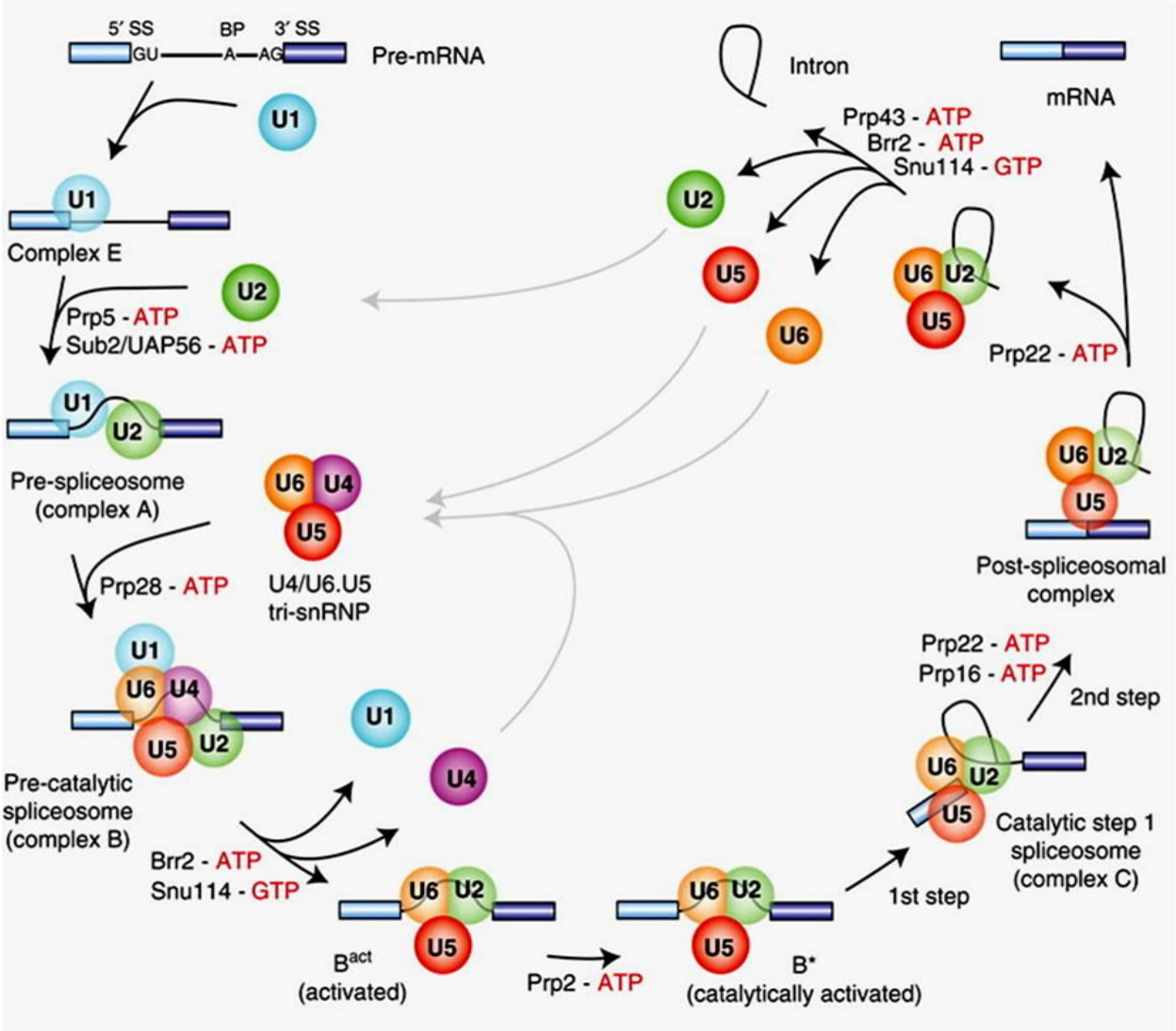

Figure 1.5: Assembly and catalysis of the spliceosome. Stepwise assembly of the five snRNPs (circles) on premRNA (exons as boxes and intron as a black line, respectively) and involvement of different proteins at each step are shown. Reprinted from (5) with the permission from CSHL press.

1.7.2.1 U snRNAs

U snRNAs are Uridine-rich RNA sequences that play a major role in spliceosome assembly and catalysis. These snRNAs provide platforms for protein binding and give discrete functions for each snRNP[44]. These U snRNAs adopt highly conserved secondary structures and consist of consensus regions that are important for the interactions with pre-mRNA. The U1, U2, U4 and U5 snRNAs are RNA polymerase II transcripts and they transport into cytoplasm to obtain 2,2,7-trimethylguanosine (m3G) cap[44]. Then the seven sm proteins bind to a short single stranded region in the U1, U2, U4 and U5 snRNAs, which contains a consensus AU(4-6)G sequence (known as the Sm site) forming the core domain of the snRNPs[45]. The survival motor neuron protein (SMN) plays an essential role in the sm core assembly[46]. The binding site for the Sm proteins is located on a single stranded region within two stem-loop structures of these snRNAs, which stabilizes the protein binding[44]. It has been suggested that the m3G cap and the Sm core may act as a recognition signal during the transportation of these snRNAs to the nucleus[44]. In the nucleus, the snRNAs are localized in the Cajal bodies (CBs), where they are assembled into the snRNPs by binding with other snRNP-specific proteins[47].

U6 snRNA, in contrast is not transported to the cytoplasm, but stays in the nucleus. Also it's an RNA polymerase III transcript which does not contain an Sm site[44]. Instead it contains a single stranded region at the 3' end which is recognized by a set of seven proteins which are similar to the Sm proteins and thus named as 'Like-Sm' (LSm)[44,45].

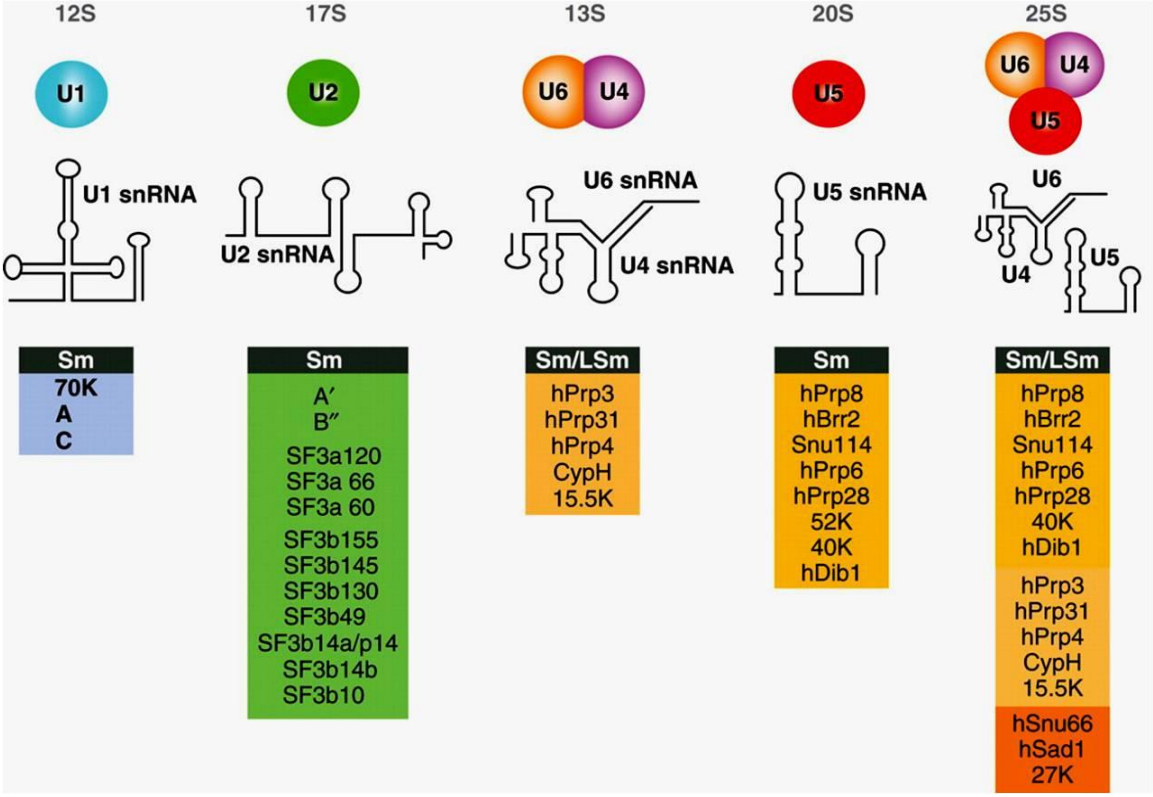

Figure 1.6: Protein composition of spliceosomal snRNPs. A cartoon of the secondary structure of each snRNP is shown and the list of proteins associated with each snRNP or snRNP complex is shown in the boxes under the corresponding structure. Sm proteins and LSm proteins complexes indicated as ‘Sm’ and ‘LSm’ respectively, on top of each box. Reprinted from (5) with the permission from CSHL press.

1.7.2.2: U snRNP Core domain

Core domains of U1, U2, U4 and U5 snRNPs consists of seven Sm proteins; SmB/B', -D1, -D2, -D3, E, F and G (in human, 8-25 kDa)[44,45,48]. These proteins are important for the assembly, transportation and stability of U snRNPs. Each of these proteins contains a conserved motif with two segments; Sm1 and Sm2 connected by a linker (Figure 1.6a)[44,45]. In the absence of snRNAs, Sm proteins are present as two heterodimers (SmD1·D2, SmD3·B/B', Figure 1.6b)[48] and a single heterotrimer (SmF·E·G). Assembly of Sm proteins to the U snRNAs occurs in the cytoplasm. First the SmF·E·G subcomplex binds with SmD1·D2 and are assembled onto the U snRNA to form a stable subcore domain followed by the binding of SmD3·B/B' to form a heptameric ring around the Sm site, resulting in the formation of the complete core structure (Figure 1.6c)[44,48]. In contrast, core domain of U6 snRNA consists of seven LSm proteins; LSm 2-8, which form a doughnut-shaped structure in the absence of U6 snRNA[49].

1.7.3: U1 snRNP

U1 snRNP is composed of U1 snRNA, seven Sm proteins and three U1-specific proteins; U1-A, U1-70K and U1-C (Figure 1.7)[44]. In total it has a molecular weight of ~240 kDa, and is the smallest snRNP among five. It is also known to be the most abundant snRNP in the cell and can be isolated in large quantities. In majority of the cases U1 is the first component loaded on to the pre-mRNA and hence marks the starting point of the spliceosomal assembly cycle (E complex). U1 plays an important role in 5' SS recognition where the 5' end of U1 snRNA base pairs with the 5' SS of pre-mRNA.

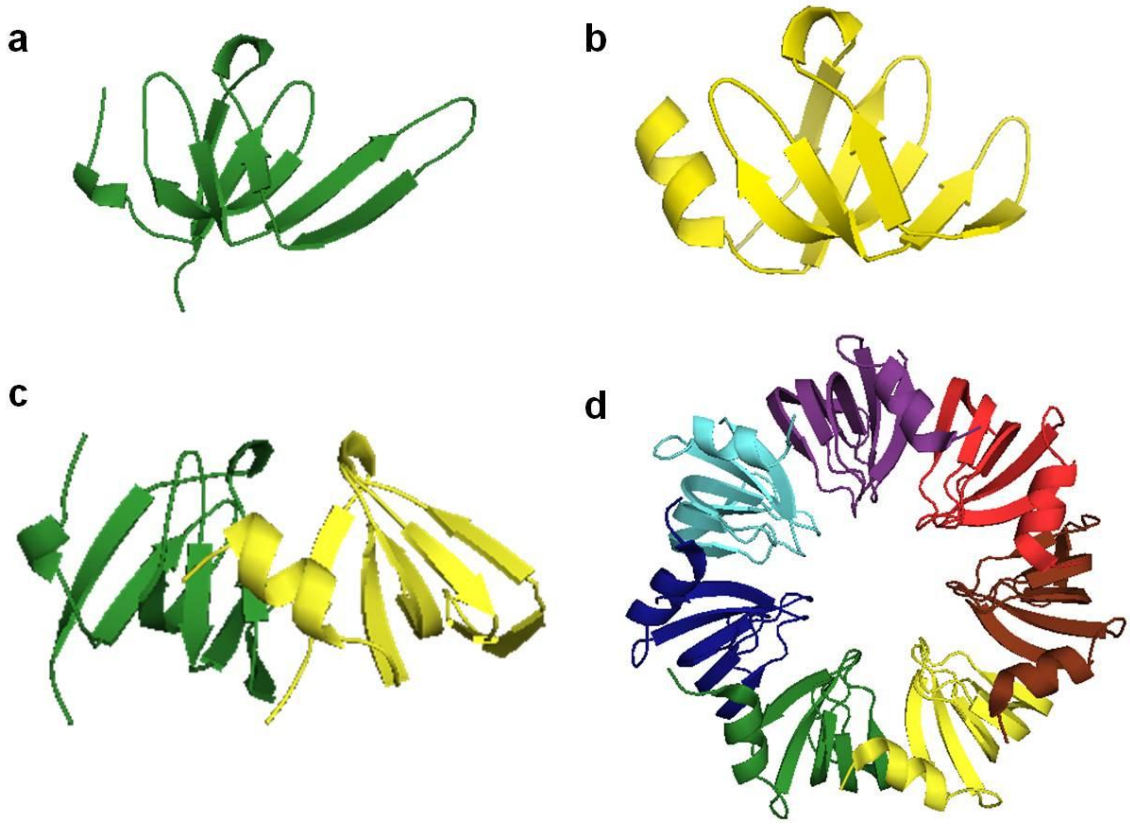

Figure 1.7: Structure of the U snRNP Sm core domain. Structure of (a) SmB and (b) SmD3 showing the Sm fold, which is common for all Sm proteins (PDB:1D3B). It consists of an N-terminal α-helix followed by a five-stranded, highly bent, antiparallel β-sheet. (c) A complex of SmD3.B, where β4 of B interacts with β5 of D3. (PDB:) (d) The heptameric ring of Sm core (PDB:1I8F). The β4-β5 interaction between adjacent Sm proteins, results in the formation of the ring structure. Figures were generated from PyMOL.

U1 snRNA consists of four stem loop structures (SL1-4) and a single helix H (Figure 1.8a)[44,50]. Sm proteins bind to the Sm site located in between SL3 and SL4 to form the core domain[44,50]. Early electron microscopy images revealed a globular core domain and two bulges represent binding of U1-A and U1-70K (Figure 1.9a)[51]. According to the recent crystal structure data, the RNA sequence 5' to the Sm site forms a four helix junction structure and SL4 is on the opposite side of the Sm ring (Figure 1.9b)[52,53]. Both SL1/SL2 and SL3/Helix H structures are coaxially stacked[44]. The U1-A and U1-70K bind to SL1 and SL2 respectively and independently from each other (Figure 1.9c)[44]. On the other hand U1-C binds to U1, only in the presence of core domain and U1-70K suggesting that U1 snRNP assembly is a hierarchical process[52]. Moreover it has been shown that U1-70K facilitates the binding of U1-C by providing an additional binding site for U1-C. Mutational studies of the zinc finger domain of U1-C have shown that 5' SS recognition is drastically affected, suggesting its vital role in the formation of E complex[52,53].

1.7.4: U2 snRNP

The U2 snRNP consists of U2 snRNA, Sm core and ~15 other proteins (Figure 1.7)[44,54]. U2 snRNA is 187 nucleotides long and has four stem-loop structures (Figure 1.8b)[45]. During spliceosomal assembly, U2 snRNA first interacts with the BS region of pre-mRNA (pre spliceosome; A complex)[5] forming a short duplex between U2 and the BS sequence, which facilitates the bulging out of the BS adenosine[55,56]. Later on, U2 snRNA is base paired with U6 snRNA to form the activated spliceosome (Bact/ B* complex)[5].

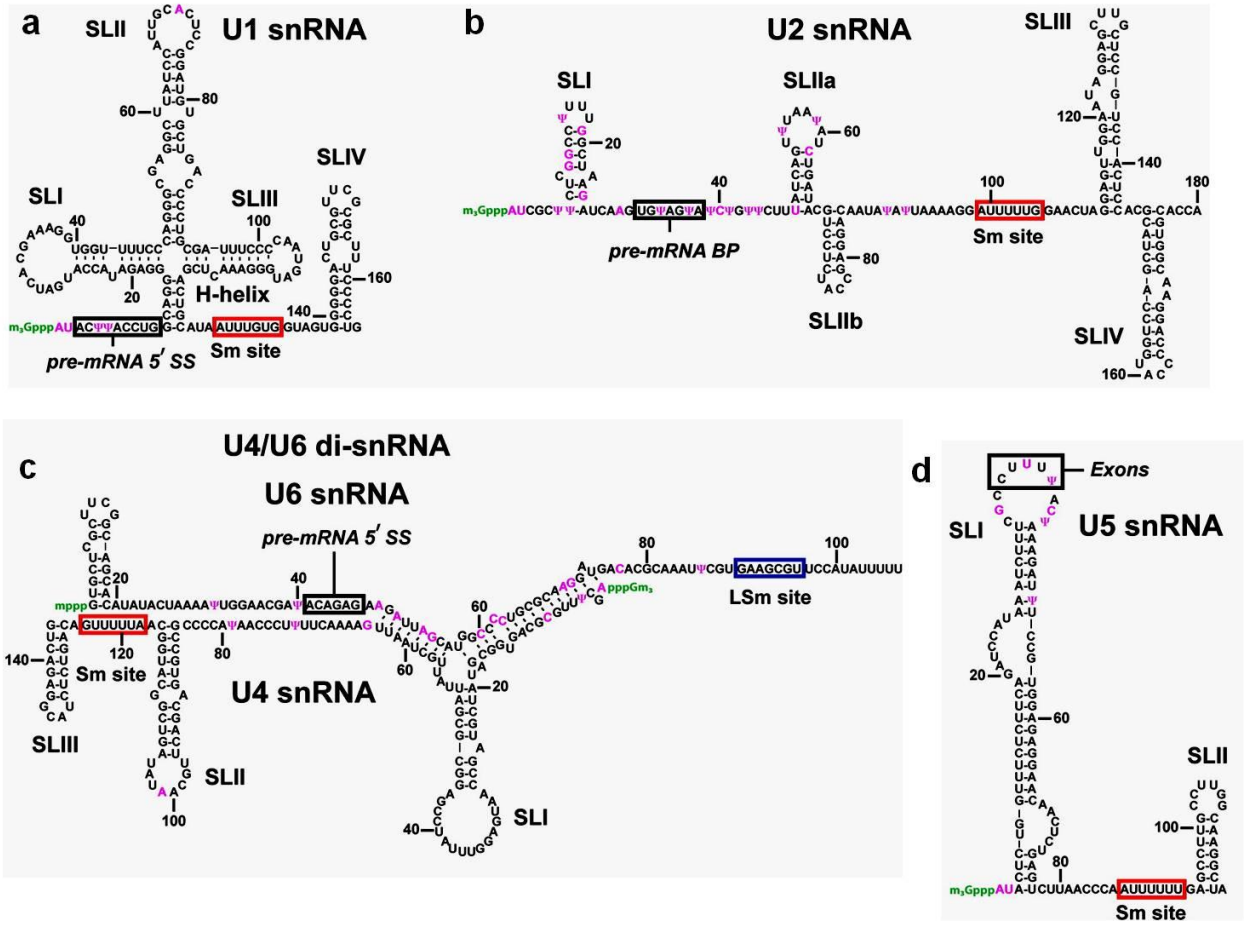

Figure 1.8: Secondary structure of U snRNAs; (a) U1, (b) U2, (c) U4/U6 and (d) U5. The Nucleotides subject to post-transcriptional modifications (Pseudo-U or 2'-O-methylated) are shown in magenta. The Sm binding sites are boxed in red where as LSm binding site is boxed in blue. Sequences that interact with premRNA 5' Splice site, exons and the branch site are highlighted with a black box. Reprinted with permission from (42). Copyright (2012) American Chemical Society.

Studies have shown that U2 undergoes intramolecular dynamics, where it adopts two stem structures; either stem IIa or stem IIc (Figure 1.10a)[5,57,58]. The stem IIa conformation facilitates the U2/BS interactions during the formation of pre-spliceosome[57,58]. On the other hand stem IIc supports the first step of splicing, followed by a change back to stem IIa which is important for the second step[57,58]. U2 snRNP associated splicing factor Cus2 is playing an important role along with the Prp5, in switching the U2 conformations from stem IIc to stem IIa[58,59].

Out of the U2 snRNP associated proteins, the majority are part of the protein subcomplexes splicing factor 3a and 3b (SF3a and SF3b)[54]. SF3a and SF3b are known to play a vital role in the stabilization of the conformation of the branch site adenosine[16,44,60]. SF3a (~197 kDa) is composed of three proteins; whereas SF3b (~450 kDa) has seven proteins (Figure 1.10b)[16,44,60,61]. These SF3b proteins bind around the branch site sequence, across 25 nucleotides upstream to 5 nucleotides downstream[16,61,62]. Chemical cross linking studies have shown a direct interaction of one SF3b component; p14 with a U2 sequence that is base paired with the branch site, after the integration of U2 snRNP to the pre-spliceosome[60,63,64]. Location of p14 within the central cavity of the SF3b indicates that pre-mRNA need to undergo some structural rearrangements to interact with U2[63]. There are two other proteins named U2A' and U2B" which play an important role in integrating U2 into the pre-spliceosome. The U2B" protein is closely related to the U1A protein and contains two RNP domains[45,65]. In contrast to U1A that binds to SL2 of U1 snRNA by itself, U2B" is associated with U2A' first and then binds to SL4 of U2 snRNP (Figure 1.10c)[45,61,65]. Also, it has been shown that SF3a associates with the U2B"- U2A' complex and the core domain[45].

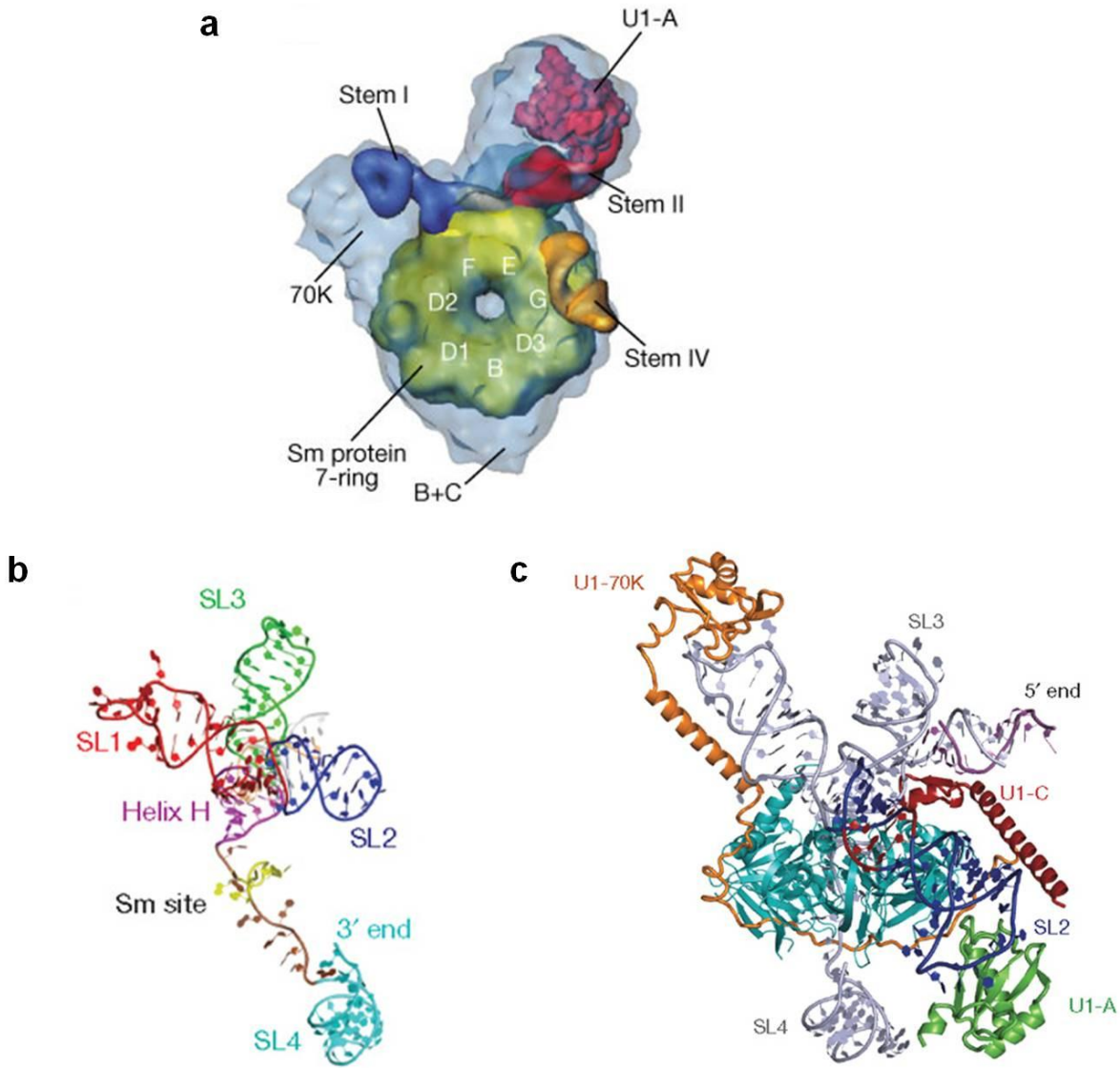

Figure 1.9: Structure of U1 snRNP. (a) Enlarged stereo view of the model for U1 snRNP based on the structural and biochemical information. Reprinted by permission from Macmillan Publishers Ltd: Nature (47), copyright (2001). (b) Crystal structure of human U1 snRNA consists of four stem-loops structures; SL1-4. Sm binding site is shown in yellow. (c) Model of the complete U1 snRNP. Figure c and d are reprinted by permission from Macmillan Publishers Ltd: Nature (48), copyright (2009).

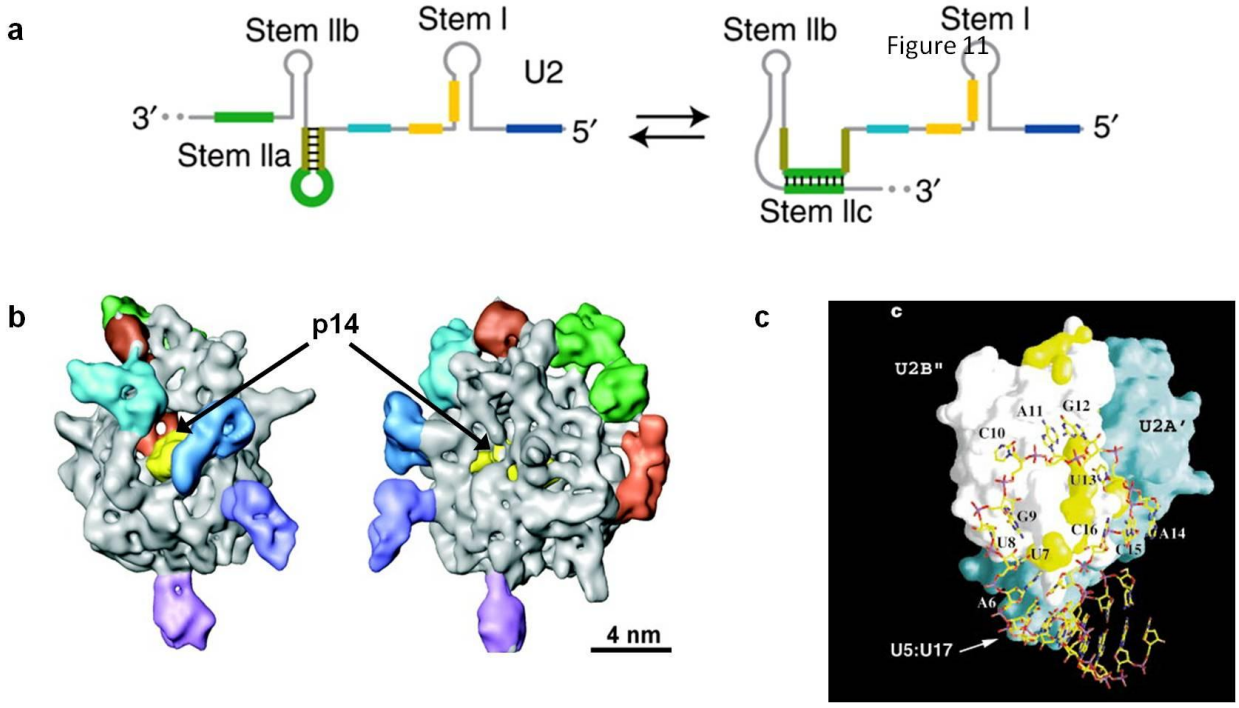

Figure 11

Figure 1.10: Structure of U2 snRNP. (a) U2 adopts two mutually exclusive stem structures; Stem IIa and Stem IIc, which may form during different steps of assembly. Reprinted from (5) with the permission from CSHL press. (b) EM structure of U2 associated protein SF3b. Reprinted from (56), with permission from AAAS. (c) Crystal structure of U2B''-U2A' protein complex bound with a fragment of U2 snRNA. Reprinted by permission from Macmillan Publishers Ltd: Nature (61), copyright (1998).

1.7.5: U5 snRNP, U4/U6 di-snRNP and U4/U6•U5 tri-snRNP

1.7.5.1: U4/U6 di-snRNP

The U4/U6 di-snRNP consists of two snRNAs; U4 and U6, a set of Sm proteins that bind to U4 and a set of LSm proteins that bind to U6 and five particle-specific proteins[66,67]. U4 and U6 base pair with each other to form a highly conserved Y shaped conformation consisting of two intermolecular helices; stem I and stem II and a 5' stem loop structure of U4 (Figure 1.11a)[66,68]. This 5' stem loop structure is an essential component, which provides a platform for the binding of U4/U6 specific proteins. Other than the Sm/LSm core, five proteins; with the molecular weight 15.5K, 20K, 60K, 61K and 90K, have been identified using biochemical assays which associate with the human U4/U6 di-snRNP[69].

Apart from the 20K protein, rest of the proteins are highly conserved over species and the orthologous proteins in yeast are named as Snu13 (15.5K), Prp4 (60K), Prp31 (61K) and Prp3 (90K) (Figure 1.11a)[69]. Snu13 and its counterpart play an important role in the formation of U4/U6 di-snRNP. Structural data have shown that Snu13 directly binds to the kink turn motif at the 5' stem loop structure of U4 snRNA (Figure 1.11b)[**70,71**]. Previous studies have suggested that Snu13 acts as a nucleating factor where it facilitates the binding of one or more of the U4/U6 specific proteins to complete the assembly of the di-snRNP[69]. Co-immunoprecipitation/pull-down studies coupled with UV-cross linking have revealed that the binding of Prp31 is greatly accelerated by the presence of Snu13 (Figure 1.11b)[69]. Similarly, this study has shown the 20/60/90K complex in human (similar to Prp3/Prp4 in yeast) can only bind with the U4/U6 in the presence of Snu13, confirming the suggested role for Snu13 as a nucleating factor[69].

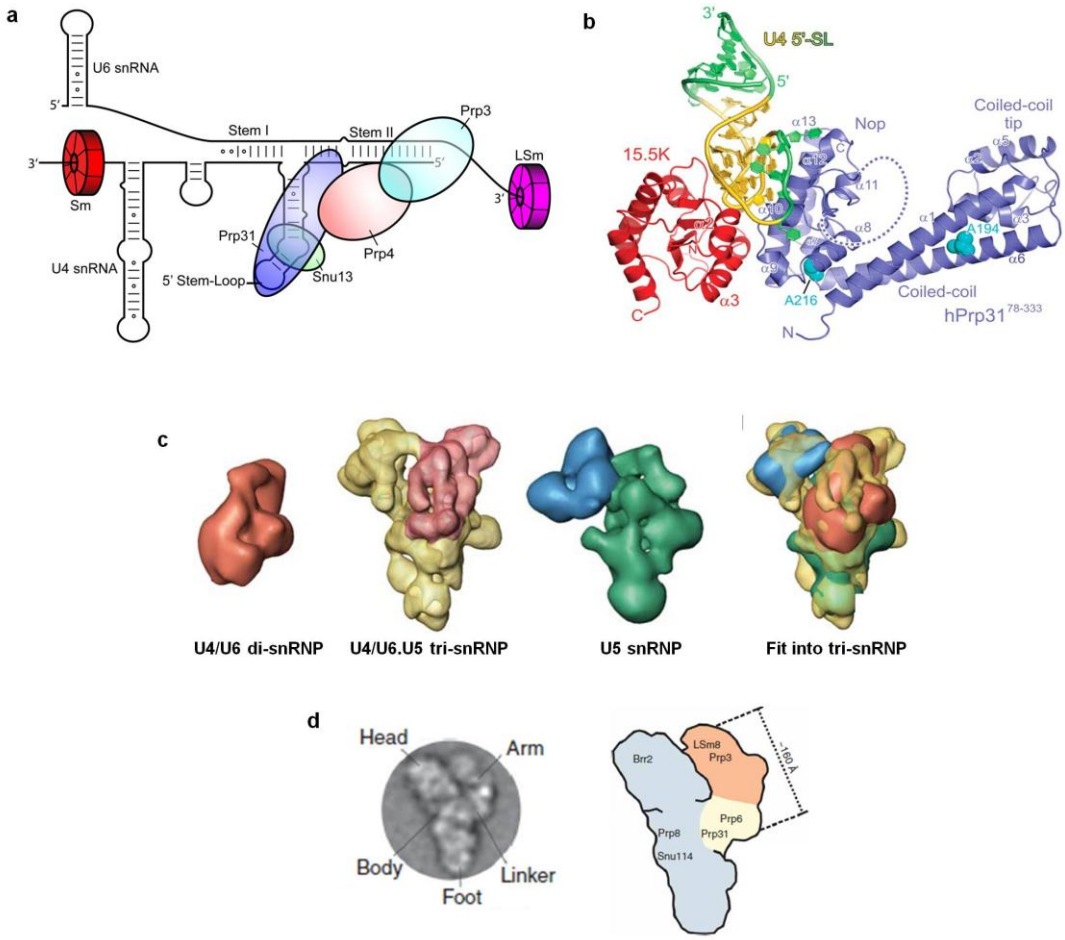

Figure 1.11: Structure of U4/U6 di-snRNP and U4/U6•U5 tri-snRNP. (a) A schematic representation of the secondary structure of the U4/U6 duplex and its associated proteins; Sm, Lsm, Snu13, Prp31 and Prp3/4. (b) A structure showing binding of hPrp31 (blue) and 15.5K (red) to the U4 5' stem-loop (gold). A disordered loop in hPrp31 is indicated with a dash line. Reprinted from (68), with permission from AAAS. (c) 3D constitution of U5, U4/U6 snRNPs and U4/U6•U5 tri-snRNP and fitting U5 and U4/U6 structures into the tri-snRNP structure. Reprinted from (73), Copyright (2006), with permission from Elsevier. (d) EM structure of yeast U4/U6•U5 tri-snRNP with the main structural domains labelled (left). A schematic representation of yeast tri-snRNP (right) illustrates the binding sites for some of the particle specific proteins. Areas corresponding to U5, U4/U6 and linker region are in grey, orange and yellow respectively Reprinted by permission from Macmillan Publishers Ltd: NSMB (81), copyright (2008).

UV-cross linking studies and NMR studies have shown that Prp31 binds to the U4 stem-loop only in the presence of Snu13[69]. A structure of a ternary complex consisting of the 5' stem-loop of U4, Prp31 and Snu13 shows that the Nop domain of Prp31 acts as an RNP interacting motif[72]. Prp3 and Prp4 are known to be the only U4/U6 di-snRNP specific proteins and they form a dimer prior to binding to the U4/U6 duplex[73-75]. Prp3 was shown to interact with a region towards the 3' end of U6 snRNA[69,75,76]. Although these studies have shown the binding of proteins to U4/U6 duplex, little is known regarding the global structure of the U4/U6 snRNP. Currently, the only global structural information comes from a low resolution (~30 Å) EM structure showing large and small domains connected by a thin bridge (Figure 1.11c) [68,77], but the relative orientation of the helices of the U4/U6 3-way junction remains uncertain.

1.7.5.2: U5 snRNP

U5 snRNP, which also plays an important role within the spliceosomal assembly, consists of U5 snRNA, Sm core and several specific proteins. The secondary structure of U5 is highly conserved among species confirming its essential role in splicing. In yeast, the structure of U5 snRNA can be divided into three main regions; a long stem-loop structure, a single stranded region containing the Sm binding site and a short stem-loop structure towards the 3' end (Figure 1.8c)[78]. The long stem-loop structure can be further divided into seven subunits including the loop I[78]. The loop I of U5 snRNA is known to play an important role in 5' and 3' exons aligning during the second step of splicing[79,80]. EM structure for U5 snRNP has revealed that the loop I is located in the centre of the particle, suggesting that other catalytic centre components interact U5 via

this central region[77]. Nine U5-specific proteins have been found in human with the molecular weights of 220, 200, 116, 110, 102, 100, 52, 40 and 15 kDa (Figure 1.7)[81]. Only four proteins have been observed in yeast namely; Prp8p (in human 220 kDa), Snu246p/Brr2 (in human 200 kDa), Snu114p (in human 116 kDa) and Prp18p (Figure 1.7)[78]. Prp18 only associates weakly with U5 and is involved in the second catalytic step[78]. Similarly the mammalian homologous of Prp18 also has a function in the second step of splicing although it is not associated with U5. Prp8 and its mammalian equivalent play an essential role in splicing *in vitro* and *in vivo*. Cross linking studies have shown that Prp8 interacts with the 5' SS, branch site and 3' SS[39,82,83]. Other than that, Prp8 also interacts with loop I of U5 and U6 snRNA, and facilitates the activation of Brr2[39,82]. Overall Prp8 brings all important components together at the catalytic centre[39]. Brr2 is a DExD/H-box helicase, which is thought to be involved in the unwinding of U4/U6 snRNA, during the formation of activated spliceosomes[84]. However, the exact mechanism of the unwinding of U4/U6 by Brr2 is not known yet. Snu114 is a GTPase, which is similar in sequence to the eukaryotic ribosomal elongation factor EF-2 (or bacterial EF-G) that is shown to be involved in modulating Brr2 activity during spliceosomal activation by switching between GTP-bound state and GDP-bound state[39,85].

1.7.5.3: U4/U6•U5 tri-snRNP

During spliceosomal assembly, the U4/U6 di-snRNP first interacts with the U5 snRNP to form the U4/U6•U5 tri-snRNP, which then binds with the pre-spliceosome to form the pre-catalytic complex or B complex[5]. This tri-snRNP is the largest spliceosomal

subcomplex, which consists of U4, U6, U5 snRNAs and ~29 distinct proteins (Figure 1.7)[54,61]. Structural rearrangements that occur within the tri-snRNP result in the removal of U4 from the tri-snRNP, which is crucial for the transition from pre-catalytic complex to the activated complex (B_{act})[5]. Visualisation of yeast U4/U6•U5 tri-snRNP by EM has revealed an elongated, triangular shaped particle with a dimension of 30-40 nm (Figure 1.11c, d)[77,86]. It has been shown that the global structure of the tri-snRNP consists of a pointed lower end ('foot'), a broader upper left structure ('head') and a smaller protruding domain ('arm'), which is connected to the centre ('body') of the particle via a linker region (Figure 1.11d, left)[86]. Whereas the images for human tri-snRNP haven't shown a clear separation of the domains as it has in yeast. U5 associated proteins; Prp8, Brr2 and Snu114 are found within the head and the central part of the particle suggesting that U5 snRNP is located in that region (Figure 11 d, right)[86]. On the other hand the U4/U6 proteins; LSm8 and Prp3 are mapped to the arm region, showing the position of U4/U6 di-snRNP among the tri-snRNP (Figure 11 d, right)[86]. Tri-snRNP specific bridging proteins such as Prp6 were found in the linker region of the particle (Figure 11 d, right)[86].

1.7.6: The proteome of spliceosomal complexes

During assembly and catalysis, the spliceosomal components undergo various conformational changes that are crucial for its activity. Other than the snRNAs, protein factors also play a crucial role in spliceosomal assembly in order to have optimum catalytic activity[5,36,39]. These proteins are involved in maintaining the structures of individual snRNPs and complexes, by engaging in the structural rearrangements within

the assembly cycle, such as unwinding, annealing, or stabilizing the RNA components in order to form the active complex[1,87]. Also, binding of some proteins may prevent the premature formation of certain conformations of snRNAs important for the catalysis or proper assembly of the spliceosome[1,5]. Although the role of majority of the spliceosomal proteins has been identified, there are many to reveal. The study of the spliceosomal proteome is difficult as the spliceosome undergoes drastic structural changes during which many proteins are added and removed from the spliceosome. During the spliceosomal assembly and catalysis, the protein composition changes dramatically. As an example, B complexes contain ~60 proteins in yeast and ~110 proteins in humans whereas C complex consists of ~50 proteins in yeast and ~110 in metazoans[5].

The spliceosomal proteome consists of various types of proteins including SR family proteins, DExD/H-box-type ATPase-dependent RNA helicases and ATPase-independent RNA chaperones. Also several proteins which are involved in the spliceosomal assembly cycle belong to the “Prp group” (Prp: pre-mRNA processing proteins), including Prp2, Prp24, Prp31, Prp43, Prp16 etc. Despite the families they belong to, spliceosomal proteins can broadly be categorized into two major groups as snRNP associated and non-snRNP associated. There are several protein factors that are the building blocks of snRNPs, which have a major role of maintaining the global structure of each of the individual snRNP or snRNP complexes. As an example Brr2, Prp8 and Snu114 are U5 associated proteins[39] whereas Prp24 is a U6-associated protein[1,39]. Brr2 is a U5 associated DExD/H-box helicase, which is composed of two helicase domains[88]. The N-terminal helicase domain carries out the unwinding activity, whereas the C-terminal helicase domain is not active and may act as a platform to

interact with other spliceosomal components[89,90]. Brr2 is involved in the unwinding of the U4/U6 duplex before the spliceosomal activation, suggesting that the role of Brr2 needs to be highly coordinated and tightly regulated in order to prevent the premature activation[1]. Brr2 has also been shown to be involved in the unwinding of the U2/U6 complex after the catalysis[91].

Prp8, another U5-associated protein is the largest and most highly conserved protein in spliceosome. Previous studies have shown that Prp8 interacts with several spliceosomal proteins, snRNAs and both 5' and 3' splice sites[39,82,83]. It contains an RNA recognition motif (RRM) in the middle of the protein, a Jab1/MPN domain (Figure 1.12a) and an RNase H like domain (Figure 1.12b) at the C-terminal region of the protein[92-94]. Prp8 plays an important role in the regulation of Brr2 function. A C-terminal portion of Prp8, containing the RNase H and Jab1/MPN domains, interact with Brr2 and facilitates its helicase activity[1,95]. Together with Prp28, Prp8 also acts as a trans-helicase, which mediates the transition of 5' SS from U1 to U6 (Figure 1.12c)[96]. Ubiquitinylation of Prp8 inhibits the helicase activity of Brr2 providing another regulatory step for the Brr2 function[1,97]. Snu114 is a GTPase which is highly homologous to translation elongation factor G (EFG)[98]. The GTPase domain of Snu114 is required for the loading of this protein to the U5 snRNP[99]. This protein also plays a crucial role in the regulation of ATPase activity and helicase activity of Brr2[5,85]. It is also required for the aligning of exons during the second step of splicing. Other than these direct roles, Snu114 is suggested to have a role in translocation of spliceosome, similar to that in eukaryotic ribosome by EF-2 (or bacterial EF-G)[98]. Prp24 is a U6 snRNP associated protein, which is composed of four RRMs[100]. This protein has been shown to function in the assembly

of U4/U6 duplex and also some studies have also suggested its role in the U4/U6 duplex unwinding[101]. Therefore Prp24 plays an important role as a recycling factor within the spliceosomal assembly pathway[102].

A large proportion of spliceosomal protein factors belong to the SR family[39]. Several snRNP associated proteins including U1-70K, Snu66 and Srm160 are part of the SR family[39,103]. On the other hand the majority of non-snRNP proteins also belong to SR family including ASF/SF2, SC35 and U2AF[39]. Proteins in this family consist of one or two copies of N-terminal RRM and a serine/arginine rich C terminal domain[39,103]. The general role for these proteins is to stimulate splicing by binding to exonic sequences, recruiting spliceosomal proteins and stabilizing RNA-RNA interactions[39]. A study done by Akusjarvi and co workers has demonstrated that the SR proteins are also function as splicing repressors[104]. As an example, during the first step of spliceosomal assembly, two SR proteins; SF1 (Figure 1.12d) and U2AF (Figure 1.12e, f) play a major role by recognizing he BS tract at the 3' SS and the polypyrimidine t respectively[36,105,106]. Interaction between U2AF and RNA occur via two modes; either the RNA interacts with RRMs from two different proteins (Figure 1.12d) or a single protein interacts with two separate RNA molecules (Figure 1.12e)[36,106].

There are at least eight DExD/H-box proteins that are highly conserved from yeast to human, and known to be involved in the various steps of spliceosomal assembly cycle such as UAP56, Prp5, Prp28, Brr2, Prp2, Prp16, Prp22 and Prp43[39,54,103]. Some of them have shown a weak helicase activity *in vitro*. Within the spliceosomal assembly pathway, these helicases play an important role in the transition of spliceosomal sub complexes by modulating the RNA-RNA, RNA-protein and protein-protein interactions.

These proteins have distinct functions at specific steps of assembly and facilitate the disruption of interactions in one sub complex and allow them to assemble with different components to form another sub complex.

UAP56 (Sub2 in yeast) and Prp5 involve in the first few steps of spliceosomal assembly. UAP56 is found to be interacting with U2AF (Mud2 in yeast) and supports the early stages of assembly in an ATP-dependent way[39,95,107]. UAP56 promotes the dissociation of U2AF and BBP (Branch point binding protein) which binds at 3' SS and BS respectively to facilitate the U2 binding[1,16,95]. The UAP56 helicase has also been shown to be important later in the assembly cycle, where it is suggested to be involved in the unwinding of U4/U6[107].

Prp5 also has a role in association of U2 to the pre-mRNA[16]. Some studies have revealed that *PRP5* genetically interacts with *CUS2*, suggesting a role in removal of Cus2 from U2 snRNP in an ATP-dependent manner[16,108]. Prp28 mediates the exchange of 5' SS interactions from the 5' end of U1 to the ACAGAG motif in U6 and thus facilitates the release of U1 during the transition from B complex to Bact/B*. Brr2 plays an important role in the unwinding of U4/U6 and also in the dissociation of U2/U6 at a later stage.

The two helicases, Prp2 and Prp16 function in the first and the second catalytic steps respectively[95,109]. In the absence of ATP, Prp2 associates with the pre-mRNA prior to the first step and leaves the mRNA upon ATP hydrolysis[110]. Another role of Prp2 is to remove the SF3a and SF3b after the activation of spliceosome[111]. Prp16 also plays a crucial role in spliceosomal assembly and catalysis. Mutations in Prp16 that affect its helicase or ATPase activity have resulted in erroneous splicing, suggesting a role in

proofreading during spliceosomal assembly[39,95]. Studies have shown that similar to Prp2, Prp16 also interacts with the pre-mRNA in the absence of ATP, and is released upon ATP hydrolysis[112]. Although it is not clear, which components are associated with Prp16, it has been suggested that Prp16 may interact with the 3' SS region of the pre-mRNA, and also mediates the remodelling of the spliceosome prior to the second step of splicing[113]. It is also involved in the rearrangement of one of the basepaired helices within the U2/U6 duplex[114,115].

Prp22 is another helicase that aids spliceosomal assembly and catalysis. This helicase binds to a region that is immediately downstream of the 3' SS in the fully spliced mRNA[116]. The role of Prp22 is to unwind mRNA/U5 interactions using its 3' to 5' helicase activity at the end of the second step of splicing, leading to the release of spliced mRNA[117]. Mutated Prp22 improves the splicing of pre-mRNAs with mutations at 3' SS, suggesting that Prp22 is involved in the regulation of fidelity of the second step of splicing[1,95]. Prp43 is another DExD/H-box RNA helicase that plays an important role in the spliceosomal assembly pathway. It associates with Ntr1/Spp382 and Ntr2 to form the NTR complex which facilitates the disassembly of the spliceosome[118]. Interaction of Ntr2 with Brr2 and U5 suggests that it may be engaged in the loading of NTR complex to the spliceosome after the completion of both steps of splicing[119]. Once the helicase activity of Prp43 is activated by Ntr1, it mediates the release of the lariat intron after the completion of splicing, which suggested its use as a quality control check point within the spliceosomal assembly[120].

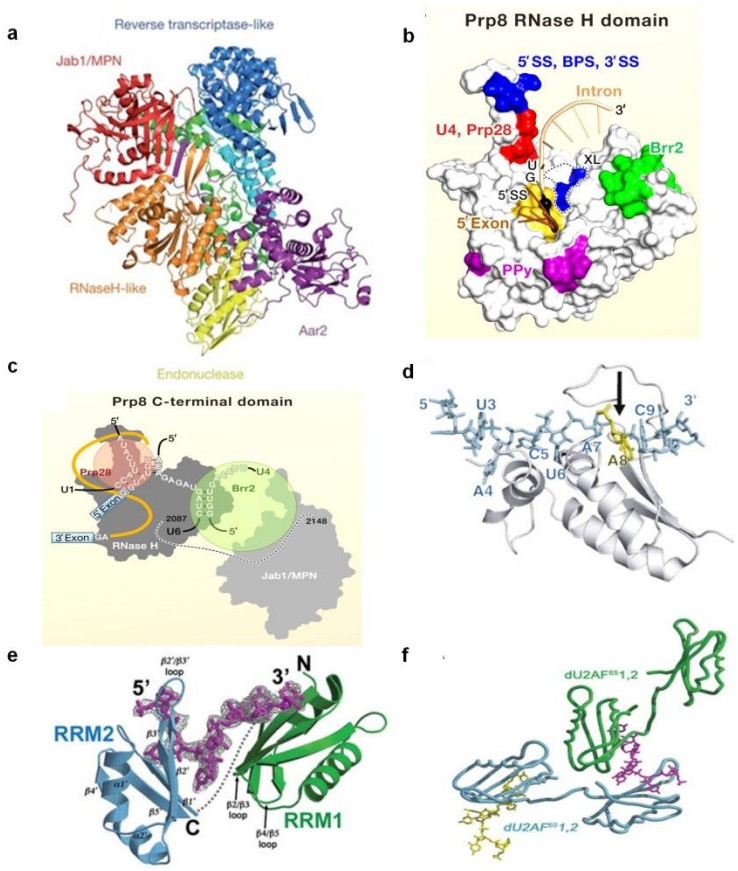

Figure 1.12: The proteome of spliceosomal complexes plays an important role in assembly and catalysis. (a) Crystal structure of a large fragment of Prp8 containing reverse transcriptase, endonuclease, MPN, and RNase H domains. Reprinted by permission from Macmillan Publishers Ltd: Nature (87), copyright (2013). (b)Surface view of RNase H like domain of yeast Prp8 interacting with the 5' SS. Other regions that interact with these two regions are shown with different colours. Reprinted from (89), copyright (2008), with permission from Wiley publishers. (c) A schematic representation of coordination activity of Prp8. Brr2, Prp28, 5' SS and exons are coordinated by RNase H domain and Jab1/MPN domains of Prp8 at the catalytic core. Reprinted from (1), Copyright (2009), with permission from Elsevier. (d) NMR structure showing binding of SF1 (grey) to the branch site sequence of RNA (cyan). Branch site adenosine is in yellow. Reprinted From (100) with permission from AAAS. (e) and (f) X-ray structure of U2AF65 RRM1 and RRM2 (grey) bound to a seven nucleotide polyuridine RNA (cyan) illustrating the early recognition of the polypyrimidine tract by U2AF. Interactions between U2AF and RNA can occur in two ways; (e) single protein interacts with two separate RNA molecules or (f) RNA interacts with RRM1 and RRM2 from separate molecules. Reprinted from (101), Copyright (2006), with permission from Elsevier

Another important protein involved in the spliceosomal assembly cycle is Prp19, which is associated with another set of 7 proteins to form Prp19/CDC5 complex in human or 8-11 proteins to form the NTC complex in yeast[121]. NTC plays an important role in the stabilization of U5, U6 interactions with the 5' SS and exonic region after dissociation of U1 and U4[39]. It also facilitates the conformational changes of U6, such as removal of LSm ring and remodelling of U6 interactions with the 5' SS[39]. Another member of NTC, namely Cwc21 is bound to Prp8 and Snu114 directly, suggesting that Cwc21 may assist the interaction of NTC with U5 snRNP[122]. On the other hand human Prp19 is associated with CDC5, which is a large complex containing ~30 proteins that plays an important role in the second step of splicing[123].

