UNIVERSITY OF CALIFORNIA
SANTA CRUZ

PROBING THE SECONDARY STRUCTURE OF THE U2 SNRNA FROM THE U2 SNRNP TO INHIBITOR-STALLED SPLICEOSOMES

A dissertation submitted in partial satisfaction
of the requirement for the degree of

DOCTOR OF PHILOSOPHY

in

MOLECULAR, CELL, AND DEVELOPMENTAL BIOLOGY

by

Veronica Kazumi Urabe

June 2021

The Dissertation of Veronica Kazumi Urabe
is approved:

Professor Melissa S. Jurica, Chair

Professor Manuel Ares, Jr.

Professor Susan Strome

Quentin Williams
Vice Provost and Dean of Graduate Studies

Table of contents

1. List of figures

2. List of Supplementary figures

Abstract

Veronica Kazumi Urabe

Probing the secondary structure of the U2 snRNA from the U2 snRNP to inhibitor-stalled spliceosomes

Human gene expression is a complicated process, and when things go wrong, disease may arise. Many diseases such as cancer arise from abnormal gene expression, so understanding the components of our cells that help facilitate normal gene expression is essential. My research focuses on the component known as the spliceosome, a very dynamic and complex molecular machine of ribonucleic acids and proteins. Splicing is a choreographed routine as there are many moving components. SF3B inhibitors target components of the spliceosome and interfere with spliceosome assembly and stability. I hypothesize that the inhibition and instability may be linked to differences in RNA arrangements in the inhibitor stalled complex and A-complex. My dissertation focuses on three goals: First, I investigated the connection between structural rearrangements in U2 snRNA and the mechanism of SF3B inhibitors that target the U2 snRNP. Second, I characterized the intrinsic capabilities of U2 snRNA to adopt different secondary structures. Last, I employed a bimolecular exon ligation reaction to test if SF3B inhibitors interfere with exon ligation. My studies have demonstrated that SF3B inhibitors do not affect U2 snRNA structure in the spliceosome and the snRNP, and the U2 snRNA has an intrinsic ability to adopt alternate and possibly novel structures than previously predicted. The basic research I conduct will ultimately inform therapeutic research to generate treatments for diseases.

Dedication

I dedicate my dissertation to my family. I would also like to dedicate my dissertation to the next generation of Chicana/Latina Scientists. Always remember, life begins at the end of your comfort zone.

Acknowledgment

"I do my best because I am counting on you counting on me."

- Maya Angelou

I am grateful to many people for their support, guidance, and their confidence in me. Thank you for counting on me! I want to acknowledge the following people: My mentor, Professor Melissa Jurica, inspired me; she gave me my first exposure to the spliceosome when I was a community college student. She has been a great advisor throughout my academic trajectory, guiding me throughout the process and pushing me to grow as a scientist. I could not have done this without her coaching and support.

For their support, safe space, and community they provided me, I want to thank past and present members of the Jurica lab, especially Dr. Kerstin Effenberger, Beth Prichard, Hannah Maul-Newby, Angela Amorello, and Dr. Janine Ilagan. You will never know the difference our interactions and gatherings made for me during the ups and downs of this journey.

Along with my advisor, my thesis committee members, Professor Susan Strome and Professor Manny Ares provided me with critical and thoughtful feedback on my dissertation. I am grateful for your time, advice, and support.

Lastly, I am immensely grateful to my family for showing up, especially when things got hard. My husband, Jesus, and my sons, Jordan and Jeremiah, you have given me so much love and support--thank you! My parents, Blanca and Ron, I am

tremendously grateful for your endless love, support, and encouragement. To my siblings for believing in me, loving me unconditionally, and for your faith in me. To my nephews and nieces, I am counting on you to be the next generation of academic scholars. To my extended family and community, my success is our collective success.

This dissertation was supported by the National Science Foundation - Graduate Research Fellowship