Other than the above mentioned protein families or groups, numerous other proteins are involved in the spliceosomal assembly cycle, such as, Spp2, Cwc25 and Yju2, Slu7 and Prp18[95]. Spp2 is a protein factor that facilitates the loading of Prp2 onto the spliceosome[95]. The G patch region of Spp2 interacts with the C-terminal region of Prp2 ensuring their association[124]. It has been shown that Spp2 binds to the spliceosome prior to the first step and both Prp2 and Spp2 are dissociated together with SF3a and 3b[111]. Another protein factor Cwc22 has also been shown to be important at this point. Although Cwc22 is not important for the binding of Prp2 to the spliceosome, in the absence of it, Prp2 is dissociated from the spliceosome upon ATP hydrolysis without releasing SF3a/b[125]. After the release of SF3a/b, another two proteins called Yju2 and Cwc25 bind to the spliceosome which promote the first step in splicing. Yju2 is shown to be interacting with NTC complex, whereas Cwc25 binds to the spliceosome after the role of Prp2 but only in the presence of Yju2[126,127]. After the first transesterification

reaction, Prp16 facilitates the removal of Yju2 and Cwc25, remodelling the spliceosome for the second step[128]. This structural rearrangement allows the binding of another set of proteins in the following order Slu7 and Prp18 and then Prp22, which promotes the second catalytic reaction in an ATP-independent manner[95]. Studies have shown that these three proteins interact directly with the 3' SS whilst Slu7 and Prp22 interact with the intron as well[113,116]. These proteins along with Prp8 and other second step factors stabilize the interactions between the U5 stem loop I and exons promoting the second catalytic reaction[95].

1.7.7: Spliceosomal sub complexes

During the spliceosomal assembly cycle, U snRNP and associated proteins form several sub complexes called E, A, B, B_{act}, B* and C. Through the formation of each of these sub complexes, other than the snRNP proteins, numerous non-snRNP associated protein factors join and release from the complexes.

The E complex is formed once the U1 snRNP is bound to the 5' SS and other non-snRNP proteins, SF1 and U2AF bind to the BS and polypyrimidine tract respectively[1]. In the E complex, all conserved regions of pre-mRNA are recognized by U1 and the other non-snRNP proteins. Upon binding of U2 snRNP to the branch site, the A complex (pre-spliceosome) will be formed. Assembly of the A complex is ATP dependent and also requires the polypyrimidine tract and a functional 3' SS in mammals and 5' SS and BS in yeast[1]. During the transition from E complex to the A complex, the E complex-bound factors need to be dissociated and ATP may have to be consumed for this.

After the formation of the A complex, the spliceosome moves on and forms the B complex. There are several intermediates that have been isolated and grouped as B complexes, although it is not clear whether all these complexes are functionally active[1,44]. The most accepted B complexes are the initial B complex, intermediate B_{act} complex (BΔU1) and the catalytically active B* complexes[44]. The B complex is formed once the U4/U6.U5 tri-snRNP is recruited on to the pre-mRNA[5]. U5 associated Prp8 protein plays a crucial role in the transition from the A complex to B complex by interacting with pre-mRNA and other proteins[82,122]. One of the major structural rearrangements that occurs within spliceosomal assembly is the transition of the B complex to the B_{act} complex[1,5,42,129]. During this transition, U1 and U4 snRNPs are dissociated from the spliceosome complex, leaving only U2, U6 and U5 on the pre-mRNA[5,42]. This exchange causes a dissociation of 35 proteins and enrolment of 12 other proteins in yeast, including entire U1 snRNP, all U4/U6 associated proteins and several U5-specific proteins (Figure 1.13)[5,129].

The B* complex is the catalytically activated form of the spliceosome. It is very similar to B_{act} in composition but several helicases needed for the structural rearrangements are present in B_{act}, but not in the B* complex[44]. One of the major structural arrangements, in B* is the positioning of U2, U6 and U5 for the catalysis. U5 is placed in a way that it can interact with 5' and 3' exons through its SLI structure[79,130]. Similarly, U6 is positioned in a way that its conserved regions can interact with the 5' SS, which allows U2 and U6 to extensively base pair with each other to form the catalytically active conformation[87]. Prp19/CDC5 (NTC in yeast) plays a major role in the transition from Bact to B* by promoting the removal of the U6 snRNP LSm ring and the

stabilization of U6 and U5 snRNAs[131,132]. Also during this conversion, U2 snRNP associated SF3a and SF3b proteins are dissociated freeing the branch site adenosine for the catalysis[133].

After the first step of splicing, the spliceosome adopts the C complex, which consists of the 5' exon and the lariat intermediate[44]. This step has fewer conformational rearrangements compared to the activation step, where only two proteins (Spp2/Prp2) are released and nine are added (Figure 13)[5,129,134]. Previously reported human C complex structures indicate that it is made up with three main domains namely; a large ovoid-shape body domain, an arm overhang from the body domain and a small domain slightly detached from the rest of the structure[135]. Purified human spliceosomal C complexes revealed the major components of this complex including Prp19/CDC5 proteins, Prp19-related factors and the U5 proteins, including Prp8 (Figure 1.13)[1,129]. It has also been shown that during C complex formation some of the SF3a and SF3b proteins are removed suggesting the disruption of U2/BS interactions[129]. Novel cryo-EM structure of yeast spliceosomal C complex at 3.6 Å resolution has shed a light on the splicing field[136]. C. Yan and co-workers have been able to resolve a spliceosomal C complex particle containing U2 and U5snRNPs, NTC, U6 snRNA and an RNA intron lariat. This structure illustrates an asymmetric conformation consisting a triangular central domain connect to a head domain and two arms; arm I and II. The head and arm II domains are similar to the head-like and arm overhang in previously published EM structure of U5.U2/U6 complex[137]. Moreover, this structure indicates that the catalytic core is located within the central triangular-shaped domain and it is ~100 Å away from the head region' either arm or the edge of the central domain which contains the 3' end

of the U5 snRNA. The distances between two arms and head and far most corner of the central body seems to be large (~320 Å) suggesting the importance of large, extended arrangement of spliceosome in proper splicing of pre-mRNAs with different lengths and sequences. The aforementioned distinct arrangements of subcomplexes along with their snRNA components are vital for spliceosomal assembly and catalysis.

1.8: Splicing by a minor class spliceosome

Two types of spliceosomes have been reported to be present in organisms: the U2-dependent major and U12-dependent minor spliceosomes[38,138]. The major spliceosome splices the highly abundant U2-type introns, whereas the minor spliceosome splices the less abundant U12-type introns[139]. As an example; in metazoans, U2-dependent major spliceosomes responsible for the splicing ~99.5% of all introns, whereas U12-dependent minor spliceosomes only splice the remaining ~0.5%[38,138,140]. This class of introns is found to be present in most eukaryotes including plants, animals and some fungi. On the other hand, some studies have shown that U12-type introns are absent in common model organisms such as *Saccharomyces cerevisiae*[38,141-144].

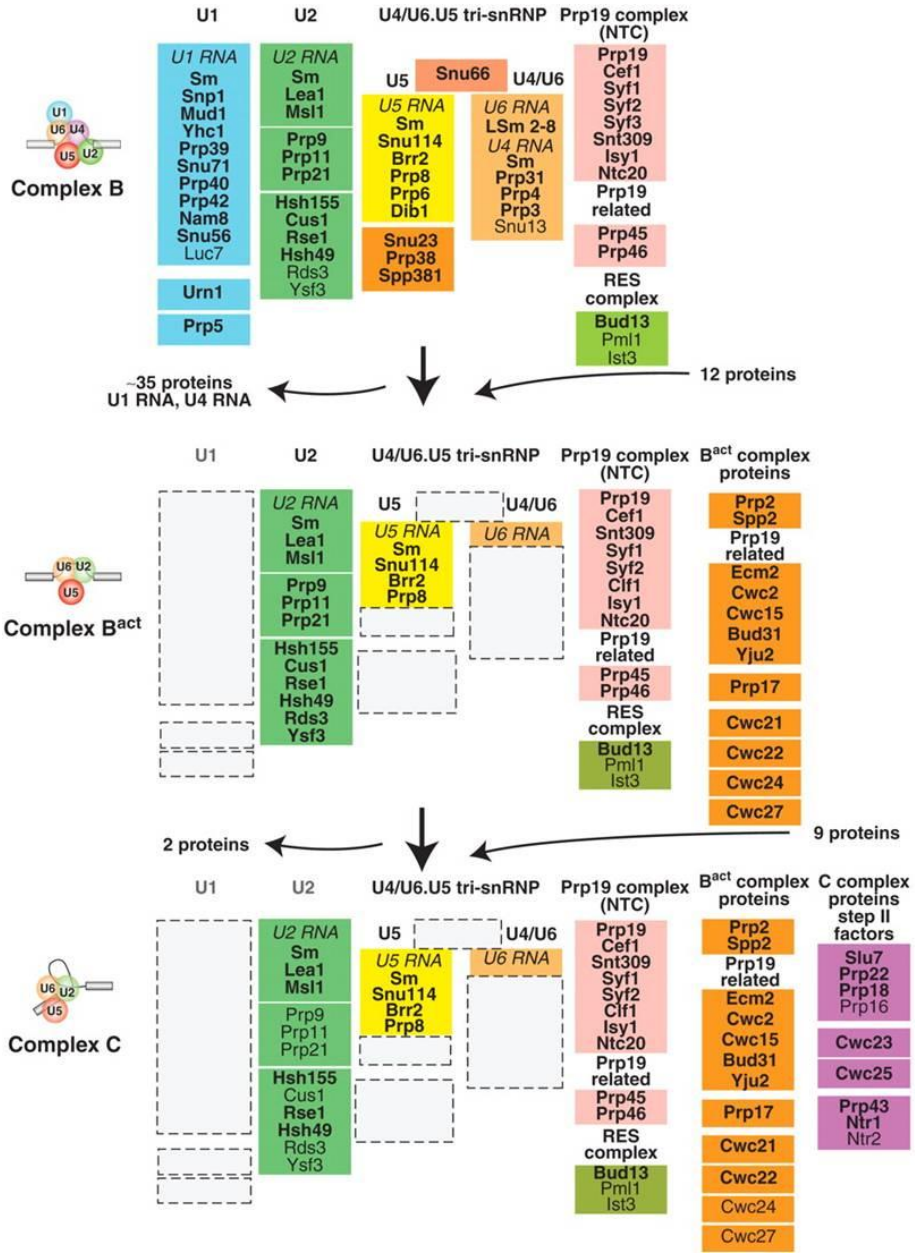

Figure 1. 13: Composition changes in the yeast spliceosomal subcomplexes during the assembly process. Protein compositions in complex B, B_{act} and C obtained from mass spectrometry are shown and are grouped according to their association with snRNP or snRNP subcomplexes. Reprinted from (124), Copyright (2009), with permission from Elsevier

Two types of spliceosomes have some characteristic differences as well as similarities. Two introns have different splice sites and branch site consensus sequences. Instead of GT and AG at the 3' and 5' ends in U2-type introns,[38,141,145] the majority of the U12-type introns have AT and AC at the 3' and 5' ends respectively (Figure 1.14a). Therefore these introns are termed as "ATAC introns"[23,141,145]. The minor spliceosome consists of four unique ribonucleoproteins named as; U11, U12, U4atac and U6atac (analogous to U1, U2, U4 and U6 respectively), and U5; which is common for both spliceosomes (Figure 1.14b)[38,146]. Minor spliceosomal assembly cycle follows the steps as same as the major spliceosomal assembly pathway, except the first step of assembly cycle (Figure 1.14c).[143] In minor spliceosome, during the formation of complex A, U11 and U12 form a di-snRNP before binding with mRNA whereas in major spliceosome they bind separately (Figure 1.14c)[38,143,145,146]. The rest of the steps in the assembly cycle as well as the intron removal mechanism are identical in both spliceosomes[38].

Previous studies have suggested that regardless of sequence differences, the analogous snRNAs in two spliceosomes can form very similar secondary structures with similar conserved regions, which suggest similar functional roles in spliceosomal assembly and catalysis (Figure 1.14b)[38,147-149]. In addition, the sub-complexes in both types of spliceosomes can adopt conformations that are very similar to each other. For example, the U12/U6atac complex is similar to the U2/U6, and the U4atac/U6atac complex similar to U4/U6 (Figure 1.14c)[143,146-148]. These structural similarities indicate that the minor spliceosomal snRNA complexes can also be highly dynamic and undergo important conformational changes during the assembly cycle same as the major

spliceosomal components. Other than sharing similar snRNA components, both spliceosomes also share most of the snRNP-associated proteins[140,150-152].

Despite its lower abundance, U12-type introns are present in many genes required for essential cellular functions such as DNA replication and repair, transcription, RNA processing and also found in vesicular transport and voltage-gated ion channel activity[140,141,143,146,153]. Surprisingly, it has been shown that U12-type introns are almost absent in genes related to metabolism and biosynthetic pathways[144,146,153]. Recently, Younis and co-workers have shed new light on the significance of minor spliceosome in cells[154]. They have shown that the minor spliceosomal U6atac is highly unstable in human cells under normal conditions, suggesting that U6atac could be a rate limiting factor for the splicing by minor spliceosome. In contrast, they have found a dramatic increase in the level of U6atac in the presence of a protein kinase, which activates in response to stress. This has shown to activate the splicing by minor spliceosome, resulting in activation of various cellular processes. With these findings, they have proposed that minor spliceosome acts like a switch that helps cells to adapt to the stress and also helps to accelerate the translation of the transcript present in the cells, which increases the protein production under the stress conditions[154].

Marking its significant involvement in cellular processes, previous studies have shown that defects in human minor spliceosomal assembly due to mutations in U4atac snRNA cause the Taybi-Linder Syndrome (TALS), which is a rare autosomal recessive development defect[155-157]. All these facts point out the importance of studying the U12-dependent spliceosome in detail.

Although many studies have been performed on the minor spliceosome, the structural rearrangements of snRNA complexes and similarities and differences between major and minor spliceosomal complexes have not yet been revealed.

1.9: Structural rearrangement of snRNA complexes during spliceosomal assembly

Five snRNAs join and release at different steps in the spliceosomal assembly cycle along with their associate proteins. The arrangement of RNA-RNA interactions within this process is highly complex and largely dynamic[5,42]. The snRNAs form well studied secondary interactions, as well as, poorly understood tertiary interactions. During spliceosomal activation and catalysis most of these interactions are restructured making it more complicated, suggesting that structural changes of snRNA complexes play an important role in assembly and catalysis[42].

Out of five snRNAs, U6 (U6atac in minor) is the most dynamic as it exhibits extensive structural rearrangements during spliceosomal assembly[158]. Before its involvement in the spliceosomal assembly, U6 snRNP can be found as a free particle (Figure 1.15)[159]. First, it base pairs with U4 and forms the U4/U6 duplex, which enters to the spliceosomal assembly cycle (Figure 1.15)[159]. Later, U4 dissociates from U6 and U6 snRNA base pairs with U2 to form the catalytically active complex (Figure 1.15)[160,161]. At the end of two catalytic steps of splicing, U6 needs to be recycled as a free snRNP, which will then be involved in another cycle of splicing (Figure 1.15). Therefore, this dynamic nature of U6 is believed to be very crucial for the spliceosomal assembly and catalysis.

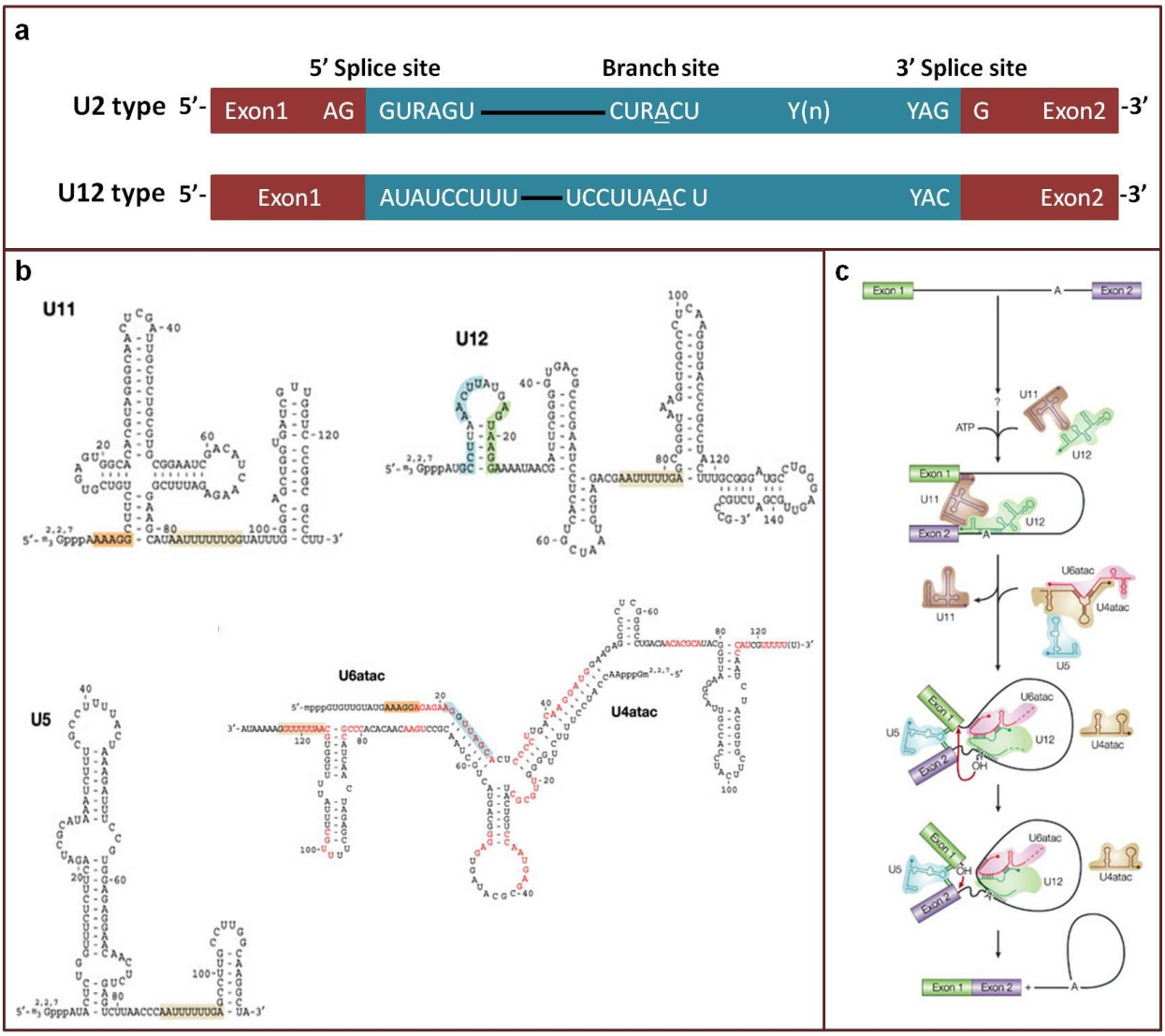

Figure 1.14: Overview of the less abundant minor spliceosome. (a) Consensus sequences at 5' SS, 3' SS and BS of U2 type (major spliceosome) and U12 type (minor spliceosome) introns are showing. Y is pyrimidine, R is purine and Y(n) is polypyrimidine tract. (b) Sequences and secondary structures of minor spliceosomal snRNAs of human; U11, U12, U5 and U4atac/U6atac. These share many similarities with the secondary structures of major spliceosomal snRNAs. (c) Assembly and catalysis pathway of minor spliceosome, which is very similar to major spliceosome. Reprinted by permission from Macmillan Publishers Ltd: Nature reviews: Molecular cell biology (136), copyright (2003).

1.9.1: U6 snRNP

Free U6 snRNP consists of U6 snRNA, seven LSm proteins (LSm2-8) bound at the 3' end and Prp24, a U6 associated chaperone[159]. It has been suggested that in yeast this LSm ring together with Prp24 act as a chaperon complex, which is involved in the conformational rearrangements of the U6 complexes[162]. The secondary structure of yeast U6 snRNA within the free U6 snRNP particle is composed of 5' stem-loop, 3' stem-loop, named as intramolecular stem-loop (ISL), a stem extension named as telestem and a large asymmetric internal loop between 3' stem-loop and telestem (Figure 1.15)[101,163]. Recently, a crystal structure of U6 has been revealed, which is showing that U6 consists of an ISL structure containing AGC triad, a telestem region which is perpendicular to ISL and a large asymmetric internal loop linking these two, similar to previous NMR structures[164]. U6 also contains a pyrimidine-rich sequence at the 3' end, which is essential for the binding the LSm proteins[49]. Although the 5' stem loop is shown unchanged and less important during the splicing cycle, the ISL region has shown to be very dynamic and contains highly conserved dynamic bulge U80. Highly conserved ACAGAGA region and the AGC triad fall within this asymmetric bulge (Figure 1.15)[165] .

1.9.1.1: U6 snRNA – the most structurally dynamic spliceosomal RNA

The U6 snRNA is highly conserved over most of the organisms in both its sequence and its size[158]. There are many structural and functional similarities between the group II intron and the RNA-based catalysis of the spliceosome.[166,167] The catalytically important domain (D5) in the group II intron has many similarities with the U6 snRNA. The

intramolecular stem loop (ISL) in U6 is similar to D5 in the group II intron (Figure 1.16a,b)[167]. They both share similar conserved residues, such as the metal ion binding bulge and the AGC triad (Figure 1.16a,b)[167,168]. Sharing common structural domains in the active site is consistent with the proposed function of the spliceosome as a ribozyme and the structural similarities with group II catalytic domain have lead to suggest that U6 is crucial for catalysis[169].

A phylogenetic study has shown that a 39 nucleotide sequence within the U6 core is conserved over almost all of the U6 snRNA sequences, whereas the flanking sequences are highly degenerated[170]. Strikingly, most of the point mutations within this region show either full or partial inhibition of splicing *in vitro* and full or partial lethality in yeast, revealing the crucial function of this core region in catalysis[170]. Interestingly, three highly conserved regions of U6; the dynamic bulge at U80, the AGC triad, and the ACAGAGA loop which are suggested to make a base-pairing network between U2 and the pre-mRNA fall within this core region, implying its importance in the formation of the catalytic core of the spliceosome (Figure 1.15 and 1.16b and d) [171,172]. Minor spliceosomal analogue, U6atac also contains these three conserved regions (Figure 1.16c)[143,148,173]. Some genetic screening studies have also shown that mutations in the ACAGAGA region and the AGC triad are lethal or temperature sensitive, suggesting they play major roles in the catalysis[87,161,174]. During the formation of the activated complex, U6 replaces U1 and interacts with the 5' splice site via the ACAGAGA loop[109,175]. The AGC triad of U6 also plays a vital role in splicing by interacting with U2 and may also be involved in the catalysis as well[176]. Previous studies have shown that mutations at the first two nucleotides of the AGC triad are lethal and the mutations at

the third position result in strong growth defects, implying its significance in splicing[87,176]. The dynamic bulge within the highly conserved ISL region contains a U nucleotide (U80) which has been shown to be important for coordinating Mg^{2+} ions necessary for the first step of catalysis[35,169,177]. An NMR study has shown that this dynamic bulge and the adjacent C67-A79 can act as a dynamic hinge, where the stabilized C67-A79 pair by protonation of A79 causes flipping out of U80 into the major grove, and deprotonated A79 keeps U80 stacked inside the ISL[177]. The binding of Mg^{2+} ions is favoured by U80 when it's stacked within the ISL, suggesting this arrangement is necessary for the catalysis[169].

1.9.2: U4/U6 complex prevents U6 from premature activation

U6 snRNA base pairing with the U4 snRNA is followed by the interaction with U5 before entering into the spliceosomal assembly cycle. U4/U6 duplex form a highly conserved Y shaped structure, which consists of two intermolecular helices namely stem I and II and highly conserved 5' stem-loop structure of U4 (Figure 1.15)[69]. During the U4/U6 duplex formation, U6 snRNA undergoes drastic structural rearrangements. Mainly the ISL region of U6 snRNA (3' stem loop) unwinds and base pairs with U4 to form stem II (Figure 1.15)[170]. Similarly the nucleotides within the internal loop region base pair with U4 to form the stem I (Figure 1.15). Overall, previous studies have shown that, all the highly conserved regions within U6 snRNA, which are suggested to be essential for the catalysis, are base paired with U4, making them unavailable for the reactions[161].

In order to be activated and to be involved in the catalysis, first, U6 needs to be unwound from U4 and then base pair with U2 to form the U2/U6 complex. Hence, it has

been proposed that U4 may act primarily as a negative regulator, sequestering U6 in a catalytically inert conformation[161]. Taken together, despite its absence in the activated complex, U4 also plays an essential role in spliceosome assembly by preventing U6 from premature activation and formation of unfavourable conformations.

The *in vitro* splicing complementation assays have shown that affinity purified U4 snRNP and U6 snRNA couldn't assemble into U4/U6 duplex without nuclear extract complementation, suggesting the requirement of other protein factor or factors for the formation of U4/U6 duplex. Also the formation of a very stable duplex in vitro (T_m ~ 53 ^oC) by extensive base pairing between U4 and U6 proposes an activity of ATP dependent RNA helicase or a RNA chaperone on the unwinding process of U4/U6 complex[178]. Moreover, an RNA binding protein may be involved in the stabilization of U4 and U6, which is energetically less favourable when compared to that of duplex[179]. A mutational analysis study indicates that the G14C mutant, where G is mutated to C at the 14th position of U4 snRNA, has shown the most dramatic cold sensitive phenotype when compared to other U4 mutations[179]. Also, this mutant has shown a T_m of ~ 37 ^oC, suggesting that the base pair G14 – C67 plays an important role on the stabilization of U4/U6 helix[179].

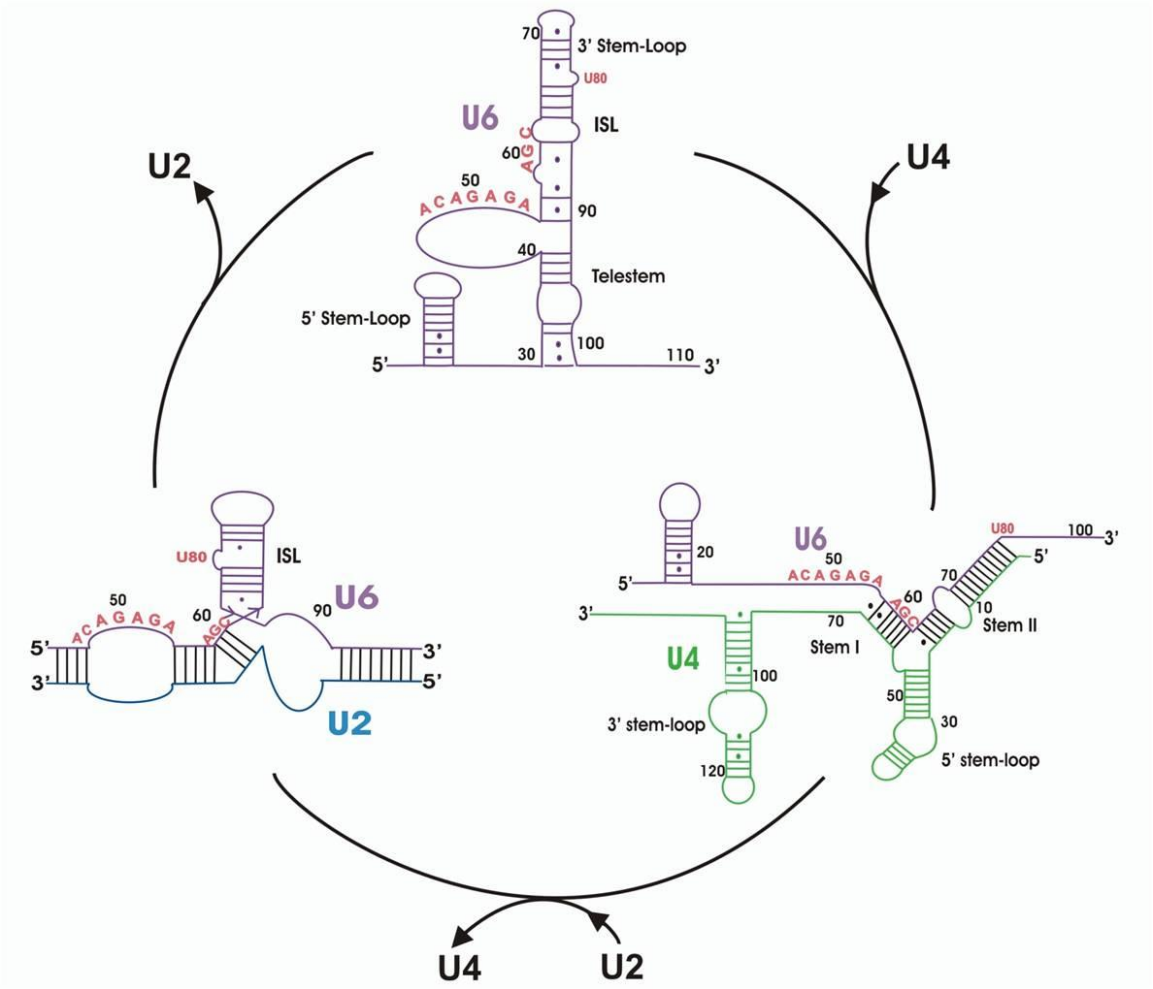

Figure 1.15: A schematic representation of current models for the secondary structure of U6 in free U6 snRNP, in U4/U6 duplex and in U2/U6 duplex. Highly conserved regions of U6 are shown in red; ACAGAGA loop, AGC triad and U80.

1.9.3: U2/U6 complex makes the catalytic core of the spliceosome

Although splicing requires the assembly of all five snRNPs, only U2, U5 and U6 have been shown to be present in the catalytically active spliceosome[5]. Previous studies have also shown that U5 loop I is functionally dispensable for both steps of splicing[79,180]. All these findings have led to the suggestion that the U2/U6 complex is responsible for the catalysis and it acts as the active site of the spliceosome. Similarly, in the minor spliceosome U12/U6atac forms the catalytic core[143,146,148]. This has been demonstrated in studies done with protein-free U2/U6 complexes, which are able to catalyze reactions similar to both steps of splicing[115,181]. U6 needs to dissociate from U4 and base pair with U2, in order to form the active spliceosomal complex. During this transition, the structure of U6 undergoes dramatic rearrangements, where the regions participate in base pairing with U4 open up and reform the 3' stem-loop (ISL) as in free U6 snRNA (Figure 1.15)[43,160,161]. The ISL domain contains AGC triad and metal ion coordinating dynamic bulge (U80), which have been shown to be important for the first step of catalysis. Sequences upstream to ISL are basepaired with the 3' end of U2 to form the helix I and helix III (may not be found in yeast) in U2/U6 complex, separated by a bulge containing highly conserved ACAGAGA domain (Figure 1.16d and e)[171,182]. This ACAGAGA loop is known to interact with the 5' SS during the spliceosomal activation (Figure 1.16d)[172]. On the other hand, the downstream sequence of U6 base pairs with U2 to form the helix II (Figure 1.16d and e)[87,182].

1.9.3.1: Secondary structure of U2/U6

The catalytically important conformation of the U2/U6 complex has been debated for many years.[5] According to genetic studies (Guthrie lab[87]), NMR studies (Butcher lab[183]) and structural probing studies (Luhrmann lab[184]), the AGC triad of U6 base pairs with a region in U2 and forms a three-helix structure consisting of helix Ia, Ib, and III (Figure 1.16d). On the other hand, mammalian genetic studies (Manley lab[182]) and NMR studies (Butcher lab[171]), have proposed a four-helix structure, in which the AGC triad base pairs with U6 itself and forms an extended ISL (Figure 1.16e). These two conformations may be present at different states of spliceosomal activation. Minor spliceosomal U12/U6atac complex also forms a similar secondary structure to U2/U6 complex, which consists of helix Ia and Ib, helix III and ISL region within U6atac corresponding to those regions in U2/U6 complex (Figure 1.16f).[38,143,146,148] However since the 5' end of U12 is truncated when compared to U2, it's unable to form the helix II structure as in U2/U6 complex (Figure 1.16f).[38,143,146]

1.9.3.2: U2/U6 adopts multiple conformations

Single-molecule fluorescence resonance energy transfer (smFRET) studies carried out in our lab have shown that the minimal U2/U6 complex from yeast can adopt at least three distinct conformations in dynamic equilibrium, corresponding to three distinct FRET states, 0.2, 0.4 and 0.6, as a function of Mg^{2+} concentration (Figure 1.17a)[185]. At zero or very low Mg^{2+} concentrations, the U2/U6 complex adopts a conformation corresponding to a high FRET state (0.6), whereas at high Mg^{2+} concentrations, it forms the low FRET (0.2) conformation[185]. The intermediate FRET state (0.4) corresponds to

an intermediate conformation. The high and low FRET states are assigned as four-helix structure and three-helix structure respectively (Figure 1.17a).

In three-helix structure, distance between ACAGAGA loop and the ISL is increased due the formation of helix IB , resulting in a low FRET state (Figure1.17a). Six-fold mutations that prevent formation of helix IB and A91G mutation which stabilizes the helix IB have shown a low FRET population, confirming three-helix structure gives rise to 0.2 FRET state (Figure 1.17b and c)[185]. On the other hand, in four-helix structure, AGC triad base pair with U6 bases to form an extended ISL and ACAGAGA loop bring closer to ISL resulting in a high FRET (Figure 1.17a).

Furthermore, studies have shown the formation of base-triple interactions within U6 snRNA, involving the ACAGAGA loop, AGC triad and U80 which brings ACAGAGA loop closer to ISL similarly in group II intron (Figure 1.16a and b)[186,187]. It has been proposed that these interactions are vital for the catalysis, since they bring all the components together. At low Mg^{2+} concentrations, U80 tends to flip out and favours the formation of the base triple interactions, which bring the ACAGAGA loop closer to U80 and results in a high FRET state (Figure 1.17a). In contrast, with higher Mg^{2+} concentrations, U80 stacks within the U6 ISL and disrupts base triples, leading to a low FRET state (Figure 1.17a).[186]

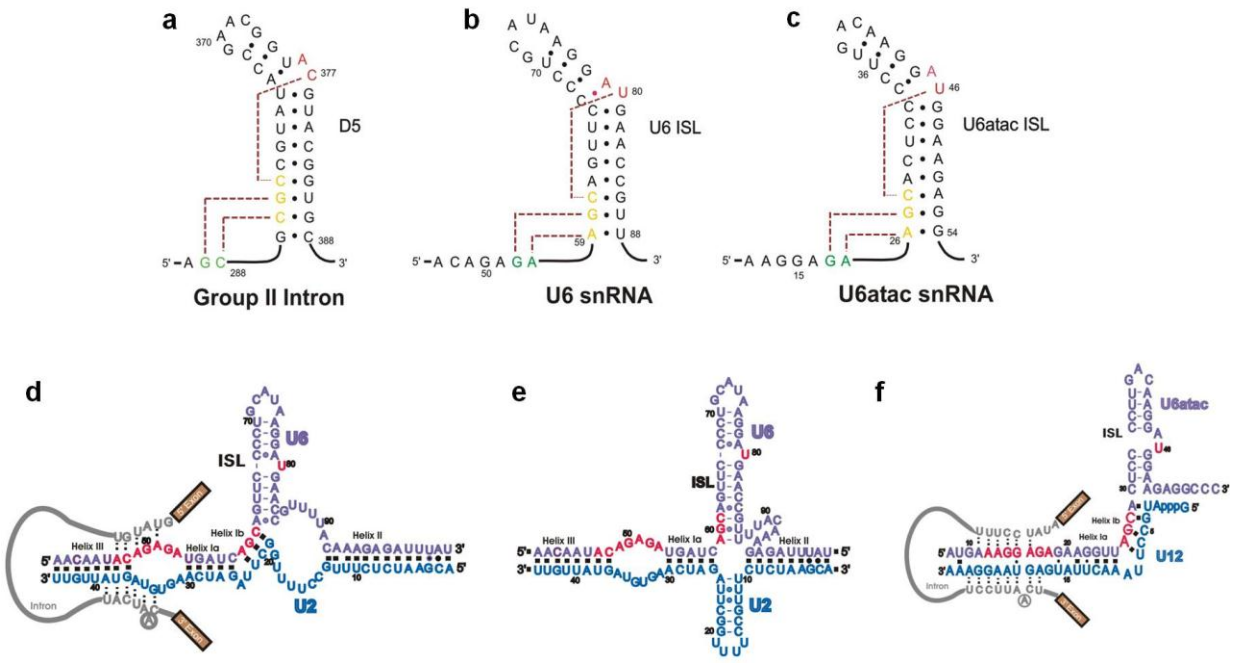

Figure 1.16: Structural similarities between Group II intron, U2/U6 complex and U12/U6atac complex. The catalytic triad and the metal ion coordinating bulge regions are present in both (a) domain 5 of group II intron, (b) ISL domain of U6 and (c) ISL domain of U6atac. Both group II intron and U6 snRNA have shown to form base triple interactions in a similar way. Similarities between U6 and U6atac structure and sequence suggest formation of base triple interactions similar to U6 as shown in (c). Major spliceosomal U2/U6 complex undergoes structural dynamics. (d) The proposed three-helix junction, where the invariant AGC triad of U6 forms intermolecular base pairs with U2. (e) The proposed four-helix junction, where AGC triad forms intramolecular base pairing with U6 resulting an extended ISL. (f) U12/U6atac complex present in the catalytic core in the minor spliceosome and shares many similarities with their correspondents; U2 and U6. This figure is adapted and reproduced from (160, 176 and 178).

These results support the notion that the 5' splice site and the branch-site adenosine are juxtaposed, and bring the metal-ion binding site closer for the catalysis.[172,180] Overall, the dynamic nature of the U2/U6 complex between different structures may play a vital role in different stages of catalysis.[5] The structural similarities between two types of spliceosomal complexes suggest that minor spliceosomal U12/U6atac complex may also adopt similar conformational dynamics and act similarly on the catalysis as in major spliceosomal U2/U6 (Figure 1.16d and f). Also, sequence and structural similarities between U6 and U6atac suggest that U6atac can also form base triple interactions similar to U6 snRNA (Figure 1.16c).

1.9.4: Structural rearrangements at the catalytic core

During the transition from the first to second step, the active site needs to be rearranged in order to remove the first step products and juxtaposes the second step components[188]. As an example, before the second step the lariat intermediate needs to be removed and the 3' SS has to be placed in the active site. Thus it has been proposed that the active site may toggle between two distinct conformations, one promotes the first step and other promotes the second step of catalysis[188]. Similar to ribosomal conformational switch between open and close states, spliceosome may also undergo transition between two structures, where some interactions are broken and reformed. Many studies have shown that the structural rearrangement occurs in the catalytic core during the activation of the spliceosome and the first step of catalysis.

Upon activation, U1 and U4 are released from the spliceosome and U2, U6 and U5 interact with the pre-mRNA to facilitate catalysis. During activation, the ACAGAGA loop of U6 interacts with the 5' SS by displacing U1, which is important for 5' SS recognition

and initiation of the first step (Figure 1.18). U2 stays bound with the BS as in the pre-spliceosome. Taken together, the base pairing network between U2 and U6 in activated spliceosome allows ACAGAGA loop and ISL to come closer and form tertiary interactions. As a result, both 5'SS and BS move closer to the metal ion binding site (U80), which is crucial for the first step of splicing (Figure 1.18) [5,172,180]

Studies have shown that U2 snRNA also exhibits structural dynamics similar to U6. The stem II of U2 adopts two conformations; stem IIa and stem IIc proposed to form at different stages of spliceosomal assembly (Figure 1.10a)[57,58]. In the pre-spliceosome U2 adopts stem IIa conformation and interacts with BS, whereas during the first step of splicing U2 interacts with BS through stem IIc structure[57,58]. Also, it has been shown that, after the first step, U2 reforms the helix IIa conformation which is involved in the second step of splicing[57,58]. U5 helps to join 5' exon to the spliceosome after the first step and form interactions with the exon nucleotides downstream to 3' SS (Figure 1.18)[78,189,190]. Furthermore, it is involved in the alignment of the two exons for the second step of catalysis (Figure 1.18)[79,190]. Structural rearrangements, how all the components interact within the second step of splicing such as how 3' SS position at the active site, where it comes from, what happen to U2/BS interaction are not yet clear and further studies need to be done.

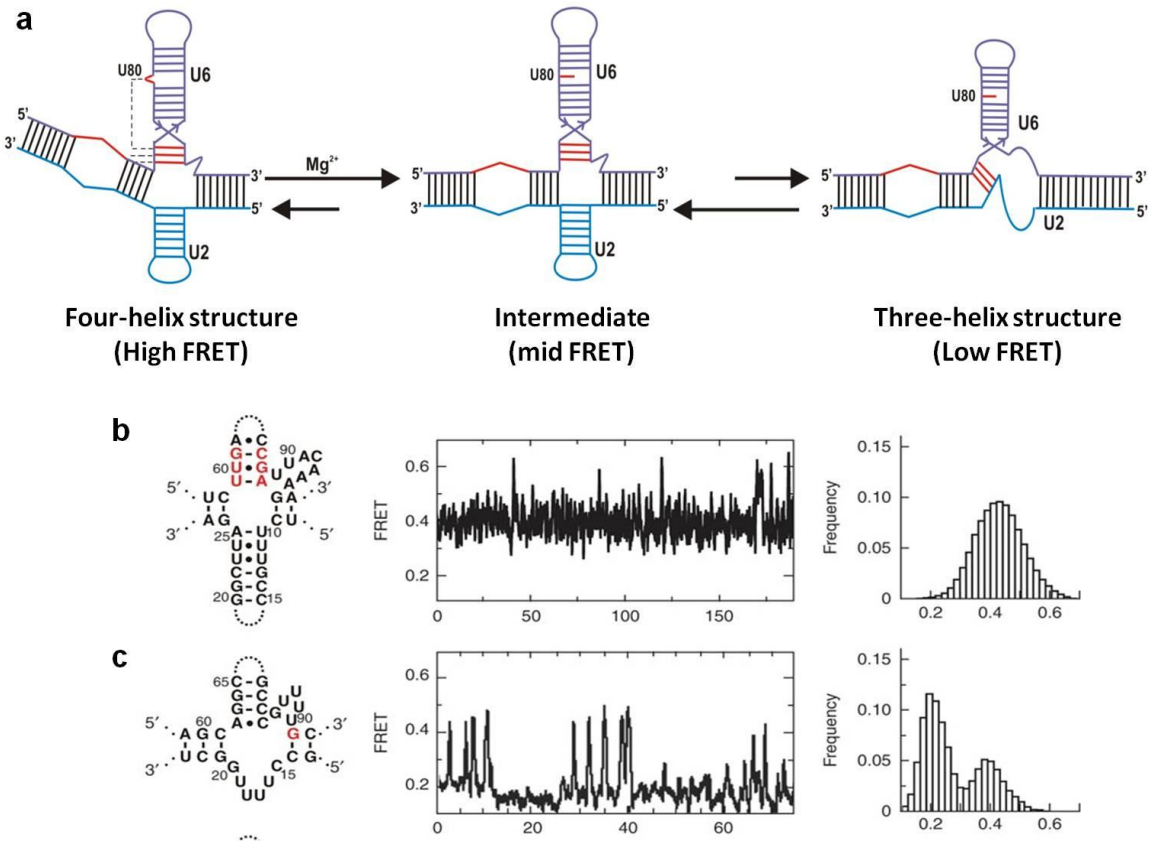

Figure 1.17: Spliceosomal U2-U6 snRNA complex adopts multiple conformations in dynamic equilibrium. (a) In the absence or low [Mg^{2+}], the U2/U6 complex adopts a high FRET (0.6) conformation through the formation of base-triple interactions (blue dash lines) between the highly conserved ACAGAGA loop, AGC triad, and U80 (red). Binding of Mg^{2+} induces a conformational change in the U2/U6 complex by disrupting the base-triple interaction and eventually stabilizes a low FRET (0.2) conformation via an intermediate (corresponding to 0.4 FRET).[185,186] U2 and U6 sequences are shown in blue and purple respectively. Schematic representation of different U6 mutations and resulting time trajectories and corresponding FRET histograms are shown. (b) Six-fold mutant of U6 snRNA which destabilises the low FRET conformation. (c) A91G mutant which stabilises the low FRET conformation. Reprinted by permission from Macmillan Publishers Ltd: NSMB (177), copyright (2009).

After the completion of exon ligation to form the mature mRNA, all the components at the catalytic core need to be dissociated from each other in order to engage in another cycle of splicing. Although previous studies have proposed a role for Brr2[85] and Prp43[120,191] in the dissociation of spliceosomal components at the end of splicing cycle, the exact mechanism of how U2/U6 and U5 dissociate from each other is still poorly understood. Furthermore, some studies have suggested that other than the aforementioned helicases there could be some other protein factors involved in the dissociation of snRNAs after the completion of splicing. One possible candidate could be Prp24; U6 snRNP associated chaperone, which is known to be promoting structural rearrangements of U6 during the assembly cycle and thus proposed as a recycling factor.

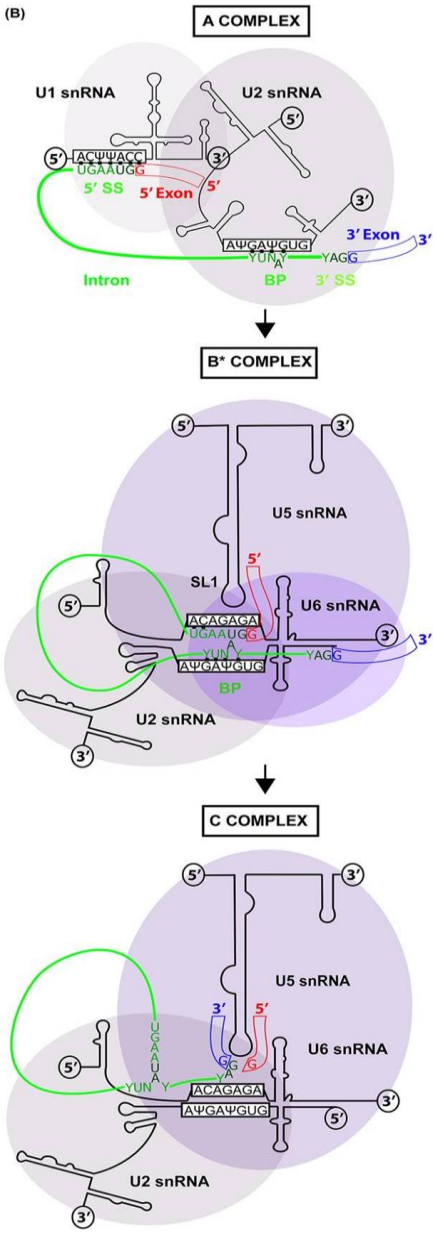

Figure 1.18: Interactions between U snRNA and premRNA in A, B* and C complexes. Protein component of U snRNPs are shown as purple spheres, U snRNAs in black lines, 5' exon in red, 3' exon in blue and intron in green. In A complex, U1 interact with 5' SS and U2 with BS. In complex B*,U2 and U6 base pair with each other. U6 and U5 recognise the 5' exon-intron junction, whereas U2 binds with BS. In C complex, U5 SLI interacts with both 5' and 3' exons facilitating the alignment of two exons for the ligation. Reprinted with permission from (42). Copyright (2012) American Chemical Society.

1.10: Prp24 is an U6 associated chaperone

The spliceosomal protein, Prp24 plays a major role in the spliceosomal assembly. Prp24 is thus an essential component of the U6 snRNP and it is considered as the RNA chaperone of U6, since it does not hydrolyze ATP[160,192-196]. Studies on Prp24 have suggested that it helps U6 to remodel the catalytically important structure[179,197]. Prp24 accelerates the annealing of U4 and U6, allowing U6 to enter into the assembly pathway[102]. Prp24 binds to U6 as well as with Lsm ('like sm') proteins, another member of the U6 snRNP[193]. It has been shown that after formation of the U4/U6 complex, Prp24 leaves the complex and helps to re-anneal two snRNAs for the next cycle of assembly, suggesting its role as a recycling factor[102,198] Some studies proposed that Prp24 returns during the activation of the spliceosome, in which it is involved in unwinding of U4, allowing U6 to bind with U2 to form the catalytically activated complex.[101,199,200] Tight binding of Prp24 to U6, but not with U4, has led to the proposal that Prp24 may also play other roles in spliceosomal assembly, such as helping to stabilize free U6 until it is ready to bind with U2 to form the active complex, after helping to dissociate U4/U6. In other words, Prp24 may prevent the premature formation of the active conformation of U6 snRNA.[101,179,199]

1.10.1: Structure of Prp24

Prp24 is a 51-kDa protein, present in eukaryotes, from yeast to human, and consists of four RNA recognition motifs (RRMs); RRM1-4 and a C terminal sequence that interacts with the LSm proteins (Figure 1.19a)[100,192]. RRMs are a special arrangement of amino acids into a four stranded antiparallel β-sheets through two conserved RNP motifs and

present in many proteins providing extensive binding sites for single-stranded RNA sequences. RRMs-containing proteins are mostly involved in the post-transcriptional gene expression processes (i.e. mRNA processing, RNA export and stability) including PTB (polypyrimidine tract binding protein, 4 RRMs)[201,202], U2AF65 (2 RRMs)[106] and SRp20 (1 RRM)[203] and Prp24 (4 RRMs)[192]. Almost all the RRM domains are made up with ~ 90 amino acids and consist of canonical topology of β_1-α_1-β_2-β_3-α_2-β_4 forming a four stranded β sheets packed along with two α helices[204]. The number of nucleotides that binds with the RRMs are also varied ranging from minimum two to maximum of eight[204]. Previous crystal and/or NMR structures have shown that the first three RRMs of Prp24 adopt the canonical topology as described earlier (Figure 1.19b, left) [192], whereas the fourth RRM4 folds non-canonically, in which it has additional flanking α-helices compared to the canonical RRM-fold, thus it was named as occluded RRM (oRRM4, Figure 1.19b, right)[205]. Based on the previous crystal structures of Prp24-N123 (doesn't contain RRM4 and C terminal), residues in or close to all four β-strands of RRM1 interact with residues in or close to the β-strands 2 and 3 of RRM2[192]. On the other hand, residues in or close to the β-strands 1, 3 and 4 of RRM2 interact with residues in α-helices 1 and 2 of RRM3[192]. These results suggest that, despite the interacting residues being buried within protein-protein interactions, RRM1 and 2 are the mandatory domains to bind with U6 snRNA. Hence, either a dramatic structural rearrangement of Prp24 is required for the association of Prp24 with U6 snRNA, or the binding of U6 occurs through non-canonical interactions. A recent crystal structure of the full length Prp24 shows that a β-sheet of RRM1 is masked by the interactions with

RRM2 and β-sheet of oRRM4 is obstructed by two flanking α-helices, suggesting only RRM2 and 3 interact with U6 snRNA in a canonical way[164].

1.10.2: Binding of Prp24 to U6 snRNA

According to gel shift assays done by Kwan and co-workers, RRM1 and 2 are important for high-affinity binding of Prp24 to U6, and RRM3 and 4 may have a function in controlling the stoichiometry of Prp24 binding[206]. Previous studies have shown that the high-affinity binding site for Prp24 is mainly located between the nucleotides 45-87 of U6 snRNA[206]. Based on the early NMR chemical shift mapping studies, RRM2 interacts sequence specifically with the GAGA region in the ACAGAGA loop, and RRM1 destabilizes a 3' downstream weakly paired region (nucleotides 54-61: 86-91) of U6 snRNA.[100,205] NMR data on the Isolated RRM3 have shown that RRM3 binds to the U6 ISL region [48] whereas oRRM4 disrupts base pairing within the bases at the bottom of the U6 ISL.[205,207] According to these findings, unwinding of base pairs in U6 by RRM1 and 4 may allow U4 to base pair with U6, in order to form the U4/U6 complex.

A recently published crystal structure of Prp24 bound to U6 with A62G mutation has helped to fill the missing pieces of information on how Prp24 binds to U6, what structural requirements need etc (Figure 1.19c)[164]. Moreover, it provides new insight in to the role of Prp24 during the spliceosomal assembly cycle. Although there are many proteins containing multi RRMs, this study is the first to reveal a crystal structure for a protein containing more than two linked RRM motifs. According to this crystal structure, the internal bulge region of U6 creates an interaction network with RRM2, 3, oRRM4 and area prior to RRM1.

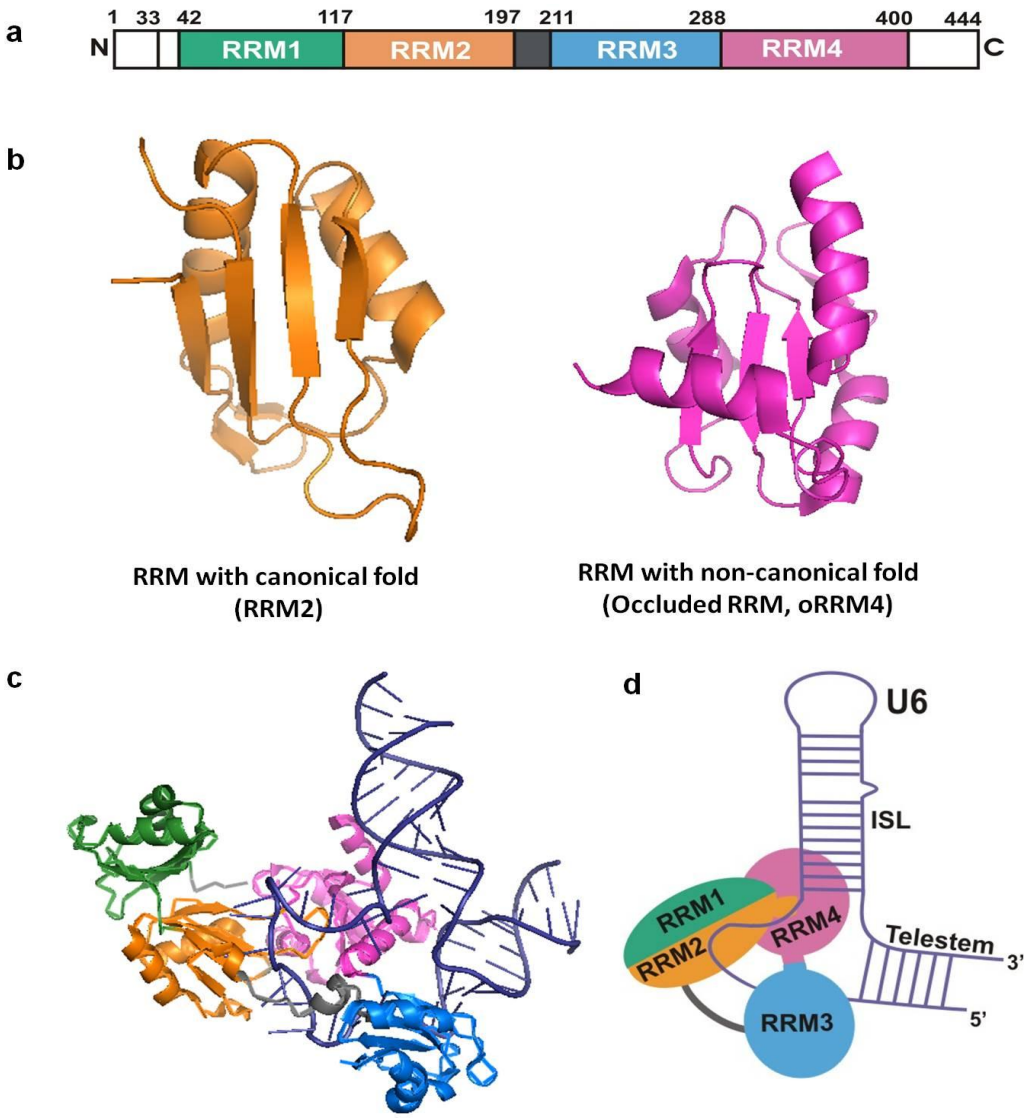

Figure 1.19: Structure of yeast U6-Prp24 complex. (a) Domain architecture of yeast Prp24. The first 33 and last 44 nucleotides (white) are absence in the construct used in the crystal structure. PDB: 4NOT (b) Prp24 consists of 3 canonical RRM motifs (RRM1, 2 and 3) and one non-canonical RRM, which were named as 'occluded RRM (oRRM4). This figure illustrates the differences between canonical and non-canonical RRM motifs. (c) Crystal structure of U6-A62G-Prp24 complex, with U6 RNA in purple and four RRMs of Prp24 which is coloured as in a. Figures were generated from PyMOL (PDB: 4NOT). (d) A cartoon of the interlocked RNP topology formed by Prp24. Colours are the same as in c. This figure is reproduced from (157).

According to this new crystal structure, RRM1 interaction with U6 belongs to an adjacent complex in the crystal. This interaction occurs via a wide electropositive groove enclosed by RRM1 as well as RRM2 and oRRM4 with U6 ISL. RRM2 binds to the nucleotides 46-58, which encloses the highly conserved ACAGAGA loop. In contrast to previous studies, RRM3 has been shown to interact with more upstream region of U6; nucleotides 39-44. Also compared to the interactions observed in the free protein, in the presence of U6, RRM3 makes 180^{o} rotation and 20 Å shift, which let RRM2 to bind with U6 (Figure 1.19c). This is in consistent with previous NMR studies showing that RRM3 does not interact with RRM2 in solution, and thus, does not interfere with RRM2 binding to RNA. In oRRM4, the occluded α-helices have been shown to interact with both ISL and telestem thus form non canonical contacts with double stranded RNA (Figure1. 19c). These differences in the binding modes could be a result of cooperativity in RNA recognition by multi RRM proteins when compared to the isolated protein domains.

More interestingly, this study has proposed an interlocked RNP topology for the Prp24 binding on U6, which has not been found for any other RNA-protein interaction. Within this arrangement, the C terminus of RRM3 wraps around the U6 bulge region in a way in which oRRM4 has moved to the opposite face of U6 (Figure 1.19d). Then, the RRM2 and oRRM4 form tertiary interactions which lead to the formation of a ring of protein residues around the asymmetric bulge (Figure 1.19d). Hydrogen bond formation by the last residue of RRM3; an aspartate with A42 and G55 within the asymmetric bulge, is shown to be involved in stabilization of the proposed Interlocked topology. This unique RNP architecture could be the reason for previously observed low nanomolar

apparent dissociation constant for Prp24 binding to U6[206], and for the stability of Prp24-U6 complex even at higher salt concentrations (2M)[179].

Analysis of A62G suppressor mutations in both Prp24 and U6 point out the relationship between the suppressor site and protein-RNA interaction, indicating that the suppression of A62G mutation is due to the destabilization of U6-Prp24 interactions. As an example, components are involved in the aspartate bridge between RRM3-RRM4 junction and the bulge, acting as suppressor substitutions. Similarly, mutations in RRM3 residues; Asn216, Asn253 and Ser283 which form hydrogen bonds with the last two nucleotides of telestem result in suppression of cold sensitive mutations. In accordance with this, previous studies have also shown that mutations within RRM3 suppress the cold-sensitive mutations in U4 (ex. G14C in U4/U6 stem II) that destabilize the U4/U6 duplex.[179]

With all of these findings, it has been proposed that, together, all four RRMs in Prp24 induce the conformational remodelling of U6 snRNA, and the protein acts as a recycling factor during spliceosomal assembly.[198,199] Nevertheless, the mechanism of how Prp24 binds with U6 and its function on the formation and dissociation of U4/U6 complex are still not clear. The recent crystal structure study has proposed how Prp24 acts as a U6 chaperone, where it facilitates the formation of U4/U6 duplex as well as dissociation of U2/U6 to release free U6 snRNP[164]. According to that study, the electropositive groove of RRM1, 2 and 4 could provide a platform for the RNA annealing and thus RRM1 could facilitate the association of U6 with U4. U6 nucleotides 54-60 which shown to be interact with RRM1-RRM2 are found along one face of electropositive groove suggesting that RRM1-RRM2 bound to U6 is an intermediate

within the U4/U6 assembly. These interactions may prevent re-annealing of U6 ISL residues in free U6 snRNP, allowing the base pairing with U4 to form U4/U6 duplex. After the release of lariat intron, Prp24 may directly interact with U2/U6 complex, where RRM2 recognises the U6 nucleotides 49-53. This leads to the dissociation of U2 from the complex, allowing U6 to form the telestem region and thus Prp24 create the interlocked topology.

1.11: Defects in splicing and spliceosomal components can be lethal

Proper assembly of the spliceosomal components is critical for its function, and thus defects in its assembly can be lethal.[14,34] According to many studies done on splicing, it has been shown that defects in splicing or in spliceosomal assembly are associated with many disease conditions such as various cancers (leukemia, ovarian cancer, etc.), neurodegenerative disorders such as Parkinson's, genetic disorders such as Cystic fibrosis and many more.[7,208,209] Studying the structural dynamics and distinct functions of snRNA complexes and the factors that affect the stability of those complexes can provide an overall idea about the structure and function of the spliceosome, which will guide us to discover novel therapeutics for splicing-related diseases.

1.12: Detection of spliceosomal dynamics

Spliceosomal assembly and catalysis have been extensively studied over many years and many techniques have been used. Several biochemical assays, genetic studies and mutational assays have been used to determine the role f each component of the spliceosome at different steps of the assembly cycle. Also various structural studies;

NMR spectroscopy, EM studies and X-ray crystallography have been used to resolve the structures of the sub complexes and dissect the components of each sub complex. Also, to understand RNA-RNA and RNA-protein interactions within the spliceosomal components and importance of such interactions in the assembly and catalysis, several methods have been used such as the electrophoresis mobility shift assays (EMSA), chemical probing and hydroxyl radical foot printing. Altogether, these studies have revealed an immense amount of information on the composition of spliceosome, structure of sub-complexes and splicing mechanism. However, structural rearrangements within spliceosomal subcomplexes, kinetics information about association/dissociation of spliceosomal components have not yet been elucidated. Although many techniques have been used in the past to understand the structural and functional details of spliceosome, all of them have certain disadvantages making them less effective in providing a quantitative measure of structural dynamics and distinct interactions of spliceosomal components at different steps. To overcome most of such shortcomings, fluorescence based techniques have been introduced, where the analyte is associated with a fluorescent dye (ex. Cyanine dyes; Cy3, Cy5) or a fluorescent protein (Green fluorescent protein; GFP).

1.12.1: Fluorescence spectroscopy

Fluorescence, a form of luminescence, is the emission of light from a substance as a result of absorption of light or other electromagnetic radiation. The compounds that emit fluorescence are known as fluorophores. When molecules absorb energy they get excited into higher energy level, and the subsequent relaxation can occur via several

processes as described in a Jablonski diagram (Figure 1.20)[210]. When a molecule absorbs light at a certain wavelength, it is excited from its ground state (S_0) to one of the vibrational levels in the first excited singlet state (S_1, Figure 1. 20, black arrows). This excitation process is radiative and emission can be radiative if it releases a photon. The relaxation occurs through different processes. The electrons in the higher vibrational levels relax to the lowest vibrational level of the excited state. This is known as vibrational relaxation (Figure 1.20, brown arrows). Internal conversion (IC, Figure 1.20, Purple curved arrows) is a process where the vibrational level of an electronically excited state is coupled to a vibrational level of a lower electronic state. Molecules can have two types of excited states, namely singlet and triplet, which differ based on the spin multiplicity. If energy levels of singlet state are overlapped with that of a triplet state, a vibrational coupling can occur between singlet excited state and triplet state named as intersystem crossing (ISC, Figure 1.20, orange curved arrow). This transition is non-radiative and which is technically forbidden. Phosphorescence, which is a slow and forbidden transition, takes place when molecules in a triplet state relax into ground state by emitting light (Figure 1.20, red arrows).

On the other hand, fluorophores absorb energy and excite, relax from singlet excited state to singlet ground state; they emit light as fluorescence (Figure 1.20, green arrows). Because of its short life time (~10^{-8} s), fluorescence is more favourable when compared to the external relaxation, intersystem crossing or phosphorescence. Fluorescence light is lower in energy and observed at longer wavelength, whereas the excitation has higher energy and shorter wavelength. Thus, fluorescence can be used as a spectrochemical method of analysis, where the analyte is attached with a fluorophore,

which is then excited by irradiation at a particular wavelength and emit radiation at different wavelength. The resulting emission spectrum provides both qualitative and quantitative information about the analyte and thus can be used to determine properties of the analyte.

Fluorescence spectroscopic techniques are powerful tools that can be used to visualize interactions among biomolecules, and more importantly, these assays can be done under biologically relevant conditions. Also, fluorescence imaging has several advantages over traditional imaging methods (such as radioisotope labelling and electrochemical detection) including, use of low volumes of nonradioactive, homogeneous samples, use of readily available instruments and detection is highly sensitive and safe[211]. Another very important advantage of these assays is that the fluorescence signal of the analyte can be modified in a way that the detection of signal is more accurate and correlates with the properties of the analytes, such as existence, orientation and concentration[211]. There are several fluorescence based assays that have been employed in spliceosome studies, including fluorescence anisotropy based, fluorescence intensity measurement based and fluorescence resonance energy transfer (FRET) based.

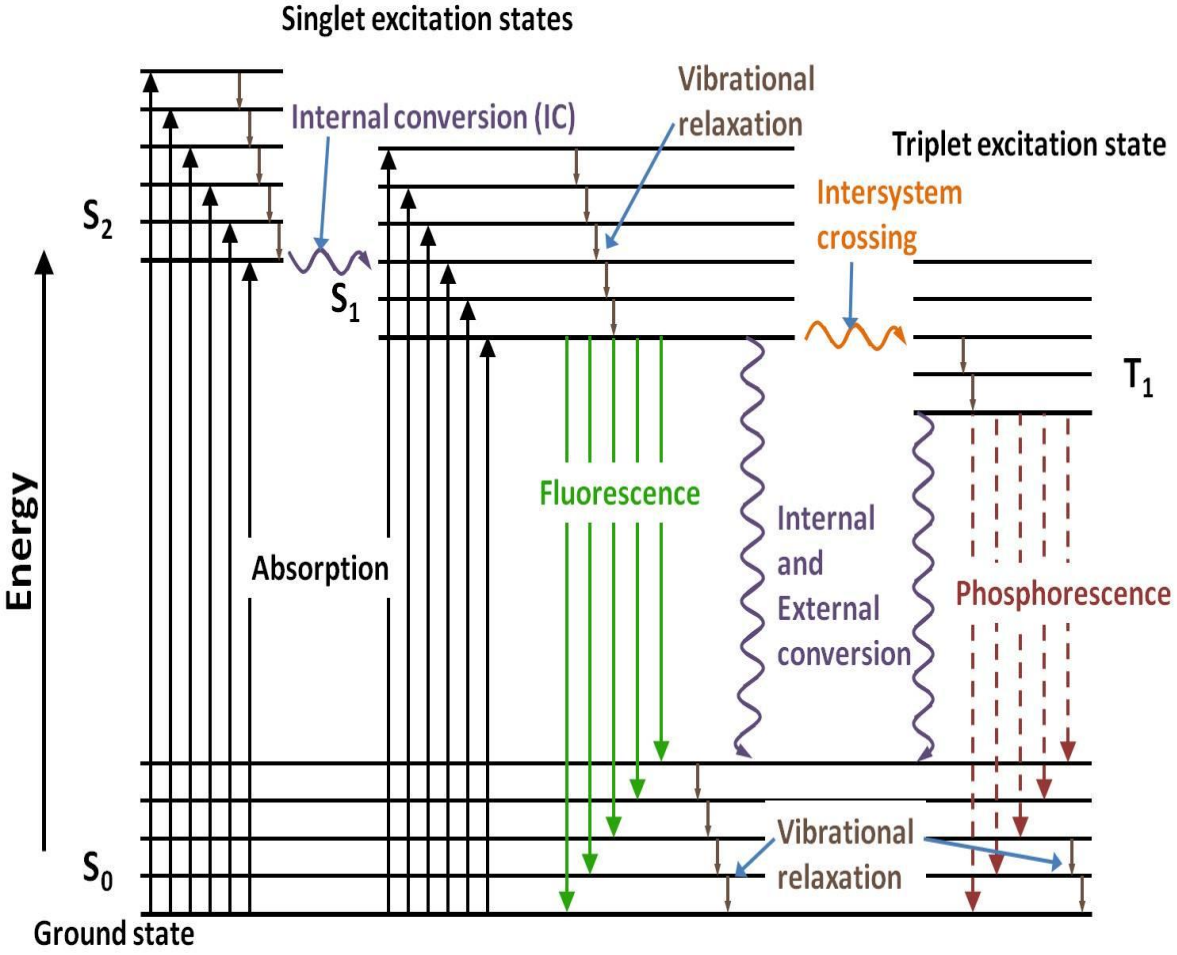

Figure 1.20: Jablonski diagram showing the electronic transition between energy levels. Once a molecule absorbs energy (black) and excite into higher energy levels from ground state (S_0). These excited molecules then come to the lowest level in the excited state (S_1), by releasing energy by vibrational relaxation (brown). Further releasing energy can occur in different ways as shown in the figure, such as internal conversion (IC, purple), intersystem crossing (orange), fluorescence (green), phosphorescence (red).

1.12.2: Fluorescence anisotropy

Fluorescence anisotropy can be used to study the interaction between nucleic acid molecules and proteins. In fluorescence anisotropy, a sample with fluorophore is excited using vertically polarized light (i.e. the electric vector is parallel to the z axis, Figure 1.21a)[210]. The emission intensity is detected through a polarizer. When the polarizer is parallel to the polarized excitation intensity then we denote it as $I_{\parallel}$ and if it's perpendicular then it's shown as $I_{\perp}$G is an instrument dependent correction factor. The anisotropy can be calculated using following equation;

$$r = \frac{I_{\parallel} - GI_{\perp}}{(I_{\parallel} + 2GI_{\perp})}$$

The idea under this technique is that, when a fluorophore is excited with polarized light, the emission light will also be polarized within a certain range of angles. If the fluorophore is rotating freely the degree of polarization of emitted light will be reduced resulting in a low anisotropy value (Figure 1.21b). In contrast, when the fluorophore attached molecule is bound to a larger molecule such as a protein, the free rotation of fluorophore is hindered, resulting in a higher degree of polarisation of emitted light giving rise to high anisotropy (Figure1.21b). Therefore this technique can be used to study the binding of Protein to the RNA.

1.12.3: Fluorescence resonance energy transfer (FRET)

Fluorescence resonance energy transfer (FRET) is a powerful technique that can be used to study inter- and intra-molecular dynamics in biological systems[212]. In this technique, energy is transferred between two fluorophores through non-radiative dipole-dipole coupling. Many fluorophore pairs such as; Cy3-Cy5, Alexa488-Cy3, Fluorescein-Tetramethylrhodamine (TAMRA) and Cyan fluorescent protein (CFP)-Yellow fluorescent protein (YFP) have been used in FRET studies. First, a fluorophore (donor, D) is excited to higher energy level and eventually relax by releasing energy. This releasing energy will then be transferred to an adjacent molecule, an acceptor (A, Figure 1.22a) via long-range dipole-dipole interactions[212,213]. The concept behind this radiationless energy transfer is that dipole oscillation of donor molecule undergoes energy exchange with an adjacent molecule (acceptor) having a similar resonance frequency by inducing a dipole oscillation of acceptor molecule[210]. To have energy transfer between donor and acceptor, certain conditions need to be fulfilled[214]. First both dyes need to have high quantum yield and absorption coefficient. Second, the dipole moments of two fluorophores need to be in proper orientation. Third, the donor emission spectrum should be overlapped with absorption spectrum of the acceptor. In figure 1.22b, absorption and emission spectrum for cy3 and cy5 pair, as donor and acceptor respectively, are showing, where emission spectrum of cy3 is overlapping with excitation spectrum of cy5. Finally, the distance between donor and acceptor should be in the range of 10-100Å. The rate of energy transfer between two fluorophores is given by[210,212,213,215],

$$k_T = \left(\frac{1}{\tau_D}\right) \times \left(\frac{R_0^6}{R^6}\right)$$

where, $\tau_D$is the fluorescence life time of donor molecule, R is the distance between the donor and the acceptor, R_0 is the Förster distance, which corresponds to the distance resulting in 50% efficiency of energy transfer. The energy transfer efficiency can be given by the Förster equation[210,212]:

$$E_{FRET} = \frac{1}{1 + \left(\frac{R}{R_0}\right)^6}$$

This shows that efficiency of energy transfer depends on the inverse sixth power of the distance between donor and acceptor. R_0 depends on the spectral overlap between donor and acceptor fluorophores and the relative orientation of their transition dipoles[216].

$$R_0^6 = 8.79 \times 10^{-25} \frac{\kappa^2 \emptyset_D J}{n^4}$$

where, κ^2 is the orientation factor (2/3 in average), $\emptyset_D$ is the quantum yield of the donor in the absence of the acceptor, n is the refractive index of the solution (1.33 for aqueous solutions) and J is the spectral overlap between donor emission and acceptor excitation.

In general, the apparent FRET value can be calculated using the following equation[210];

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

In which, I_A is the acceptor intensity and I_D is the donor intensity. The figure 1.22c illustrates the energy transfer efficiency as a function of distance between two fluorophores, where it shows a sharp change in efficiency within the distance range of 25 – 75 Å. Since, most of the biomolecular interactions fall within this distance range; FRET can be used as a 'molecular ruler' to study the interactions and dynamics of biomolecules, which cannot be obtained using other light microscopy methods.

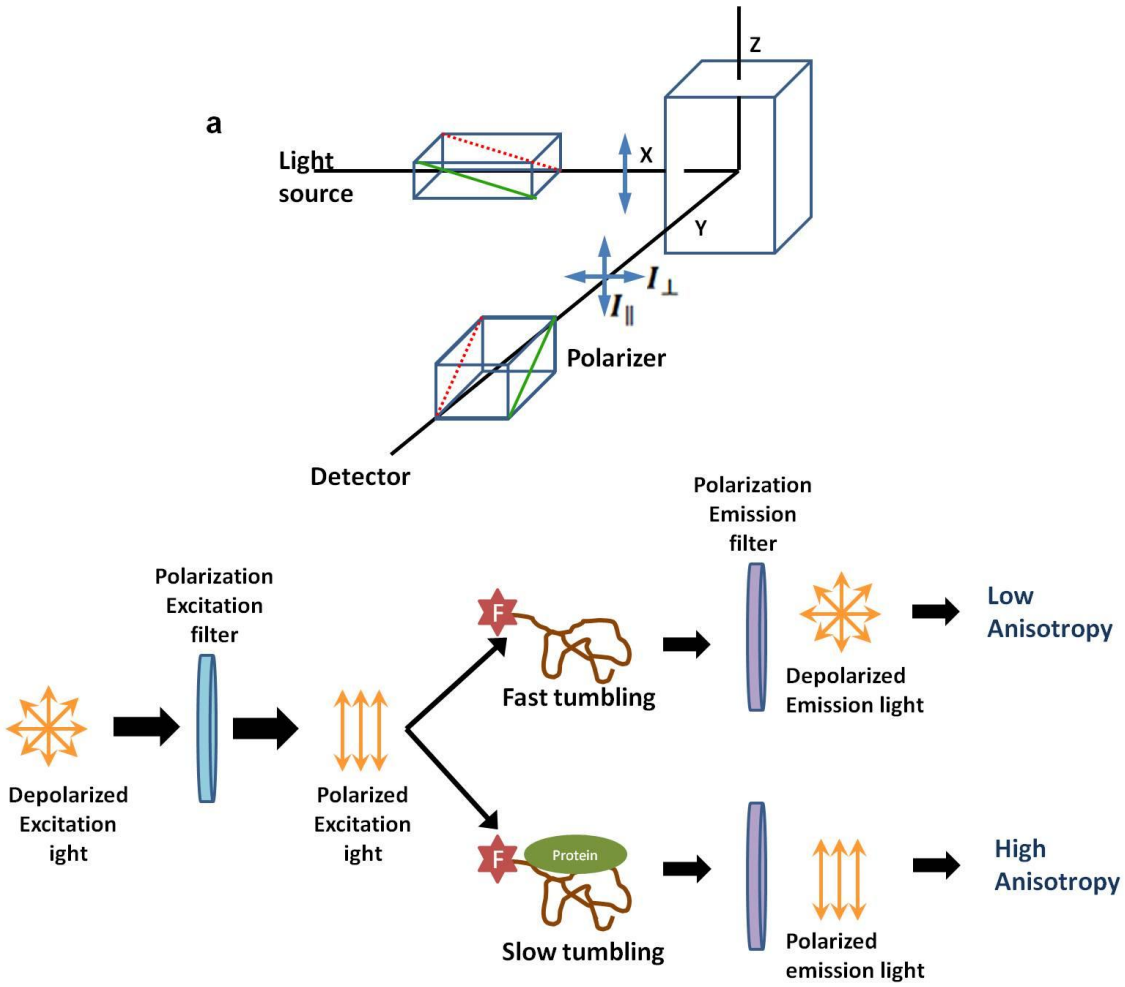
a
Light source
X
Z
Y
$I_{\perp}$
$I_{\parallel}$
Polarizer
Detector
Polarization Excitation filter
Depolarized Excitation ight
Polarized Excitation ight
F
Fast tumbling
Slow tumbling
Protein
Polarization Emission filter
Depolarized Emission light
Polarized emission light
Low Anisotropy
High Anisotropy

Figure 1.21: Fluorescence anisotropy to study RNA-protein interactions. (a) A schematic representation of fluorescence anisotropy measurement. A sample with fluorophore is excited using a vertically polarised light and the electric vector is parallel to the z axis. The emission intensity is detected through a polarizer. When the polariser is parallel to the polarised excitation intensity then we denote it as $I_{\parallel}$ and if it's perpendicular then it's shown as $I_{\perp}$. (b) A schematic diagram illustrating effect of larger compound (protein) binding on the fluorescence anisotropy of a fluorophore attached to a small molecule (RNA or DNA). When a fluorophore is attached to a small molecule like RNA it can move freely (fast tumbling) and hence when it excites with a polarized light, the degree of polarisation of the emitted light will be reduced (more depolarised) resulting low anisotropy. On the other hand, when a larger molecule like a protein is bound to the small molecule with the fluorophore attached, the degree of free rotation of the fluorophore is reduced (slow tumbling) resulting in an increase in the polarised emitting light, which gives rise to high anisotropy. Adapted from (199).**1.12.3: Single-molecule fluorescence microscopy**

In biological complexes, the components can form a wide range of interactions with each other, adopt multiple conformations at different stages and undergo significant changes in the composition during their activity. Several conventional methods have been used to study the structural details and interactions of biological molecules, such as NMR spectroscopy, X-ray crystallography, cryo-EM and chemical probing. However, each of these techniques has some limitations; moreover they fail to give quantitative understanding about the biological systems, their interactions and dynamics. Also, most of the conventional methods provide an ensemble or bulk measurements of a sample,

which represent the average of all the molecules present in the sample. In a sample at a given time, molecules can behave differently, can have different dynamics and can adopt different conformations. But in a bulk study, all these different properties are summed together and provide an average result, which could hide information about the behaviour of each individual molecule in real time. Although some methods such as EM, X-ray crystallography and atomic force microscopy can be used to reveal structural information of individual molecules, the molecules need to be fixed in a particular conformation. Therefore these techniques cannot provide any structural dynamics and kinetic information of the molecule of interest. In order to provide a better picture about the biomolecular interactions, conformational dynamics and kinetics in biological systems, scientists have developed single-molecule microscopic approach.

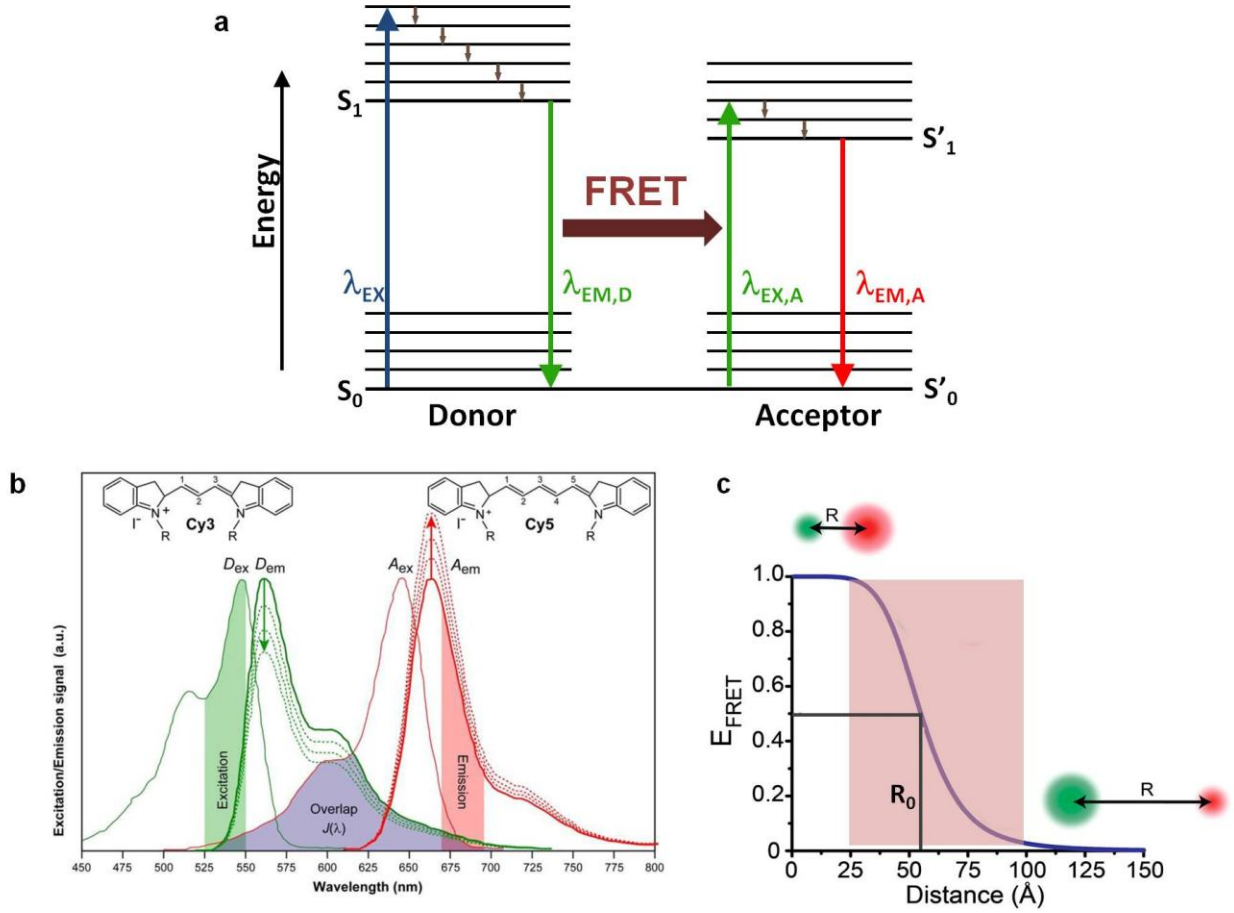

Figure 1.22: Fluorescence resonance energy transfer (FRET). (a) Energy diagram showing the electronic transitions during FRET. Energy release from one fluorophore is absorbed by an adjacent fluorophore, so that it will excite and eventually release energy. This transfer of energy from one molecule to another is known as FRET. (b) Spectral overlap between donor and acceptor fluorophores. Absorption and emission spectrum of cy3 (donor) and cy5 (acceptor) are in green and red respectively. The overlap region within the emission spectrum of cy3 and the absorption spectrum of cy5 is shaded in purple. (c) Efficiency of FRET as a function of distance is shown. This indicates that the energy transfer efficiency is inversely proportional to the distance between two fluorophores. Within the region between 25-100Å (shaded in pink), energy transfer efficiency is more sensitive to the distance. Adapted by permission from Macmillan Publishers Ltd: Nature methods (204), copyright (2008).

The major advantage of single-molecule microscopy is its ability to distinguish different populations present in a heterogeneous sample[217]. As an example, when a sample is composed of different populations and the distribution of each population can change rapidly, the conventional ensemble methods will give information by averaging all the molecules in the sample, and thus, show presence of only one population. In contrast, single-molecule approach differentiates each individual member of a heterogeneous sample, and characterise each population quantitatively[217]. Also this is an ideal method to study the structural dynamics of biological systems as well as to recognise interactions (association/dissociation) between components at different stages of both time dependent and independent reaction mechanisms[217]. With the advances of single molecule microscopy, the molecules can be detected with sub-nanometer spatial resolution, which is important in detecting inter and intra molecular interactions and dynamics. Since single-molecule approaches have better time resolution; mostly in milliseconds to nanoseconds, it can be used to detect the transient or short-lived intermediates that are present in the reaction pathways. This method also provides kinetic information on transitions of molecules between different states/conformations, rate of reaction mechanisms including shift to/from intermediates. Another advantage of this approach is small sample size. Mostly this technique requires only μM to pM concentrations and thus it is very useful when the isolation of some biomolecules is challenging and resulting in very low yield.

Single-molecule techniques have been used extensively to study the biological systems and all these approaches can be divided into two major groups; fluorescence based and force based. Fluorescence based single molecule studies include FRET,

localization, polarization, life time and intensity. On the other hand the force based studies are optical and magnetic tweezers and AFM. Optical tweezers is a technique where the molecule of interest (eg. nucleic acid) is attached to a bead/microsphere via avidin-biotin interaction, which is attached to a surface from the other side and trapped in a laser beam[218,219]. This can be used to manipulate the beads by applying a small force in the piconewton range, which in turn measures displacement of biomolecules in the nanometer range and the amount of force need to break a bond or associated length changes. In AFM, a nanometer size tip on a cantilever is attached to a biomolecule and a force is applied to unfold the molecule and study the properties of that molecule.

Single-molecule studies based on fluorescence have developed many reporter systems during last few years. Colocalization Single Molecule Spectroscopy (CoSMoS) provides information about the localization of molecules within the cell, which indicates the interactions between different molecules[220]. Similarly in fluorescence anisotropy, single-molecule fluorescence polarization anisotropy uses polarized light and gives information about interactions or the binding of molecules with others. Single-molecule fluorescence life time imaging (SM-FLIM) is used to quantify the fluorescence lifetime, which is a reflection of constitution and surrounding of molecules[221]. The time-resolved fluorescence microscopy/ time-correlated single photon counting (TCSPC) measures the intensity of fluorophore decay which can be used to monitor the distance distribution of molecules. In the fluorescence recovery after photobleaching (FRAP) approach, the equilibrium of fluorophores are disturbed by photobleaching and then the increase in intensity on the initially bleached sites are measured[221]. FIONA, or ‘fluorescence

imaging with one-nanometer accuracy', is another single-molecule fluorescence approach[218]. This method detects the location of molecules with ~1.5nm accuracy. Analysis of fluorescence image and fit it into a Gaussian distribution will result in such higher resolution. Among all these approaches, the single molecule assay combined with FRET (sm-FRET) is known to be the most widely used method.

This method can be used to determine the distance, conformational dynamics, and kinetic properties of individual molecules within biological systems. Some of the advantages of sm-FRET over other single-molecule approaches are less prone to environmental noise since it is a ratiometric technique and easy data acquisition.

Two main setups to measure the fluorescence at single molecule level have been developed, namely, confocal microscopy and total internal reflection fluorescence microscopy (TIRF)[218,219,222].

In confocal microscopy, only a small volume is excited by focusing a laser through the objective lens of the microscope as well as it uses much diluted samples[218,219]. This helps to reduce the background signal. Also in this setup, a pinhole barrier is used in between objective and detector, to reduce the out-of-focus fluorescence emission (Figure 1.23a). These characteristic features increase the signal-to-noise ratio and the sensitivity that leads to detect single fluorophores. One of the advantages of confocal microscopy is that it gives better time resolution, with an order of microseconds. This setup can be used for the samples with free diffusing molecules as well as for the molecules that are tethered to a surface.

In TIRF, the excitation sample volume is confined to a small area by placing it in a channel made in between a quartz slide and a cover slip[218]. The laser beam use in this

setup arrives at the interface between the quartz slide and the aqueous solution with an incident angle larger than its critical angle, achieving a total internal reflection of the laser. Although the laser is reflected completely, an evanescent wave is created at the slide-solution interface which penetrates and excites the molecules closer to the slide surface, typically within the range of ~200 nm. This prevents the penetration of laser through the sample, which in turn minimizes the background fluorescence and increases the sensitivity for single molecule detection. The emitted fluorescence from the individual biomolecules is then captured by the objective and detected through a charge-coupled device (CCD) camera. The TIRF setup allows detecting signals from hundreds of molecules at the same time, permitting rapid data acquisition. One major limitation of the TIRF is its time resolution, which depends on the time resolution of CCD camera, mostly between ten to hundred milliseconds, and thus it cannot detect faster dynamics. Based on how the laser beam directed to the sample TIRF can be categorised into two systems; prism-based and objective-based TIRF.

In objective-based TIRF the evanescent wave is created through the objective, hence the sample can be manipulated easily[223]. Also inexpensive glass slides can be used for this setup. However, alignment of laser beam in the objective in order to achieve a total internal reflection is challenging. Also the excitation beam needs to be passed parallel to the vertical axis (Figure 1.23b)[223]. In this setup, first the laser beam needs to be expanded by using lenses and then the beam is aligned into the microscope using mirrors. Another lens needs to be used to focus the beam onto the back focal plane (BFP). A dichroic mirror (DM) is used to reflect the shorter wavelength excitation light and transmit the longer wavelength emission light.

As the name implies the prism-based TIRF uses a prism to reflect the laser beam entirely and thus the incident angle can be much larger when compared to the objective based setup (Figure 1.23c) [223]. Advantage of having a larger incident angle is that the evanescent wave will only penetrate to a small area of the sample, which helps to minimize the sample volume and improve signal-to-noise ratio. It also allows separating the excitation and emission paths completely. A few major limitations of prism-based TIRF include, most of the sample chamber is covered by a prism, making it difficult to manipulate, requirement of expensive and fragile quartz slides, emission wave passing through the sample to reach to the detector, resulting in background scattering. Despite these disadvantages, the prism-based setup has been used widely when compared to the objective-based TIRF. Hence, smFRET using prism-based TIRF is a very powerful tool to detect interactions between molecules and their conformational changes.

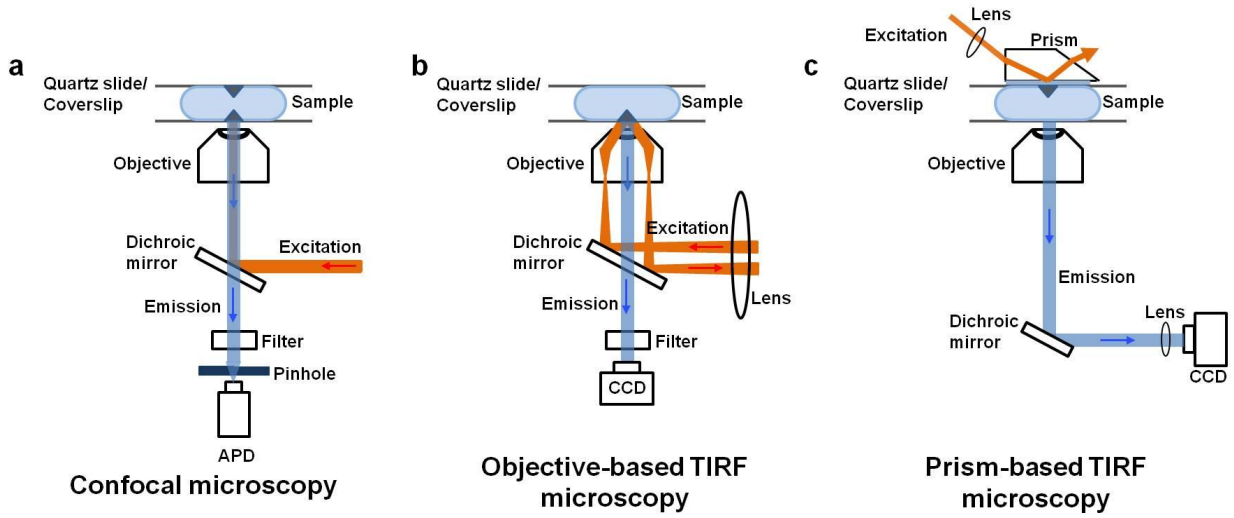

Figure 1.23: Schematic representation of different types of fluorescence microscopy. (a) A schematic representation of confocal microscopy. In this setup, laser beam is focused through the objective to excite only a small volume of a sample, which minimizes the background signals. Also the pinhole barrier between objective and detector helps to reduce the out-of-focus fluorescence emission. (b) A schematic representation of objective-based TIRF microscopy. In this setup, first the laser beam needs to expand by using lenses and then the beam is aligned into the microscope using mirrors. Another laser is used to focus the beam onto the back focal plane (BFP). A dichroic mirror is used to reflect the shorter wave length fluorescence (excitation) and transmit the longer wave length fluorescence (emission). (c) A schematic diagram of prism-based TIRF. In this setup a prism is used to bring the laser beam onto the slide-solution interface and reflect totally. This creates an evanescent wave which only excites the molecules closer to the surface. An inverted objective is used to collect the emitted wave of donor and acceptor, which will then send to the light-tight box through a slit. Inside the light-tight box, donor and acceptor signals are separated and detected side by side via the CCD camera. This figure is adapted and reproduced from (207 and 211).

1.12.4: Detection of spliceosomal assembly at single molecule level

(This section is adopted from Warnasooriya C. and Rueda D., Biochem.Soc.Trans. (2014), 42)

Spliceosomal assembly is a highly dynamic process that consists of hundreds of components, which associate/ dissociate at different stages in the assembly cycle, making it difficult to study the dynamics of these components using traditional bulk studies. Hence, single molecule fluorescence studies are ideal to study the dynamics of spliceosomal components. During the past few years many single molecule approaches have been made to study the spliceosomal assembly dynamics and catalysis.

Single molecule FRET studies on protein-free yeast U2/U6 complex revealed that it adopts at least three conformations which are in Mg^{2+}- dependent dynamic equilibrium[185]. These three conformations are assigned as three-way junction structure, four-way junction structure in consisting with previous studies and a novel intermediate conformation. Furthermore, it suggested that in the four way junction structure, ACAGAGA loop and U80 are in close proximity allowing 5'SS and metal binding sites to juxtapose with each other, facilitating the splicing reaction. Similarly, single molecule studies with protein-free human U2/U6 also revealed the structural and dynamics similarities between yeast and human supporting previous studies[224]. Moreover, authors demonstrated the role of post-transcriptional modifications on the stability of human U2/U6 complex.

Other than spliceosomal components, pre-mRNA itself undergoes structural rearrangements during splicing. Several studies have been done to monitor the pre-mRNA dynamics at single molecule level. One of such assays was developed to track

dynamics and removal of introns in whole cell extract (WCE) with single-molecule resolution, using colocalization single molecule spectroscopy (CoSMoS) technique. This study suggests that free pre-mRNA adopts multiple conformations which results in multiple FRET states[225]. Furthermore, binding of U1snRNA stabilizes a low FRET conformation of pre-mRNA in which 5'SS and BS are far apart, preventing any chemical reaction prior to the spliceosomal activation. Higher FRET states observed after addition of NTC complex reveal that binding of NTC stabilizes the interactions between U5, U6 and pre-mRNA which facilitates formation of activated complex to carryout both splicing reactions. Hence this study proposed the importance of dynamic nature of pre-mRNA. Another approach was taken to determine the distance changes between 5'SS and BS using smFRET[226]. Similar to the CoSMoS study, this study has also shown that the pre-mRNA is maintained a conformation where the 5' SS and the BS are away from each other until the formation of catalytically active spliceosome, which prevents the premature catalysis. Moreover, this study has also shown pre-mRNA adopts several conformations which are present in dynamic equilibrium and upon completion of first step of splicing, the equilibrium shift towards a high FRET conformation.

Another CoSMoS study has been carried out to monitor the spliceosomal assembly pathway, more specifically to observe the order of spliceosomal components assembly[227]. Over many years it has been considered as U1 binds first to the pre-mRNA followed by U2. This new study, strikingly, has shown that either U1 or U2 can binds first, followed by the other snRNP, suggesting that these alternative assembly pathways can be useful in alternative splicing.

In addition to these in vitro studies, several approaches have been made to visualise splicing and assembly dynamics in living cells by tracking labelled pre-mRNAs, which have provided new insight into the splicing machinery. Nevertheless, many studies have been done using single-molecule approaches both *in vitro* and *in vivo*, to obtain information regarding spliceosomal assembly and catalysis processes, a vast number of important questions remain unanswered. For example, the machinery of how spliceosomal components recycle after completion of splicing and what factors are involved, how spliceosomal sub-complexes assemble, the global conformation of the sub-complexes and how binding of proteins affect the overall conformation, and what are the structural and functional similarities/ differences between major and minor spliceosomes, what role do protein factors and regulatory motifs play in spliceosomal assembly, how do these factors affect related biological processes such as alternative splicing, and co- and post-transcriptional splicing in cells, what is the structure-function relationship between most of the spliceosomal components and what are the differences in splicing kinetics among different organisms. Since single-molecule microscopy is a powerful tool to study the structural dynamics of large, complex biological systems, the single molecule approaches including its novel advances; ex. super resolution microscopy, can be used to solve most of the un-answered questions related to spliceosomal assembly and catalysis.

1.13: Objective of this study

Studying the structural dynamics and distinct functions of snRNA complexes, as well as the factors that affect the stability of those complexes provides an overall idea regarding

the structure and function of the spliceosome. Defects in splicing machinery, spliceosomal components or in cis/trans-acting splicing factors are associated with many disease conditions. A thorough understanding of splicing machinery and assembly is important for advancement in diagnosis and treatment of these diseases. Although the spliceosome has been studied extensively, many aspects of its structure and function are yet to be known. Some of such information that needs to be unveiled are; How snRNPs are dissociated from each other at the end of splicing cycle? What factors involve in this process? How spliceosomal sub complexes are assembled? How the associated components affect on the global structure of each of these sub complexes? What are the structural and functional similarities among major and minor spliceosomal components?.

Conventional bulk studies that cannot provide information on transient intermediates and population heterogeneity may not be a better approach to explore aforementioned aspects. Hence, the aim of this study is employing smFRET technique to monitor the structural dynamics and assembly of snRNA complexes and the effect of protein factors on those dynamics with single molecule resolution. This could eventually contribute for a much clearer picture of spliceosomal components rearrangements within the assembly cycle and catalysis, which lay another step towards the progress of understanding the structure and function of spliceosome. Following are the three specific aims that I'm addressing in this thesis work to achieve the main goal.

1.13.1: Specific Aim 1: Elucidating the role of Prp24 in U2 and U6 snRNP recycling

Prp24 is a U6 snRNP associated chaperone which plays an important role in spliceosomal assembly as a recycling factor. Previous studies have proposed that Prp24 involves in association/dissociation of U4/U6 and helps U6 to interchange its conformations at different stages of assembly cycle. Hence, we suggest that it may also be involved in unwinding of U2/U6 and allowing the reformation of U4/U6. I have used sm-FRET to test our hypothesis; Prp24 may unwind the U2/U6 complex after the catalysis helping these snRNAs to recycle.

1.13.2: Specific Aim 2: Study the assembly and global structure of U4/U6 complex

Although U4 is not present in the activated complex, the formation of U4/U6 complex is important for spliceosomal assembly. Binding of U4 with U6 is necessary to prevent U6 from premature activation and formation of unfavourable conformations. Several proteins are involved in the assembly of U4/U6 di-snRNP, which has been suggested to play an important role in the formation of global structure. The aim of this study is to characterize the relative orientation and the global structure of U4/U6 duplex in the presence of its associated proteins (Snu1 3, Prp31 and Prp3/4) using single-molecule spectroscopy and electrophoretic mobility shift assays (EMSA).

1.13.3: Specific Aim 3: Study the structural dynamics of U12/U6atac snRNA complex in minor spliceosome.

Two spliceosomes with distinct composition have been identified; major (U2-dependent) and minor (U12-dependent)[5,38,138]. The major spliceosome is present in all eukaryotes

and splices the highly abundant U2-type introns[38,139]. The minor spliceosome splices less abundant U12-type introns. Previous studies have shown that, despite its low abundance, the minor spliceosome plays an important role in processing of many genes. As an example, defects in minor spliceosome assembly due to mutations in U4atac snRNA cause the Taybi-Linder Syndrome (TALS)[155-157]. According to early studies U12/U6atac forms a complex similar to U2/U6 and forms the catalytic core[150,228,229]. The aim of this study is to monitor the structural dynamics of U12/U6atac complex and compare the similarities and differences between major and minor complexes using single molecule FRET.

CHAPTER 2: Material and Methods

2.1 Materials

Table 2.1: List of chemicals used in this study

Chemical Name	Company
Acetic acid	Sigma-Aldrich Company Ltd, Dorset, UK
Acetone	Sigma-Aldrich Company Ltd, Dorset, UK
Acetonitrile	Sigma-Aldrich Company Ltd, Dorset, UK
Acylamide: Bis acrylamide 40%	
Alconox (VWR International Inc.)	VWR International Inc., Leicestershire, UK
3-aminopropyltriethoxysilane (VECTABONDTM reagent)	Vector Laboratories, Inc., Burlingame, CA
Ammonium hydroxide (ACS grade)	Mallinckrodt Chemicals, USA
Ammonium persulfate (APS)	Fisher Scientific UK Ltd, Loughborough, UK
Biotin-BSA	Fisher Scientific UK Ltd, Loughborough, UK
Boric acid	Sigma-Aldrich Company Ltd, Dorset, UK
Bromophenol blue	Fisher Scientific UK Ltd, Loughborough, UK
Biotin Polyethylene Glycol Succinimidyl Carboxymethyl (BIO-PEG-SCM, MW 5000)	Laysan Bio. Inc, AL, USA
Cy3 and Cy5 NHS esters	GE healthcare, Little Chalfont Bucks, UK
Dimethyl sulfoxide	Fisher Scientific UK Ltd, Loughborough, UK
Ethanol	Sigma-Aldrich Company Ltd, Dorset, UK
EDTA	Sigma-Aldrich Company Ltd, Dorset, UK

Glycerol	Fisher Scientific UK Ltd, Loughborough, UK
Hydrochloric acid (HCl)	Sigma-Aldrich Company Ltd, Dorset, UK
Hydrogen peroxide (H2O2, 30 %)	Fisher Scientific UK Ltd, Loughborough, UK
Magnesium Chloride	Sigma-Aldrich Company Ltd, Dorset, UK
Magnesium acetate	Fisher Scientific UK Ltd, Loughborough, UK
Methanol	Sigma-Aldrich Company Ltd, Dorset, UK
Methoxy Polyethylene Glycol Succinimidyl Carboxymethyl (m-PEG-SCM, MW 5000)	Laysan Bio. Inc, AL, USA
Protocatechuic acid (PCA) and Protocatechuate dioxygenase (PCD)	Sigma-Aldrich Company Ltd, Dorset, UK
Sodium acetate (NaOAc)	Fisher Scientific UK Ltd, Loughborough, UK
Sodium bicarbonate and Sodium carbonate	Fisher Scientific UK Ltd, Loughborough, UK
Sodium chloride	Sigma-Aldrich Company Ltd, Dorset, UK
Sodium hydroxide	Sigma-Aldrich Company Ltd, Dorset, UK
Streptavidin	Invitrogen, Life technologies Ltd, Paisley, UK
Triethylammine	Fisher Scientific UK Ltd, Loughborough, UK
Triethylammine trihydrofluoride (TEA.3F)	Sigma-Aldrich Company Ltd, Dorset, UK
Tris base	Sigma-Aldrich Company Ltd, Dorset, UK
Trolox	Fisher Scientific UK Ltd, Loughborough, UK
Xylene cylanol	Sigma-Aldrich Company Ltd, Dorset, UK

2.2: Methods-Single-molecule FRET study

Single-molecule fluorescence resonance energy transfer (sm-FRET) can be used to detect interactions between molecules and their conformational changes. Sm-FRET is a very important technique in which we can focus on one molecule at a time.[230,231] This technique provides an idea about how individual molecules behave in a bulk solution, revealing their structural dynamics and heterogeneity in the system. Also, this is a very important tool to understand and identify transient intermediate states that cannot be revealed from bulk experiments.[185,186,216,231,232]

2.2.1: Surface preparation- cleaning of slides and coverslips

In sm-FRET studies, a quartz slide is used to make the microfluidic channel. Two holes were drilled on the quartz slide using a hand drill and a diamond drill bit with a diameter of 1mm. Slides need to be passivated before introducing the sample and the following steps were done to passivate the slides as explained in previous study[233].

As the first step of slide preparation, slides are needed to be cleaned well, since impurities can increase the back ground fluorescence. First the reaction containers (staining jars/ coplin jars) were cleaned well and dried using Nitrogen gas. The Quartz slides were cleaned with the detergent; Alconox to remove any debris remaining on the slide. First a small amount of Alconox mixed with water to make a paste. Then each slide was scrubbed for few minutes with this paste and washed well with distilled water to remove the detergent completely. Next, the slides were cleaned well by scrubbing several minutes with distilled water, ethanol and distilled water respectively. Then the cleaned slides were placed in a cleaned beaker and filled with the “basic piranha

solution" (100 ml distilled water, 20 ml of 30% Hydrogen peroxide and 20 ml of Ammonium hydroxide) and boiled for about 30-40 minutes. Next, the slides were washed thoroughly with distilled water and flame dried using a Bunsen burner. These two steps help to remove any organic compound stuck on to the slide surface and restore the silanol groups on the surface.

2.2.2: Aminosilanisation of slides and coverslips

As the next step, the slides and coverslips are needed to be activated prior to the PEGylation process. During this step an amino group is attached to the slide surfaces which will later react with the N-hydroxysuccinimide (NHS) ester group on the PEG molecules we use for PEGylation (Figure 2.1a). First, reaction containers (coplin jars/ staining jars) were cleaned and dried well. Then the cleaned and dried slides from previous step and new cover slips were placed in the coplin jars separately. These jars were then filled with 1 M KOH and sonicated for > 20 minutes. Next, the KOH solution was removed and slides were rinsed with distilled water followed by methanol. Later the jars were filled again with methanol and sonicated for another > 20 minutes. Finally, the methanol solution was discarded and the slides were rinsed again with new batch of methanol solution. For the aminosilanisation of the slide surface, first a reaction mixture was made with 100ml of methanol, 5ml of glacial acetic acid and 1ml of 3-aminopropyltriethoxysilane (VECTABOND^{TM}, Vector Laboratories, Inc., Burlingame, CA). Slides and coverslips were incubated with this solution for 10 minutes, sonicated for 2 minutes followed by 10 minutes incubation. Then the reaction solution was removed from the reaction jars and the slides were rinsed with methanol, distilled water

and methanol respectively. Finally the slides and coverslips were dried well with nitrogen gas. Now the slides are ready for the PEGylation.

2.2.3: Surface PEGylation

Passivation of slide surfaces with a polymer is important for the prevention of non-specific binding or the adsorption of proteins to the slide surface. Linear polyethylene glycol (PEG) is the most common polymer used in slide passivation (Figure 2.1a). A 10% biotin polyethylene glycol succinimidyl carboxymethyl (BIO-PEG-SCM, MW 5000) solution was made by mixing biotin-PEG with methoxy polyethylene glycol succinimidyl carboxymethyl (m-PEG-SCM, MW 5000) and dissolve in PEGylation buffer (100mM sodium bicarbonate, pH 8.4). For five slides, 50 mg of m-PEG/ biotin-PEG mixture (total 10% biotin-PEG) was dissolved in 360 µl of PEGylation buffer, vortex gently and centrifuged at ~10,000 rcf for few minutes. Then the cleaned and dried slides from previous step are placed in a reaction chamber (a pipette box) filled ~10% with distilled water. Next 70 µl of PEGylation reaction mixture was added on to each slide and a coverslip was placed carefully on the top of each slide and reaction solution without making any bubbles. Then the slides were incubated with the reaction solution in dark for overnight. The next day, slides and coverslips were carefully separated out, rinsed well with distilled water and dried with nitrogen gas.

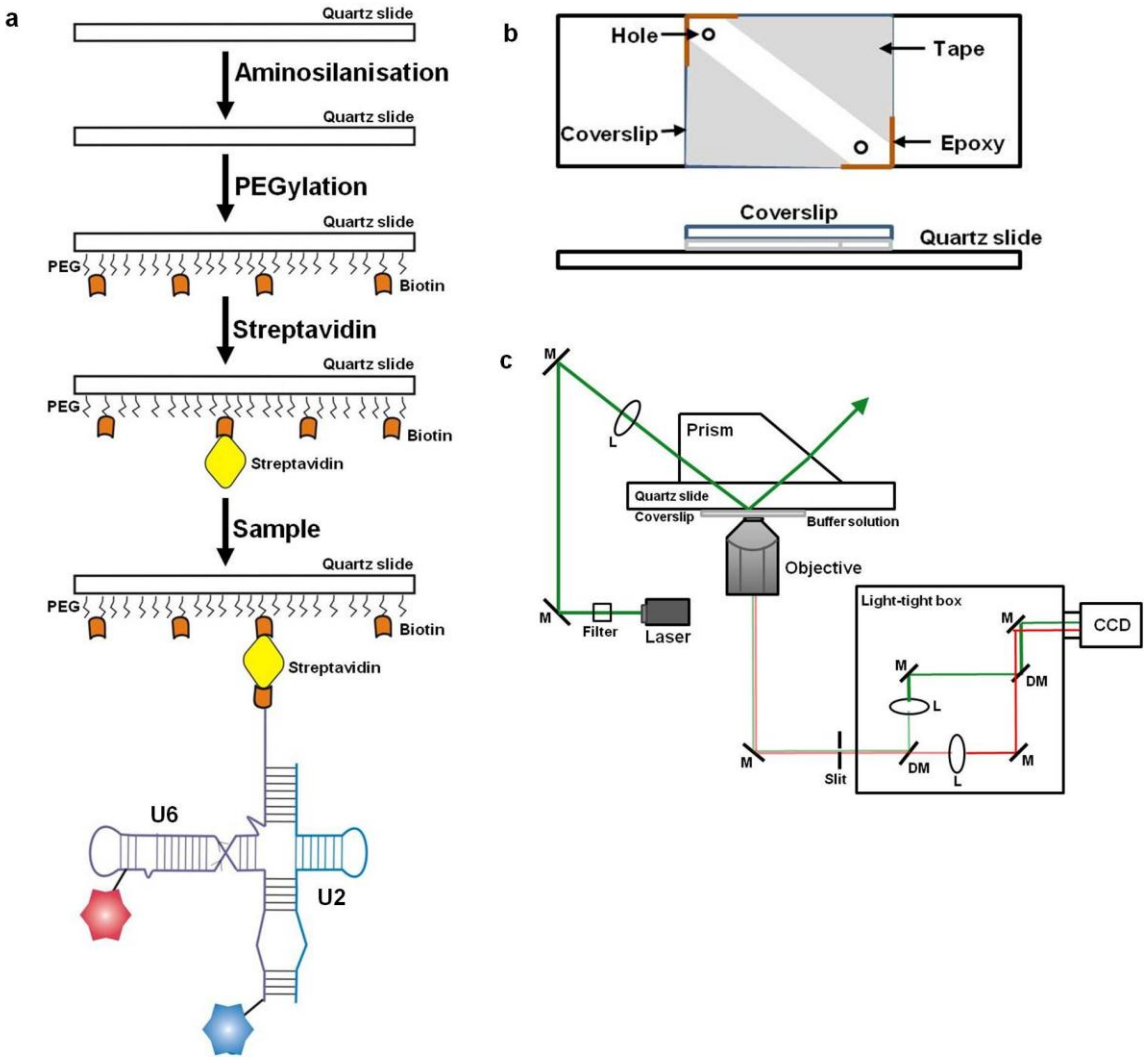

Figure 2.1: Slide preparation, sample immobilization and Imaging in single-molecule experiments. (a) Schematic representation of stepwise process of slide passivation and sample immobilization. (b) Preparation of flow channel using double sided tape between the quartz slide and coverslip (top), and the top view of an assembled slide. (c) Schematic representation of microscope for single-molecule experiments.This figure is adapted and reproduced from (226) with the permission from Elsevier.

2.2.4: Preparation of flow channel

The cleaned slides and coverslips are then assembled to make a flow channel, which will be used to immobilize molecules and flow buffers. The slides were kept on a rack placing the PEGylated side up. Then two layers of double-sided sticky tape were applied to each side of the holes, making a microfluidic channel with ~ 200 μm deep and ~5 mm wide as shown in Figure 2.1b. Then a coverslip was placed carefully on each slide covering the two holes and the channel made. The extra tapes were removed from the coverslip edges by cutting with a razor blade. The epoxy glue (5min Epoxy) was applied on the edges of the coverslips to seal the channels and let them completely set for ~20 minutes (Figure 2.1b). Finally the assembled slides were stored in covered 50 ml falcon tubes under nitrogen gas. Now the slides are ready for experiments.

2.2.5: Sample purification

The samples (DNA, RNA or Protein) that are used in the single-molecule FRET experiments need to consist of two fluorophores to obtain a fluorescence energy transfer, which corresponds to conformational changes between the molecules. Also one of the molecules needs to be biotinylated for surface immobilization. Sample preparation procedure for single-molecule experiments is as follows.

The RNA samples used in this thesis work were purchased from the Keck Foundation Resource Laboratory at the Yale University School of Medicine unless otherwise stated. The 2'-hydroxyl protective groups were removed and the RNAs were purified as described previously[215,232]. In summary, dried RNA sample was dissolved in

200 µl DMSO and take only half of the solution for deprotection reaction while storing other half in -80 ^{o}C. Next, 125ul Triethylamine trihydrofluoride (TEA.3HF) was added to the tube containing RNA dissolved in DMSO and thoroughly mixed by vortexing. Then the reaction mixture was mixed at 65 ^{o}C and at 650 rpm speed for 2.5 hours followed by a brief cooling in the freezer for few minutes. Then 25 µl 3 M sodium acetate (pH5.2) and 1 ml of butanol were added to the RNA solution, mixed well and kept in -20 ^{o}C over night or in -80 ^{o}C 2-3 hours. Next, the RNA solution was centrifuge at 8900 rpm for 10 minutes to precipitate RNA and poured out the butanol solution. The pellet was then rinsed with 0.75 ml 70% Ethanol twice and dried in vaccum dryer. Finally the dried RNA pellet was dissolved in distilled water. Then RNA was purified by denaturing gel electrophoresis (20% wt/vol polyacrylamide and 8 M urea), cut out the desired band and eluted against 4ml elution buffer (0.5 M Sodium acetate and 0.1 mM EDTA, 10 mM Tris.HCl) overnight at 4 °C. The eluted solution containing RNA was then mixed thoroughly with chloroform: isoamyl alcohol (25:1) solution, centrifuged at 1000 rpm for 10 minutes and the aqueous layer was separated into a new tube. Next, a 1/10th volume of 3 M Sodium Acetate and 2-2.5x volume of 100% Ethanol were added to the tube, shake well and placed in -20^{o}C overnight. Then the solution was centrifuged at 12500 rpm, supernatant was discarded and the pellet was washed with 70% ethanol twice. Finally the pellet was dried in vaccum dryer. The RNA samples that are not used for further labelling were then dissolved in HPLC buffer (50 % Acetonitrile, 14 % Triethylammine in distilled water, pH 7.0) and purified using C8 reverse-phase (column-Supelco, PA, USA) HPLC (Shimadzu, Kyoto, Japan).

2.2.6: Fluorophore labelling of sample

The substrates use in single-molecule FRET experiments required to be labelled with a donor and an acceptor fluorophore.

2.2.6.1: Fluorophores

A fluorophore, which is bright (with a higher quantum yield) and photostable is ideal for single-molecule FRET studies. Also it needs to be a small, water-soluble molecule containing some bio-conjugation groups. A good FRET fluorophore pair requires a well separated emission spectra, similar quantum yields and similar detection efficiencies, in order to obtain a better signal[215]. Several pairs of fluorophores have commonly been used in FRET studies. Fluorescein and Tetra-methyl rhodamine (TAMRA) pair has extensively been used in bulk experiments, whereas cyanine dyes have been known as a good dye pair for single-molecule FRET studies (Table 2.2, Figure 2.2). In this study I have used Cy3 and Cy5 as donor and acceptor fluorophores respectively (Table 2.2, Figure 2.2).

Table 2.2: Photophysical properties of some common fluorophores use in FRET studies

Fluorophore	Excitation (nm)	Emission (nm)	Extinction coefficient (ε) ($cm^{-1}M^{-1}$)	Quantum yield
Fluorescein	490	520	83,000	0.7
Tetramethylrhodamine (TAMRA)	554	580	95,000	0.2-0.5
Cyanine 3 (Cy3)	550	570	150,000	>0.15
Cyanine 5 (Cy5)	650	670	250,000	>0.28

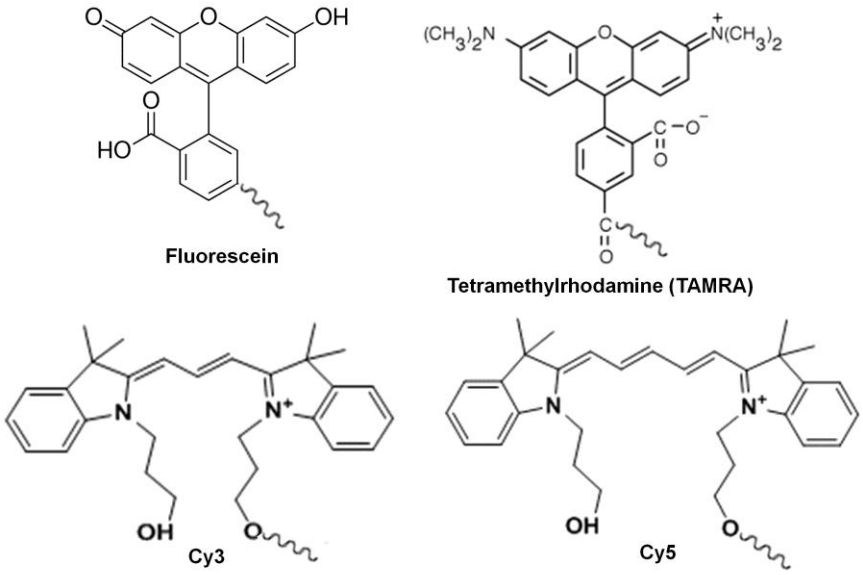

Figure 2.2: Common fluorophores use in FRET studies. Fluorescein and TAMRA are commonly used in bulk studies where as Cy3 and Cy5 are extensively used in single-molecule FRET studies.

2.2.6.2: Sample labelling

The fluorophores need to be attached at a specific position on the RNA, DNA or protein samples to detect the inter- and intra-molecular conformational changes. A fluorophore can be added as a dye-phosphoramidite to an oligonucleotide during the chemical synthesis process. Also oligonucleotides can be synthesised with an amino linker at a desired position, on a dT residue. Succinimidyl ester group attached to a fluorophore reacts with the amino linker attached to the RNA or DNA. On the other hand, proteins are modified to have only one cysteine residue, which then can be labelled with fluorophores containing mono-functional maleimide group. During RNA labelling, first, dried mono reactive N-Hydroxysuccinimide (NHS) ester form of cy3 or cy5 dye is mixed in 14 µl of DMSO, vortexed and dissolved completely. Then the desired RNA sample (~ 50-100 µg) is dissolved in labelling reaction solution (7 µl dye mixture, 20 µl 100 mM Sodium Carbonate, pH 8.5 and 73 µl water) and incubated overnight at 25^{o}C with mixing at 350 rpm. Following day, 10 µl 3 M Sodium Acetate (pH 5.2) and ~ 390 µl of 100 % Ethanol are added to the reaction mixture, mixed well and kept at -20^{o}C overnight. Then the reaction mix was centrifuged and discarded the supernatant, followed by washing with 70% ethanol twice. Finally the pellet is dried and dissolved in 100 µl of HPLC buffer and purified in HPLC to separate the labelled RNA from unlabelled RNA. The RNA concentrations of all samples are measured by UV-Vis absorbance at 260 nm.

2.2.7: Sample immobilization

The RNA samples in standard buffer prepared in saturated Trolox solution is heated at 94 °C for 45 s and annealed by cooling to room temperature over 20 min. First, 100 µl

streptavidin (0.2 mg/ml in 50 mM Tris.HCl, pH 7.5 and 50 mM NaCl) is injected to a PEGylated slide containing ~10% biotin PEG and incubated 10 minutes allowing the formation of streptavidin-biotin interaction (Figure 2.1b). Unbound streptavidin is washed out with T50 buffer (50 mM Tris-HCl, pH 7.5 and 50 mM NaCl). The annealed, biotinylated, fluorophore-labelled complex is diluted to 25 pM and then injected to the slide to immobilize via biotin-streptavidin interaction to generate a surface density of ~0.1 molecules/μm^2 and incubated 10 minutes (Figure 2.1b). Finally, 200 µl oxygen scavenging system (OSS) is injected and after several minutes the slide is ready for imaging.

2.2.8: Prevention of photobleaching and blinking

Two of the major limitations that occur in fluorescent based techniques are photobleaching and blinking. Oxygen molecules present in the solution can cause faster photobleaching of the fluorophores under the laser excitation. To obtain single-molecule movies for longer time without photobleaching of fluorophores, it is important to deplete the oxygen molecules present in the microfluidic chamber. An oxygen-scavenging system (OSS) plays an important role in single-molecule studies by removing oxygen. A solution mixture containing 5 mM Protocatechuic acid (PCA) and 0.1 µM Protocatechuate 3, 4-dioxygenase (PCD) is a commonly used OSS system in smFRET studies to diminish the oxygen level in the solution and reduce photo-bleaching[234]. Protocatechuate 3, 4-dioxygenase (PCD) is an enzyme, which uses 1 moles of oxygen to produce two moles of protons by converting 1 moles of PCA to β-carboxy-cis, cis-muconic acid (Figure 2.3a).

The intersystem crossings of fluorophores from singlet state to triplet state (dark state) could result in short intermittencies in fluorescence emission, which is known as blinking. The time that fluorescence emission stays in the dark state can vary from microseconds to tens of milliseconds depending on the molecule. To quench the triplet state blinking, a saturated solution of Trolox (6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid, a vitamin E analogue, Figure 2.3b) is used to prepare all the imaging buffers solutions used in smFRET experiments. This also acts as an antibleaching reagent in the experiments. Although the actual working mechanism of trolox has not been found, a study using a combination of an oxidizing and reducing system (ROXS) has shown that the antiblinking property of trolox is a result of quinoid derivatives of Trolox, which arise from the reaction with molecular oxygen.

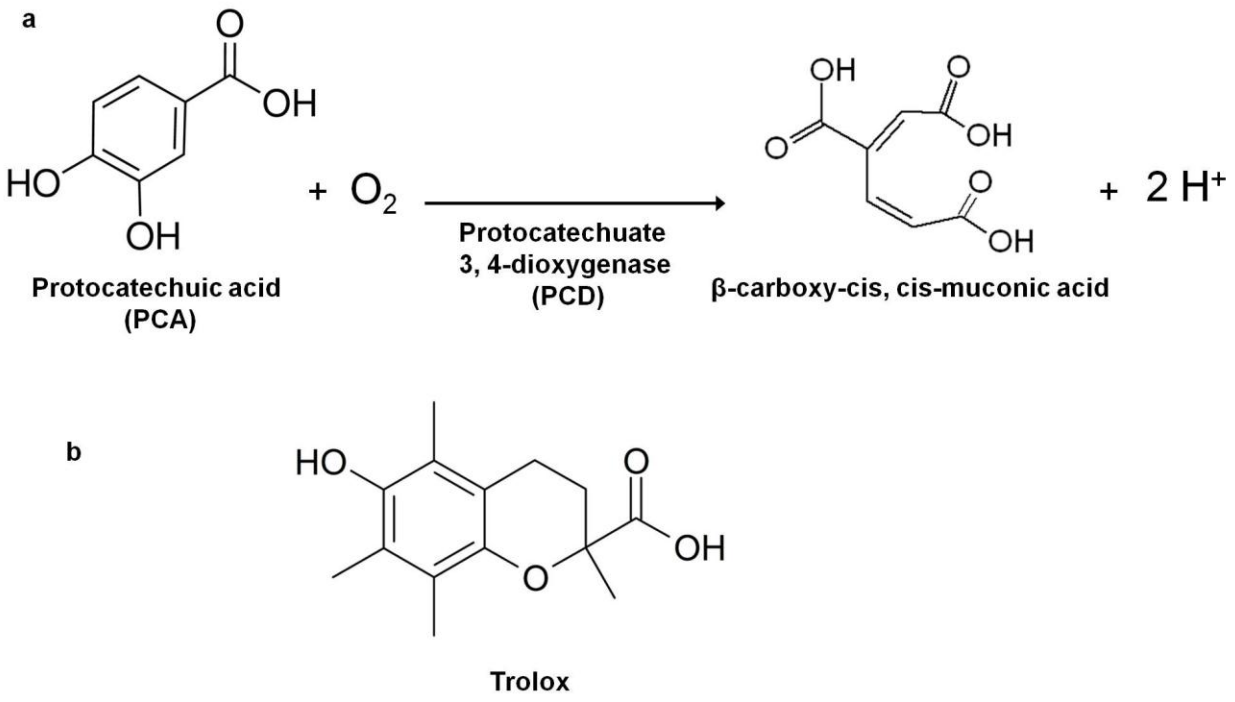

Figure 2.3: Chemicals commonly use in smFRET to minimize photobleaching and blinking. (a) Chemical reaction for PCA/PCD system; a commonly use oxygen scavenging system (OSS) in smFRET studies, illustrating how PCD consumes oxygen molecules to convert PCA to β-carboxy-cis, cis-muconic acid. (b) Chemical structure of Trolox, a commonly use antiblinking agent.

2.2.9: Data acquisition and analysis

The donor fluorophores are excited using a home-built prism-based, total internal reflection (TIRF) microscope and a 532 nm laser (3 mW, Spectra-Physics Excelsior) as described (Figure 2.1c)[223]. An evanescent wave is created by reflecting incident laser beam completely, which will only excite the molecules in the environment near the surface (~100 nm)[223],[235]. The excited signals were captured through a water immersion objective (60X) as single molecules. The Labview software was used to record the data with 33 ms exposure time. Resulting donor and acceptor emissions were separated using appropriate dichroic mirrors (610DCXR, Chroma) and recorded as two side-by-side images on a back-illuminated electron-multiplied CCD camera (Andor I-Xon Ultra 897). To match the signals from each half, a sample with bright fluorescent beads was used as a reference, which maps the corresponding spots in two channels. These smFRET data were then analysed as explained in previous studies[215,223].

First, for mapping the spots in two channels corresponding to the signal from donor and acceptor, a sample of fluorescent beads (0.2 μm modified red microsphere, Invitrogen, Carlsbad, CA) was immobilized and taken a short movie. Then the IDL software is used to map spots in two channels, which will then be used as a reference to analyse data from each single experiment (Figure 2.4a). Another IDL script is used to find the donor and acceptor molecules in each half based on the bead map. The individual donor (I_D) and acceptor (I_A) intensity time trajectories for each molecule were obtained by integrating their relative spot intensities. Then a home written MATLAB script is used to select real single molecules (Figure 2.4b). When picking up real molecules, there are few factors that need be considered such as continuous

fluorescence emission and single step photobleaching of both fluorophores and anti-correlated relationship between donor and acceptor emission intensities. Next, the intensity traces are converted to the apparent FRET time trajectories using the equation;

$$E_{FRET} = \frac{1}{1 + \left(\frac{R}{R_0}\right)^6}$$

The resulting FRET time trajectories are time binned to obtain population distribution histograms for each RNA sample using IGOR program (Figure 2.4c and d). Each peak in the histogram represents different conformation of the RNA sample and their ratio, which illustrates the conformational dynamics of these samples. The corresponding FRET value indicates the distance between two fluorophores which in turn provides information regarding the conformation of the RNA sample used.

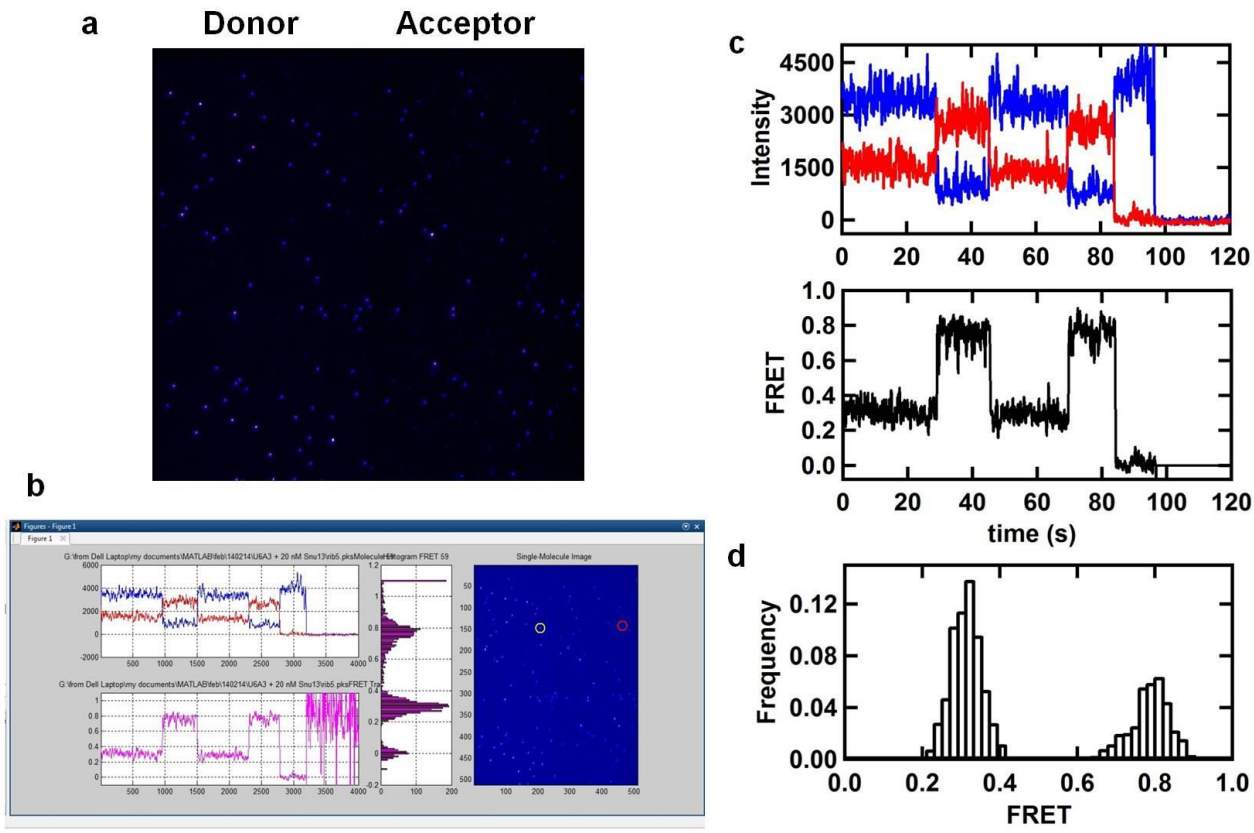
a
Donor
Acceptor
b
c
Intensity
4500
3000
1500
0
0
20
40
60
80
100
120
FRET
1.0
0.8
0.6
0.4
0.2
0.0
time (s)
d
Frequency
0.12
0.08
0.04
0.00
0.0
0.2
0.4
0.6
0.8
1.0
FRET

Figure 2.4: Single-molecule data analysis. (a) A side-by-side image recorded on CCD camera illustrating emission of donor and acceptor signals. (b) Image of MATLAB analysis program, illustrating intensity and FRET trajectory for corresponding molecule selected. (c) Redrawn images of intensity and FRET trajectories and (d) corresponding distribution histogram obtained using IGOR program.

2.3: Methods: Elucidating the role of Prp24 in U2 and U6 snRNP recycling

2. 3.1: Sample preparation

RNA samples for single-molecule experiments were purchased from the Keck Foundation Resource Laboratory at the Yale University School of Medicine. U6 was ordered with a 5'-Cy3, an internal amino C6 dT and 3' – biotin (U6; Table 2.3). U2 was ordered with or without 3' – C7 amino linker (U2 and U2-cy5, respectively; Table 2.2). All the Protein samples (full length and truncated Prp24) were obtained from the Butcher's lab in University of Wisconsin-Madison. Protein samples came in the storage buffer containing 50% glycerol, 25mM Tris, 125 mM NaCl, 1.25 mM DTT and 0.2 mM EDTA.

2.3.2: Sample purification and fluorophore labeling

The 2'-hydroxyl protective groups were removed and the RNAs were purified as described earlier (Section 2.2.5).[215,232] The C6 amino linker in U6 and C7 amino linker in U2-cy5 were labelled with Cy5 (GE Healthcare) as described earlier (Section 2.2.6).

Table 2.3: RNA sequences used in this study

Name	Modifications	Sequence (5' -3')
U6	P=dT with C6 amino linker for Cy5 labelling	Cy3-AUACAGAGAUGAUCAGCAGUUCCCC**P**GCAU AAGGAUGAACCGUUUUACAAAGAGAU-(3'-Biotin)
U6-FI	None	Fluorescein-AUACAGAGAUGAUCAGCAGUUCCCCU GCAUAAGGAUGAACCGUUUUACAAAGAGAU
U2	None	UAUGAUGUGAACUAGAUUCGGUUUUCCGUUUCUC UA
U2-cy5	Q=dT with C6 amino linker for Cy5 labelling	UAUGAUGUGAACUAGAUUCGGUUUUCCGUUUCUC UA**Q**
U4-full length	None	AUCCUUAUGCACGGGAAAUACGCAUAUCAGUG AGGAUUCGUCCGAGAUUGUGUUUUUGCUGGUUGA AAUUUAAUUAUAAACCAGACCGUCUCCUCAUGGUC AAUUCGGUGUUCGCUUUUGAAUACUUCAAGACUAU GUAGGGAAUUUUUGGAAUACCUUU
U6-full length	None	GUUCGCGAAGUAACCCUUCGUGGACAUUUGGUC AAUUUGAAACAAUACAGAGAUGAUCAGCAGUUCCC CUGCAUAAGGAUGAACCGUUUUACAAAGAGAUUUA UUUCGUUUU
U4	X=dT with C6 amino linker for Cy5 labeling	AUCCUUAUGCACGGGAAAUACGCAUAUCAGUGAG GA**X**CGUCCGAGAUUGUGUUUUUGCUGGUUGAAAU UU-(3'-Biotin)
U6-I	None	Cy3-UGAUCAGCAGUUCCCCUGCAUAAGGAU
U6-II	Y=dT with C6 amino linker for Cy3 labeling	UGAUCAGCAGUUCCCCUGCAUAAGGAU**Y**
U6-III	Z=dT with C6 amino linker for Cy5 labeling	Cy3-UGAUCAGCAGUUCCCCUGCAUAAGGAU**Z**
U6atac	R=dT with C6 amino linker for Cy5 labelling	Cy3-AUGAAAGGAGAGAAGGUUAGCACUCCCCU**R**GA C AAGGAUGGAAGAGGCCC-(3'-Biotin)
U12	None	GAUGCCUUAAACUUAUGAGUAAGGAAA

2.3.3: MALDI-MS experiments

Matrix-assisted laser desorption ionization-mass spectrometry was carried out to confirm the position of the modified nucleotide used for cy5 labelling in U6 RNA as described.[236] The U6 RNA (300 pmol) were RNase T1 (~100 units) digested for 15 minutes in order to produce smaller fragments. The molecular weights of the fragments were calculated by using the Mongo Oligo Mass Calculator v2.06. After RNase T1 digestion, the reaction was stopped in dry-ice, dried and dissolved in 1ul of distilled water. 1 ul of MALDI matrix (3-hydroxypicolinic acid (HPA) in 50% acetonitrile), 0.5 µl of 100 mM ammonium citrate and 1µl of RNA sample were mixed on the MALDI plate in the given order. The spot was dried for about 30min and used in MALDI experiment. Representative spectrum for Cy3-U6 with internal amino linker sample is shown in Figure 2.5. The fragment that appears at 2654.716 is the fragment with the modified nucleotide used for Cy5 labelling in U6. In addition to the fragment with the modified nucleotide, all the other digested fragments were also observed within a reasonable error (Table 2.4).

Table 2.4: The expected and observed masses from MALDI-MS experiments for the RNaseT1 digested U6 RNA. All the masses are observed within a 5% error. * Not observed.

Fragment	Sequence	Expected mass	Expected mass (without 3' phosphate)	Observed mass
A1:G6	pAUACAGp	3267.597	3249.579	N/O*
A7:G8	AGp	692.433	674.418	693.985
A9:G11	AUGp	998.602	980.587	999.348
A12:G16	AUCAGp	1632.995	1614.980	1633.940
C17:G19	CAGp	997.617	979.602	999.348
U20:G27	UUCCCCUGp	2654.708	2636.693	2654.716
C27:G32	CAUAAGp	1962.204	1944.189	1964.155
G33:G33	Gp	363.224	345.209	N/O*
A34:G36	AUGp	998.602	980.587	999.348
A37:G41	AACCGp	1632.010	1613.995	1633.940
U42:G51	UUUUACAAAGp	3209.919	3191.904	3211.360
A52:G53	AGp	692.433	674.418	693.985
A54:U59	AUUUAU	2390.739		2392.434

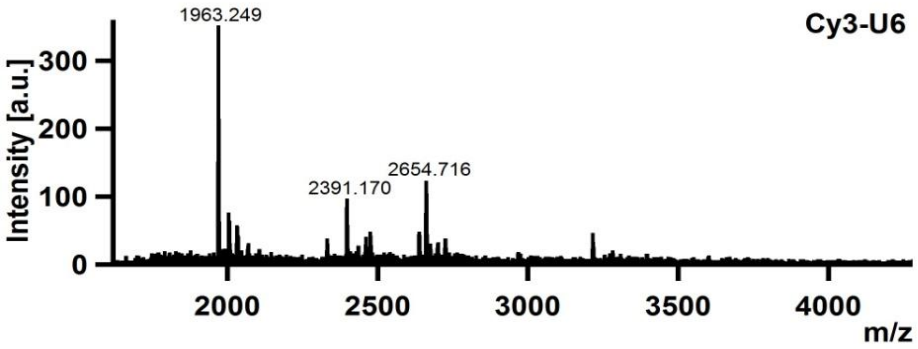

Figure 2.5: MALDI result reveals the mass of the fragment that is used for Cy5 labelling in U6.The peak at 2654.716 is corresponding to the fragment containing the modified nucleotide for Cy5 labelling in U6. The other two peaks at 1963.249 and 2391.170 are two of other fragments resulting from RNase T1 digestion.

2.3.3: Fluorescence electrophoresis mobility shift assay (EMSA)

U2/U6 complex was made by mixing U2 (4 μM)/ U6 (2 μM) or U2-cy5 (2 μM)/U6-cy3 (2 μM) in buffer (50 mM Tris-Cl, pH 7.5, 100 mM NaCl, 10 mM $MgCl_2$) and heating at 94 °C for 45 s, followed by 20 min incubation at room temperature for annealing. Prp24 (full length and truncated-first RRM is removed, denoted as 234C; 4 μM each) was added and incubated for ~10 min. After incubation, an equal volume of 40% glycerol was added to the reaction mixture. Non-denaturing (29:1 acrylamide:bisacrylamide ratio, 15%) gel electrophoresis was performed in 50 mM Tris-acetate, pH 7.5, 100 mM sodium acetate, 10 mM magnesium acetate, and 5 mM DTT at 4 °C for >18 hrs at 100 V. The gel was scanned using a Typhoon 9210 Variable Model Imager (GE Healthcare) and analyzed with ImageQuant software (Amersham Bioscience).

2.3.4: Fluorescence anisotropy experiments

Anisotropy experiments were carried out using a spectrofluorometer with automated polarizers (Varian, Carry Eclipse). The U6 RNA with 5' fluorescein (U6-Fl, 25 nM, Table 2.3) and U2 RNA (50 nM) were mixed in a standard buffer (50 mM Tris-HCl, pH 7.5, 100 mM NaCl, and 10 mM $MgCl_2$) and heated at 94 °C for 45 s and then annealed at room temperature for 20 min. Anisotropy was measured for Prp24 concentrations of 0 to 1 μM. Fluorescein was excited at 490 nm (5 nm bandwidth), and parallel ($I_{\parallel}$) and perpendicular ($I_{\perp}$) emission intensities were measured at 520 nm (5 nm bandwidth).

The resulting anisotropy values were plot into a binding curve as a function of Prp24 concentration and the data was fitted into quadratic equation to obtain the dissociation constant (K_D) for Prp24 binding.

$$f(Prp24) = r_O + (r_{max} - r_O)\left[\frac{K_D + [RNA_O] + [Prp24] - \sqrt{(K_D + [RNA_O] + [Prp24])^2 - 4[RNA_O][Prp24]}}{(2 * RNA_O)}\right]$$

In the equation above, r_o is lowest anisotropy value, r_{max} is highest anisotropy value, [Prp24] is the corresponding protein concentration, [RNA] is the concentration of U6-fl used and K_D is the dissociation constant.

2.3.5: Single molecule study

2.3.5.1: Sample immobilization

Single molecule experiments were performed as described.[232, 223] Two RNA strands (2 μM U6 and 4 μM U2) in standard buffer (50 mM Tris.HCl, pH 7.4, 100 mM NaCl, and 10 mM $MgCl_2$) were heated at 94 °C for 45 s and annealed by cooling to room temperature over 20 min. The annealed, biotinylated, fluorophore-labelled complex was then diluted to 25 pM and immobilized on a PEGylated surface of Quartz slides via a biotin-streptavidin interaction.

In concentration-based experiments, the measurements were obtained under variable Protein concentrations (0 – 100 nM full length and truncated Prp24, 234C) at room temperature in the presence of OSS solution.

In time-dependent experiments, cy3-U6 (2 μM) and U2-cy5 RNA (2 μM) strands (Table 2.2) were used and samples were prepared as described earlier. The measurements were done every five minutes and histograms were drawn for each time point. Data acquisition is carried out similarly as explained in previous section.

2.3.5.2: Data analysis

In concentration-based experiments resulting time trajectories were separated based on the FRET value and time binned to draw FRET histograms, which represent the frequency of the population at a particular FRET value. A cut off FRET value of 0.3 was used to distinguish between high and low FRET conformations. To determine dissociation constant (K_D) values for binding of full length Prp24 and truncated proteins; the fraction of molecules with low FRET state was plotted as a function of the concentration of protein and fitted to a modified Hill equation:

$$f(Prp24) = f_0 + (f_{max} - f_0)\frac{[Prp24]^n}{{K_D}^n + [Prp24]^n}$$

in which f_0 is the initial fraction of molecules at low FRET state, f_{max} is the final fraction of molecules at low FRET state, n is the Hill coefficient, K_D is the dissociation constant and [Prp24] is the concentration of protein .

In time-dependent experiments, a cut off FRET value of 0.14 was used to distinguish the FRET states corresponding to the conformations with and without U2. To determine the rate constant (K_{obs}) for dissociation of U2, the fraction of molecules with zero FRET state was plotted as a function of time and fitted to an exponential fit.

2.4: Methods- Study the assembly and global structure of U4/U6 complex

2. 4.1: EMSA

2.4.1.1: U4/U6 duplex formation

For substoichiometric electrophoretic mobility shift assays, in-vitro transcribed, fluorescein labelled U4 snRNA was mixed with unlabelled U6 snRNA (Table 2.3) to a final concentration of 1 µM and 2 µM respectively in 10 mM K-HEPES (pH 7.5), 100 mM KCl. The mixture was heated to 90 °C and slow-cooled to 4 °C at -0.03 °C/s. The RNA duplex was then gel purified on a native gel at 4 °C, band excised, RNA eluted by the "crush and soak" method into 10 mM K-HEPES (pH 7.5), 100 mM KCl, and concentrated in an ultra-centrifugation filtration device (Amicon).

2.4.1.2: Substoichiometric assembly analysis

Direct binding EMSA experiments were performed with samples containing 2 nM fluorescein labelled RNA and protein typically within the range ~0.3 pM to 2.5 µM in an EMSA sample buffer consisting of 10 mM K-HEPES (pH 7.5), 100 mM KCl, 0.01% NP-40, 20 µg *E. coli* tRNA in a volume of 100 µL. For step-wise assembly (Table 1) protein concentrations for pre-assembled components were Snu13 (200 nM), Prp31 (120 nM), Sm proteins (64 nM) and LSm proteins (240 nM). Reactions were allowed to equilibrate on ice for 60 min before loading on native polyacrylamide gels (4% at 37.5:1 acrylamide:bis-acrylamide) run in 0.5x TBE buffer at 4 °C. Gels were imaged on a Typhoon variable-mode scanner and the signals in the gel bands corresponding to protein bound and unbound RNA were integrated. Parameters in the following function were fit to the data for fraction of RNA bound versus protein concentration:

$$\theta = \left[\frac{a-b}{1+\left(\frac{Kd}{[protein]}\right)^{n}}\right] + b$$

where θ is fraction bound, Kd is the apparent dissociation constant, a is the upper baseline and b is the lower baseline, and n is the Hill coefficient. At least two gel shifts were performed for each sample and associated error is reported as one standard deviation from the mean (Table 4.1). The shifts in Table 1 are for entirely wild-type components with the exception of the LSm complex wherein LSm 4 was truncated (amino acids 1-114) so as to remove the C-terminal region absent from the human homolog and predicted to be disordered (DISOPRED) and a truncated Sm complex was used (SmB 1-105).

Concentrations used in figures are as follows: Supplementary Figure 2a had U4 (4 nM), U6 (4 nM), U4/U6 snRNA duplex (4 nM), Snu13 (200 nM), Prp31 (120 nM), and the shift with LSm was conducted with 2-fold dilutions and a maximum concentration of 1 μM. Supplementary Figure 2b had H46 hybrid RNA (10 nM), Snu13 (100 nM), Prp31 (150 nM) and the shift with LSm was conducted with 2-fold dilutions and a maximum concentration of 500 nM. Figure 1b had U4 (4 nM), U6 (4 nM), U4/U6 snRNA duplex (4 nM), Snu13 (180 nM), Prp31 (120 nM), Sm proteins (64 nM), LSm proteins (240 nM), and Prp3/4 (500 nM).

2.4.2: Single-molecule assay

2.4.2.1: Sample preparation

Three U6 RNA strands and one U4 RNA strand were utilized for the single-molecule experiments to study the orientation of three helices (Table 2.3). The U4 and U6-II strands were purchased from Dharmacon, whereas U6-I and U6-III strands were purchased by Keck Foundation Resource Laboratory at the Yale University School of Medicine (Table 2.3).

2.4.2.2: Sample purification and fluorophore labelling

The 2'-hydroxyl protective groups were removed and the RNAs were purified as described (section 2.2.5). The C6 amino linker in U6-II was labelled with Cy3 (GE Healthcare) and the C6 amino linker in U6-III was labelled with Cy5 (GE Healthcare) as described earlier (section 2.2.6). Also, a sample of Prp31 was modified for cy5 labelling. The naturally occurring cysteine residues within the Prp31 were removed and new cysteine residue was introduced. Then the protein was labelled with cy5 as explained (section 2.2.6).

2.4.2.3: Sample immobilization

Single molecule experiments were performed as described[223,232]. Two RNA strands (2 μM U4 and 2 μM U6-I, U6-II or U6-III) in standard buffer (10 mM Tris.HCl, pH 7.4, 100 mM NaCl) were heated at 94 °C for 45 s and annealed by cooling to room temperature over 20 min. The annealed, biotinylated, fluorophore-labelled complex was then diluted to 10 pM and immobilized on a quartz slide via a biotin-streptavidin interaction. In

protein binding experiments, the measurements were obtained in the presence of Snu13, Prp31 or Prp3/4.

In single molecule experiments using labelled Prp31, unlabeled U4, cy3 labelled U6 and cy5 labelled Prp31 were used. Two RNA strands (unlabelled U4, 2 μM and cy3 labelled U6, 2 μM) in standard buffer (10 mM Tris-HCl, pH 7.4, 100 mM NaCl) were heated at 94 °C for 45 s and annealed by cooling to room temperature over 20 min. The annealed, complex was then diluted to 10 pM and immobilized on a quartz slide via a biotin-streptavidin interaction. Then cy5 labelled Prp31 in the presence of Snu13 and Prp3/4 is injected along with oxygen-scavenging system (OSS) in standard buffer. Before imaging excess cy5-labelled Prp31 protein was washed with the standard buffer. The sample was first imaged through 532nm (Green) laser for ~ 2-3 minutes. Then switch the laser to 637nm (red) and imaged for another ~2min, followed by imaging again with 532nm (green) laser for 2 min. Similar experiment was carried out with unlabelled Prp31 as a control. Data acquisition is carried out similarly as explained in previous section.

2.4.2.4: data analysis

In this study, the resulting time trajectories; RNA alone and in the presence of proteins were separated based on the FRET value and time binned to draw FRET histograms, which represent the frequency of the population at a particular FRET value. Average FRET values for each population were determined by fitting the histograms to Gaussians. These values then used to fit the peaks in overall histograms which combine all the single molecule trajectories of each RNA complex. Percentage of each

population is calculated by collecting all individual molecules at each population as a fraction of total number of molecules. The percentage of population 1 under different conditions (RNA alone, with different proteins) was then illustrated as a bar graph.

2.5: Methods-Minor spliceosome

2.5.1: Sample preparation

RNA samples for single-molecule experiments were purchased from the Keck Foundation Resource Laboratory at the Yale University School of Medicine. U6atac was ordered with a 5'-Cy3, an internal amino C6 dT and 3' – biotin (U6atac; Table2.3). U12 was ordered without any modification (Table 2.3).

2.5.2: Sample purification and fluorophore labelling

The 2'-hydroxyl protective groups were removed and the RNAs were purified as described (section 2.2.5). The C6 amino linker in U6atac was labelled with Cy5 (GE Healthcare) and purified using HPLC (section 2.2.6).

2.5.3: Single-molecule study

2.5.3.1: Sample immobilization

U6atac RNA strand was diluted to 25pM in standard buffer (50 mM Tris.HCl, pH 7.4, 100 mM Sodium chloride, and 10 mM Magnesium chloride) and heated at 94 °C for 1.5 min before flash cooling on ice. This flash cooling was done to avoid homodimerization. The U6atac RNA was then immobilized to the quartz slides via a biotin-streptavidin interaction and then U12 RNA in standard buffer was introduced to the slide.

In concentration-based experiments with U12, the measurements were obtained under variable U12 concentrations (0 – 50 nM). In the studies with varying Mg^{2+}, the standard buffer was made by adding corresponding Mg^{2+} (0 - 50nM) concentration. Data acquisition is carried out similarly as explained in previous section.

2.5.3.2: Data analysis

In this study, the resulting time trajectories were separated based on the FRET value and time binned to draw FRET histograms, which represent the frequency of the population at a particular FRET value. Average FRET value was determined by fitting the histograms to Gaussians, which is 0.2 FRET.

To determine dissociation constant (K_D) values for binding of U12; the fraction of molecules with 0.2 FRET state was plotted as a function of the concentration of U12 and fitted to a modified Hill equation:

$$f(U12) = f_0 + (f_{max} - f_0)\frac{[U12]^n}{{K_D}^n + [U12]^n}$$

in which f_0 is the initial fraction of molecules at 0.2 FRET state, f_{max} is the final fraction of molecules at 0.2 FRET state, n is the Hill coefficient, K_D is the dissociation constant and [U12] is the concentration of U12 .

In Mg^{2+} titration, K_{Mg} is obtained in a similar way as explained in previous paragraph, where the fraction of molecules stay at 0.2 FRET has plotted against the concentration of Mg^{2+} ions, and fitted to the modified hill equation;

$$f(Mg^{2+}) = f_0 + (f_{max} - f_0)\frac{[Mg^{2+}]^n}{{K_{Mg}}^n + [Mg^{2+}]^n}$$

in which f_0 is the initial fraction of molecules at 0.2 FRET state, f_{max} is the final fraction of molecules at 0.2 FRET state, n is the Hill coefficient, K_{Mg} is the dissociation constant for Mg and $[Mg^{2+}]$ is the concentration of Mg^{2+}.

CHAPTER 3: Elucidating the role of Prp24 in U2 and U6 snRNP recycling

3.1: Objective

The objective of specific aim 1 is to understand how snRNP complexes dissociate from each other after the completion of splicing and what are the factors involved in this process. To address these questions, I selected to study the role of Prp24 on the dissociation of U2/U6.

U6 snRNP associated chaperon Prp24 has been studied by many groups and it was determined to function as a recycling factor. Prp24 is also found to be involved in U4/U6 complex association and proposed involvement in U4/U6 unwinding prior to the spliceosomal activation[102,164,192,197,206]. Considering the tight association of Prp24 with U6 and its role as a recycling factor, I hypothesize that after catalysis, Prp24 may facilitate the dissociation of U2/U6 complex allowing U2 and U6 to participate in another round of splicing. Hence the main objective of this study is to determine the role of Prp24 in unwinding U2/U6 complex.

To address this, I employed Fluorescence based techniques including, electrophoretic mobility shift assay (EMSA), anisotropy, and sm-FRET. The EMSA and anisotropy data reveal the binding of Prp24 to the U2/U6 complex. The sm-FRET data show that the binding of Prp24 affects the conformational changes of the U2/U6 complex and stabilizes a low FRET conformation. Furthermore, EMSA and sm-FRET data show that the binding of Prp24 facilitates the removal of U2 from U6, revealing its role in U2/U6 unwinding. Moreover, the removal of Prp24 RRM1 has shown a lesser

effect on the stabilization of low FRET conformation and the removal of U2, suggesting its requirement for a high affinity binding.

3.2: Experimental design

Throughout this study, I have used a minimal U6 strand (nucleotides from 45-104) and a minimal U2 strand (nucleotides from 1-44), which together catalyze both steps of splicing[237]. These two strands were heated to 94^{o}C for 45 s and annealed for ~20 min in the reaction buffer to form the duplex which is denoted in the thesis as 'U2/U6 complex', unless otherwise stated (Table 2.3, Figure 1.18a, b). Alternative labelling of U6 and U2 strands were used for different experiments. For the anisotropy experiments U6 was labelled with fluorescein (yellow) at the 5' end and U2 was not labelled (Figure 3.1a), this labelling strategy results in higher anisotropy upon binding with protein. For the single-molecule studies and EMSA studies, doubly labelled U6 was used containing cy3 (blue) and cy5 (red) at the 5' end and near ISL loop (U70) respectively (Figure 3.1b). The 3' end of U6 was attached with biotin-BSA (orange) for surface immobilization. For the unlabelled U2 strand annealed opposite a labelled U6 strand, different FRET states were observed based on the conformational changes within the U2/U6 complex. For experiments with U6 labelled with cy3 at the 5' end and U2 labelled with cy5 at the 3' end a high FRET (1.0) state is expected for the U2/U6 complex formation and a low or zero FRET state upon U2 removal (Figure 3.1c). Two proteins were used in this study namely, a full length Prp24 and a truncated Prp24; where RRM1 and the N-terminus were removed (denoted in the thesis as 'Prp24' and '234C' respectively, unless otherwise stated, Figure 3.1d).

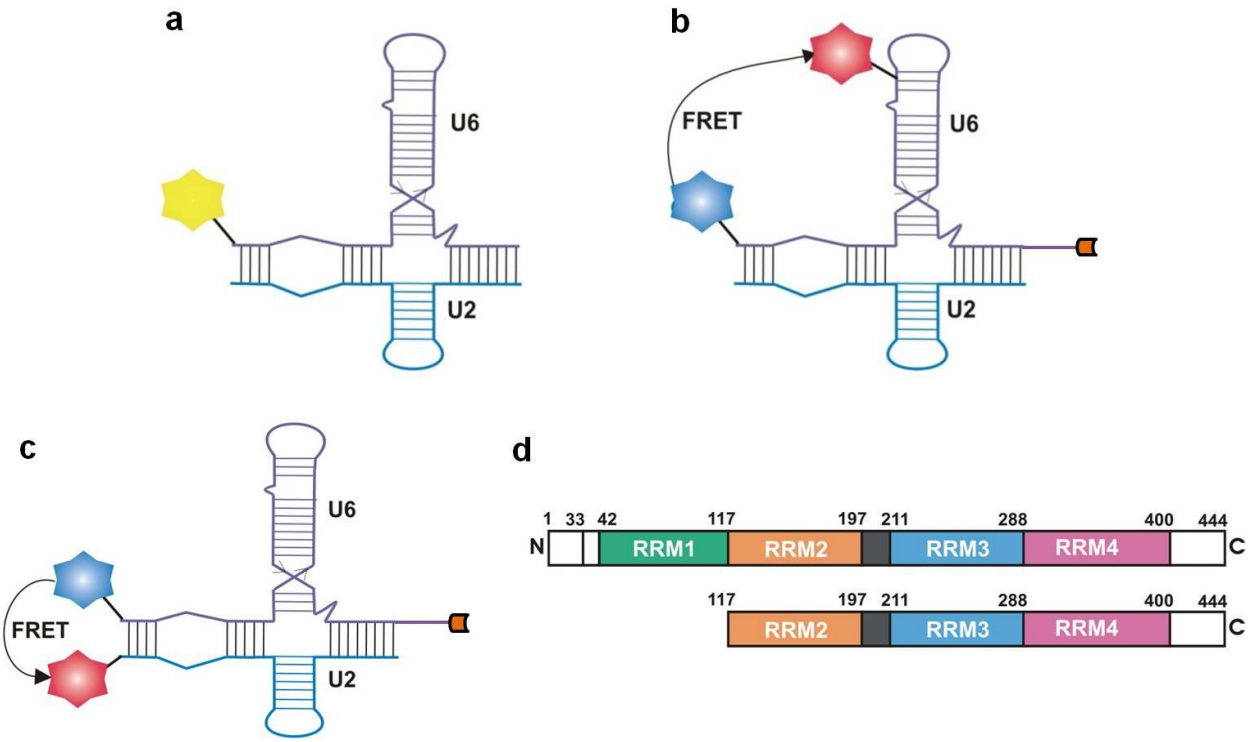

Figure 3.1: Schematic representation of RNA strands and proteins used in this study. (a) U2/U6 complex used in the fluorescence anisotropy study, where U6 is labelled with fluorescein (yellow) at the 5' end and U2 is unlabelled. (b) U2/U6 complex used in the first part of the study, where U6 is labelled with cy3 (blue) at the 5' end, cy5 (red) around ISL loop and biotin (orange) at the 3' end. U2 is unlabelled. (c) U2/U6 complex used in the second part of the study, where U6 is labelled with cy3 at the 5' end and biotin at the 3' end. U2 is labelled with cy5 at the 3' end. In (a), (b) and (c) U6 snRNA strand is in purple and U2 snRNA strand is in blue colour. (d) Two proteins used; full length Prp24 (Prp24, upper) and a truncated Prp24 where first RRM and the N-terminus region have been removed (234C, bottom).

3.3: Results

3.3.1: EMSA and anisotropy measurements reveal binding of Prp24 with the U2/U6 complex

To characterize binding of Prp24 to U2/U6 complex, I have first used EMSA by incubating Proteins (Prp24 and 234C, 4 μM each) with preformed U2 (4μM)/U6 (2 μM) complex with doubly labelled U6 (Figure 3.2a). Figure 3.2a, lane 1 has U6 only, represented by the high intensity band. The low intensity band, migrating slower than the main band in lane 1, may correspond to randomly folded U6, given that it was not observed in a denaturing gel. The shift in the band in the second lane corresponds to formation of the U2/U6 complex. The slow migrating band in the 3rd lane corresponds to Prp24 binding with U2/U6 complex. Interestingly, the band corresponding to the protein-RNA complex seems to have a lower FRET state (yellow) compare to the U2/U6 complex (orange), suggesting that Prp24 binding induces a rearrangement in the U2/U6 complex, resulting in a low FRET value. The 4th lane has a very faint band that corresponds to the protein-RNA complex, suggesting that the binding of truncated protein 234C is decreased in comparison to that of the full-length protein, in agreement with the previously published data[206]. In the 3rd and 4th lanes, some of the samples were retained in the wells, which could be due to formation of higher order complexes with RNA and proteins.

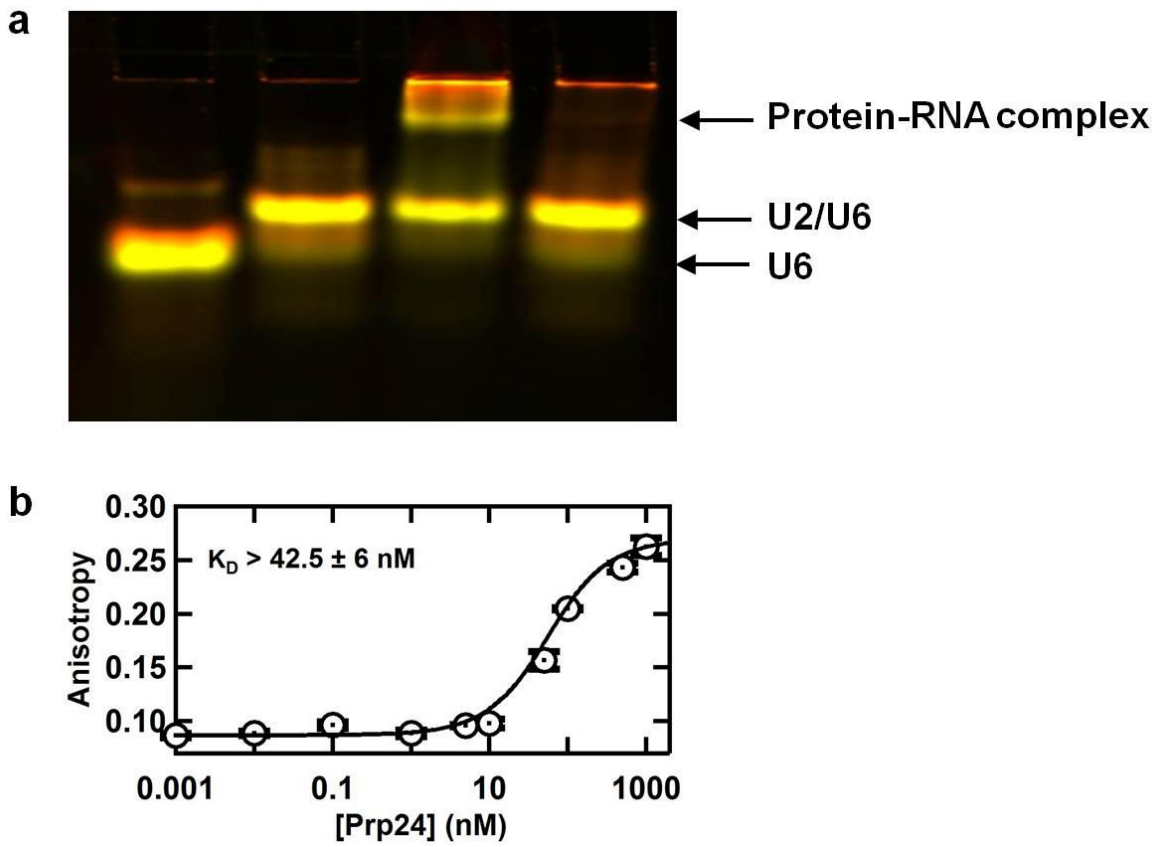

Figure 3.2: Prp24 binds to U2/U6 complex. (a) Gel-shift assay with Prp24 (full length and truncated; 234C, 4 μM each) and the pre-formed U2 (4 μM) /U6(2 μM) complex in which U6 is labelled with both cy3 and cy5 . Gel was scanned by exciting with 532 nm and 635 nm lasers to detect Cy3-U6 and Cy5-U4, respectively. Resulting image is an overlay of both channels. Lane 1: U6 only, Lane 2: U2/U6 complex, Lane 3: U2/U6 with Prp24, Lane 4: U2/U6 with 234C. (b) Fluorescence anisotropy experiment with pre-formed U2 (50 nM)/U6 (25 nM) complex at different Prp24 concentrations. Anisotropy data was plotted against Prp24 concentration and fitted to a quadratic binding equation.

For further characterization of Prp24 binding to the U2/U6 complex, fluorescence anisotropy experiments was performed. For these studies, a pre-formed U2 (50 nM)/U6 (25 nM) complex were utilized, in which the 5' end of U6 was labelled with a fluorescein dye. A relatively low anisotropy value (r= 0.087, Figure 3.2b) was obtained for the U2/U6 complex under standard conditions (50mM Tris-HCl, pH 7.5, 100mM NaCl, 10mM $MgCl_2$). In the presence of Prp24, the anisotropy value was increased, indicating the binding of protein to the complex, as a result of the hindrance of free rotation of the fluorescein dye attached to the U6 strand. This result demonstrates that the anisotropy value has gradually increased with increasing concentration of Prp24 and saturated nearly 1μM (anisotropy value of ~ 0.26), yielding a dissociation constant (K_D) of 42 ± 6 nM (Figure 3.2b). Since, in fluorescence-anisotropy experiments RNA concentration is limited (at least 25nM fluorescein-labelled RNA needs to be used for a better signal), the result in binding affinity is not the actual but an apparent value. The apparent K_D value obtained from the anisotropy experiments is in accordance with the previously published gel shift data with an apparent K_D value of 43 ± 11 nM[206].

3.3.2: Binding of Prp24 affects the structural dynamics of U2/U6 complex

To determine the influence of Prp24 binding on the conformational dynamics of the U2/U6 complex, sm-FRET experiments were done in the absence and presence of Prp24. Preformed U2/U6 complexes were diluted and then immobilized on a quartz slide and correspondingly the protein was introduced. The sample was then excited using a 532 nm laser and the resulting intensities were captured through the objective and detected through a CCD camera (Figure 3.3a). The data shows that the U2/U6

complex has different conformations as three colours of dots; green, yellow and red, represented in the single molecule image (Figure 3.3b). The resulting trajectories and the corresponding histograms (Figure 3.3c,d and e) illustrate that the U2/U6 complex adopts three different FRET populations; 0.2, 0.4 and 0.6, similar to a previous study done by our lab, which proposed that U2/U6 is in a dynamic equilibrium with at least three distinct conformations (Figure 1.17)[185].

In order to compare the results I obtained with the previous study, single molecule experiments were performed using the construct used in the previous study, where the U6 ISL loop is removed (named as three-piece construct; Figure 3.3f). With this construct, I have observed three interchanging FRET populations (Figure 3.3g) consistent with the previous study with a similar construct and with the construct containing the ISL loop, validating that the presence of ISL loop has no effect on the dynamic nature of the complex observed[185].

In the presence of 100nM Prp24, mostly green dots appeared in the single molecule image (Figure3.4a), suggesting that the Prp24-bound U2/U6 complex adopts a single conformation compared to the three dynamic FRET states observed for RNA alone. However, the resulting FRET trajectories and the histograms show that the Prp24-bound U2/U6 complex has two static FRET populations; 0.2 and 0.4 (Figure 3.4b and c). Upon analysis of the total number of molecules in the presence of 100nM Prp24 (n=156), a majority of the molecules were in the static 0.2 state (~85%), indicating that the binding of Prp24 affects the structural dynamics of U2/U6 complex (Figure 3.4b). Only about 15% of the molecules were in the static 0.4 state, suggesting that this minor population can be a misfolded population.

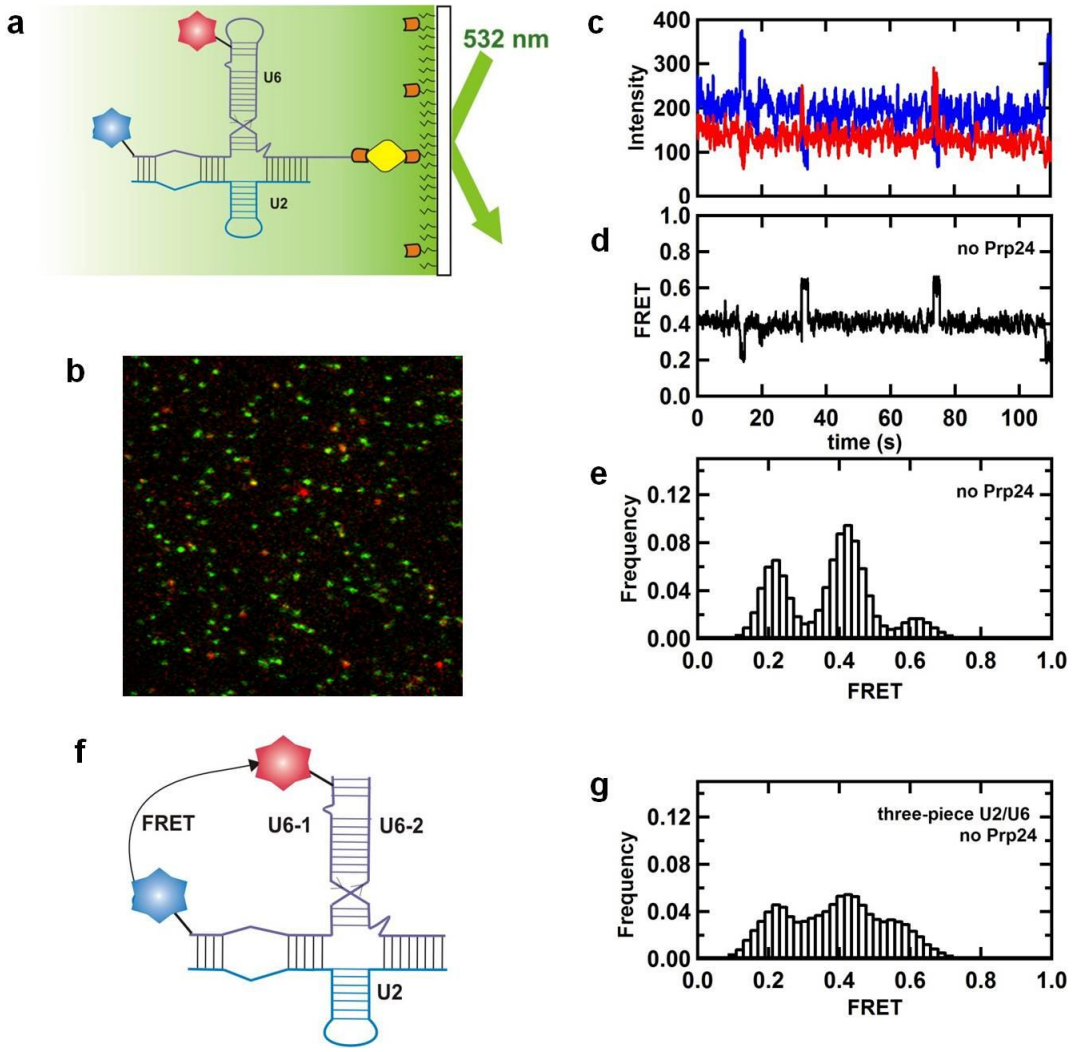

Figure 3.3: U2/U6 adopts three distinct conformations. (a) Single molecule experimental setup for U2/U6 complex. The U2/U6 complex is immobilized on quartz slide via streptavidin-biotin interactions. Donor molecules (cy3, blue) are excited by a 532 nm laser and energy transfer from the donor to the acceptor (cy5, red) molecules are measured and corresponding FRET values are calculated. (b) Single-molecule movie obtained for the U2/U6 RNA duplex alone, showing that the molecules adopt three colours; green, orange and red. (c) Intensity and (d) FRET time trajectories and (e) the corresponding histogram for the RNA duplex alone showing three interchanging FRET states naming 0.2, 0.4 and 0.6. (f) The RNA construct used in the previous study where the U6 ISL loop was removed and (g) resulting histogram for the single-molecule experiment using the three-piece construct (U6-1, U6-2 and U2) indicating that U6 ISL loop has no effect on the dynamic nature of U2/U6 complex. U2 and U6 are colour coded as same as previous.

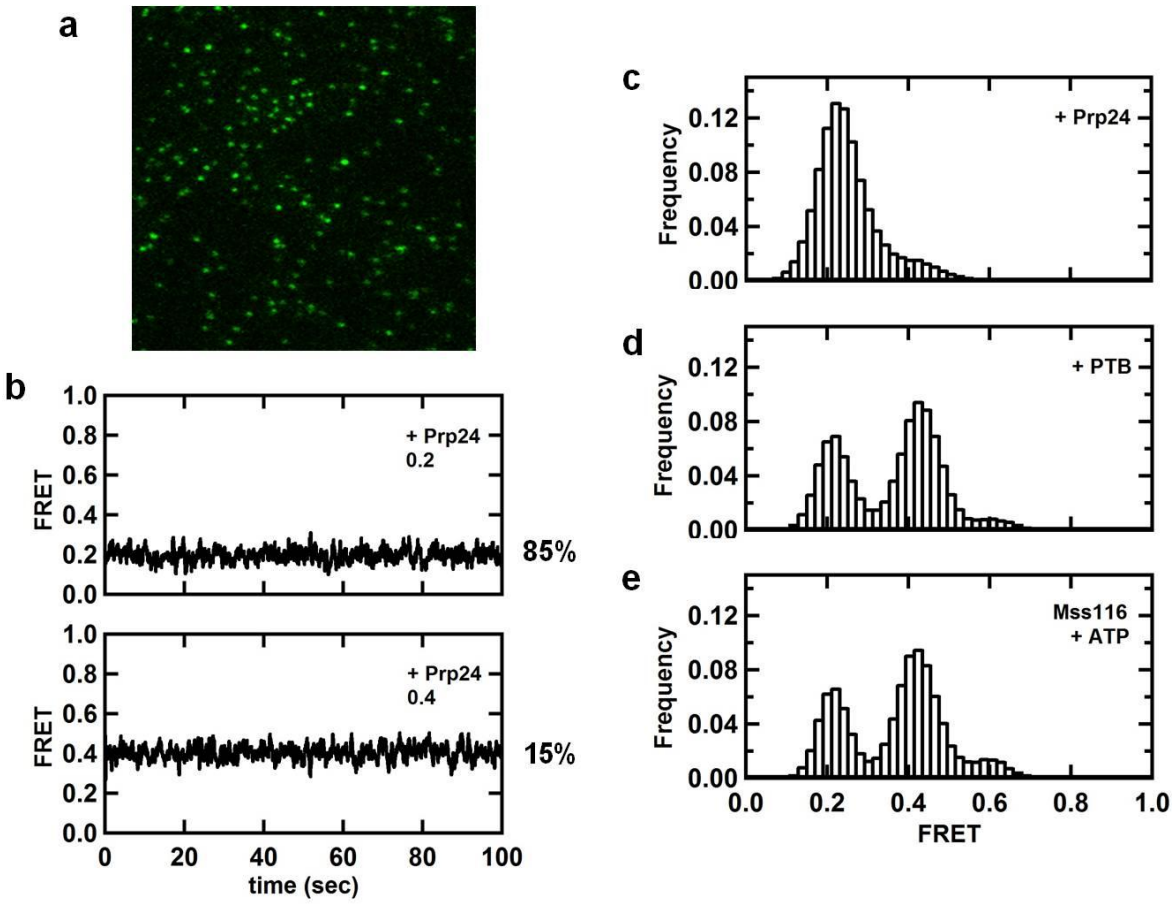

Figure 3.4: Binding of Prp24 affect the dynamic nature of the U2/U6 complex. (a) A single-molecule movie obtained in the presence of 100nM Prp24, showing majority of the molecules are green in colour. (b) Single-molecule FRET time trajectories and (c) the corresponding histogram in the presence of Prp24 showing two static populations; 0.2 and 0.4. Binding of Prp24 has shifted the dynamic population to a static population, revealing a conformational change in the U2/U6 complex. Histograms for the single-molecule control experiments with (d) 100nM PTB and (e) 100nM Mss116 proteins reveal that conformational change in U2/U6 is specific for Prp24.

In order to confirm that the binding and stabilizing low FRET state of U2/U6 complex is Prp24-specific, single-molecule FRET experiments with PTB (polypyrimidine tract binding) protein and Mss116 at 100 nM concentrations were done. PTB protein is a RNA binding protein with RNA recognition motifs similar to the Prp24[192], whereas, Mss116 is an ATP-dependent RNA helicase, which unwinds RNA helices[238]. Interestingly, resulting histograms (Figure 3.4d, e) show that there is no change in the FRET populations even at a higher concentration of both PTB and Mss116. This suggests that the stabilization of low FRET state is corresponding to Prp24-specific conformational changes in the U2/U6 complex.

3.3.3: Effect of RRMs on Prp24 binding with U2/U6 complex

To characterize the effect of Prp24 binding on the dynamic nature of U2/U6, single-molecule experiments at various concentrations of Prp24 and 234C were performed[206]. As shown in the histogram (Figure 3.5a, top), in the absence of Prp24, molecules were dynamic between three FRET states (0.2, 0.4 and 0.6). On the other hand, with increasing concentrations of Prp24, molecules tended to stay in the 0.2 FRET state compare to the other two states. Also at 5nM or higher concentrations, more molecules were present at a static 0.2 state, suggesting that binding of Prp24 stabilizes a low FRET conformation of the duplex (Figure 3.5a). Moreover, the 0.6 FRET population decreased upon the increase in Prp24 concentration, whereas at concentration above 5 nM it has completely disappeared (Figure 3.5a).

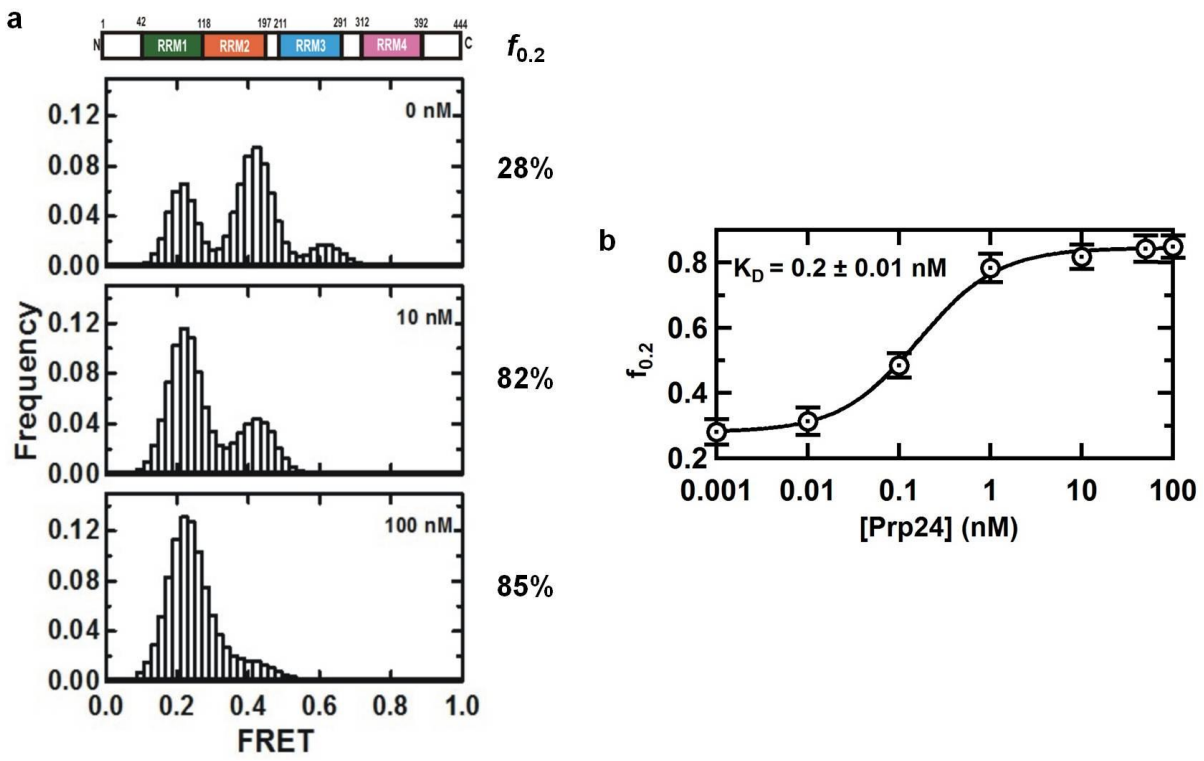

Figure 3.5: Prp24 binding to U2/U6 stabilizes a low FRET (0.2) conformation. (a) Effect of Prp24 on the conformational dynamics of U2/U6. Each panel corresponds to an average smFRET histogram from >100 single-molecule trajectories at the indicated full-length protein concentrations. The fraction of molecules in the 0.2 FRET state at each protein concentration was calculated and demonstrated as a percentage. (b) The fraction of molecules in the 0.2 FRET state is plotted against Prp24 concentrations and fitted to a Langmuir equation to yield the K_D for the protein binding.

Although the 0.4 population has also decreased with the increasing concentration of Prp24, a minor fraction of molecules (~15%) showed a static 0.4 FRET state even at 100nM of Prp24. The fraction of molecules remaining at 0.2 FRET state was then plotted against Prp24 concentration to determine the binding affinity of Prp24 to the U2/U6 complex (Figure 3.5b). This clearly shows that increases in Prp24 concentration leads to an increase in the 0.2 population, which eventually reaches a plateau around 5 nM of Prp24 and yields a K_D of ~0.2 nM. This indicates a tight binding of Prp24 with U2/U6 and stabilizing a conformation that corresponds to a FRET of 0.2. This could be a conformation that contains either U2/U6 or U6 only. Further experiments were done to determine whether binding of Prp24 unwinds U2 or not as explained in later section. Previous studies have shown that the first RNA recognition motif of Prp24 is important for the high affinity binding of the protein with U6[206]. To characterize the function of RRM1 in Prp24 binding and its effect on U2/U6 unwinding by Prp24, I have done single-molecule experiments using 234C protein in which the RRM1 was truncated. Experiments were done in the presence of different concentrations of 234C, where an increase in the 0.2 FRET state was observed with increasing concentrations of the 234C protein. However, the increase in the 0.2 population in the presence of 234C is comparatively lower than that of Prp24, where even at 100nM 234C concentration only ~58% of molecules were found to be in 0.2 FRET state (Figure 3.6a). Binding curve drawn with the fraction of molecules stay at 0.2 FRET state as a function of 234C concentration yields a K_D of ~0.4nM, which is a two fold increase than that of Prp24 (Figure 3.6b). This suggests that RRM1 is important for high affinity binding of Prp24 with U2/U6 complex and thus removal of RRM1 has lesser effect on the protein binding

to the complex, which in turn lowers the effect on stabilization of low FRET conformation.

3.3.4: Binding of Prp24 facilitates the unwinding of U2

Once Prp24 unwinds the U2/U6 duplex it could remain bound to U6. This bound U6-Prp24 complex can form a different conformation compared to that of the U2/U6 complex. This new conformation of U6 may correspond to the 0.2 FRET population observed with higher concentrations of Prp24. To determine whether the Prp24 facilitates the removal of U2 and stabilizes a U6 conformation by remaining bound to it, EMSA studies followed by single-molecule experiments were done. For this set of experiments, the same U2/U6 construct was used, however with a different labelling pattern, where U6 is labelled with cy3 and U2 is labelled with cy5 (Figure 3.1c). In the resulting EMSA gel, the green and red bands in lane 1 and 2 correspond to U6 and U2 RNA respectively. The shifted band in the lane 3 corresponds to the U2/U6 complex (Figure 3.7a). Based on the labelling pattern, the reddish colour of the shifted band in lane 3 represents a higher FRET state, which indicates the formation of U2/U6 complex. The slow migrating band in lane 4 corresponds to the binding of Prp24. Interestingly, an increase in free U2 can be seen in the lane 4 compared to lane 3 containing U2/U6 duplex, suggesting that the binding of Prp24 has resulted in removal of U2 from the complex. Hence, the slow migrating green band in lane 4 possibly corresponds to the free U6 (labelled with cy3), after the removal of U2 strand. Altogether, these results suggest that the 0.2 FRET population observed in the previous set of experiments results from the formation of a U6 conformation upon U2 removal. The very faint band in

the 5th lane containing the truncated protein (234C), corresponds to the RNA-protein complex, illustrating that the removal of RRM1 causes a lesser effect on the U2 removal compared to the full length protein.

To determine the role of Prp24 in U2/U6 unwinding, single-molecule flow experiments were performed, in which Prp24 was introduced to the slide in real time. In this experiment, a single-molecule movie was taken for a few minutes in the absence of Prp24 and then Prp24 was introduced while the movie was recording. The resulting FRET trajectory shows that in the absence of Prp24, the molecule stayed in the high FRET state (FRET = 1.0), whereas, upon addition of 50nM Prp24, the FRET state changed from one to zero via an intermediate state of 0.2 (Figure 3.7b). In order to validate this result, a control experiment was done by introducing the buffer solution without Prp24. The resulting trace shows that the molecules remain at a higher FRET state, even after the buffer flow was introduced (Figure 3.5c). This suggests that the observed zero FRET state correspond to the U2 dissociation upon binding of Prp24. The short lived 0.2 FRET state may corresponds to an intermediate conformation with partially unwound U2/U6 complex. Characterization of this intermediate has shown that ~ 15% of total molecules are going through the 0.2 FRET state during the transition from high FRET to zero FRET, with an average life time of 14s.

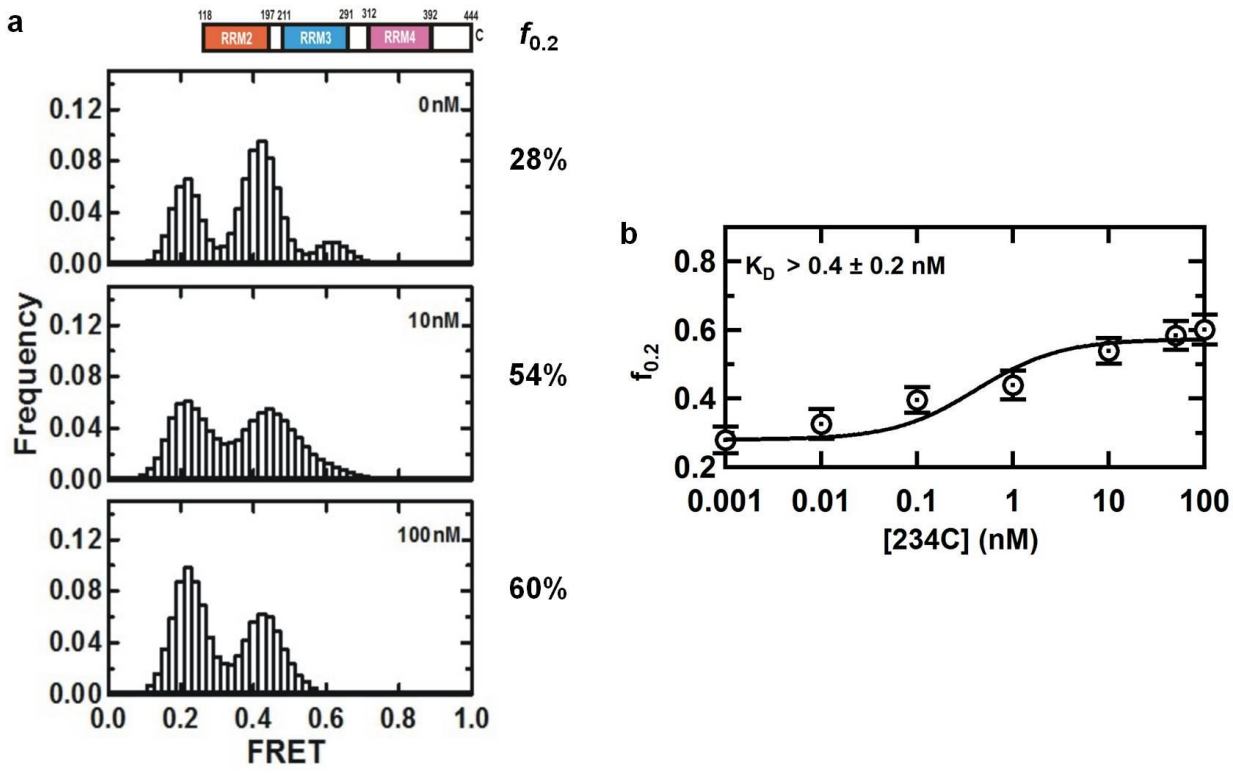

Figure 3.6: Removal of RRM1 affects the binding of Prp24 to the U2/U6 complex. (a) Effect of 234C on the conformational dynamics of U2/U6. Each panel corresponds to an average smFRET histogram from >100 single-molecule trajectories at the indicated full-length protein concentrations. (b) The fraction of molecules in the 0.2 FRET state is plotted against 234C concentrations and fitted to a Langmuir equation to yield the K_D for the protein binding.

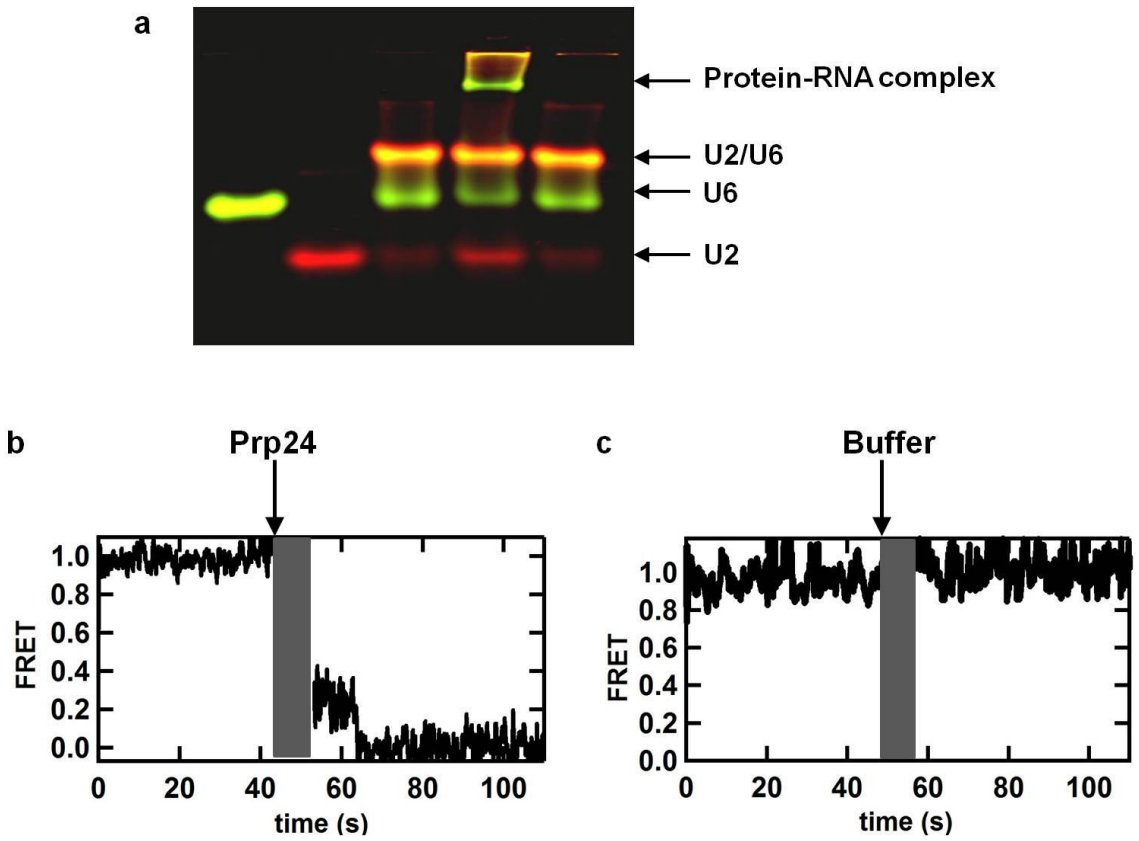

Figure 3.7: Prp24 facilitates the U2/U6 unwinding. (a) EMSA with Prp24 (full length and truncated; 234C, 4μM each) and pre-formed U2 (2μM)/U6 (2μM) complex, in which the U2 is labelled with cy5 and U6 with cy3. Gel was scanned by exciting with 532 nm and 635 nm lasers to detect Cy3 and Cy5, respectively. Resulting image is an overlay of both channels. In the image each lane contains; lane 1: U6-cy3, lane 2: U2-cy5, lane 3: U2/U6 complex, lane 4: U2/U6 with Prp24, lane 5: U2/U6 with 234C. Characteristic FRET trajectories for single-molecule flow experiment (b) with Prp24 and (c) without Prp24, illustrating the introduction of Prp24 results in a FRET change from one to zero through an intermediate at 0.2, as a result of U2 removal.

Next, single-molecule time dependent experiments were done at different concentrations of Prp24 and 234C to verify the zero FRET state observed. Upon addition of 1nM Prp24, the zero FRET population was increased over time and at 45 min ~99% molecules were in zero FRET state (Figure 3.8a). Whereas, in the presence of 234C, the zero FRET population was relatively low (Figure 3.8b). As a control, similar sets of experiments were done with the RNA complex without protein. The observed zero FRET state in the absence of Protein represent the fluorophore photobleaching (Figure 3.8c). Each time point was corrected for photobleaching by considering the difference between two zero FRET fractions; in the absence and in the presence of Prp24. The corrected value corresponds to the actual change in the zero FRET population as a result of U2 removal upon Prp24 binding. These data were then fitted to a mono-exponential function, indicating that the zero FRET population was increased over time and saturated nearly 25 minutes, yielding a U2 removal rate (k_{obs}) of ~ 0.15 min^{-1} (Figure 3.9a). Similar experiments were done in the presence of 1nM 234C, which has shown a decrease in the U2 removal rate by four fold (k_{obs} ~ 0.8 min^{-1}, Figure 3.9b). This indicates that the truncated protein has a lesser effect on the U2 removal, suggesting that RRM1 is important for the binding of Prp24.

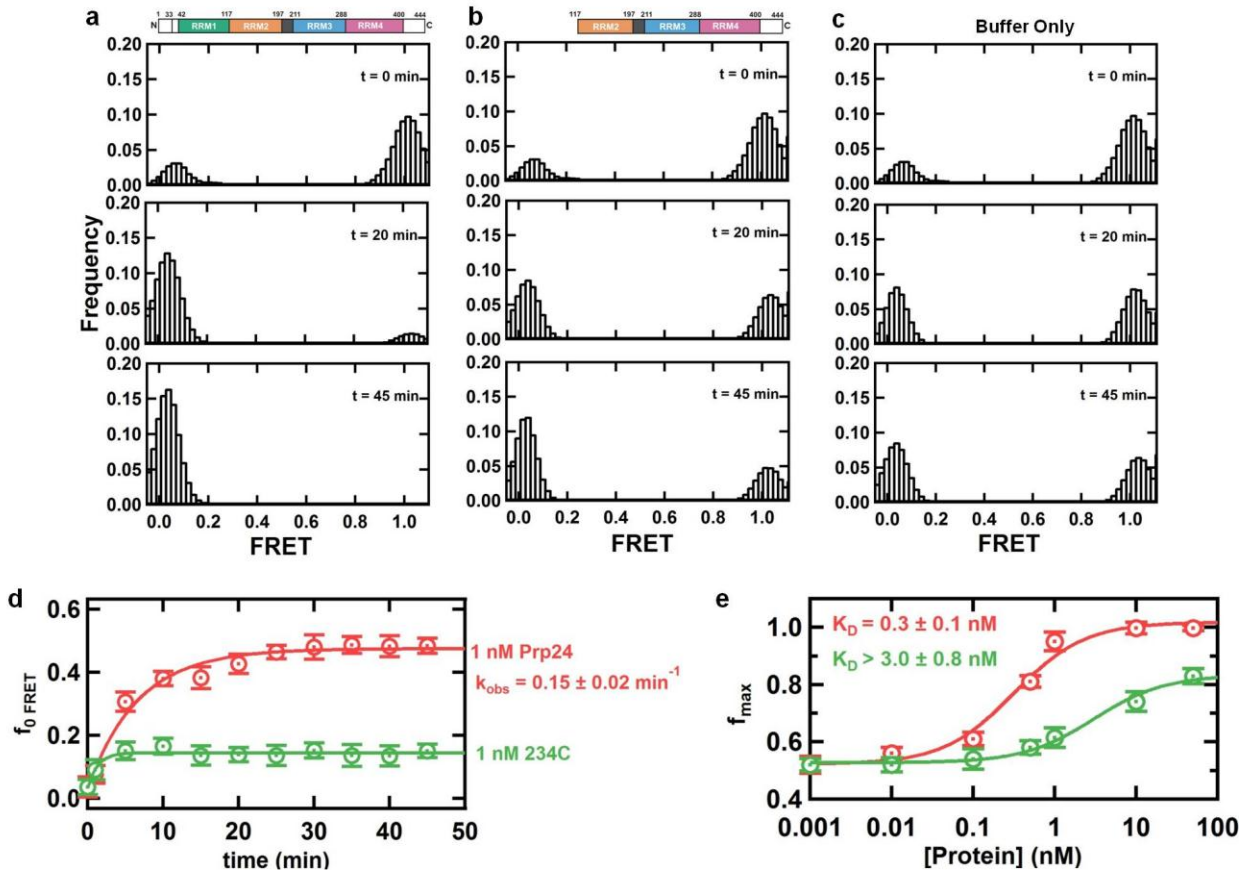

Figure 3.8: Binding of Prp24 facilitates the removal of U2. Histograms correspond to single molecule time dependent experiments with (a) Prp24, (b) 234C and (c) buffer only, showing the change in zero FRET state along with the time. (d) Fraction of molecules at zero FRET state as a function of time was plotted to obtain the rate of U2removal in the presence of Prp24 (red) and 234C (green). (e) Fraction of molecules at zero FRET state as a function of Prp24 concentration (red) and 234C concentration (green) was plotted to obtain the dissociation constant for U2 removal. In both (b) and (c), each point has corrected for the photobleaching rate.

Likewise, the fraction of molecules at zero FRET state corresponds to the last time point (45 min) that was plotted as a function of Prp24 concentration and fitted into a modified hill equation. The resulting plot demonstrates an increase in the zero FRET population over Prp24 concentration and yields a K_D of ~ 0.3 nM (Figure 3.9c). Contrary, in the presence of 234C the binding affinity was reduced by ~ 10-fold compared to that of Prp24 (K_D ~ 3.0 nM, Figure 3.9d). These results suggest that Prp24 facilitates the dissociation of U2 from U6 as a function of time and Prp24 concentration. On the other hand, the removal of RRM1 reduced the Prp24 binding and thus showed a reduced effect on the U2 removal rate.

With these data we proposed a model for the role of Prp24 in the removal of U2 from a minimal U2/U6 complex. When Prp24 binds with the minimal U2/U6 complex, first the RRM2 interacts with nucleotides 46-58, containing the ACAGAGA loop and then disrupts the base pairing between U2 and U6 in helix I. According to a recent crystal structure, RRM3 undergoes a 180^{o} rotation and threads through U6. During this rearrangement, RRM3 interacts with the U6 nucleotides at the 5' end, which results in an unwinding of helix III of the complex. Furthermore, the rearrangement of RRM3 pushes RRM4 towards the opposite side of U6, which eventually interacts with the lower region of U6 ISL and destabilizes the U6 ISL base region. These disruptions in the helix I, III and lower region of ISL, result in the formation of partially unwound complex as shown in Figure 3.9, which could correspond to the intermediate FRET population (0.2 FRET) that was observed in the flow experiments. These sequential events will ultimately lead to a destabilization of helix II, facilitating the complete removal of U2, as proposed in the recent crystal structure study. The conformation adopted by the free U6

binding to Prp24 after the U2 removal could represent the low FRET population obtained with the doubly labelled U6 and the zero FRET population obtained with the singly labelled U6. On the other hand, when RRM1 is removed, the truncated protein, 234C cannot bind to the U2/U6 complex as tight as full-length protein and thus it cannot destabilize the base pairing as efficiently as full-length protein, hence resulting in a lower rate and a lower dissociation constant than that of full-length protein.

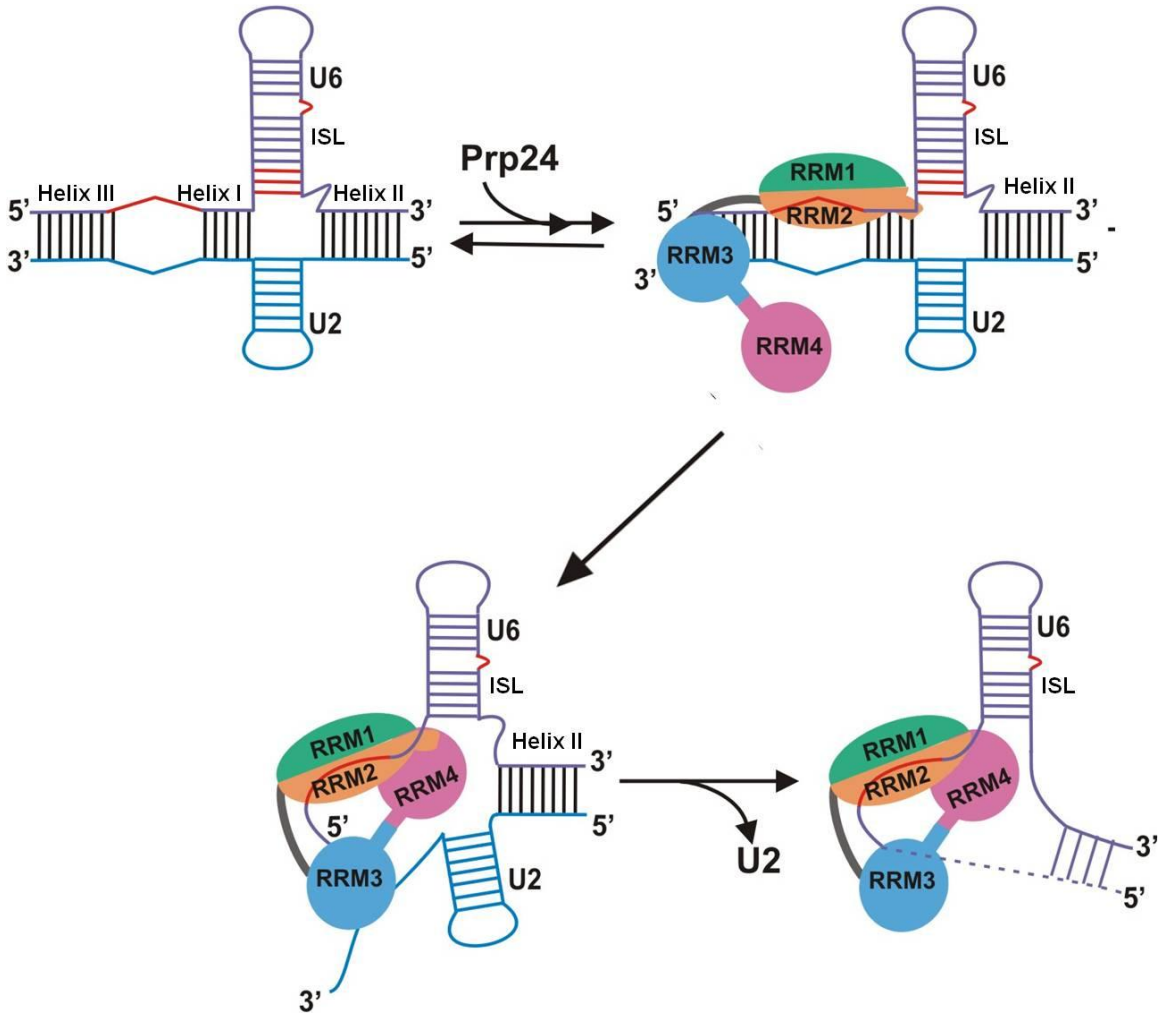

Figure 3.9: The proposed model for the role of Prp24 in U2/U6 complex unwinding. In the presence of Prp24, base pairing between U2 and U6 RNAs are disrupted and U2 is unwound from U6. The RNA recognition motifs of Prp24, U2 and U6 snRNA strands are colour coded as same as previous. The dash lines indicate the 5' nucleotides of U6 present in the crystal structure study but absent in our study.

CHAPTER 4: Study the assembly and global structure of U4/U6 complex

[This study has done as collaboration with the Kiyoshi Nagai Lab, MRC-LMB, Cambridge. The EMSA studies with the full length construct were done by Dr. John Hardin and Dr. Yasushi Kondo]

4.1: Objective

The objective of specific aim 2 of my research is to understand the conformational changes and formation of global structure of spliceosomal sub-complexes during the spliceosomal assembly cycle and the role of particle-specific proteins on these structural arrangements.

One of the important spliceosomal sub-complexes is U4/U6.U5 tri-snRNP, which consists of U4/U6 di-snRNP and U5 snRNP. Little is known about the U4/U6 di-snRNP structure and the only information available is a cryoEM structure of U4/U6.U5 tri-snRNP as a whole at ~30 nm resolution[77]. The relative orientation of the three helices of the U4/U6 duplex, the conformational dynamics of duplex upon protein binding, and the global structure of di-snRNP remains structurally unresolved.

To reveal aforementioned structural information and understand the assembly process of U4/U6 di-snRNP assembly, I have used single-molecule approach. First, EMSA studies were performed by the Nagai Lab to determine the order and affinities with which the proteins bind to the U4/U6 duplex. This allowed us to obtain step-wise reconstitution of the U4/U6 di-snRNP *in vitro,* and provided a protocol for complete

assembly of di-snRNP, which can be used for further studies, even for the assembly of other complexes with minor alterations. The single-molecule data then revealed that the two stems of the duplex; stem I and II are coaxially stacked. Moreover, it also shown that the U4/U6 di-snRNP adopts a pre-formed rigid structure, which does not change upon protein binding.

4.2: Experimental design

A full length U4 and U6 strands were used in the EMSA studies, where RNAs were labelled with fluorescein at the 3' end of U4 (Table 2.3, Figure 4.1a). For single-molecule FRET studies, I have used a minimal U6 strand (nucleotides from 54-80) and a minimal U4 strand (nucleotides from 1-70, Table 2.3, Figure4.1b). The U4 strand was labelled internally with cy5 (at the 5’ stem-loop) and at the 3’ end with biotin-BSA for surface immobilization (Figure 4.1b). To study the relative orientation of the helices, three U6 strands were used with different labelling schemes but with the same sequence (Figure 4.1b). All proteins (Snu13, Prp31, Prp3, Prp4, Sm proteins and LSm proteins) were expressed and purified by the Nagai Lab.

4.3: Results

4.3.1: *In-vitro* reconstitution of the U4/U6 di-snRNP

To characterize the step-wise assembly of the U4/U6 di-snRNP, EMSA studies were done by incubating purified proteins of the U4/U6 snRNP complex (Snu13, Prp31, Prp3/Prp4, Sm proteins and LSm proteins, Figure 1.11a) with a pre-formed full length

U4/U6 RNA complex. Binding was monitored as changes in electrophoretic mobility. The apparent binding affinities (K_D) are summarized in Table 4.1.

First each U4/U6 associated protein or protein sub-complex was titrated against the pre-formed U4/U6 snRNA duplex. The fraction of protein bound to U4/U6 duplex was measured as a function of protein concentration to obtain apparent binding affinities (K_D). Since previous studies have shown that Snu13 facilitates binding of other proteins, the U4/U6 RNA duplex was first titrated with different concentrations of Snu13, resulting in a K_D = 17 ± 1 nM (Figure 4.2a). The tight binding of Snu13 with U4/U6 duplex is in accordance with the previous studies[70]. Similarly, the snRNA duplex was titrated with Prp3/4, which gives an apparent K_D of 57 ± 2 nM (Figure 4.2b). A previous study on human spliceosomal components has shown that binding of human Snu13 ortholog to the U4/U6 duplex is required for further binding of hPrp3/4[69], which suggests that yeast Prp3/4 binding may utilize a different pathway than human Prp3/4. Similarly, Sm proteins exhibit a K_D of 89 ± 4 nM (Figure 4.2c). The binding affinity of Prp31 to the duplex is very low (K_D = 243 ± 16 nM, Figure 4.2d), which is in accordance with the previous studies showing that binding of Snu13 to the duplex facilitates the Prp31 binding[72]. On the other hand, titration of LSm2-8 complex with the full-length U4/U6 complex exhibits a very high affinity (K_D = 5.0 ± 0.2 nM, Figure 4.2e). Titrations with individual proteins have shown that Snu13 exhibits very tight binding, and the smallest shift in the gel is due to its small size. This suggests that Snu13 could be the starting point for the further assembly of other proteins, which also in agreement with the previous findings[72]. Therefore, we have determined the binding affinity of each protein (Prp31, Sm, LSm, and Prp3/Prp4) with a fully formed U4/U6 +Snu13 complex.

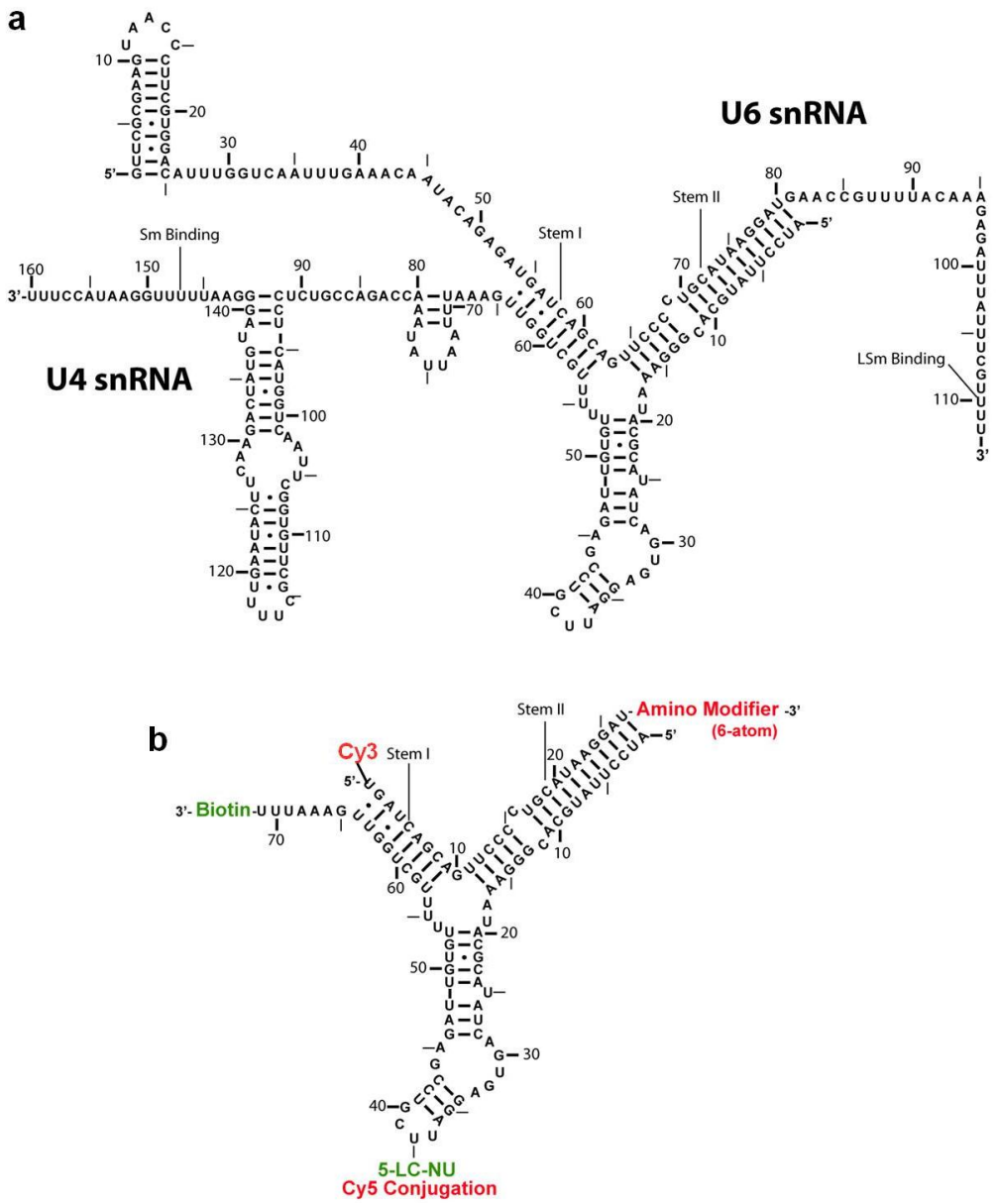

Figure 4.1: U4/U6 snRNA duplex. (a) Sequence of full length U4 and U6 used in the EMSA studies. (b) Minimal U4 and U6 sequences used in the single-molecule studies, which don't contain the single-stranded regions of U4 and U6 known to be the binding sites for Sm and LSm proteins, respectively. The U6 strand is labelled with Cy3 at the 5' end, and a C6 amino modifier is attached at the 3' end. In stem I-II construct, Cy5 is attached to this amino modifier. The U4 strand consists of an amino modifier at the 5' stem-loop to attach Cy5 and Biotin is attached to the 3' end.

Table 4.1: Apparent K_d ($K_{d, app}$) measurements for step-wise *in vitro* assembly of the *Saccharomyces cerevisiae* U4/U6 di-snRNP as determined by EMSA studies. Associated error is reported as one standard deviation from the weighted mean (The binding affinities were kindly provided by Dr. J. Hardin and Dr. Y. Kondo).

Components	Titrated Component	K_D
U4/U6	Snu13	17 ± 1 nM
U4/U6	Prp3/4	57 ± 2 nM
U4/U6	Sm	89 ± 4 nM
U4/U6	Prp31	243 ± 16 nM
U4/U6	LSm	5 ± 0.2 nM
U4/U6/Snu13	Prp31	50 ± 4 nM
U4/U6/Snu13	Sm	92 ± 9 nM
U4/U6/Snu13	LSm	26 ± 1 nM
U4/U6/Snu13	Prp3/4	88 ± 8 nM
U4/U6/Snu13/Prp31	Sm	108 ± 9 nM
U4/U6/Snu13/Prp31	LSm	98 ± 34 nM
U4/U6/Snu13/Prp31	Prp3/4	417 ± 43 nM
U4/U6/Snu13/Prp31/Sm	LSm	154 ± 15 nM
U4/U6/Snu13/Prp31/Sm	Prp3/4	557 ± 75 nM
U4/U6/Snu13/Prp31/Sm/LSm	Prp3/4	20 ± 1 nM

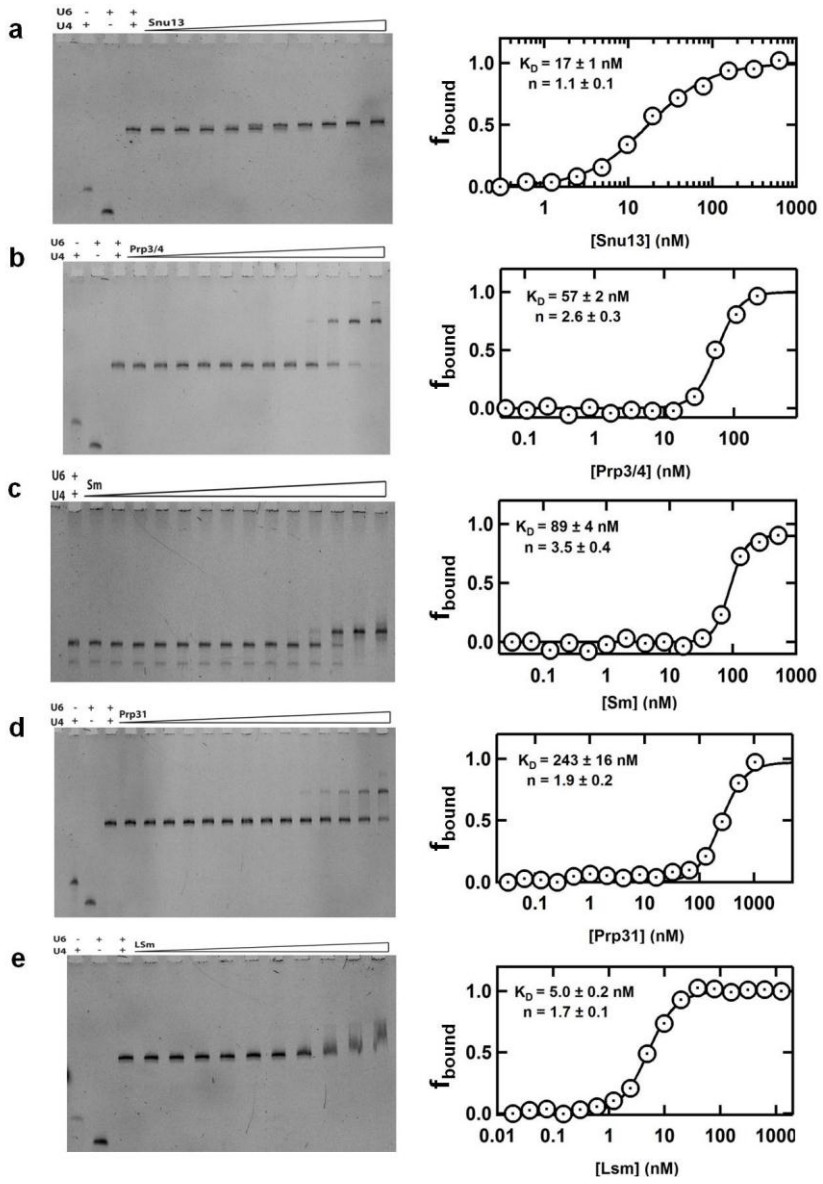

Figure 4.2: Gel shifts and binding curves for the interaction between the pre-formed U4/U6 duplex RNA and proteins. The pre-formed U4/U6 duplex RNA is titrated with varying concentrations of (a) Snu13, (b) Prp3/4, (c) Sm proteins, (d) Prp31, and (e) LSm proteins. The concentration of pre-formed U4/U6 duplex is 2 nM. The concentration of proteins ranges from 0 to ~2 μM in two-fold increments. The apparent K_D and estimated Hill coefficient (n) is given for a single representative binding curve (The gel pictures and binding data were kindly provided by Dr. J. Hardin and Dr. Y. Kondo).

The binding curve indicates that in the presence of Snu13, binding of Prp31 has greatly increased (K_D = 50 ± 4 nM, Figure 4.3a). On the other hand, the binding of Sm proteins shows no increase in the presence of Snu13 (K_D= 92 ± 9 nM, Figure 4.3b), indicating there is no direct interaction between Sm proteins and Snu13. Also, LSm binding to the pre-formed U4/U6 +Snu13 complex remains high (K_D = 26 ± 1 nM, Figure 4.3c), despite the ~5 fold decrease in the affinity compared to that for naked duplex RNA. Similarly, the binding affinity of Prp3/4 for the pre-formed U4/U6 +Snu13 complex is slightly reduced (K_D = 88 ± 8 nM, Figure 4.3d) from that observed for binding to the naked U4/U6 duplex RNA.

As the next assembly step, we have used pre-formed U4/U6 +Snu13 +Prp31 complex to examine the binding affinities of each of the remaining components (Sm, LSm, and Prp3/Prp4). The binding affinity for Sm proteins shows no significant change from what has observed for the binding of these proteins to the naked U4/U6 duplex (K_D = 108 ± 9 nM, Figure 4.4a). On the other hand, LSm proteins show a substantially lower binding affinity (K_D = 98 ± 34 nM, Figure 4.4b). Prp3/4 exhibits significantly weaker binding to the pre-formed U4/U6 +Snu13 +Prp31 complex than that observed for the naked RNA duplex (K_D = 417 ± 43 nM, Figure 4.4c).

Next, we have examined the binding of LSm and Prp3/4 proteins to the fully formed U4/U6 +Snu13 +Prp31 +Sm complex. Both LSm proteins and Prp3/4 show drastic reduction in apparent binding affinities (K_D = 152 ± 15 nM and K_D = 557 ± 75 nM respectively; Figure 4.4d and e) compare to the naked U4/U6 duplex.

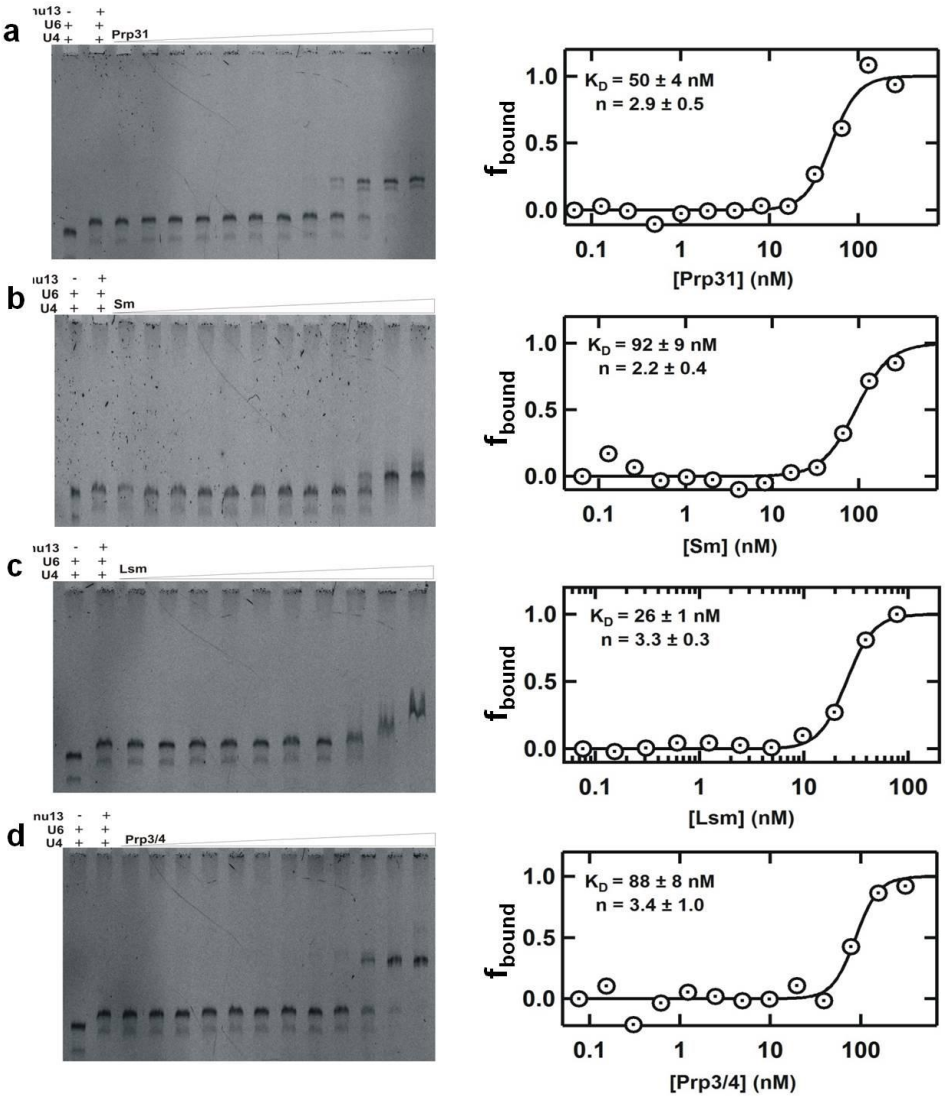

Figure 4.3: Gel shifts and binding curves for the interaction between the pre-formed U4/U6/Snu13 complex and proteins. The pre-formed U4/U6 duplex RNA is titrated with varying concentrations of (a) Prp31, (b) Sm proteins, (c) LSm proteins, and (d) Prp3/4. The concentrations of component used are; 2 nM pre-formed U4/U6 duplex and 200 nM Snu13. The concentration of proteins ranges from 0 to ~2 μM in two-fold increments. The apparent K_D and estimated Hill coefficient (n) is given for a single representative binding curve (The gel pictures and binding data were kindly provided by Dr. J. Hardin and Dr. Y. Kondo).

Finally, we have examined binding of Prp3/4 to the pre-formed U4/U6 +Snu13 +Prp31 +Sm +LSm complex. Interestingly, when all other U4/U6 components are present, the Prp3/4 binding affinity increases dramatically (K_D = 20 ± 2 nM, Figure 4.5a). This may be due to the presence of alternative conformations of the U6 3'-end, which may hinder the binding of Prp3/4. However the presence of LSm proteins may minimize these alternative conformations, clearing the path for Prp3/4 binding.

In summary, we have showed the stepwise assembly of U4/U6 di-snRNP complex (Figure 4.5b). To our knowledge, this is the first *in vitro* reconstitution of complete U4/U6 di-snRNP, and the first time of step-wise assembly of any snRNP with all its components *in vitro*.

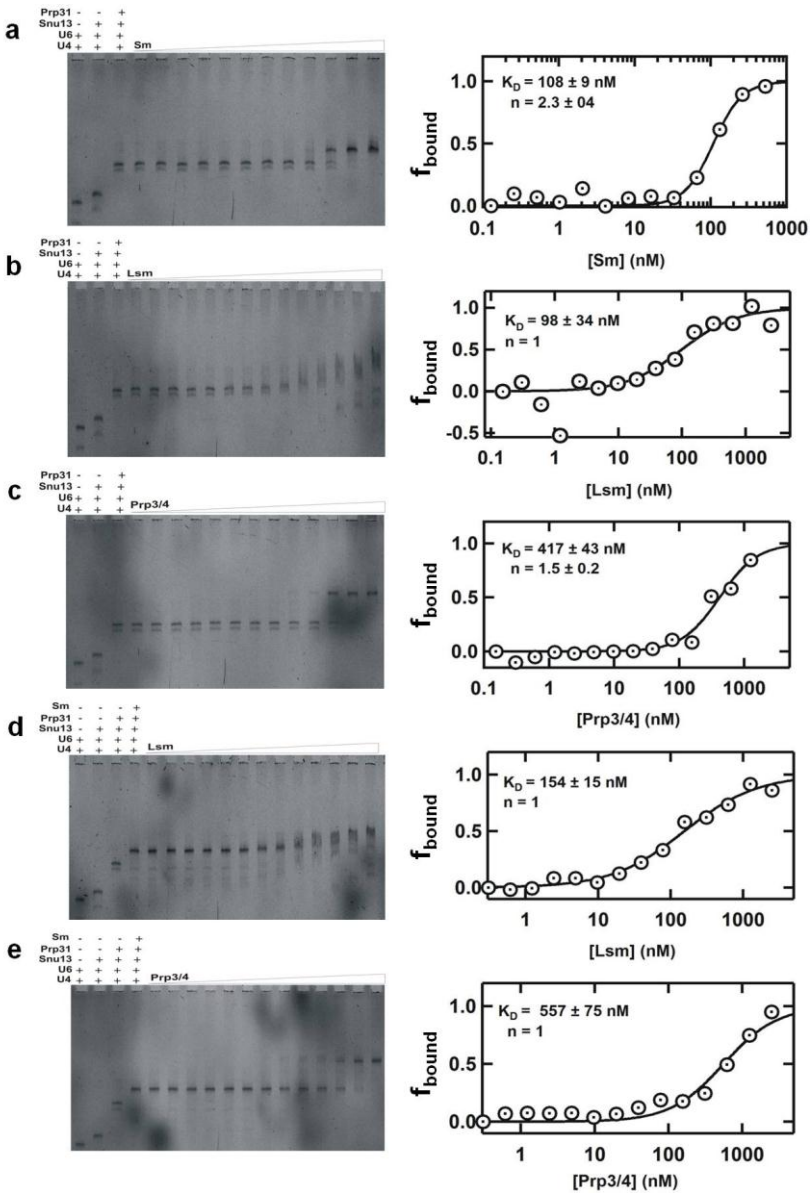

Figure 4.4: Gel shift assays and binding curves for the interaction between the pre-formed U4/U6/Snu13/ Prp31complex and proteins. The pre-formed U4/U6 duplex RNA is titrated with varying concentrations of (a) Sm proteins, (b) LSm proteins, and (c) Prp3/4. Gel shift assays and binding curves for the interaction between the pre-formed U4/U6/Snu13/ Prp31/Sm complex and (d) LSm proteins, and (e) Prp3/4. The concentrations of the component used are; 2 nM pre-formed U4/U6 duplex, 200 nM Snu13 and 120 nM Prp31 in d and e panels. Protein concentration ranges from 0 to ~2 μM in two-fold increments. The apparent K_D and estimated Hill coefficient (n) is given for a single representative binding curve (The gel pictures and binding data were kindly provided by Dr. J. Hardin and Dr. Y. Kondo).

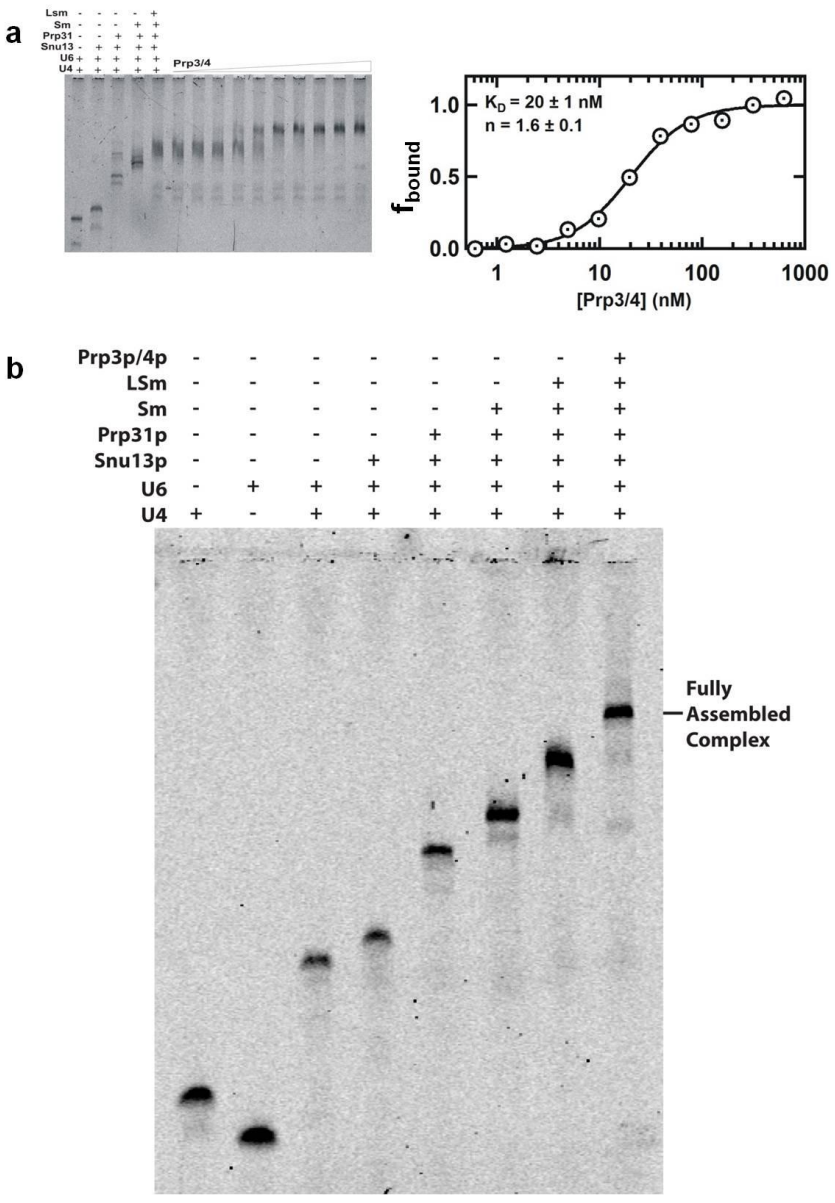

Figure 4.5: Assembly of complete U4/U6 di-snRNP. (a) Gel shift assays and binding curves for the interaction between the pre-formed U4/U6/Snu13/ Prp31/Sm/LSm complex and Prp3/4 proteins. The concentrations of the components used are; 2 nM pre-formed U4/U6 duplex, 200 nM Snu13, 120 nM Prp31, 64 nM Sm proteins, and 240 nM LSm proteins. **(b)** Step-wise assembly of complete U4/U6 di-snRNP. The binding of each protein shows a complete shift **(The gel pictures and binding data were kindly provided by Dr. J. Hardin and Dr. Y. Kondo)**.

4.3.2: SmFRET studies revealed that the stems I and II are coaxially stacked

Single-molecule FRET experiments were performed (as described in chapter 2) to characterize the relative orientation of the three helical arms and the global structure of the U4/U6 snRNA 3-way junction. Hence, a minimal U4 and a minimal U6 strands were used, in which the Sm and LSm binding sites are removed, the constructs were biotinylated, and fluorophore labelled in order to monitor the position of each helix relative to the others (Figure 4.1b)[239,240]. First, I have done an EMSA study with this truncated construct to confirm the formation of U4/U6 duplex and the binding of proteins (Figure 4.6). Next, I have investigated the relative orientation of stem I and II. For this experiment, I have used a doubly labelled minimal U6 strand with Cy3 at the 5' end and Cy5 at the 3' end, along with unlabelled U4 is biotinylated at the 3' end (Figure 4.7a). Pre-formed U4/U6 duplex was immobilized on a quartz slide and fluorophores were excited with 532 nm laser. The single-molecule trajectories reveal a single static confirmation around 0.2 FRET (Figure 4.7b and c, n= 108) for this construct. This suggests that stem I and II adopts a rigid position, with no structural dynamics. Using the Forster's equation and R_0 = 60 Å for the Cy3/Cy5 pair[241], I have estimated the distance between two fluorophores corresponds to the aforementioned 0.2 FRET state, which is ~76 Å (Figure 4.7d). I have also calculated the length of each helix based on the size of each stem assuming them as A-form helices. The stem I (~10 bp) is ~28 Å long and stem II (~17 bp) is ~48 Å long, which adds up to a total length of 76 Å (Figure 4.7d), suggesting that stem I and stem II are coaxially stacked in solution (Figure 4.7d).

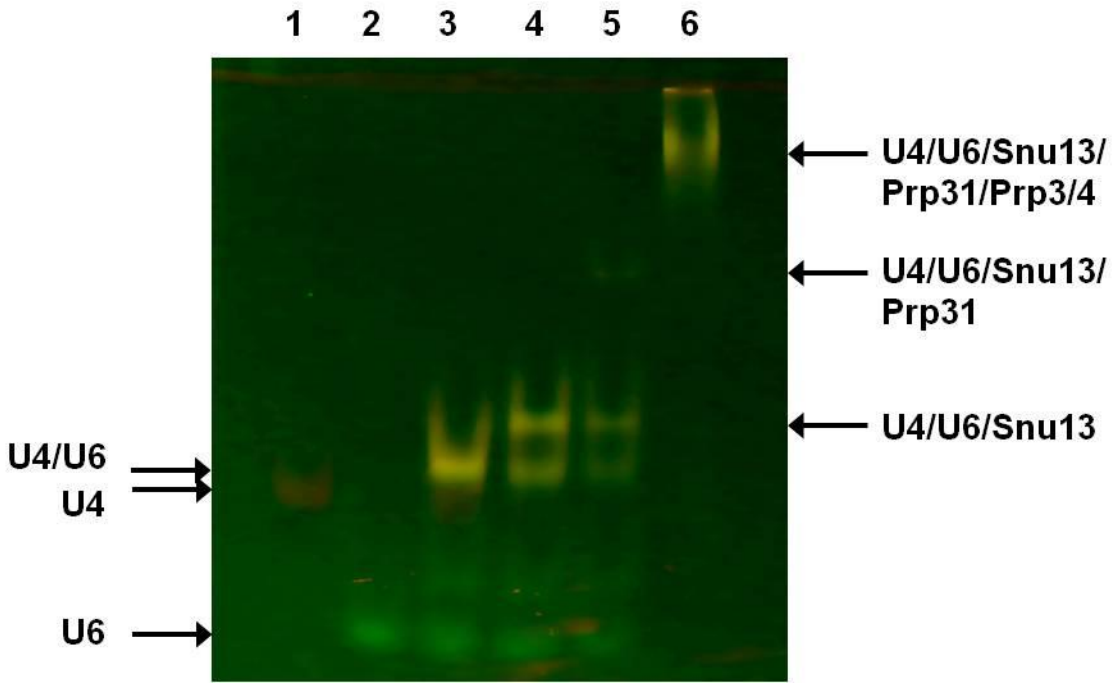

Figure 4.6: Minimal fluorophore labelled U4/U6 snRNA duplex binds with proteins. Gel was scanned by exciting with 532 nm and 635 nm lasers to detect Cy3-U6 and Cy5-U4, respectively. Resulting image is an overlay of both channels. Lane 1: Cy5-U4 snRNA (red band). Lane 2: Cy3-U6 snRNA (green band). Lane 3: U4/U6 snRNA duplex (yellow band). Lane 4: U4/U6 snRNA/Snu13 (shifted yellow band). Lane 5: U4/U6 snRNA/Snu13/Prp31 (further shifted band). Lane 6: U4/U6 snRNA/Snu13/Prp31/Prp3/4 (top band). The complete snRNP complex forms in presence of all proteins.

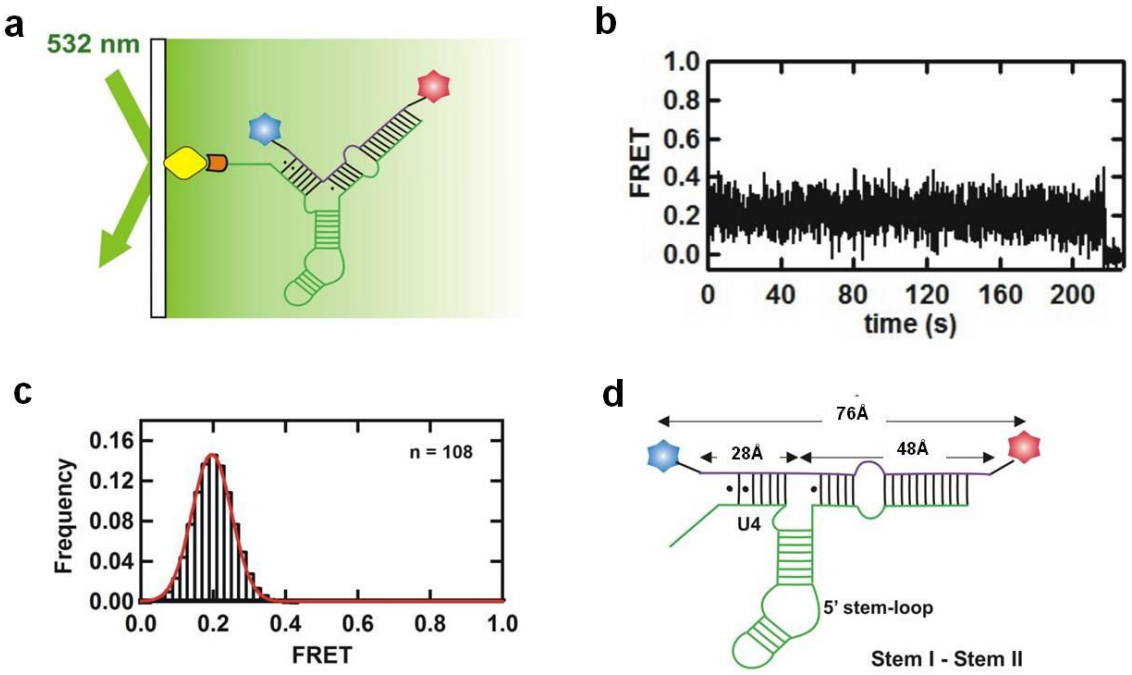

Figure 4.7: Stems I and II are coaxially stacked. (a) Single molecule setup for the U4/U6 snRNA complex, in which U6 is labelled with Cy3 and Cy5 at the 3' and the 5' ends respectively, and U4 is biotinylated at the 3' end. Pre-formed U4/U6 duplex is immobilized on a quartz slide using biotin-streptavidin interaction and the fluorophores are excited with a 532 nm laser. (b) FRET time trajectory and (c) histogram for the stem I and stem II construct shows a single static FRET state at 0.2. (d) The calculated distance between two fluorophores based on the FRET value (76 Å) is similar to the total length of two helices (stem I and stem II are 28 Å and 48 Å, respectively), suggesting that these two stems are coaxially stacked.

4.3.3: The U4/U6 3-way junction is static

Next, the orientations of stem I and II relative to the 5' stem-loop was examined. First, the orientation of stem II relative to the 5' stem-loop was tested by using a U4/U6 duplex, where 5' end of U6 was labelled with Cy3 (Figure 4.8a). Three static and non-interconverting conformations at 0.2 (44 ± 7%), 0.3 (45 ± 7%), and 0.4 (11 ± 3 %) FRET were observed for these experiments (Figure 4.8b and c, n= 102). These results indicate that the 5' stem-loop adopts multiple conformations compare to the coaxially stacked stem I and II, which are non-interconverting on the timescale of the experiment, which is on the order of minutes. Then the orientation of stem I relative to 5' stem-loop was studied to support above results (Figure 4.9a). These results indicate that two helices adopt three static populations with FRET values of 0.3 (41 ± 6%), 0.4 (51 ± 6%) and 0.5 (8 ± 3%) similar to what observed for stem II, confirming the static heterogeneity for 3-way Junction (Figure 4.9b and c, n= 105). Based on the FRET values observed for both constructs, these populations were assigned to different orientations of 5' stem-loop relative to the stems I and II. Population 1 is likely to be a conformation where the 5' stem-loop is closer to the stem I. On the other hand, population 2 is where the 5' stem-loop is closer to the stem II. Population 3 could be a minor population such as a misfolded state that is always present. This minor population could correspond to the small fraction that migrating between U6 and U4/U6 duplex in the fluorescence EMSA (Figure 4.6).

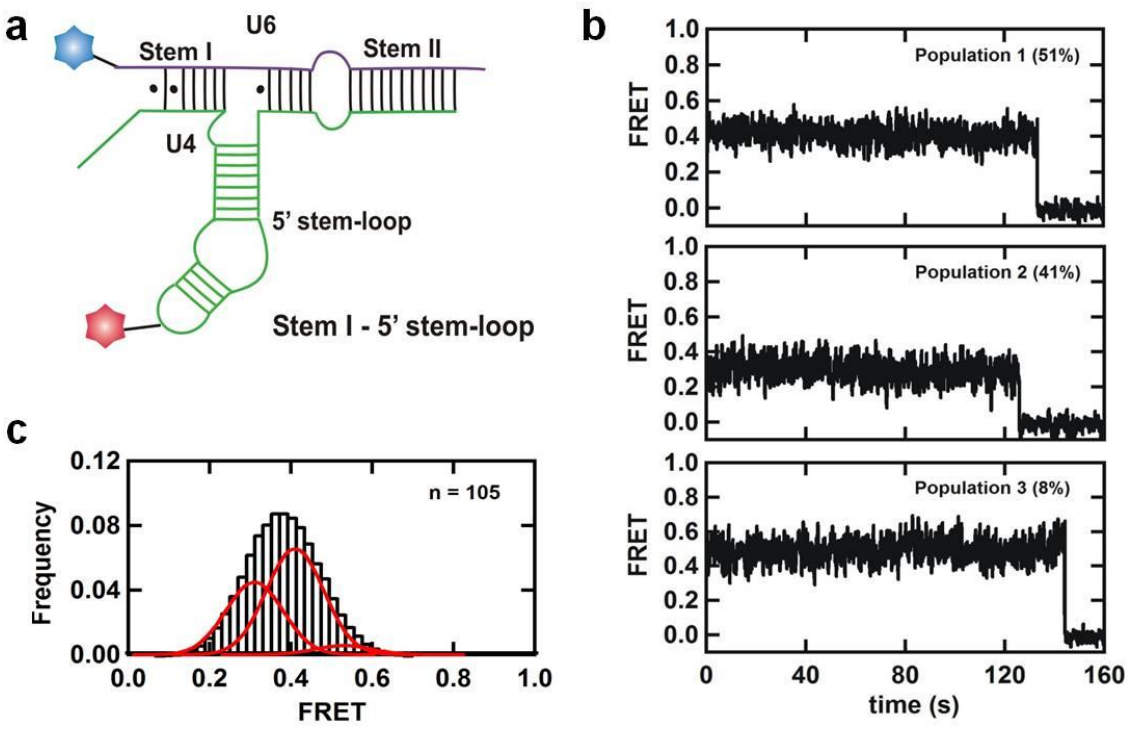

Figure 4.8: Stem II and 5' stem-loop of U4 adopts three static relative orientations. (a) Schematic representation of the U4/U6 snRNA duplex used to study the orientation of stem II relative to the 5' stem-loop of U4. U6 is labelled with Cy3 (blue) at the 3' end stem II. U4 is biotinylated at the 3' end and labelled with Cy5 (red) within the stem-loop as shown. (b) Time trajectories indicating three static, non-interconverting FRET populations observed for the stem II - 5' stem-loop. (c) FRET histogram showing three distinct FRET peaks (n= 102).

In order to determine the effect of divalent metal ions on the different FRET populations observed, then the single-molecule experiments were done in the presence of different concentrations of Magnesium ions along with 100 mM NaCl, which also results in no observable FRET changes or relative populations (Figure 4.10a). These observations are in contrast with what have been observed for other RNA complexes[185,239,240,242,243]. A possible explanation for these conformations could be due to the rotation at the kink-turn motif in the 5' stem-loop upon its stabilization [71,244,245], or rotation around three-way junction as observed for the Hammerhead ribozyme and the VS ribozyme [240,242,246]. Then experiments were done in the presence of particle-specific proteins (Snu13, Prp31 and Prp3/4) to test these two suggestions. First a fluorescence scanning experiment were done to determine the effect of protein binding on the fluorescence emission (Figure 4.11). Resulting emission spectrum and the relative intensity change for Cy3 (Figure 4.11a) and Cy5 (Figure 4.11b) in the absence and presence of proteins suggest that binding of proteins to the fluorophore labelled U4/U6 duplex does not affect the fluorescence properties of Cy3 and Cy5.

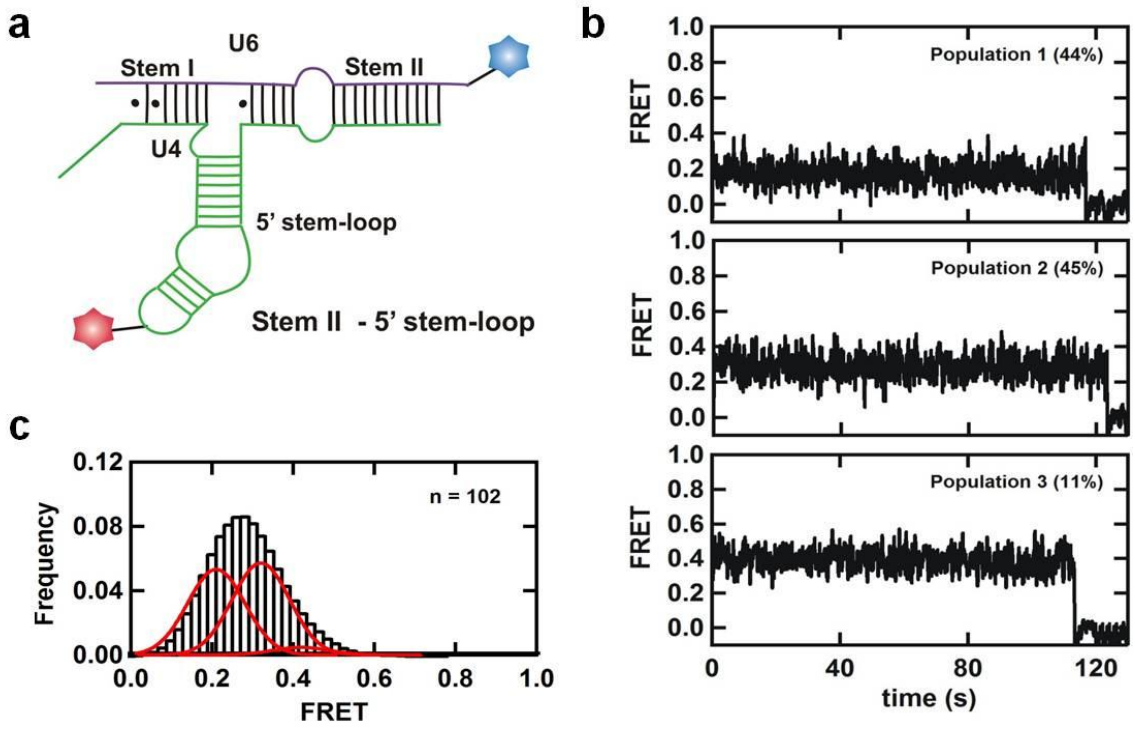

Figure 4.9: Stem I and 5' stem-loop of U4 adopts three static relative orientations. (a) Schematic representation of the U4/U6 snRNA duplex used to study the orientation of stem I relative to the 5' stem-loop of U4. U6 is labelled with Cy3 (blue) at the 5' end of stem I. U4 is biotinylated at the 3' end and labelled with Cy5 (red) within the stem-loop as shown. (b) Time trajectories indicating three static, non interconverting FRET populations observed for the stem I - 5' stem-loop. (c) FRET histogram showing three distinct FRET peaks (n= 105).

4.3.4: The kink-turn is pre-formed to accelerate Snu13 binding

To test the hypothesis that different orientations of the kink-turn motif correspond to the observed multiple conformations; I have introduced Snu13 to the immobilized U4/U6 duplex, which is expected to stabilize only one conformation of the kink-turn upon binding[247,248]. Single-molecule FRET trajectories for each construct show that binding of Snu13 to the 5' stem-loop does not cause any change in the observed FRET populations relative to either stem II or stem I compared to RNA alone (Figure 4.12a, b and 4.13a, b, n> 100). Similar experiments were done at saturating concentrations of Snu13 (200nM) also show no significant change in the populations (Figure 4.14a, b). However, previous study with L7Ae protein, an archaeal Snu13 homolog [244,249], has shown that the Snu13 homolog stabilizes a single conformation of a minimal kink-turn construct [247]. This suggests that the multiple orientations of the kink-turn motif in the 5' stem-loop are not responsible for the observed different FRET populations. Furthermore, the conformation of stems I and II remain coaxially stacked even in the presence of Snu13 (Figure 4.12c and Figure 4.14c). Taken together, these results raise the interesting possibility that under these experimental conditions, the U4/U6 kink-turn is pre-formed and thus promotes further snRNP assembly.

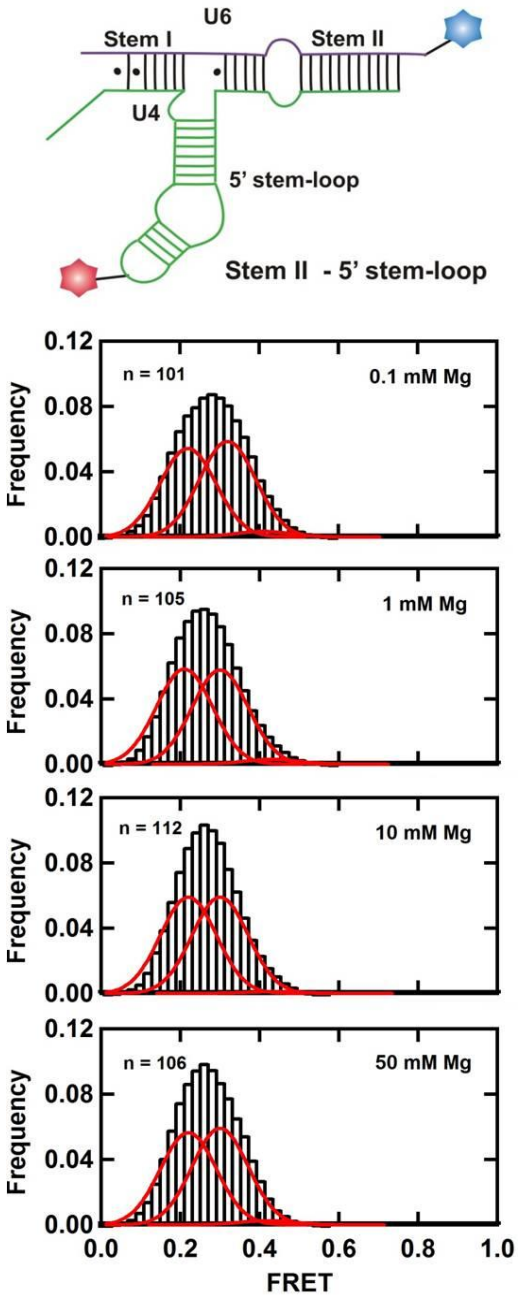

Figure 4.10: Multiple conformations of U4/U6 duplex are Mg2+ independent. FRET histograms for single-molecule experiments with the stem II-5' stem-loop construct in the presence of 0.1 -50 mM Mg^{2+}, indicating that Mg^{2+} ions have no effect on the multiple FRET populations observed.

4.3.5: Prp31 preferentially binds to one of two internal stem-loop conformations

Since kink-turn orientation is not the cause for the observed multiple FRET populations, experiments were done to test whether these conformations result from different orientations of the three-way junction. Prp31 protein, which interacts with the 5' stem-loop and stem II (Figure 1.13a) was used to test this and an effect on the relative orientation of the three-way junction was expected. First, the orientation of stem II relative to the 5' stem-loop (Figure 4.12a) was examined. Single-molecule FRET trajectories in the presence of Prp31 show a significant decrease in population 1(from 44 ± 7% to 28 ± 5%, $p < 0.05$, Figure 4.13a, n= 107). This result indicates that Prp31 binding stabilizes the three-way junction in a conformation, where the 5' stem-loop is farther from stem I and closer to stem II. To confirm this interpretation, the orientation of 5' stem-loop relative to stem I (Figure 4.12b) was tested. In the presence of Prp31 this construct also shows a decrease in the population 1 (from 51± 6% to 41 ± 6%, Figure 4.13b, n= 104), consistent with previous results. As expected, stem I-stem II shows no FRET change in the presence of Prp31 (Figure 4.12c), further supporting the idea that these two helices are rigidly and coaxially stacked even in the presence and absence snRNP proteins. Taken together, these data show that Prp31 preferentially binds and stabilizes a conformation of the U4/U6 3-way junction that brings the 5'stem-loop closer to stem II and further from stem I.

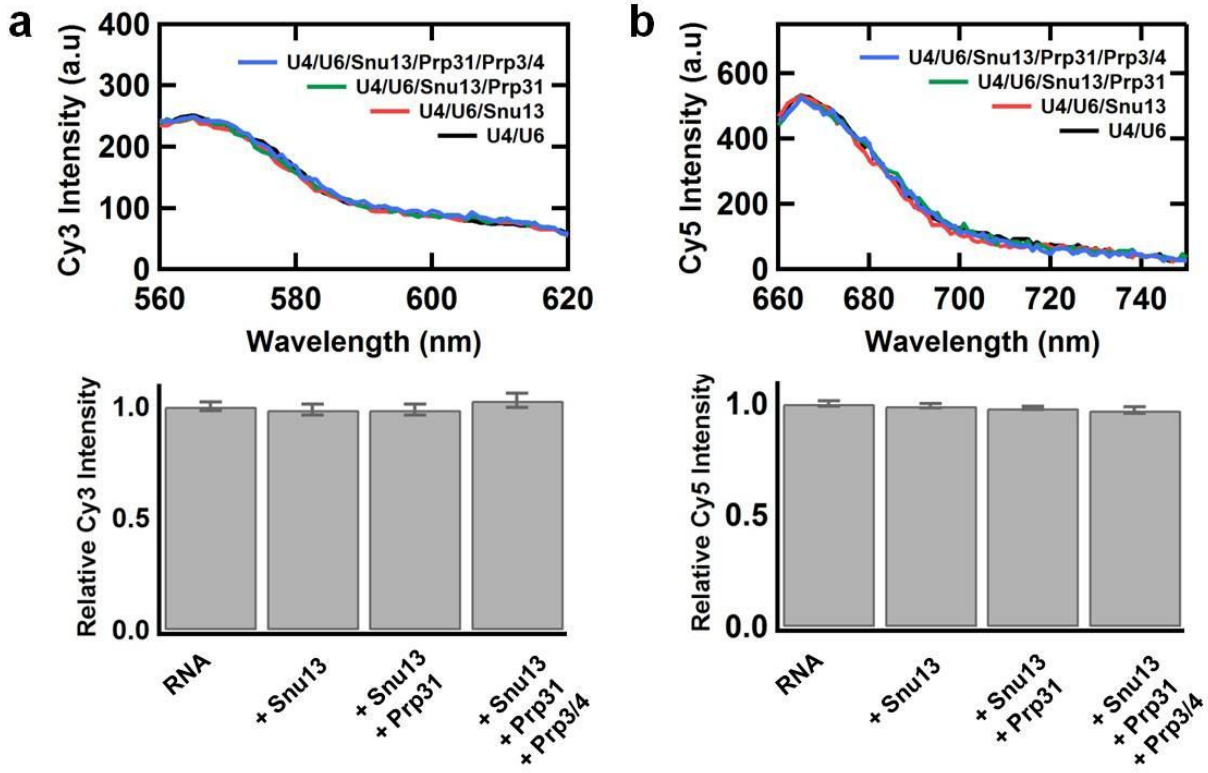

Figure 4.11: Fluorescent intensities of Cy3 and Cy5 do not change in the presence of protein. Fluorescence emission spectra of (a) Cy3 and (b) Cy5 the U4/U6 RNA duplex (black) and in presence of Snu13 (red), Snu13 + Prp31 (green), Snu13 + Prp31 + Prp3/4 (blue). The spectra remain unchanged, indicating that the proteins do not affect the fluorescence properties of Cy3 and Cy5.

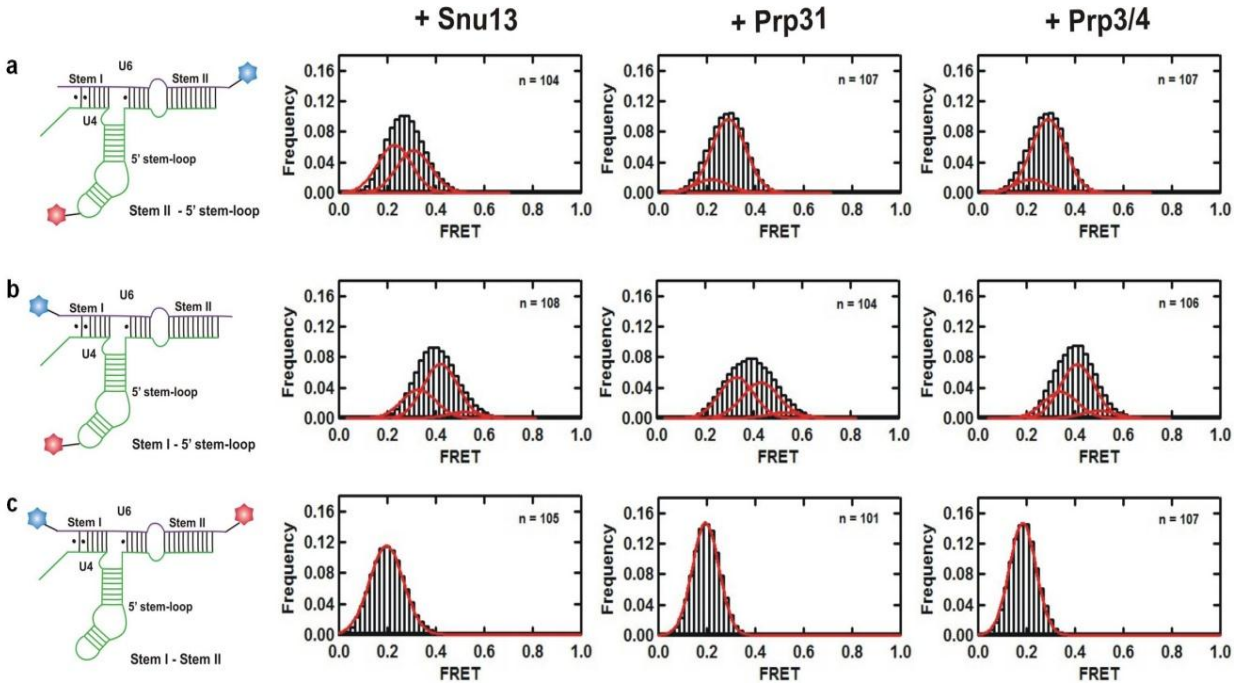

Figure 4.12: Single molecule assembly for individual proteins onto the U4/U6 duplex. Binding of Snu13, Prp31 and Prp3/4 individually to the (a) stem II-5' stem-loop construct, (b) stem I-5' stem-loop construct, and (c) stem I- stem II construct. Presence of individual proteins has no effect on the stem I-stem II complex, further confirming that these two arms are coaxially stacked. The presence of Snu13 also has no effect on the multiple populations observed with the naked RNA duplex. Prp31 stabilizes a conformation where 5' stem-loop is closer to stem II, where as Prp3/4 stabilizes a conformation where 5' stem-loop is closer to stem I.

4.3.6: Prp3/4 preferentially binds to the alternative internal stem-loop conformation

Furthermore, to study the effect of Prp3/4 binding on the orientation of the three-way junction, the experiments were done in the presence of Prp3/4. Previous studies have shown that Prp3/4 binds around stem II and the 5' stem-loop, which is expected to cause changes in the U4/U6 orientation (Figure 1.13a). First the orientation of stem II relative to the 5' stem-loop (Figure 4.12a) was examined. In contrast to Prp31 binding, the FRET trajectories show a significant increase in the population 1 (from 44 ± 7% to 58 ± 7%, $p < 0.05$, Figure 4.13a, n= 105). Next, the orientation of the stem I relative to 5' stem-loop (Figure 4.12b) was tested. This also shows an increase in population 1 (from 51 ± 7% to 58 ± 7%, Figure 4.13b, n = 106) consistent with previous result, indicating that binding of Prp3/4 stabilizes the three-way junction in a conformation where the 5' stem-loop is farther from stem II and closer to stem I. As expected, the relative orientation of Stems I and II remains unchanged in the presence of Prp3/4, confirming that these two helices remain coaxially stacked in presence of any of the proteins (Figure 4.12c). Taken together, these data indicate that binding of individual proteins Prp31 and Prp3/4 favours the stabilization of two different orientations of the 5' stem-loop relative to the coaxially stacked stems I and II.

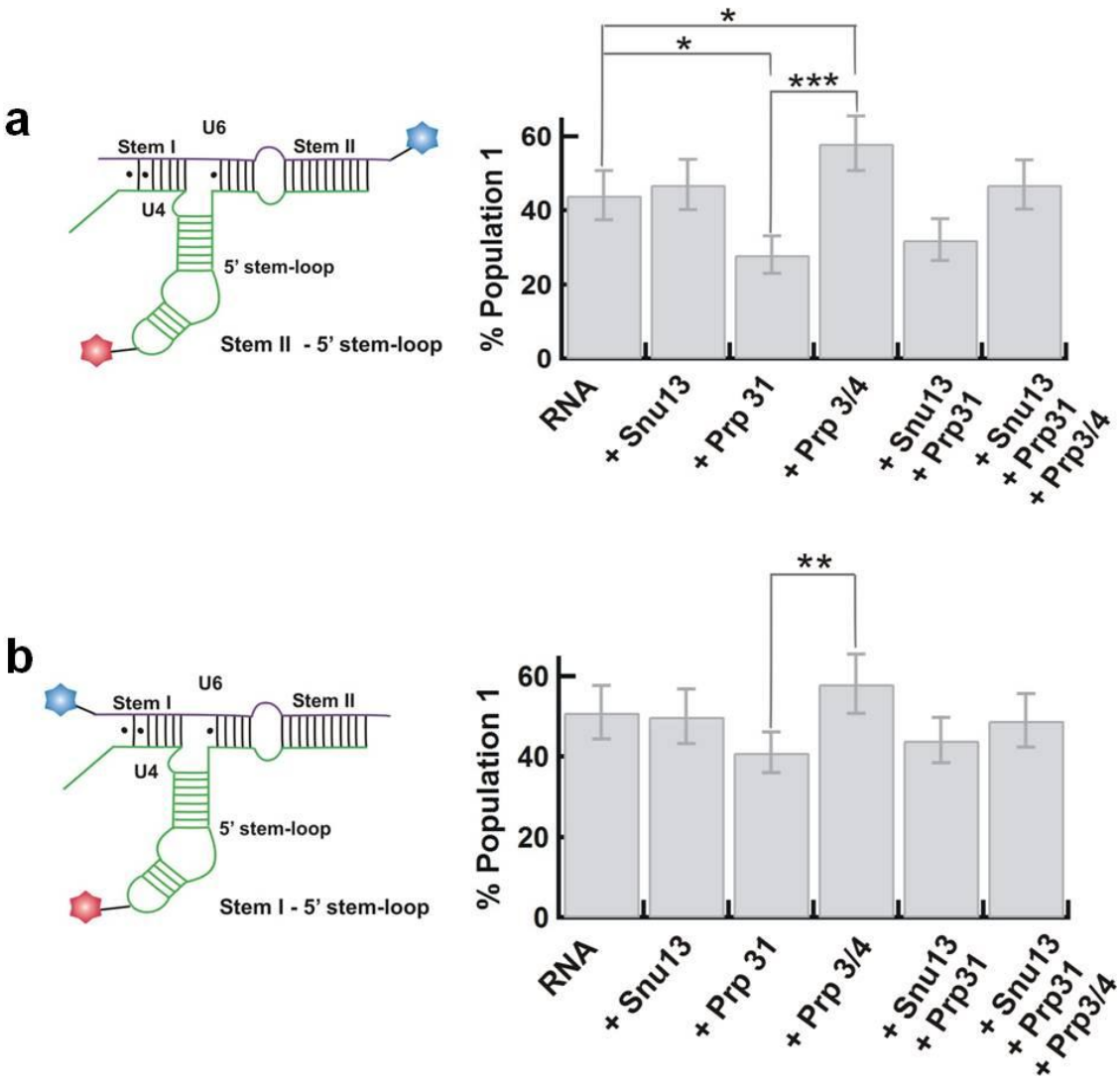

Figure 4.13: Binding of proteins affect on the local conformation of U4/U6 duplex. Fraction of molecules at population 1 in the presence of each protein or protein sub complex is shown for (a) stem II-5' stem-loop construct, (b) stem I-5' stem-loop construct. This illustrate that binding of Prp31 reduces the fraction of population 1 whereas Ppr3/4 increases the fraction of population 1. The p values are calculated using t-test and represent as * - $p<0.05$, ** - $p<0.01$ and *** - $p<0.001$.

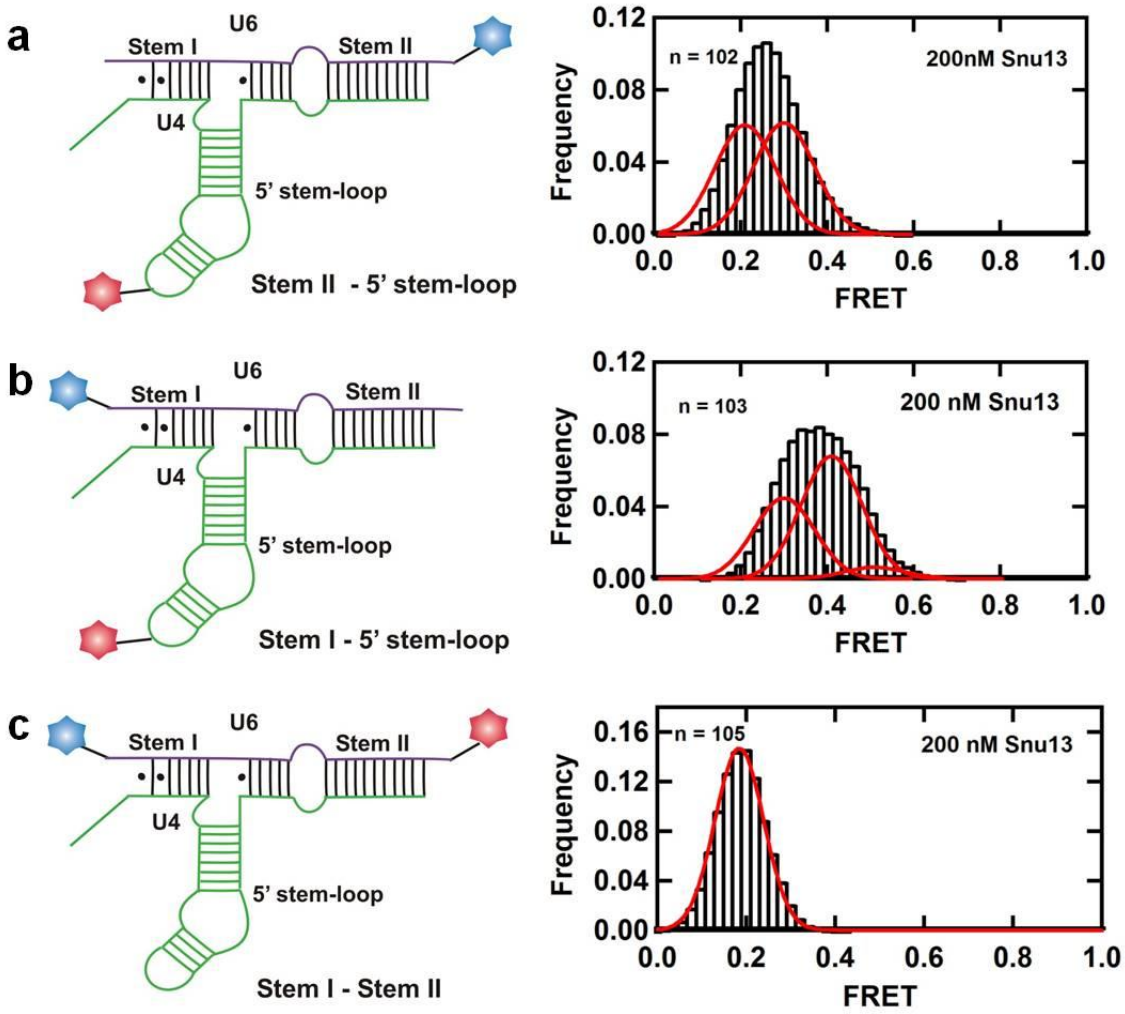

Figure 4.14: Binding of Snu13 has no effect on the multiple conformations of U4/U6 duplex. Histograms for (a) stem II-5' stem-loop construct (b) stem I-5' stem-loop construct and (c) stem II- stem I construct in the presence of saturating concentration of Snu13 (200nM), illustrating the FRET populations are similar to that observed for 20nM Snu13 and naked RNA duplex.

4.3.7: Single molecule assembly of multiple proteins onto the U4/U6 3-way junction RNA

Next, I have examined the possible conformational changes upon simultaneous binding of multiple proteins to the U4/U6 duplex. First, I have tested the orientation of stem II relative to the 5' stem-loop, in the presence of both Snu13 and Prp31 proteins (Figure 4.15a). Upon binding of both proteins, the percentage of population 1 has reduced compare to the percentage obtained for RNA only (from 44 ± 6% to 32 ± 5%, Figure 4.13a, n= 103). These data are consistent with the previous observation that binding of Prp31 with the U4/U6 duplex results in a shift in the 5' stem-loop towards the stem II and away from the end of stem I, whereas Snu13 doesn't cause any change to the duplex conformation. To further confirm this result, I have then examined the orientation of stem I relative to the 5' stem-loop (Figure 4.15b), which also shows a decrease in the population 1, favouring the suggestion that Prp31 binding to the U4/U6/Snu13 complex facilitates a conformation where stem I is further from the 5' stem-loop relative to the coaxially stacked stems I and II (from 51 ± 7% to 44 ± 6%, Figure 4.13b, n= 102). Testing the effect of multiple proteins binding to the orientation of stems I and II has therefore shown that none of the snRNP proteins affect the observed FRET distributions, indicating that these two stems remain coaxially stacked for the duration of entire snRNP assembly (Figure 4.15c). I have then examined the subsequent binding of the proteins; Snu13, Prp31 and Prp3/4, to the U4/U6 duplex, and tested the conformational changes of the three helices. The orientation of stem II relative to the 5' stem-loop was studied first, which has shown that binding of all proteins results in FRET population distributions where the percentage of population 1

is 47 ± 6 % (Figure 4.13a and 4.15a), which is almost similar to what is observed for the naked RNA (44 ± 6 %, Figures 4.7 and 4.11a). Similarly, testing the orientation of stem I relative to the 5' stem-loop, have shown that the percentage of population 1 in the presence of all proteins (49 ± 7 %, Figures 4.13b and 4.15b) is comparable with RNA alone results (51 ± 7 %, Figures 4.8 and 4.11b). As expected, stems I and II and remain coaxially stacked upon binding of all proteins (Figure 1.15c). Moreover, experiments were done in the presence of all the protein and data were taken for longer time (> 30 minutes) at 500 ms time resolution (Figure 4.16). I have observed three static FRET populations (Figure 4.16a and b) similar to what have observed before (short movies with 30 ms time resolution). These data reveal that proteins remain bound to the U4/U6 duplex for a long time, and do not show any association/dissociation dynamics within this experimental time window.

Overall, these results suggest that complete assembly of U4/U6 di-snRNP (except Sm and LSm proteins) stabilizes the U4/U6 in a conformation that closely resembles the conformation for naked U4/U6 (Figure 4.17). This indicates that binding of both Prp31 and Prp3/4 together prevents the relative stabilization of the alternative conformations observed upon binding of individual proteins.

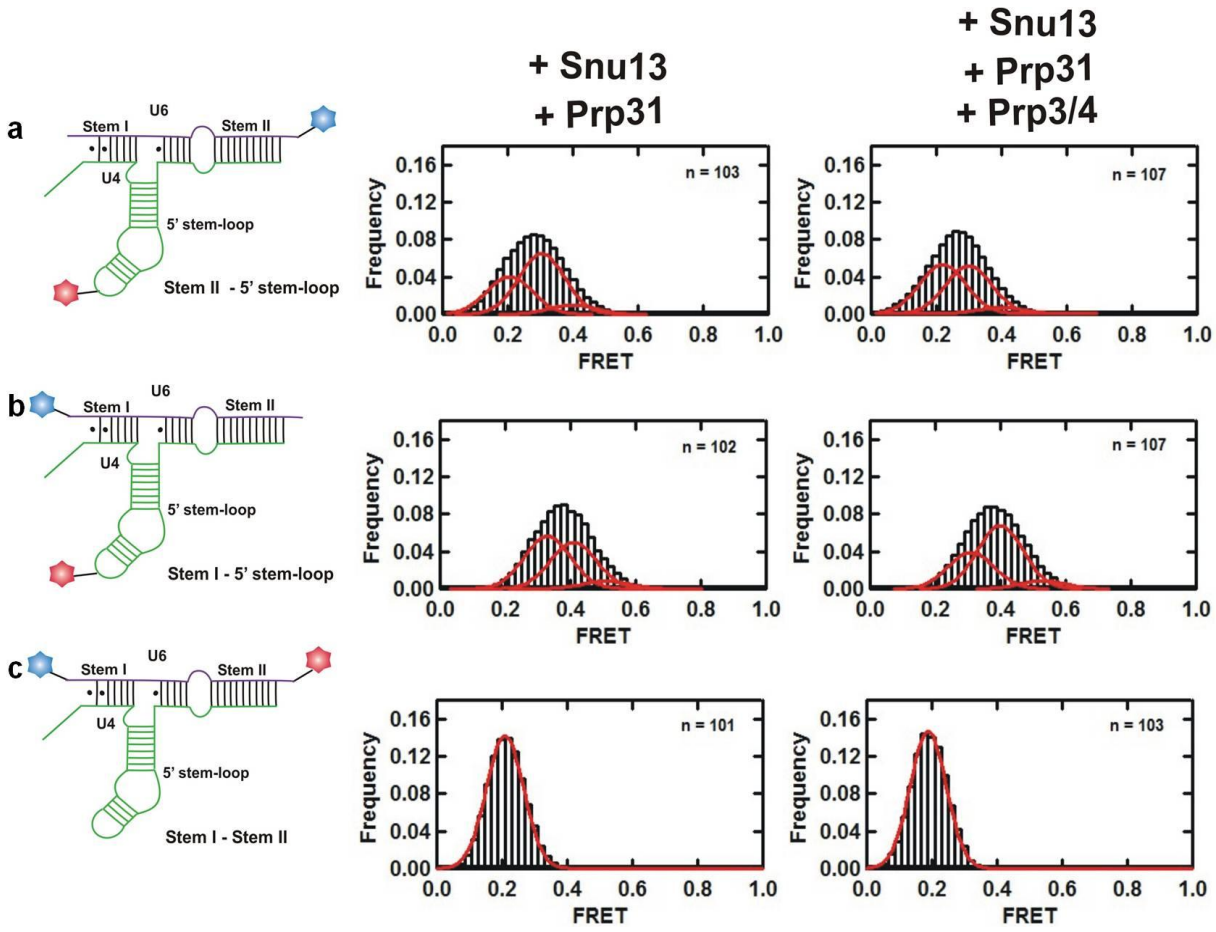

Figure 4.15: Assembly of multiple proteins onto the U4/U6 duplex maintains a rigid conformation. Binding of multiple proteins to the (a) stem II-5' stem-loop construct, (b) stem I-5' stem-loop construct, and (c) stem I- stem II construct. The presence of multiple proteins has no effect on the stem I-stem II complex, further confirming that these two arms are coaxially stacked. Similar to individual binding, Snu13+Prp31 stabilizes a conformation where the 5' stem-loop is closer to stem II, whereas binding of Snu13+Prp31+Prp3/4 returns U4/U6 duplex to a conformation similar to RNA alone.

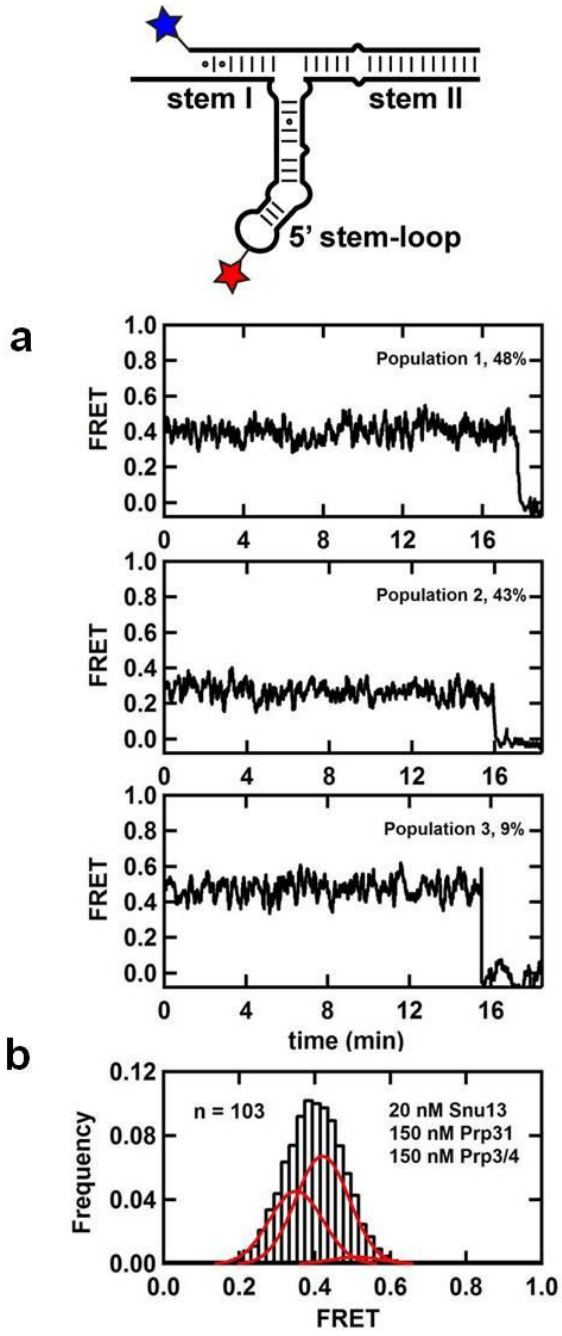

Figure 4.16: The snRNP complex remains static for long times. (a) Single-molecule FRET time trajectories and (b) corresponding histogram for experiment recorded for >15 min at 500 ms time resolution with the surface immobilized U4/U6 snRNA duplex in presence of 20 nM Snu13, 150 nM Prp31 and 150 nM Prp3/4. The complex does not exhibit any dynamics within this experimental time window.

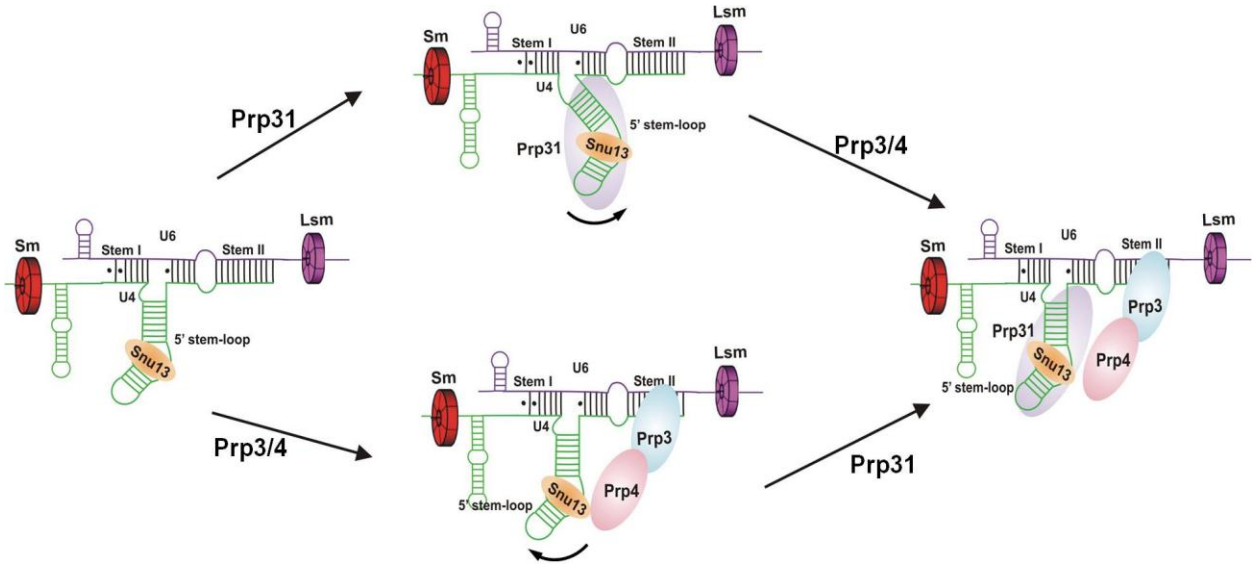

Figure 4.17: U4/U6 adopts a rigid, pre-formed global conformation. Stepwise assembly of U4/U6 di-snRNP is shown here. Based on our EMSA data and previous studies, we propose that Sm and LSm proteins are pre-bound to the U4/U6 complex. During the assembly process, Snu13 binds first to the K-turn motif at the 5' stem-loop of U4. Then, based on the binding affinities we obtained, either Prp31 or Prp3/4 can bind to the complex followed by the other protein (either Prp3/4 or Prp31). Prp31 preferentially binds to a conformation where the 5' stem-loop closer to the stem II. Prp3/4 preferentially binds to a conformation where the 5' stem-loop closer to the stem I. Assembly of all proteins to the complex brings U4/U6 conformation back to the same as RNA alone, suggesting that U4/U6 maintains a rigid conformation throughout the assembly process. During the assembly process stems I and II remain coaxially stacked without being affected by protein binding.

4.3.8: Visualization of Prp31 binding to the U4/U6 duplex

To confirm protein binding to the U4/U6 construct, and that Prp31 stays bound upon addition of Prp3/4, I have performed single molecule experiments using labelled Prp31 in the presence of Snu13 and Prp3/4. In this experiment, I have used biotinylated U4, Cy3 labelled U6 and Cy5 labelled Prp31. Single molecule experiments were conducted by switching the lasers from 532 nm (green) to 637 nm (red). In the presence of Cy5 labelled Prp31, a low FRET value was observed (Figure 4. 18a top and bottom panels).The direct excitation of Cy5 fluorophores by 637 nm laser results in an increase in the cy5 signal, indicating the binding of Cy5 labelled Prp31 to the U4/U6 duplex (Figure 4.18a, middle panel). In contrast, a sample containing all components as above but with unlabelled Prp31 shows zero Cy5 signal with both lasers (Figure 4.18b). These results confirm Prp31 binding to the U4/U6 duplex, which remains bound with the addition of Prp3/4.

Taken together, the EMSA data illustrate a step-wise assembly of U4/U6 di-snRNP and report binding affinities for individual and sequential binding of proteins. Furthermore, the single-molecule data suggest that stems I and II are coaxially stacked and remain stacked throughout the whole di-snRNP assembly process (Figure 4.17). These results also demonstrate that despite local structural changes upon binding of individual proteins, U4/U6 duplex maintains a rigid global conformation during the assembly process (Figure 4.17).

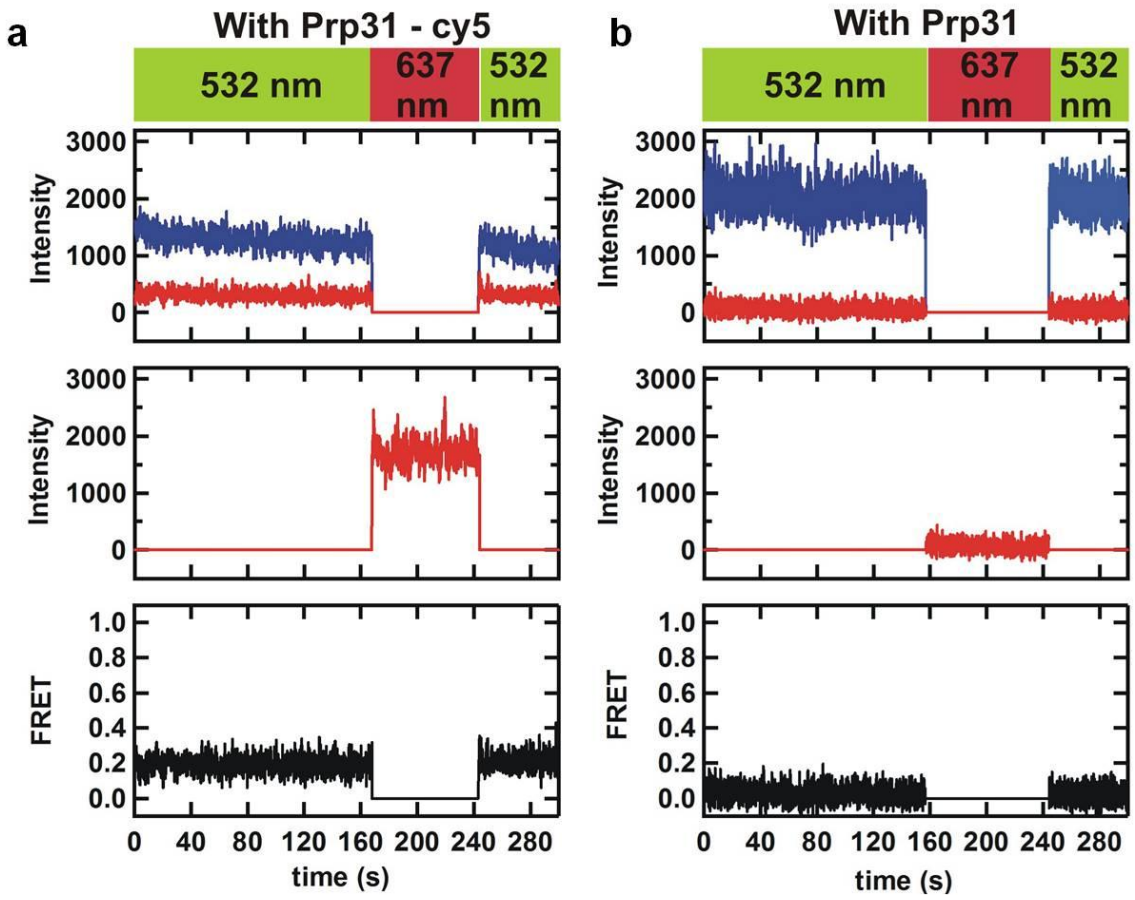

Figure 4.18: Prp31 protein stays bound to the U4/U6 complex even in the presence of Snu13 and Prp3/4. Intensities of the excited fluorophores at 532 nm (upper panel) and at 637 nm (middle panel) and the corresponding FRET trajectories (lower panel) are shown for the experiments in the (a) presence of cy5-Prp31 and (b) absence of labelled Prp31.

CHAPTER 5: Study the structural dynamics of U12/U6atac snRNA complex in minor spliceosome

5.1: Objective

The objective of specific aim 3 in my research is to understand the structural and functional similarities between major and minor spliceosomal snRNA complexes. Other than the highly abundant major spliceosome, which is present in almost all eukaryotes, a less abundant, second spliceosome has been discovered, referred to as the minor spliceosome. The minor spliceosome consists of five snRNPs: U11, U12, U4atac, U6atac and U5, which are analogous to their major spliceosomal counterparts. Despite their sequence differences, snRNA complexes from both spliceosomes share structural similarities. For example, the U4atac/U6atac and U12/U6atac complexes are similar in structure to the U4/U6 and U2/U6 complexes, respectively. Also both spliceosomes share a similar splicing mechanism. Although several studies have been done on the minor spliceosome, the conformational dynamics and assembly of snRNA complexes remain poorly understood.

The minor spliceosomal U12/U6atac complex is analogous to the major spliceosomal U2/U6 complex, which is known to be present in the catalytic core. The structural details of U12/U6atac complex will shed light on understanding the assembly process of the minor spliceosome and the role of each component. Hence, this work focuses on structural analysis of the U12/U6atac complex.

To examine the conformational dynamics of the minor spliceosomal U12/U6atac complex, I have employed single-molecule FRET. The resulting preliminary data shows

the formation of U12/U6atac complex. The data also have shown that the U12/U6atac complex is stabilized by Mg^{2+} ions, which is similar to what has been observed for the U2/U6 complex from a previous study done by our lab[185]. Furthermore, these results suggest that U12/U6atac complex adopts a conformation similar to the three-helix junction structure of U2/U6.

5.2: Experimental design

In this study, I have used a minimal U12 and U6atac snRNA sequences (Table 2.3, Figure 5.1a). Similar to the U2 and U6 snRNAs used in the first part of my work, U12 is unlabelled and U6atac is labelled with Cy3 at the 5' end, Cy5 internally, and biotinylated at the 3' end. U6atac snRNA was heated to 94 ^{o}C and freeze cooled on ice for 20 min. Then diluted U6atac was immobilized on a quartz slide and U12 was introduced. The Cy3 fluorophores were then excited using 532 nm laser and resulting intensities were captured through the objective and detected through the CCD camera (Figure 5.1b).

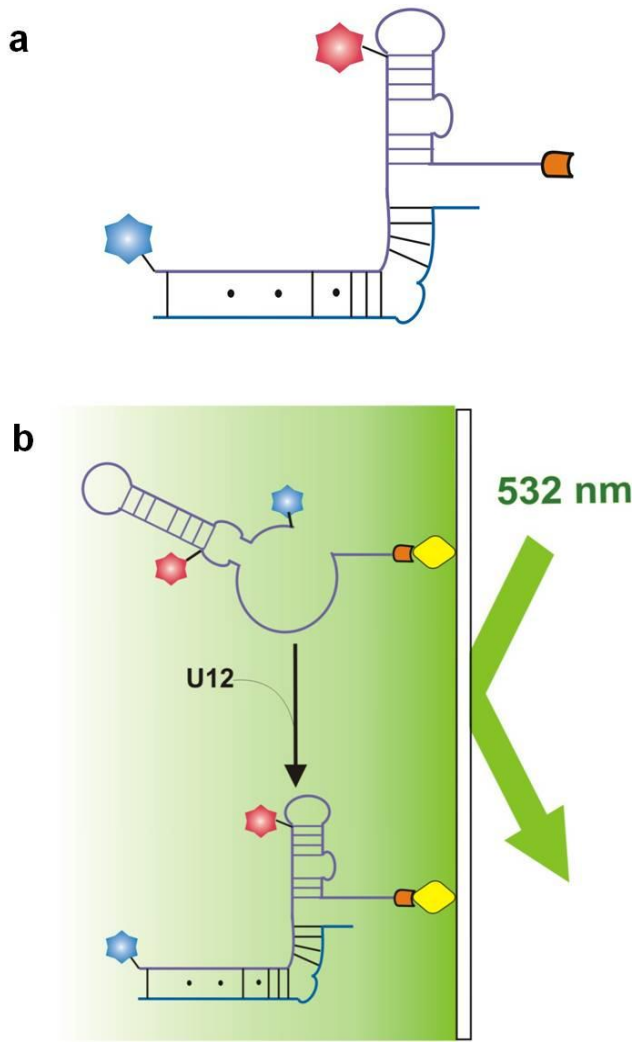

Figure 5.1: Single-molecule studies on the minor spliceosomal U12/U6atac complex. (a) Schematic representation of the minimal U12 and U6atac construct used in this study. U6atac is labelled with Cy3, Cy5 and biotin. (b) Single-molecule experimental setup showing that labelled U6atac is immobilized on to a quartz slide via streptavidin-biotin interactions. Fluorescent molecules were excited using 532 nm laser, and the resulting signal is captured through the objective and detected through CCD camera. Upon addition of U12, the conformation of U6atac alone changes due to the formation of U12/IU6atac.

5.3: Results

5.3.1: U12/U6atac complex stabilizes a low FRET conformation

First, single-molecule experiments were done in the presence of varying concentrations of U12 in order to confirm the formation of the U12/U6atac complex. In the presence of U6atac only, almost all the molecules are at 1.0 FRET (Figure 5.2a), suggesting that U6atac may form a secondary structure with a high FRET value (Figure 5.1b). On the other hand, addition of U12 results in an appearance of a peak at the 0.2 FRET state, and molecules show dynamic transitions from 1.0 to 0.2 FRET (Figure5.2a). At lower concentrations of U12, very frequent transitions were observed, whereas at higher concentrations, molecules tend to stay more at the 0.2 FRET state (Figure 5.2a). The resulting histograms (Figure 5.2b) and the binding curve plotted with the fraction of 0.2 FRET population against U12 concentration (Figure 5.2c) illustrate that addition of U12 stabilizes a 0.2 FRET conformation, which corresponds to the formation of U12/U6atac complex (Figure 5.1b). When comparing the secondary structures of U2/U6 and U12/U6atac, it's clear that the minor complex cannot form the four-helix conformation due to the truncation of U12 at 5' end (cannot form the intramolecular U2 helix or intermolecular U2/U6 helix II). Therefore it only can form the three-helix structure, which previously has been assigned to 0.2 FRET state[185]. The short length of U12 snRNA explains the requirement of higher U12 concentration for the complex formation. The fewer number of base pairs between U12 and U6atac may result in easy dissociation of U12 from U6atac or weak interaction with U6atac, which gives rise to the dynamic nature observed.

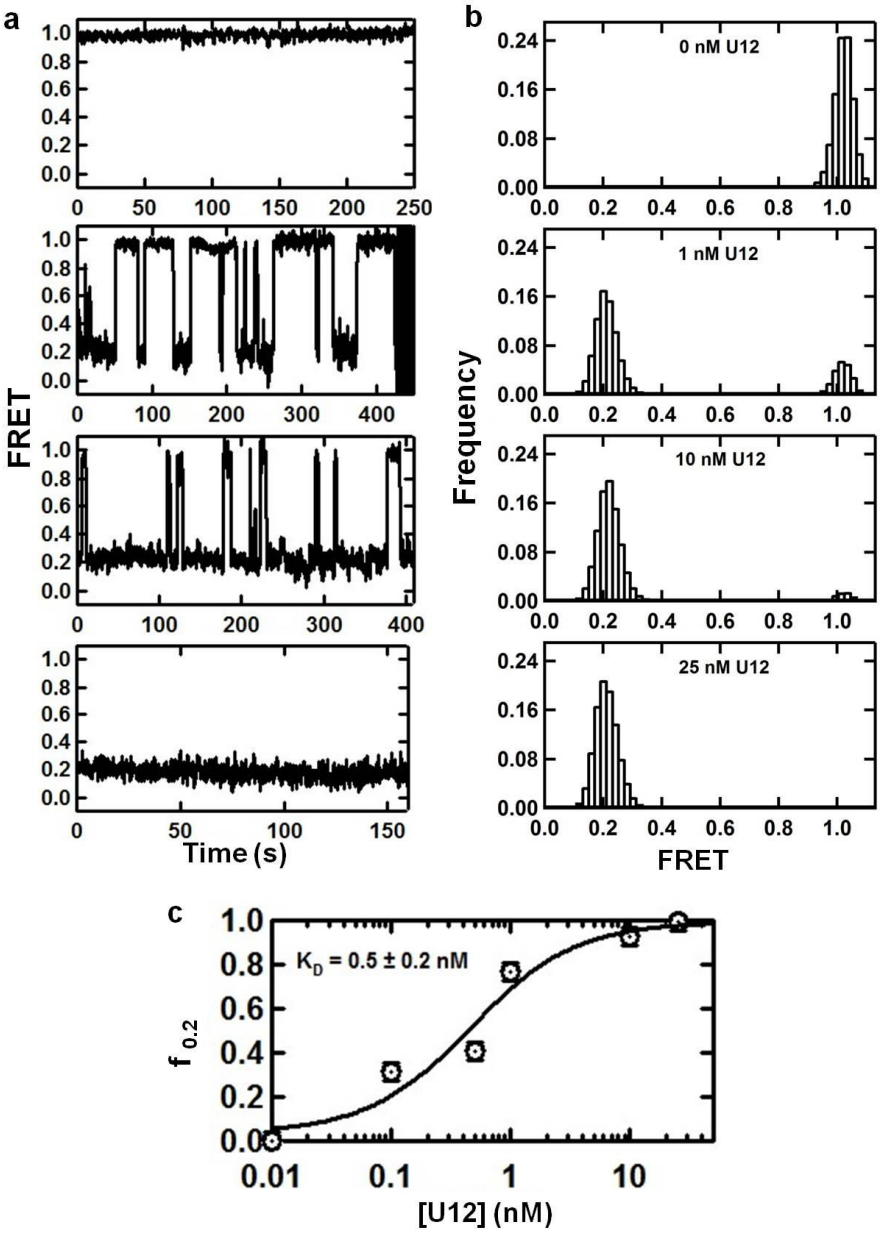

Figure 5.2: U12/U6atac complex adopts a low FRET conformation. (a) Single-molecule time trajectories and (b) histograms (n >100) for varying concentrations of U12, indicating that U6atac only forms a high FRET (1.0) conformation and the U12/U6atac forms a low FRET (0.2) conformation. (c) The binding curve obtained by plotting the fraction of molecules at 0.2 FRET state as a function of U12 concentration. The dissociation constant (K_D) was obtained by fitting the data into the Hill equation.

5.3.2: Mg^{2+} ions stabilize the U12/U6atac complex

A previous single-molecule study on the U2/U6 complex has shown that Mg^{2+} ions stabilize a low FRET (0.2) conformation of the U2/U6 complex, which then was assigned to three-helix junction structure[185]. Based on the structural similarities between U2/U6 and U12/U6atac complexes, Mg^{2+} may cause a similar effect on the minor spliceosomal complex. To characterize the Mg^{2+} ion requirement for the stability of the U12/U6atac complex, I have done experiments with varying Mg^{2+} concentrations. Trajectories showing a static 1.0 FRET state corresponding to the U6atac alone, were removed during data analysis in order to ensure that all the trajectories represent U12/U6atac complex. In the absence of Mg^{2+}, the single molecules are more dynamic resulting in a broader peak ranging from 0.4 to 1.0 FRET (Figure 5.3a). In the presence of Mg^{2+} ions, the population of high and intermediate FRET states decreases and a new FRET state at 0.2 appears. At elevated concentrations of Mg^{2+} (10 mM and higher), a further increase in the 0.2 FRET population and a decrease in the 1.0 FRET state is observed, while intermediate FRET states have disappeared. Similar results have been obtained for the previous single-molecule studies with U2/U6 complex, where in the absence of, or in the presence of, lower concentrations of Mg^{2+}, a broader peak was observed ranging through higher FRET states. On the other hand, in higher Mg^{2+} concentrations, the U2/U6 complex has shown an increase in the 0.2 FRET population[185]. These results indicate that Mg^{2+} ions stabilize the U12/U6atac complex in a low FRET conformation that is similar to the three-helix junction structure of U2/U6 complex.

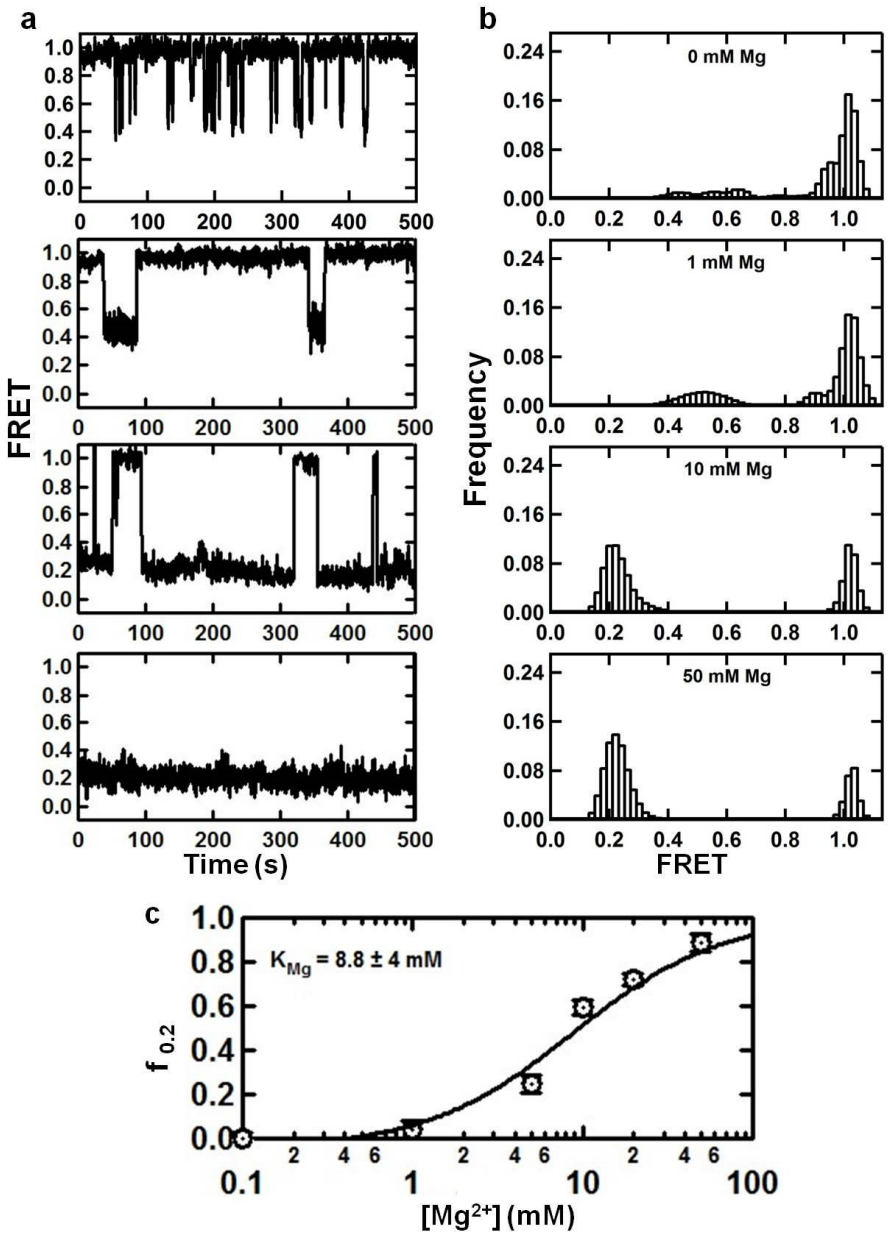

Figure 5.3: Mg^{2+} ions stabilize the U12/U6atac complex in a low FRET conformation. (a) Single-molecule time trajectories and (b) histograms (n >100) for varying concentrations of Mg^{2+}, indicating that in the absence of Mg2+ ions, U12/U6atac complex structure is highly dynamic, where as in higher Mg^{2+} concentrations U12/U6atac complex stabilizes at a low FRET (0.2) conformation. (c) The binding curve obtained by plotting the fraction of molecules at 0.2 FRET state as a function of Mg^{2+} concentration. The K_{Mg} was obtained by fitting the data to the Hill equation.

Overall, my preliminary data reveal that minor spliceosomal U12/U6atac adopts a conformation similar to the three helix conformation of U2/U6 complex. Moreover, Mg^{2+} ions stabilize U12/U6atac complex in a conformation with a 0.2 FRET state, similar to what has been observed for U2/U6 complex. This suggests that both complexes require Mg^{2+} ions for their stability in a similar manner. Taken together, these results indicate that the major and minor spliceosomal complexes, U2/U6 and U12/U6atac share structural similarities, which may represent their functional similarities.

CHAPTER 6: Discussion

The main goal of my research is to examine the spliceosomal snRNA complexes at the single-molecule level in order to obtain structural dynamic information that has not been revealed yet, such as assembly and conformational dynamics of snRNA complexes, factors involved in the dissociation of snRNP complexes after the completion of splicing, effect of associated proteins on the conformational rearrangements of spliceosomal sub-complexes and structural similarities between snRNA complexes of major and minor spliceosomes representing their functional relationships. To accomplish this goal, I have done single-molecule studies on three different snRNA complexes addressing three distinct aims; elucidating the role of Prp24 in U2 and U6 snRNP recycling, assembly and global structure of U4/U6 di-snRNP and structural comparison of minor U12/U6atac snRNA complex with major U2/U6 complex.

6.1: Elucidating the role of Prp24 in U2 and U6 snRNP recycling

To address the first question; how snRNA complexes get recycled, I have studied the effect of U6 snRNP associated protein Prp24 in U2/U6 complex dissociation.

First of all, I have done EMSA and fluorescence-anisotropy experiments to check binding of Prp24 to the yeast minimal U2/U6 construct used in this study. Both studies have shown that Prp24 binds with the RNA construct used in this study. The resulting gel shifts and the binding affinity (K_D) of Prp24 to the RNA complex used are in good agreement with previously published gel-shift results (Figure 3.2)[206].

One of the limitations of the EMSA and the anisotropy experiments is that the experiments are limited by the RNA concentration used, thus an apparent binding constant is derived and not the actual value. As an example, in fluorescence-anisotropy experiments, the RNA concentration is limited to ~25 nM (in order to obtain a better signal) and thus the resulting binding constant will be an apparent K_D value rather than the actual K_D value. The apparent K_D value for Prp24 binding to U6 was reported by previous gel shift assays and is similar to what I have obtained from these anisotropy experiments. Kwan and co-authors have determined the actual K_D value by titrating higher concentrations of Prp24 until the saturation of U6 binding is achieved. The corrected binding affinity is reported as ~2 nM, which suggests that the binding of Prp24 is much tighter than expected[206].

To overcome this issue, I have employed single-molecule studies. In single-molecule FRET experiments, since the RNA molecules are immobilized, the RNA concentration is considered as zero. Hence the resulting binding affinities are not limited by the RNA concentration. Also, sm-FRET allows us to detect binding of one protein molecule to one RNA molecule, which avoids averaging errors and gives rise to the actual K_D values.

Single-molecule studies in the presence and absence of Prp24 clearly indicate that binding of Prp24 causes a conformational change within the U2/U6 complex, results in a static low FRET conformation (0.2) in contrast to the three dynamic FRET states obtained for the U2/U6 complex alone (Figure 3.3 and 3.4). Another set of experiments in the presence of varying concentrations of Prp24 indicate that binding of Prp24 stabilizes this 0.2 FRET conformation. Here, the higher the protein concentration the

higher the fraction of molecules that stay at the 0.2 state. Moreover, these experiments result in an actual K_D of 0.2 nM, which is ~10 fold tighter than the corrected K_D value obtained from the EMSA studies (2 nM)[206]. The difference between the two K_D values may be due to the use of single-molecule experiments, where we consider one molecule at a time compare to the gel-shift experiments where they report an average change. Other than the major 0.2 FRET population, I also observed a minor population of 0.4 static FRET state, even at 100 nM of Prp24, which could be a misfolded state of the complex since it only appeared at elevated concentrations of Prp24.

As shown in previous studies, the four RRMs of Prp24 have different functions in formation and destabilization of the U4/U6 complex[101,179]. Previous studies have shown that RRM1 interacts with the phosphate backbone of the 3' downstream GAUCA sequence via electrostatic interactions[100]. Gel shift binding assays demonstrated that RRM1 and RRM2 are important for the high affinity binding of U6 snRNA and removal of RRM1 results in a five-fold reduction in the binding of Prp24[206]. Consistent with these results, in EMSA studies, I also have observed a less intense band corresponding to the 234C binding compare to the band corresponding to Prp24 binding, indicating that the truncated protein results in a reduction in protein binding to the complex. Similarly, results from single-molecule experiments have demonstrated a two-fold difference between the K_D for 234C (0.4 nM) and Prp24 (0.2 nM). These results suggest that the removal of RRM1 weakens protein binding, resulting in a less efficient stabilization of the low FRET state even at high concentrations of 234C.

To test if Prp24 plays a role in unwinding U2 from the complex, I used an RNA construct with Cy3 on U6 and Cy5 on U2. Single-molecule flow experiments with this

construct illustrate that removal of U2 from U6 upon binding of Prp24 occurs through an intermediate conformation. This intermediate conformation, with FRET value 0.2, could be a result of partial unwinding of U2 from the complex. Data from the single-molecule time dependent experiments shows that upon addition of Prp24, the zero FRET state becomes populated over time exhibiting a U2 removal rate of 0.15 min^{-1}, whereas when 234C is added, an increase in the zero FRET state is comparable to the photo bleaching rate. Results obtained for the protein titration have given rise to binding affinities of 0.3 nM and >3 nM for Prp24 and 234C respectively, illustrating that rate of U2 dissociation has been affected by the removal of RRM1, similar to what I have observed with previous single-molecule experiments. The gel shift assay with this construct has also shown that addition of Prp24 increases the level of free U2, suggesting that binding of Prp24 facilitates the removal of U2 from the complex.

In summary, these results suggest that Prp24 plays an important role in U2 and U6 snRNA recycling by dissociating the U2/U6 complex. After U2 dissociates, Prp24 may remain bound to the U6 and could stabilize U6 in an intermediate conformation, preventing the premature activation of U6 and later binding with U4 to form the U4/U6 complex and join with another spliceosomal assembly cycle.

There are two possible pathways that Prp24 could be involved in the dissociation of full length U2/U6 complex. Prp24 may solely be involved in the complete unwinding of U2/U6 complex or may only initiate the disruption of base pairs within the U2/U6 complex, which allows another protein factor to bind and facilitates the complete removal of U2. The recent crystal structure of Prp24 bound to a U6 sequence has also illustrated that Prp24 binding can disrupt the base pairs within helix I, III and base of

ISL, which could result in a partially opened complex. However, disruption of helix II upon binding of Prp24 is not clear. The results obtained in this study clearly illustrate that Prp24 promotes a complete dissociation of two snRNAs from each other, though this could be as a result of the short helical lengths of the minimal U2/U6 complex used. Hence, I suggest that at the cellular level where full-length U2 and U6 snRNAs are present, it could be possible that other protein factors such as some helicases may also be involved in this process. Previous studies have proposed that the ATP-dependent, DExD/H-box family protein Prp43 is involved in releasing U2, U5 and U6 after the completion of catalysis[2]. Brr2; another DExD/H-box family protein, which is known to be involved in unwinding the U4/U6 complex, has also been proposed to have a role in U2/U6 dissociation[250]. Hence, in light of these results, I propose that Prp24 helps to unwind the minimal U2/U6 complex of yeast, similar to its function in dissociation of U4/U6. However in cellular conditions, in which full-length U2/U6 is associated with U5 at the catalytic core, Prp24 may initiate the disruption of base pairs at 5' end of U6 and form a partially unwound complex, but Prp43 and/or Brr2 may be involved in the complete unwinding, which results in dissociation of all three snRNAs from each other.

6.1.1: Future directions

To further investigate the role of Prp24 and specific binding sites of RRMs, the following experiments have been proposed. One experiment is to use fluorescently labelled Prp24 in single-molecule experiments. Using labelled protein will confirm the presence of protein after the removal of U2. Also, having fluorophores attached to each of the RRMs separately will provide a better idea about how each RRM contributes to the

overall function of Prp24 by illustrating the regions in U6 that each RRM is interacting with. Another important study would involve single-molecule experiments under cellular conditions with spliceosomal extracts. This would help to understand the exact role of Prp24, Prp43 and Brr2 and whether they act cooperatively in releasing U2, U5 and U6 after completion of catalytic reactions. This study can be done by using labelled proteins to see whether the U2 removal is solely carried out by Prp24 by sequestering U6 from the U2 or if Prp24 initiates the disruption of base pairs at the 5' region of U6 followed by Prp43 and/or Brr2 moving along the partially unwound U2/U6 complex using their ATPase activity to completely unwind U2 from U6. These results will provide novel insight to understand the spliceosomal assembly cycle and how components are recycled after the completion of splicing.

6.2: Study the assembly and global structure of U4/U6 complex

To address the second question; how snRNPs assemble and effect protein binding on their global structures, we have studied the U4/U6 di-snRNP assembly using single-molecule FRET and EMSA studies.

Our EMSA studies revealed that the U4/U6 associated proteins are assembled onto the RNA duplex efficiently in a stepwise manner. Based on the structural and biochemical studies, it has been known that, Snu13 preferably binds to the K-turn motif at the 5' stem-loop of U4[65-67,233], where it has a role as a nucleating factor for the binding of other proteins[69]. Similarly our results indicate that Snu13 binds the U4/U6 duplex very tightly and it enhances the binding of Prp31 significantly. Consistent with the previous studies[69,138], similar binding affinities were observed for the binding of

Prp31 and Prp3/4 to the U4/U6/Snu13 complex, suggesting that each of these proteins can bind to the duplex independently. Furthermore our binding affinities suggest that Sm proteins and LSm proteins may be pre-bound to the U4 and U6 snRNAs respectively before the duplex formation. Previous studies have been shown that U6 snRNA, which is transcribed by Pol III, assembled first with LSm proteins and remains in the nucleus[49,251]. On the other hand U4 snRNA is exported to the cytoplasm where Sm proteins are assembled onto U4 and then transported back to the nucleus where it assembles with U6 and other proteins[252,253]. These findings suggest that Sm and LSm proteins are bound to individual snRNAs prior to their assembly into a duplex, which also in accordance with our results. All together, we propose that the assembly of U4/U6 di-snRNP occurs in a step wise manner, where Snu13 binds first to the U4/U6 duplex containing Sm and LSm rings, and then either Prp31 or Prp3/4 can bind to the complex followed by the other to form the fully assembled di-snRNP.

Prp3/4 binds to naked U4/U6 duplex snRNA tightly, but interestingly binding becomes substantially weaker when all other components, except LSm proteins, are pre-bound. Prior to the binding of LSm proteins, single stranded region in the 3' end of U6 could form multiple secondary structures. Assembly of Prp3/4 before the assembly of LSm could be weaker due to the steric hindrance of these multiple conformations on the 3' end of U6. Consistent with this, our single-molecule data also shows that in the presence of a minimal U4/U6 duplex, without the 3' single stranded region, Prp3/4 binds to the duplex with a higher affinity even without LSm proteins. Taken together, these results illustrate that binding of LSm is needed for the proper assembly of Prp3/4 on to the duplex. Moreover, tight binding of Prp3/4 in the presence of all proteins suggest that

it could be the last protein assembled onto the di-snRNP assembly. Consistent with our results, previous studies have shown that among the U4/U6 associated proteins, only Prp3 and Prp4 are U4/U6 duplex specific proteins whereas other proteins (Snu13, Prp31, LSm and Sm) are individual snRNP specific[67,235,236]. Hence assembly of Prp3/4 required prior binding of other proteins, giving rise to a tighter K_D for Prp3/4 in the presence of all other components.

Although previous studies have provided some basic low-resolution information about the structure of the U4/U6 complex and how it associates within the tri-snRNP, the relative orientation and the global structure of the U4/U6 duplex in the presence of its associated proteins remains structurally unresolved. Also, the mechanism by which Brr2 disrupts the U4/U6 duplex is poorly understood. To understand the relative orientation of helices and the effect of protein binding on the global structure of the U4/U6 di-snRNP, I have employed single-molecule FRET.

Single-molecule experiments were done with the U4/U6 duplex labelled at different helices to observe the relative orientation of the three helices. The single-molecule data revealed that the U4/U6 duplex adopts and maintains a rigid or static global structure throughout the formation of di-snRNP. Interestingly, the experimentally determined inter-fluorophore distance within stem I and II from the single-molecule data is identical to the calculated total length of the two helices (assuming ideal A-form helices), suggesting that these two helices are coaxially stacked, forming a family A three way junction conformation, which is in contrast with previous studies[86,254]. Interestingly, a new EM structure of U4/U6.U5 tri-snRNP deduced by T. Nguyen et al., has also shown that the stems I and II are coaxially stacked[255].

Despite the static global conformation of the U4/U6 duplex, I have observed some heterogeneity within the population distribution of FRET states, where the two constructs designed to study the orientation of the 5' stem-loop relative to stems I and II adopt two major, non-interconverting conformations. The minor high FRET population that was observed consistently in the absence as well as in the presence of proteins (~10%) likely corresponds to a misfolded conformation, which was also observed in the fluorescence EMSA study.

Additionally, experiments were carried out to visualise the assembly of proteins at the single-molecule level. Single-molecule FRET and molecular dynamics simulation assays have shown that the K-turn motif in the 5' stem-loop adopts two conformations (extended and compact structures), and binding of Snu13 stabilizes the latter conformation[256]. Furthermore, studies have reported on salt dependent conformational changes of the K-turn motif into the compact structure[247,256]. Based on these studies the K-turn is folded into a more tight arrangement in the presence of 20 to 100 mM Na^+ or ~ 0.1 to 1 mM Mg^{2+}. Our single-molecule data in the presence of 100 mM Na^+ has indicated that the population percentages for the two conformations do not change upon binding of Snu13. This result indicates that under the experimental conditions used, the K-turn in the 5' stem-loop of U4 is already folded into a compact conformation, which is in accordance with previous studies[247,256]. Hence, binding of Snu13 would not cause further folding of the K-turn, which suggests that the observed population heterogeneity has no relationship with K-turn dynamics.

Furthermore, our single-molecule data shows that binding of Prp31 stabilizes one observed population whereas Prp3/4 stabilizes another population, indicating that

binding of these two proteins to the U4/U6 duplex results in conformational changes within the duplex. Prp31 has been shown to interact with the 5' stem-loop of U4[69,72] and Prp3/4 binds to the stem II[86]. Thus, based on our results we assigned the two populations to two conformations; population 1 where the 5' stem-loop is closer to stem I and population 2 where the 5' stem-loop is closer to stem II. No change was seen for the complex with and without Snu13, suggesting that although it facilitates the binding of Prp31, Snu13 doesn't affect the overall structural changes occurring within the duplex upon binding of this protein. Similar to our gel shift data, single-molecule results have also shown that either Prp31 or Prp3/4 can bind to the U4/U6/Snu13 complex first, followed by the other. Interestingly, the fully assembled complex adopts a conformation similar to what we observed for RNA alone. Hence we propose that the binding of individual proteins or partial assembly of di-snRNP can cause some local structural rearrangement, mainly in the 5' stem-loop of U4, whereas stem I and II orientations remain unchanged. Furthermore, our results indicate that binding of one protein (either Prp31 or Prp3/4) changes the 5' stem-loop conformation in a way that facilitates the binding of the other protein (either Prp3/4 or Prp31). Otherwise, the fully assembled complex upholds a static global conformation.

Maintaining a rigid conformation of U4/U6 may have a role in the function of U4/U6 duplex as a negative regulator, which prevents U6 from premature activation. Also, the structure of U4/U6 may play an important role in Brr2 activity. As shown in previous studies, Brr2 plays an important role in the unwinding of U4 from U6 allowing U6 to base pair with U2 to form the catalytic core[35,84,89,187]. Nevertheless, the mechanism of Brr2 unwinding of these two snRNAs is not yet clear. Previous studies have also

proposed that U4/U6 associated proteins may also play an important role in the stabilization/destabilization of this duplex[69]. During the unwinding process, U6 needs to be free of proteins in order to be available for efficient pairing with U2 while U4 is released with U4 associated proteins. From our structural data, we deduce a possible mechanism for Brr2 function, where we propose that having a static conformation of U4/U6 duplex may favour the Brr2 activity by letting it move through the duplex easily, which in turn facilitates the unwinding of U6 from the complex, as a free snRNA while keeping most of the proteins with U4. Releasing with most of the proteins bound could be an efficient way of recycling U4 after its release from spliceosome.

Taken together, our EMSA data reveal a step-wise assembly of proteins onto the RNA duplex and we provide a protocol for full assembly of the di-snRNP, which could be useful in the reconstitution of other snRNP particles as well. Our single-molecule data revealed interesting information concerning the relative orientation of the three helices of the U4/U6 duplex, where we propose that stem I and stem II are coaxially stacked with each other and the overall di-snRNP maintains a rigid conformation. This static structure may promote Brr2 unwinding activity, by allowing it to remove U6 easily from U4. Overall, the resulting rigid structure for the U4/U6 complex suggests that U4/U6 adopts a preformed conformation and acts as a scaffold for protein binding**,** while preventing U6 from premature activation.

6.2.1: Future directions

To further understand and characterize the assembly of U4/U6 di-snRNP the following experiments are proposed. Since we have now found that the U4/U6 di-snRNP

maintains a rigid global structure, the immediate question arising is what will happen to this rigid conformation upon formation of U4/U6.U5 tri-snRNP. To answer this question, further experiments can be done using labelled U5 snRNP. These experiments will provide information about how U4/U6 duplex interacts with the U5 snRNA, how these snRNAs are rearranged and which conformation they adopt upon formation of tri-snRNP. Also, as a further advancement of this project, CoSMoS studies can also be performed to visualise the colocalization of all components of the tri-snRNP; U4/U6 di-snRNP and the U5 snRNP, revealing the formation of tri-snRNP at the single-molecule level. To understand the effect of other components present in the spliceosome on the assembly of U4/U6 di-snRNP, single-molecule experiments can be done with spliceosomal extracts. For this study, the endogenous di-snRNP needs to be depleted and the exogenous complex with fluorophore labels needs to be introduced. This would provide more information on how other spliceosomal components affect the global structure of U4/U6 di-snRNP and provide a clearer picture on the overall assembly process. Taken together, these experiments will help to elucidate the assembly of tri-snRNP, determine the global structure of U4/U6 duplex within the tri-snRNP and reveal the relative orientations and interactions between U4/U6 duplex and U5 snRNA. Along with the recent cryo-EM structure of U4/U6.U5 tri-snRNP, these results may lay a further step on the progress of understanding the overall structure of U4/U6.U5 tri-snRNP, hence understanding the overall picture of the spliceosome.

6.3: Study the structural dynamics of U12/U6atac snRNA complex in minor spliceosome

To address the third aim; understand the structural and functional similarities and differences of major and minor spliceosomal snRNA complexes, I have studied the conformational dynamics of U12/U6atac complex using single-molecule FRET.

Previous studies have shown that minor spliceosomal U12/U6atac complex forms a secondary structure similar to the three helix junction structure of major spliceosomal U2/U6 complex[147]. Hence, I have carried out single-molecule experiments to determine whether minor U12/U6atac complex also undergoes conformational dynamics as observed for U2/U6 complex. The results from single-molecule data indicate that formation of a minimal U12/U6atac complex produces a FRET value of 0.2. Also, this study has revealed that U6atac adopts a static high FRET conformation, whereas in the addition of increasing concentrations of U12, molecules were oscillating between FRET 1.0 and 0.2 and with higher U12 concentrations molecules adopting a static low FRET conformation (0.2). These complex dynamics could be due to the small size of U12, which may result in partial binding with U6atac. Because of the unstable binding, U12 could easily dissociate allowing for free U6atac to revert to the high FRET conformation. Furthermore, presence of only one FRET state (0.2) upon formation of U12/U6atac complex suggests that due to the short length of U12 snRNA, the U12/U6atac complex cannot undergo any conformational dynamics, as we observed for the U2/U6. Upon analysis of the sequence, it clearly shows that U12/U6atac cannot form the helix III structure as present in the major spliceosomal complex and the complex adopts a conformation similar to the three helix junction structure of U2/U6.

Furthermore, the single-molecule experiments were done to determine the effect of Mg^{2+} ions on the stability of U12/U6atac complex. The results indicate that in the addition of varying concentrations of Mg^{2+} the FRET trajectories change from static 1.0 FRET state to a static 0.2 FRET state, through a dynamic state fluctuate between 1.0 and 0.2 FRET states. These results suggest that the Mg^{2+} ions stabilize the U12/U6atac complex formed at a conformation similar to the three helix junction structure of U2/U6.

As a whole, these data with the minor spliceosomal complex reveal the formation of minimal U12/U6atac complex and the conformation of this complex is similar to a three helix junction structure like its major spliceosomal counterpart. Also, these data illustrate that the presence of Mg^{2+} stabilizes this conformation.

6.3.1: Future directions

As this study is on-going, there are additional experiments that need to be done in order to reveal the structural and functional relationship between minor and major spliceosomal snRNA complexes. For further investigation of the structural dynamics and the significance of minor spliceosomal U12/U6atac complex as the catalytic core, more single-molecule FRET studies can be done with U12/U6atac complex. Similar to major U6, U6atac also has three highly conserved regions; AAGGAGA loop, AGC triad, U46 (U80 in Major U6) and ISL structure, that have been shown to be important for the complex formation and the catalytic activity[147]. Also, it has been shown that RNA-RNA interactions at the catalytic core of the minor spliceosome are markedly similar to that of the U2/U6 complex[228]. Similar to its counterpart, these conserved regions within the minor spliceosomal U6atac are important for the formation of the catalytic core, interact

with pre-mRNA and rearrange the active site facilitating the catalysis. Hence, obtaining information related to the orientation and interactions between U12/U6atac duplex and pre-mRNA, conformational dynamics within the complex upon activation can enlighten the overall picture of the assembly and machinery of the minor spliceosome, which has not been revealed yet. The single-molecule experiments with U6atac containing mutations at highly conserved regions (ACAGAGA loop and AGC triad) can be done to determine the effect of those regions on the formation of U12/U6atac complex, as well as in the catalysis. Similarly, singe-molecule experiments can be done to study the effect of U46, which is known to be important for the metal ion coordination as its major U6 snRNA counterpart, U80. U6atac snRNA with mutated U46 to A, G or C or deleted U46 can be used in these assays to reveal the effect of these changes on the formation of the complex, similar to the study done with U80 in major U6[185]. The experiments with these mutations provide information about the importance and specificity of these regions in assembly and catalysis. Other than U6atac, U12 have also shown to contain some important regions. Previously a photochemical study has shown that an upstream 5' exon region is necessary for stabilizing the RNA-RNA interactions in the catalytic core[257]. Therefore, single-molecule experiments can be done to study the effect of these highly conserved regions on the U12/U6atac complex formation, simply by mutating those regions within U12[258]. Another important study that can be done is determine the conformation of pre-mRNA bound U12/U6atac complex and elucidate the conformational dynamics within the snRNA helices upon pre-mRNA binding. It has been shown that U6atac interacts with the 5' SS in a same manner as U6 and the 5' SS mutations can be rescued using some compensatory mutations in U6atac[228]. To obtain

this information, single-molecule experiments can be done in the presence of short RNA strand consisting 5’ and 3’ splice sites (resembling pre-mRNA), with and without mutations at 5’ SS and U6atac strand. These experiments will evaluate the effect of pre-mRNA binding on the overall conformation of U12/U6atac complex. Altogether the results from these experiments will reveal more information related to the conformation of U12/U6atac complex, importance of the conserved regions within the two snRNAs on the active site formation and how U12/U6atac interacts with other components at the catalytic core, which eventually leads to understand how this structural information is related to its function. Other than the U12/U6atac complex, similar experiments can be done to study the structural information about other minor spliceosomal sub-complexes such as U4atac/U6atac di-snRNP. As a whole, these studies provide more information regarding minor spliceosome, which has not studies in extend, though it plays an important role in regulation of many important genes.

REFERENCES

1. Wahl, M.C., Will, C.L. & Luhrmann, R. The spliceosome: design principles of a dynamic RNP machine. *Cell* **136**, 701-18 (2009).
2. Proudfoot, N.J., Furger, A. & Dye, M.J. Integrating mRNA processing with transcription. *Cell* **108**, 501-12 (2002).
3. Rogozin, I.B., Carmel, L., Csuros, M. & Koonin, E.V. Origin and evolution of spliceosomal introns. *Biol Direct* **7**, 11 (2012).
4. Toor, N., Keating, K.S. & Pyle, A.M. Structural insights into RNA splicing. *Curr Opin Struct Biol* **19**, 260-6 (2009).
5. Will, C.L. & Luhrmann, R. Spliceosome structure and function. *Cold Spring Harb Perspect Biol* **3**(2011).
6. Hoskins, A.A. & Moore, M.J. The spliceosome: a flexible, reversible macromolecular machine. *Trends Biochem Sci* **37**, 179-88 (2012).
7. Kalnina, Z., Zayakin, P., Silina, K. & Line, A. Alterations of pre-mRNA splicing in cancer. *Genes Chromosomes Cancer* **42**, 342-57 (2005).
8. Bonnal, S., Vigevani, L. & Valcarcel, J. The spliceosome as a target of novel antitumour drugs. *Nat Rev Drug Discov* **11**, 847-59 (2012).
9. Moore, M.J. & Proudfoot, N.J. Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell* **136**, 688-700 (2009).
10. Valadkhan, S. snRNAs as the catalysts of pre-mRNA splicing. *Curr Opin Chem Biol* **9**, 603-8 (2005).
11. Warf, M.B. & Berglund, J.A. Role of RNA structure in regulating pre-mRNA splicing. *Trends Biochem Sci* **35**, 169-78 (2010).
12. Black, D.L. Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem* **72**, 291-336 (2003).
13. Saltzman, A.L., Pan, Q. & Blencowe, B.J. Regulation of alternative splicing by the core spliceosomal machinery. *Genes Dev* **25**, 373-84 (2011).
14. Wang, G.S. & Cooper, T.A. Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat Rev Genet* **8**, 749-61 (2007).
15. Roy, S.W. & Gilbert, W. The evolution of spliceosomal introns: patterns, puzzles and progress. *Nat Rev Genet* **7**, 211-21 (2006).
16. Brow, D.A. Allosteric cascade of spliceosome activation. *Annu Rev Genet* **36**, 333-60 (2002).
17. Pozzoli, U. et al. Comparative analysis of the human dystrophin and utrophin gene structures. *Genetics* **160**, 793-8 (2002).
18. Penny, D., Hoeppner, M.P., Poole, A.M. & Jeffares, D.C. An overview of the introns-first theory. *J Mol Evol* **69**, 527-40 (2009).
19. Rodriguez-Trelles, F., Tarrio, R. & Ayala, F.J. Origins and evolution of spliceosomal introns. *Annu Rev Genet* **40**, 47-76 (2006).
20. Irimia, M. & Roy, S.W. Origin of spliceosomal introns and alternative splicing. *Cold Spring Harb Perspect Biol* **6**(2014).
21. Sharp, P.A. On the origin of RNA splicing and introns. *Cell* **42**, 397-400 (1985).
22. Haugen, P., Simon, D.M. & Bhattacharya, D. The natural history of group I introns. *Trends Genet* **21**, 111-9 (2005).
23. Fedorova, O. & Zingler, N. Group II introns: structure, folding and splicing mechanism. *Biol Chem* **388**, 665-78 (2007).

24. Hedberg, A. & Johansen, S.D. Nuclear group I introns in self-splicing and beyond. *Mob DNA* **4**, 17 (2013).
25. Cech, T.R., Herschlag, D., Piccirilli, J.A. & Pyle, A.M. RNA catalysis by a group I ribozyme. Developing a model for transition state stabilization. *J Biol Chem* **267**, 17479-82 (1992).
26. Saldanha, R., Mohr, G., Belfort, M. & Lambowitz, A.M. Group I and group II introns. *FASEB J* **7**, 15-24 (1993).
27. Tanner, N.K. Ribozymes: the characteristics and properties of catalytic RNAs. *FEMS Microbiol Rev* **23**, 257-75 (1999).
28. Adams, P.L., Stahley, M.R., Kosek, A.B., Wang, J. & Strobel, S.A. Crystal structure of a self-splicing group I intron with both exons. *Nature* **430**, 45-50 (2004).
29. Golden, B.L., Kim, H. & Chase, E. Crystal structure of a phage Twort group I ribozyme-product complex. *Nat Struct Mol Biol* **12**, 82-9 (2005).
30. Guo, F., Gooding, A.R. & Cech, T.R. Structure of the Tetrahymena ribozyme: base triple sandwich and metal ion at the active site. *Mol Cell* **16**, 351-62 (2004).
31. Roitzsch, M., Fedorova, O. & Pyle, A.M. The 2'-OH group at the group II intron terminus acts as a proton shuttle. *Nat Chem Biol* **6**, 218-224 (2010).
32. Toor, N., Keating, K.S., Taylor, S.D. & Pyle, A.M. Crystal structure of a self-spliced group II intron. *Science* **320**, 77-82 (2008).
33. Faustino, N.A. & Cooper, T.A. Pre-mRNA splicing and human disease. *Genes Dev* **17**, 419-37 (2003).
34. Cooper, T.A., Wan, L. & Dreyfuss, G. RNA and disease. *Cell* **136**, 777-93 (2009).
35. Fica, S.M. et al. RNA catalyses nuclear pre-mRNA splicing. *Nature* **503**, 229-34 (2013).
36. Ritchie, D.B., Schellenberg, M.J. & MacMillan, A.M. Spliceosome structure: piece by piece. *Biochim Biophys Acta* **1789**, 624-33 (2009).
37. Hesselberth, J.R. Lives that introns lead after splicing. *Wiley Interdiscip Rev RNA* **4**, 677-91 (2013).
38. Will, C.L. & Luhrmann, R. Splicing of a rare class of introns by the U12-dependent spliceosome. *Biol Chem* **386**, 713-24 (2005).
39. Valadkhan, S. & Jaladat, Y. The spliceosomal proteome: at the heart of the largest cellular ribonucleoprotein machine. *Proteomics* **10**, 4128-41 (2010).
40. Warkocki, Z. et al. Reconstitution of both steps of Saccharomyces cerevisiae splicing with purified spliceosomal components. *Nat Struct Mol Biol* **16**, 1237-43 (2009).
41. Hoskins, A.A. et al. Ordered and dynamic assembly of single spliceosomes. *Science* **331**, 1289-95 (2011).
42. Makarov, E.M. et al. Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. *Science* **298**, 2205-8 (2002).
43. Dunn, E.A. & Rader, S.D. Secondary structure of U6 small nuclear RNA: implications for spliceosome assembly. *Biochem Soc Trans* **38**, 1099-104 (2010).
44. van der Feltz, C., Anthony, K., Brilot, A. & Pomeranz Krummel, D.A. Architecture of the spliceosome. *Biochemistry* **51**, 3321-33 (2012).
45. Kambach, C., Walke, S. & Nagai, K. Structure and assembly of the spliceosomal small nuclear ribonucleoprotein particles. *Curr Opin Struct Biol* **9**, 222-30 (1999).
46. Kolb, S.J., Battle, D.J. & Dreyfuss, G. Molecular functions of the SMN complex. *J Child Neurol* **22**, 990-4 (2007).
47. Stanek, D. & Neugebauer, K.M. The Cajal body: a meeting place for spliceosomal snRNPs in the nuclear maze. *Chromosoma* **115**, 343-54 (2006).
48. Kambach, C. et al. Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. *Cell* **96**, 375-87 (1999).

49. Achsel, T. et al. A doughnut-shaped heteromer of human Sm-like proteins binds to the 3'-end of U6 snRNA, thereby facilitating U4/U6 duplex formation in vitro. *EMBO J* **18**, 5789-802 (1999).
50. Weber, G., Trowitzsch, S., Kastner, B., Luhrmann, R. & Wahl, M.C. Functional organization of the Sm core in the crystal structure of human U1 snRNP. *EMBO J* **29**, 4172-84 (2010).
51. Stark, H., Dube, P., Luhrmann, R. & Kastner, B. Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle. *Nature* **409**, 539-42 (2001).
52. Pomeranz Krummel, D.A., Oubridge, C., Leung, A.K., Li, J. & Nagai, K. Crystal structure of human spliceosomal U1 snRNP at 5.5 A resolution. *Nature* **458**, 475-80 (2009).
53. Kondo, Y., Oubridge, C., van Roon, A.M. & Nagai, K. Crystal structure of human U1 snRNP, a small nuclear ribonucleoprotein particle, reveals the mechanism of 5' splice site recognition. *Elife* **4**(2015).
54. Zhang, L., Li, X. & Zhao, R. Structural analyses of the pre-mRNA splicing machinery. *Protein Sci* **22**, 677-92 (2013).
55. Query, C.C., Moore, M.J. & Sharp, P.A. Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model. *Genes Dev* **8**, 587-97 (1994).
56. Sashital, D.G., Venditti, V., Angers, C.G., Cornilescu, G. & Butcher, S.E. Structure and thermodynamics of a conserved U2 snRNA domain from yeast and human. *RNA* **13**, 328-38 (2007).
57. Hilliker, A.K., Mefford, M.A. & Staley, J.P. U2 toggles iteratively between the stem IIa and stem IIc conformations to promote pre-mRNA splicing. *Genes Dev* **21**, 821-34 (2007).
58. Perriman, R.J. & Ares, M., Jr. Rearrangement of competing U2 RNA helices within the spliceosome promotes multiple steps in splicing. *Genes Dev* **21**, 811-20 (2007).
59. McKay, S.L. & Johnson, T.L. An investigation of a role for U2 snRNP spliceosomal components in regulating transcription. *PLoS One* **6**, e16077 (2011).
60. Golas, M.M., Sander, B., Will, C.L., Luhrmann, R. & Stark, H. Molecular architecture of the multiprotein splicing factor SF3b. *Science* **300**, 980-4 (2003).
61. Stark, H. & Luhrmann, R. Cryo-electron microscopy of spliceosomal components. *Annu Rev Biophys Biomol Struct* **35**, 435-57 (2006).
62. Gozani, O., Feld, R. & Reed, R. Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A. *Genes Dev* **10**, 233-43 (1996).
63. Query, C.C., Strobel, S.A. & Sharp, P.A. Three recognition events at the branch-site adenine. *EMBO J* **15**, 1392-402 (1996).
64. Will, C.L. et al. Characterization of novel SF3b and 17S U2 snRNP proteins, including a human Prp5p homologue and an SF3b DEAD-box protein. *EMBO J* **21**, 4978-88 (2002).
65. Price, S.R., Evans, P.R. & Nagai, K. Crystal structure of the spliceosomal U2B"-U2A' protein complex bound to a fragment of U2 small nuclear RNA. *Nature* **394**, 645-50 (1998).
66. Bringmann, P. et al. Evidence for the existence of snRNAs U4 and U6 in a single ribonucleoprotein complex and for their association by intermolecular base pairing. *EMBO J* **3**, 1357-63 (1984).
67. Will, C.L. & Luhrmann, R. Spliceosomal UsnRNP biogenesis, structure and function. *Curr Opin Cell Biol* **13**, 290-301 (2001).
68. Kastner, B., Bach, M. & Luhrmann, R. Electron microscopy of U4/U6 snRNP reveals a Y-shaped U4 and U6 RNA containing domain protruding from the U4 core RNP. *J Cell Biol* **112**, 1065-72 (1991).
69. Nottrott, S., Urlaub, H. & Luhrmann, R. Hierarchical, clustered protein interactions with U4/U6 snRNA: a biochemical role for U4/U6 proteins. *EMBO J* **21**, 5527-38 (2002).

70. Nottrott, S. et al. Functional interaction of a novel 15.5kD [U4/U6.U5] tri-snRNP protein with the 5' stem-loop of U4 snRNA. *EMBO J* **18**, 6119-33 (1999).
71. Vidovic, I., Nottrott, S., Hartmuth, K., Luhrmann, R. & Ficner, R. Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment. *Mol Cell* **6**, 1331-42 (2000).
72. Liu, S. et al. Binding of the human Prp31 Nop domain to a composite RNA-protein platform in U4 snRNP. *Science* **316**, 115-20 (2007).
73. Horowitz, D.S., Kobayashi, R. & Krainer, A.R. A new cyclophilin and the human homologues of yeast Prp3 and Prp4 form a complex associated with U4/U6 snRNPs. *RNA* **3**, 1374-87 (1997).
74. Anthony, J.G., Weidenhammer, E.M. & Woolford, J.L., Jr. The yeast Prp3 protein is a U4/U6 snRNP protein necessary for integrity of the U4/U6 snRNP and the U4/U6.U5 tri-snRNP. *RNA* **3**, 1143-52 (1997).
75. Ayadi, L. et al. Functional and structural characterization of the prp3 binding domain of the yeast prp4 splicing factor. *J Mol Biol* **284**, 673-87 (1998).
76. Lauber, J. et al. The human U4/U6 snRNP contains 60 and 90kD proteins that are structurally homologous to the yeast splicing factors Prp4p and Prp3p. *RNA* **3**, 926-41 (1997).
77. Sander, B. et al. Organization of core spliceosomal components U5 snRNA loop I and U4/U6 Di-snRNP within U4/U6.U5 Tri-snRNP as revealed by electron cryomicroscopy. *Mol Cell* **24**, 267-78 (2006).
78. Dix, I., Russell, C.S., O'Keefe, R.T., Newman, A.J. & Beggs, J.D. Protein-RNA interactions in the U5 snRNP of Saccharomyces cerevisiae. *RNA* **4**, 1675-86 (1998).
79. O'Keefe, R.T., Norman, C. & Newman, A.J. The invariant U5 snRNA loop 1 sequence is dispensable for the first catalytic step of pre-mRNA splicing in yeast. *Cell* **86**, 679-89 (1996).
80. O'Keefe, R.T. & Newman, A.J. Functional analysis of the U5 snRNA loop 1 in the second catalytic step of yeast pre-mRNA splicing. *EMBO J* **17**, 565-74 (1998).
81. Will, C.L., Behrens, S.E. & Luhrmann, R. Protein composition of mammalian spliceosomal snRNPs. *Mol Biol Rep* **18**, 121-6 (1993).
82. Turner, I.A., Norman, C.M., Churcher, M.J. & Newman, A.J. Dissection of Prp8 protein defines multiple interactions with crucial RNA sequences in the catalytic core of the spliceosome. *RNA* **12**, 375-86 (2006).
83. Collins, C.A. & Guthrie, C. Allele-specific genetic interactions between Prp8 and RNA active site residues suggest a function for Prp8 at the catalytic core of the spliceosome. *Genes Dev* **13**, 1970-82 (1999).
84. Raghunathan, P.L. & Guthrie, C. RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2. *Curr Biol* **8**, 847-55 (1998).
85. Small, E.C., Leggett, S.R., Winans, A.A. & Staley, J.P. The EF-G-like GTPase Snu114p regulates spliceosome dynamics mediated by Brr2p, a DExD/H box ATPase. *Mol Cell* **23**, 389-99 (2006).
86. Hacker, I. et al. Localization of Prp8, Brr2, Snu114 and U4/U6 proteins in the yeast tri-snRNP by electron microscopy. *Nat Struct Mol Biol* **15**, 1206-12 (2008).
87. Madhani, H.D. & Guthrie, C. A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. *Cell* **71**, 803-17 (1992).
88. Lauber, J. et al. The HeLa 200 kDa U5 snRNP-specific protein and its homologue in Saccharomyces cerevisiae are members of the DEXH-box protein family of putative RNA helicases. *EMBO J* **15**, 4001-15 (1996).
89. Kim, D.H. & Rossi, J.J. The first ATPase domain of the yeast 246-kDa protein is required for in vivo unwinding of the U4/U6 duplex. *RNA* **5**, 959-71 (1999).
90. van Nues, R.W. & Beggs, J.D. Functional contacts with a range of splicing proteins suggest a central role for Brr2p in the dynamic control of the order of events in spliceosomes of Saccharomyces cerevisiae. *Genetics* **157**, 1451-67 (2001).

91. Hahn, D. & Beggs, J.D. Brr2p RNA helicase with a split personality: insights into structure and function. *Biochem Soc Trans* **38**, 1105-9 (2010).
92. Galej, W.P., Oubridge, C., Newman, A.J. & Nagai, K. Crystal structure of Prp8 reveals active site cavity of the spliceosome. *Nature* **493**, 638-43 (2013).
93. Nguyen, T.H. et al. Structural basis of Brr2-Prp8 interactions and implications for U5 snRNP biogenesis and the spliceosome active site. *Structure* **21**, 910-19 (2013).
94. Pena, V., Rozov, A., Fabrizio, P., Luhrmann, R. & Wahl, M.C. Structure and function of an RNase H domain at the heart of the spliceosome. *EMBO J* **27**, 2929-40 (2008).
95. Chen, H.C. & Cheng, S.C. Functional roles of protein splicing factors. *Biosci Rep* **32**, 345-59 (2012).
96. Pena, V., Liu, S., Bujnicki, J.M., Luhrmann, R. & Wahl, M.C. Structure of a multipartite protein-protein interaction domain in splicing factor prp8 and its link to retinitis pigmentosa. *Mol Cell* **25**, 615-24 (2007).
97. Bellare, P. et al. A role for ubiquitin in the spliceosome assembly pathway. *Nat Struct Mol Biol* **15**, 444-51 (2008).
98. Fabrizio, P., Laggerbauer, B., Lauber, J., Lane, W.S. & Luhrmann, R. An evolutionarily conserved U5 snRNP-specific protein is a GTP-binding factor closely related to the ribosomal translocase EF-2. *EMBO J* **16**, 4092-106 (1997).
99. Brenner, T.J. & Guthrie, C. Assembly of Snu114 into U5 snRNP requires Prp8 and a functional GTPase domain. *RNA* **12**, 862-71 (2006).
100. Martin-Tumasz, S., Reiter, N.J., Brow, D.A. & Butcher, S.E. Structure and functional implications of a complex containing a segment of U6 RNA bound by a domain of Prp24. *RNA* **16**, 792-804 (2010).
101. Vidaver, R.M., Fortner, D.M., Loos-Austin, L.S. & Brow, D.A. Multiple functions of Saccharomyces cerevisiae splicing protein Prp24 in U6 RNA structural rearrangements. *Genetics* **153**, 1205-18 (1999).
102. Raghunathan, P.L. & Guthrie, C. A spliceosomal recycling factor that reanneals U4 and U6 small nuclear ribonucleoprotein particles. *Science* **279**, 857-60 (1998).
103. Kramer, A. The structure and function of proteins involved in mammalian pre-mRNA splicing. *Annu Rev Biochem* **65**, 367-409 (1996).
104. Kanopka, A., Muhlemann, O. & Akusjarvi, G. Inhibition by SR proteins of splicing of a regulated adenovirus pre-mRNA. *Nature* **381**, 535-8 (1996).
105. Liu, Z. et al. Structural basis for recognition of the intron branch site RNA by splicing factor 1. *Science* **294**, 1098-102 (2001).
106. Sickmier, E.A. et al. Structural basis for polypyrimidine tract recognition by the essential pre-mRNA splicing factor U2AF65. *Mol Cell* **23**, 49-59 (2006).
107. Shen, H. et al. Distinct activities of the DExD/H-box splicing factor hUAP56 facilitate stepwise assembly of the spliceosome. *Genes Dev* **22**, 1796-803 (2008).
108. Perriman, R., Barta, I., Voeltz, G.K., Abelson, J. & Ares, M., Jr. ATP requirement for Prp5p function is determined by Cus2p and the structure of U2 small nuclear RNA. *Proc Natl Acad Sci U S A* **100**, 13857-62 (2003).
109. Staley, J.P. & Guthrie, C. Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell* **92**, 315-26 (1998).
110. Teigelkamp, S., McGarvey, M., Plumpton, M. & Beggs, J.D. The splicing factor PRP2, a putative RNA helicase, interacts directly with pre-mRNA. *EMBO J* **13**, 888-97 (1994).
111. Lardelli, R.M., Thompson, J.X., Yates, J.R., 3rd & Stevens, S.W. Release of SF3 from the intron branchpoint activates the first step of pre-mRNA splicing. *RNA* **16**, 516-28 (2010).

112. Schwer, B. & Guthrie, C. PRP16 is an RNA-dependent ATPase that interacts transiently with the spliceosome. *Nature* **349**, 494-9 (1991).
113. Umen, J.G. & Guthrie, C. Prp16p, Slu7p, and Prp8p interact with the 3' splice site in two distinct stages during the second catalytic step of pre-mRNA splicing. *RNA* **1**, 584-97 (1995).
114. McPheeters, D.S. & Muhlenkamp, P. Spatial organization of protein-RNA interactions in the branch site-3' splice site region during pre-mRNA splicing in yeast. *Mol Cell Biol* **23**, 4174-86 (2003).
115. Mefford, M.A. & Staley, J.P. Evidence that U2/U6 helix I promotes both catalytic steps of pre-mRNA splicing and rearranges in between these steps. *RNA* **15**, 1386-97 (2009).
116. McPheeters, D.S., Schwer, B. & Muhlenkamp, P. Interaction of the yeast DExH-box RNA helicase prp22p with the 3' splice site during the second step of nuclear pre-mRNA splicing. *Nucleic Acids Res* **28**, 1313-21 (2000).
117. Schwer, B. A conformational rearrangement in the spliceosome sets the stage for Prp22-dependent mRNA release. *Mol Cell* **30**, 743-54 (2008).
118. Tsai, R.T. et al. Spliceosome disassembly catalyzed by Prp43 and its associated components Ntr1 and Ntr2. *Genes Dev* **19**, 2991-3003 (2005).
119. Tsai, R.T. et al. Dynamic interactions of Ntr1-Ntr2 with Prp43 and with U5 govern the recruitment of Prp43 to mediate spliceosome disassembly. *Mol Cell Biol* **27**, 8027-37 (2007).
120. Tanaka, N., Aronova, A. & Schwer, B. Ntr1 activates the Prp43 helicase to trigger release of lariat-intron from the spliceosome. *Genes Dev* **21**, 2312-25 (2007).
121. Grote, M. et al. Molecular architecture of the human Prp19/CDC5L complex. *Mol Cell Biol* **30**, 2105-19 (2010).
122. Grainger, R.J., Barrass, J.D., Jacquier, A., Rain, J.C. & Beggs, J.D. Physical and genetic interactions of yeast Cwc21p, an ortholog of human SRm300/SRRM2, suggest a role at the catalytic center of the spliceosome. *RNA* **15**, 2161-73 (2009).
123. Ajuh, P. et al. Functional analysis of the human CDC5L complex and identification of its components by mass spectrometry. *EMBO J* **19**, 6569-81 (2000).
124. Silverman, E.J. et al. Interaction between a G-patch protein and a spliceosomal DEXD/H-box ATPase that is critical for splicing. *Mol Cell Biol* **24**, 10101-10 (2004).
125. Yeh, T.C. et al. Splicing factor Cwc22 is required for the function of Prp2 and for the spliceosome to escape from a futile pathway. *Mol Cell Biol* **31**, 43-53 (2011).
126. Liu, Y.C., Chen, H.C., Wu, N.Y. & Cheng, S.C. A novel splicing factor, Yju2, is associated with NTC and acts after Prp2 in promoting the first catalytic reaction of pre-mRNA splicing. *Mol Cell Biol* **27**, 5403-13 (2007).
127. Chiu, Y.F. et al. Cwc25 is a novel splicing factor required after Prp2 and Yju2 to facilitate the first catalytic reaction. *Mol Cell Biol* **29**, 5671-8 (2009).
128. Tseng, C.K., Liu, H.L. & Cheng, S.C. DEAH-box ATPase Prp16 has dual roles in remodeling of the spliceosome in catalytic steps. *RNA* **17**, 145-54 (2011).
129. Fabrizio, P. et al. The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. *Mol Cell* **36**, 593-608 (2009).
130. Wyatt, J.R., Sontheimer, E.J. & Steitz, J.A. Site-specific cross-linking of mammalian U5 snRNP to the 5' splice site before the first step of pre-mRNA splicing. *Genes Dev* **6**, 2542-53 (1992).
131. Deckert, J. et al. Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions. *Mol Cell Biol* **26**, 5528-43 (2006).
132. Hogg, R., McGrail, J.C. & O'Keefe, R.T. The function of the NineTeen Complex (NTC) in regulating spliceosome conformations and fidelity during pre-mRNA splicing. *Biochem Soc Trans* **38**, 1110-5 (2010).

133. Bessonov, S., Anokhina, M., Will, C.L., Urlaub, H. & Luhrmann, R. Isolation of an active step I spliceosome and composition of its RNP core. *Nature* **452**, 846-50 (2008).
134. Coltri, P., Effenberger, K., Chalkley, R.J., Burlingame, A.L. & Jurica, M.S. Breaking up the C complex spliceosome shows stable association of proteins with the lariat intron intermediate. *PLoS One* **6**, e19061 (2011).
135. Jurica, M.S., Sousa, D., Moore, M.J. & Grigorieff, N. Three-dimensional structure of C complex spliceosomes by electron microscopy. *Nat Struct Mol Biol* **11**, 265-9 (2004).
136. Yan, C. et al. Structure of a yeast spliceosome at 3.6-angstrom resolution. *Science* **349**, 1182-91 (2015).
137. Ohi, M.D., Ren, L., Wall, J.S., Gould, K.L. & Walz, T. Structural characterization of the fission yeast U5.U2/U6 spliceosome complex. *Proc Natl Acad Sci U S A* **104**, 3195-200 (2007).
138. Schultz, A., Nottrott, S., Hartmuth, K. & Luhrmann, R. RNA structural requirements for the association of the spliceosomal hPrp31 protein with the U4 and U4atac small nuclear ribonucleoproteins. *J Biol Chem* **281**, 28278-86 (2006).
139. Lin, C.F., Mount, S.M., Jarmolowski, A. & Makalowski, W. Evolutionary dynamics of U12-type spliceosomal introns. *BMC Evol Biol* **10**, 47 (2010).
140. Chen, W. & Moore, M.J. The spliceosome: disorder and dynamics defined. *Curr Opin Struct Biol* **24**, 141-9 (2014).
141. Burge, C.B., Padgett, R.A. & Sharp, P.A. Evolutionary fates and origins of U12-type introns. *Mol Cell* **2**, 773-85 (1998).
142. Bartschat, S. & Samuelsson, T. U12 type introns were lost at multiple occasions during evolution. *BMC Genomics* **11**, 106 (2010).
143. Patel, A.A. & Steitz, J.A. Splicing double: insights from the second spliceosome. *Nat Rev Mol Cell Biol* **4**, 960-70 (2003).
144. Russell, A.G., Charette, J.M., Spencer, D.F. & Gray, M.W. An early evolutionary origin for the minor spliceosome. *Nature* **443**, 863-6 (2006).
145. Tarn, W.Y. & Steitz, J.A. Pre-mRNA splicing: the discovery of a new spliceosome doubles the challenge. *Trends Biochem Sci* **22**, 132-7 (1997).
146. Turunen, J.J., Niemela, E.H., Verma, B. & Frilander, M.J. The significant other: splicing by the minor spliceosome. *Wiley Interdiscip Rev RNA* **4**, 61-76 (2013).
147. Tarn, W.Y. & Steitz, J.A. Highly diverged U4 and U6 small nuclear RNAs required for splicing rare AT-AC introns. *Science* **273**, 1824-32 (1996).
148. Shukla, G.C. & Padgett, R.A. Conservation of functional features of U6atac and U12 snRNAs between vertebrates and higher plants. *RNA* **5**, 525-38 (1999).
149. Tarn, W.Y. & Steitz, J.A. A novel spliceosome containing U11, U12, and U5 snRNPs excises a minor class (AT-AC) intron in vitro. *Cell* **84**, 801-11 (1996).
150. Schneider, C., Will, C.L., Makarova, O.V., Makarov, E.M. & Luhrmann, R. Human U4/U6.U5 and U4atac/U6atac.U5 tri-snRNPs exhibit similar protein compositions. *Mol Cell Biol* **22**, 3219-29 (2002).
151. Tarn, W.Y., Yario, T.A. & Steitz, J.A. U12 snRNA in vertebrates: evolutionary conservation of 5' sequences implicated in splicing of pre-mRNAs containing a minor class of introns. *RNA* **1**, 644-56 (1995).
152. Will, C.L., Schneider, C., Reed, R. & Luhrmann, R. Identification of both shared and distinct proteins in the major and minor spliceosomes. *Science* **284**, 2003-5 (1999).
153. Wu, Q. & Krainer, A.R. AT-AC pre-mRNA splicing mechanisms and conservation of minor introns in voltage-gated ion channel genes. *Mol Cell Biol* **19**, 3225-36 (1999).
154. Younis, I. et al. Minor introns are embedded molecular switches regulated by highly unstable U6atac snRNA. *Elife* **2**, e00780 (2013).

155. Pessa, H.K. & Frilander, M.J. Genetics. Minor splicing, disrupted. *Science* **332**, 184-5 (2011).
156. Edery, P. et al. Association of TALS developmental disorder with defect in minor splicing component U4atac snRNA. *Science* **332**, 240-3 (2011).
157. Shukla, G.C., Cole, A.J., Dietrich, R.C. & Padgett, R.A. Domains of human U4atac snRNA required for U12-dependent splicing in vivo. *Nucleic Acids Res* **30**, 4650-7 (2002).
158. Vidal, V.P., Verdone, L., Mayes, A.E. & Beggs, J.D. Characterization of U6 snRNA-protein interactions. *RNA* **5**, 1470-81 (1999).
159. Karaduman, R., Fabrizio, P., Hartmuth, K., Urlaub, H. & Luhrmann, R. RNA structure and RNA-protein interactions in purified yeast U6 snRNPs. *J Mol Biol* **356**, 1248-62 (2006).
160. Wolff, T. & Bindereif, A. 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**, 1377-89 (1993).
161. Madhani, H.D., Bordonne, R. & Guthrie, C. Multiple roles for U6 snRNA in the splicing pathway. *Genes Dev* **4**, 2264-77 (1990).
162. Mayes, A.E., Verdone, L., Legrain, P. & Beggs, J.D. Characterization of Sm-like proteins in yeast and their association with U6 snRNA. *EMBO J* **18**, 4321-31 (1999).
163. Ryan, D.E., Stevens, S.W. & Abelson, J. The 5' and 3' domains of yeast U6 snRNA: Lsm proteins facilitate binding of Prp24 protein to the U6 telestem region. *RNA* **8**, 1011-33 (2002).
164. Montemayor, E.J. et al. Core structure of the U6 small nuclear ribonucleoprotein at 1.7-A resolution. *Nat Struct Mol Biol* **21**, 544-51 (2014).
165. Ryan, D.E. & Abelson, J. The conserved central domain of yeast U6 snRNA: importance of U2-U6 helix Ia in spliceosome assembly. *RNA* **8**, 997-1010 (2002).
166. Jacquier, A. Self-splicing group II and nuclear pre-mRNA introns: how similar are they? *Trends Biochem Sci* **15**, 351-4 (1990).
167. Keating, K.S., Toor, N., Perlman, P.S. & Pyle, A.M. A structural analysis of the group II intron active site and implications for the spliceosome. *RNA* **16**, 1-9 (2010).
168. Konforti, B.B. et al. Ribozyme catalysis from the major groove of group II intron domain 5. *Mol Cell* **1**, 433-41 (1998).
169. Huppler, A., Nikstad, L.J., Allmann, A.M., Brow, D.A. & Butcher, S.E. Metal binding and base ionization in the U6 RNA intramolecular stem-loop structure. *Nat Struct Biol* **9**, 431-5 (2002).
170. Fortner, D.M., Troy, R.G. & Brow, D.A. A stem/loop in U6 RNA defines a conformational switch required for pre-mRNA splicing. *Genes Dev* **8**, 221-33 (1994).
171. Sashital, D.G., Cornilescu, G., McManus, C.J., Brow, D.A. & Butcher, S.E. U2-U6 RNA folding reveals a group II intron-like domain and a four-helix junction. *Nat Struct Mol Biol* **11**, 1237-42 (2004).
172. Rhode, B.M., Hartmuth, K., Westhof, E. & Luhrmann, R. Proximity of conserved U6 and U2 snRNA elements to the 5' splice site region in activated spliceosomes. *EMBO J* **25**, 2475-86 (2006).
173. Shukla, G.C. & Padgett, R.A. The intramolecular stem-loop structure of U6 snRNA can functionally replace the U6atac snRNA stem-loop. *RNA* **7**, 94-105 (2001).
174. McPheeters, D.S. Interactions of the yeast U6 RNA with the pre-mRNA branch site. *RNA* **2**, 1110-23 (1996).
175. Madhani, H.D. & Guthrie, C. Dynamic RNA-RNA interactions in the spliceosome. *Annu Rev Genet* **28**, 1-26 (1994).
176. Hilliker, A.K. & Staley, J.P. Multiple functions for the invariant AGC triad of U6 snRNA. *RNA* **10**, 921-8 (2004).
177. Reiter, N.J., Blad, H., Abildgaard, F. & Butcher, S.E. Dynamics in the U6 RNA intramolecular stem-loop: a base flipping conformational change. *Biochemistry* **43**, 13739-47 (2004).

178. Brow, D.A. & Guthrie, C. Spliceosomal RNA U6 is remarkably conserved from yeast to mammals. *Nature* **334**, 213-8 (1988).
179. Shannon, K.W. & Guthrie, C. Suppressors of a U4 snRNA mutation define a novel U6 snRNP protein with RNA-binding motifs. *Genes Dev* **5**, 773-85 (1991).
180. Segault, V. et al. 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**, 2782-90 (1999).
181. Valadkhan, S., Mohammadi, A., Jaladat, Y. & Geisler, S. Protein-free small nuclear RNAs catalyze a two-step splicing reaction. *Proc Natl Acad Sci U S A* **106**, 11901-6 (2009).
182. Sun, J.S. & Manley, J.L. A novel U2-U6 snRNA structure is necessary for mammalian mRNA splicing. *Genes Dev* **9**, 843-54 (1995).
183. Burke, J.E., Sashital, D.G., Zuo, X., Wang, Y.X. & Butcher, S.E. Structure of the yeast U2/U6 snRNA complex. *RNA* **18**, 673-83 (2012).
184. Anokhina, M. et al. RNA structure analysis of human spliceosomes reveals a compact 3D arrangement of snRNAs at the catalytic core. *EMBO J* **32**, 2804-18 (2013).
185. Guo, Z., Karunatilaka, K.S. & Rueda, D. Single-molecule analysis of protein-free U2-U6 snRNAs. *Nat Struct Mol Biol* **16**, 1154-9 (2009).
186. Guo, Z. Single molecule studies of spliceosomal snRNAs u2-u6. in *Department of Chemistry* Vol. Doctor of Phylosophy (Wayne state Univedrsity, Detroit, Michigan, 2010).
187. Fica, S.M., Mefford, M.A., Piccirilli, J.A. & Staley, J.P. Evidence for a group II intron-like catalytic triplex in the spliceosome. *Nat Struct Mol Biol* **21**, 464-71 (2014).
188. Smith, D.J., Query, C.C. & Konarska, M.M. "Nought may endure but mutability": spliceosome dynamics and the regulation of splicing. *Mol Cell* **30**, 657-66 (2008).
189. Umen, J.G. & Guthrie, C. A novel role for a U5 snRNP protein in 3' splice site selection. *Genes Dev* **9**, 855-68 (1995).
190. Newman, A.J. & Norman, C. U5 snRNA interacts with exon sequences at 5' and 3' splice sites. *Cell* **68**, 743-54 (1992).
191. Arenas, J.E. & Abelson, J.N. Prp43: An RNA helicase-like factor involved in spliceosome disassembly. *Proc Natl Acad Sci U S A* **94**, 11798-802 (1997).
192. Bae, E. et al. Structure and interactions of the first three RNA recognition motifs of splicing factor prp24. *J Mol Biol* **367**, 1447-58 (2007).
193. Karaduman, R. et al. Structure of yeast U6 snRNPs: arrangement of Prp24p and the LSm complex as revealed by electron microscopy. *RNA* **14**, 2528-37 (2008).
194. Pontius, B.W. & Berg, P. Rapid assembly and disassembly of complementary DNA strands through an equilibrium intermediate state mediated by A1 hnRNP protein. *J Biol Chem* **267**, 13815-8 (1992).
195. Portman, D.S. & Dreyfuss, G. RNA annealing activities in HeLa nuclei. *EMBO J* **13**, 213-21 (1994).
196. Furger, A., O'Sullivan, J.M., Binnie, A., Lee, B.A. & Proudfoot, N.J. Promoter proximal splice sites enhance transcription. *Genes Dev* **16**, 2792-9 (2002).
197. Jandrositz, A. & Guthrie, C. Evidence for a Prp24 binding site in U6 snRNA and in a putative intermediate in the annealing of U6 and U4 snRNAs. *EMBO J* **14**, 820-32 (1995).
198. Brow, D.A. & Guthrie, C. Splicing a spliceosomal RNA. *Nature* **337**, 14-5 (1989).
199. Strauss, E.J. & Guthrie, C. A cold-sensitive mRNA splicing mutant is a member of the RNA helicase gene family. *Genes Dev* **5**, 629-41 (1991).
200. Ghetti, A., Company, M. & Abelson, J. Specificity of Prp24 binding to RNA: a role for Prp24 in the dynamic interaction of U4 and U6 snRNAs. *RNA* **1**, 132-45 (1995).
201. Oberstrass, F.C. et al. Structure of PTB bound to RNA: specific binding and implications for splicing regulation. *Science* **309**, 2054-7 (2005).

202. Lamichhane, R. et al. RNA looping by PTB: Evidence using FRET and NMR spectroscopy for a role in splicing repression. *Proc Natl Acad Sci U S A* **107**, 4105-10 (2010).
203. Hargous, Y. et al. Molecular basis of RNA recognition and TAP binding by the SR proteins SRp20 and 9G8. *EMBO J* **25**, 5126-37 (2006).
204. Clery, A., Blatter, M. & Allain, F.H. RNA recognition motifs: boring? Not quite. *Curr Opin Struct Biol* **18**, 290-8 (2008).
205. Martin-Tumasz, S., Richie, A.C., Clos, L.J., 2nd, Brow, D.A. & Butcher, S.E. A novel occluded RNA recognition motif in Prp24 unwinds the U6 RNA internal stem loop. *Nucleic Acids Res* (2011).
206. Kwan, S.S. & Brow, D.A. The N- and C-terminal RNA recognition motifs of splicing factor Prp24 have distinct functions in U6 RNA binding. *RNA* **11**, 808-20 (2005).
207. Hashimoto, C. & Steitz, J.A. U4 and U6 RNAs coexist in a single small nuclear ribonucleoprotein particle. *Nucleic Acids Res* **12**, 3283-93 (1984).
208. Ward, A.J. & Cooper, T.A. The pathobiology of splicing. *J Pathol* **220**, 152-63 (2010).
209. Licatalosi, D.D. & Darnell, R.B. Splicing regulation in neurologic disease. *Neuron* **52**, 93-101 (2006).
210. Lakowicz, J.R. *Principles of Fluoresence Spectroscopy*, (Springer, 2006).
211. Terai, T. & Nagano, T. Fluorescent probes for bioimaging applications. *Curr Opin Chem Biol* **12**, 515-21 (2008).
212. Stryer, L. Fluorescence energy transfer as a spectroscopic ruler. *Annu Rev Biochem* **47**, 819-46 (1978).
213. Clegg, R.M. Fluorescence resonance energy transfer. *Curr Opin Biotechnol* **6**, 103-10 (1995).
214. Periasamy, A. Fluorescence resonance energy transfer microscopy: a mini review. *J Biomed Opt* **6**, 287-91 (2001).
215. Roy, R., Hohng, S. & Ha, T. A practical guide to single-molecule FRET. *Nat Methods* **5**, 507-16 (2008).
216. Klostermeier, D. & Millar, D.P. Time-resolved fluorescence resonance energy transfer: a versatile tool for the analysis of nucleic acids. *Biopolymers* **61**, 159-79 (2001).
217. Ha, T. Single-molecule fluorescence resonance energy transfer. *Methods* **25**, 78-86 (2001).
218. Cornish, P.V. & Ha, T. A survey of single-molecule techniques in chemical biology. *ACS Chem Biol* **2**, 53-61 (2007).
219. Tinoco, I., Jr. & Gonzalez, R.L., Jr. Biological mechanisms, one molecule at a time. *Genes Dev* **25**, 1205-31 (2011).
220. Hoskins, A.A., Gelles, J. & Moore, M.J. New insights into the spliceosome by single molecule fluorescence microscopy. *Curr Opin Chem Biol* **15**, 864-70 (2011).
221. Ishikawa-Ankerhold, H.C., Ankerhold, R. & Drummen, G.P. Advanced fluorescence microscopy techniques--FRAP, FLIP, FLAP, FRET and FLIM. *Molecules* **17**, 4047-132 (2012).
222. Haustein, E. & Schwille, P. Single-molecule spectroscopic methods. *Curr Opin Struct Biol* **14**, 531-40 (2004).
223. Zhao, R. & Rueda, D. RNA folding dynamics by single-molecule fluorescence resonance energy transfer. *Methods* **49**, 112-7 (2009).
224. Karunatilaka, K.S. & Rueda, D. Post-transcriptional modifications modulate conformational dynamics in human U2-U6 snRNA complex. *RNA* **20**, 16-23 (2014).
225. Crawford, D.J., Hoskins, A.A., Friedman, L.J., Gelles, J. & Moore, M.J. Single-molecule colocalization FRET evidence that spliceosome activation precedes stable approach of 5' splice site and branch site. *Proc Natl Acad Sci U S A* **110**, 6783-8 (2013).
226. Krishnan, R. et al. Biased Brownian ratcheting leads to pre-mRNA remodeling and capture prior to first-step splicing. *Nat Struct Mol Biol* **20**, 1450-7 (2013).

227. Shcherbakova, I. et al. Alternative spliceosome assembly pathways revealed by single-molecule fluorescence microscopy. *Cell Rep* **5**, 151-65 (2013).
228. Incorvaia, R. & Padgett, R.A. Base pairing with U6atac snRNA is required for 5' splice site activation of U12-dependent introns in vivo. *RNA* **4**, 709-18 (1998).
229. Padgett, R.A. & Shukla, G.C. A revised model for U4atac/U6atac snRNA base pairing. *RNA* **8**, 125-8 (2002).
230. Walter, N.G. Structural dynamics of catalytic RNA highlighted by fluorescence resonance energy transfer. *Methods* **25**, 19-30 (2001).
231. Aleman, E.A., Lamichhane, R. & Rueda, D. Exploring RNA folding one molecule at a time. *Curr Opin Chem Biol* **12**, 647-54 (2008).
232. Rueda, D. & Walter, N.G. Fluorescent energy transfer readout of an aptazyme-based biosensor. *Methods Mol Biol* **335**, 289-310 (2006).
233. Lamichhane, R., Solem, A., Black, W. & Rueda, D. Single-molecule FRET of protein-nucleic acid and protein-protein complexes: surface passivation and immobilization. *Methods* **52**, 192-200 (2010).
234. Aitken, C.E., Marshall, R.A. & Puglisi, J.D. An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments. *Biophys J* **94**, 1826-35 (2008).
235. Klostermeier, D. & Millar, D.P. RNA conformation and folding studied with fluorescence resonance energy transfer. *Methods* **23**, 240-54 (2001).
236. Prescott, E.M. & Proudfoot, N.J. Transcriptional collision between convergent genes in budding yeast. *Proc Natl Acad Sci U S A* **99**, 8796-801 (2002).
237. Valadkhan, S. & Manley, J.L. Splicing-related catalysis by protein-free snRNAs. *Nature* **413**, 701-7 (2001).
238. Karunatilaka, K.S., Solem, A., Pyle, A.M. & Rueda, D. Single-molecule analysis of Mss116-mediated group II intron folding. *Nature* **467**, 935-9 (2010).
239. Rueda, D., Wick, K., McDowell, S.E. & Walter, N.G. Diffusely bound Mg2+ ions slightly reorient stems I and II of the hammerhead ribozyme to increase the probability of formation of the catalytic core. *Biochemistry* **42**, 9924-36 (2003).
240. Penedo, J.C., Wilson, T.J., Jayasena, S.D., Khvorova, A. & Lilley, D.M. Folding of the natural hammerhead ribozyme is enhanced by interaction of auxiliary elements. *RNA* **10**, 880-8 (2004).
241. Murphy, M.C., Rasnik, I., Cheng, W., Lohman, T.M. & Ha, T. Probing single-stranded DNA conformational flexibility using fluorescence spectroscopy. *Biophys J* **86**, 2530-7 (2004).
242. Rueda, D. et al. Single-molecule enzymology of RNA: essential functional groups impact catalysis from a distance. *Proc Natl Acad Sci U S A* **101**, 10066-71 (2004).
243. Pereira, M.J. et al. Single VS ribozyme molecules reveal dynamic and hierarchical folding toward catalysis. *J Mol Biol* **382**, 496-509 (2008).
244. Wang, J. et al. Single-molecule observation of the induction of k-turn RNA structure on binding L7Ae protein. *Biophys J* **103**, 2541-8 (2012).
245. Lilley, D.M. The structure and folding of kink turns in RNA. *Wiley Interdiscip Rev RNA* **3**, 797-805 (2012).
246. Lafontaine, D.A., Norman, D.G. & Lilley, D.M. Structure, folding and activity of the VS ribozyme: importance of the 2-3-6 helical junction. *EMBO J* **20**, 1415-24 (2001).
247. Goody, T.A., Melcher, S.E., Norman, D.G. & Lilley, D.M. The kink-turn motif in RNA is dimorphic, and metal ion-dependent. *RNA* **10**, 254-64 (2004).
248. McPhee, S.A., Huang, L. & Lilley, D.M. A critical base pair in k-turns that confers folding characteristics and correlates with biological function. *Nat Commun* **5**, 5127 (2014).
249. Huang, L. & Lilley, D.M. The molecular recognition of kink-turn structure by the L7Ae class of proteins. *RNA* **19**, 1703-10 (2013).

250. Maeder, C., Kutach, A.K. & Guthrie, C. ATP-dependent unwinding of U4/U6 snRNAs by the Brr2 helicase requires the C terminus of Prp8. *Nat Struct Mol Biol* **16**, 42-8 (2009).
251. Cooper, M., Johnston, L.H. & Beggs, J.D. Identification and characterization of Uss1p (Sdb23p): a novel U6 snRNA-associated protein with significant similarity to core proteins of small nuclear ribonucleoproteins. *EMBO J* **14**, 2066-75 (1995).
252. Hamm, J. & Mattaj, I.W. Monomethylated cap structures facilitate RNA export from the nucleus. *Cell* **63**, 109-18 (1990).
253. Hamm, J., Darzynkiewicz, E., Tahara, S.M. & Mattaj, I.W. The trimethylguanosine cap structure of U1 snRNA is a component of a bipartite nuclear targeting signal. *Cell* **62**, 569-77 (1990).
254. Lescoute, A. & Westhof, E. Topology of three-way junctions in folded RNAs. *RNA* **12**, 83-93 (2006).
255. Nguyen, T.H. et al. The architecture of the spliceosomal U4/U6.U5 tri-snRNP. *Nature* **523**, 47-52 (2015).
256. Wozniak, A.K. et al. Detecting protein-induced folding of the U4 snRNA kink-turn by single-molecule multiparameter FRET measurements. *RNA* **11**, 1545-54 (2005).
257. Frilander, M.J. & Steitz, J.A. Dynamic exchanges of RNA interactions leading to catalytic core formation in the U12-dependent spliceosome. *Mol Cell* **7**, 217-26 (2001).
258. Sikand, K. & Shukla, G.C. Functionally important structural elements of U12 snRNA. *Nucleic Acids Res* **39**, 8531-43 (2011).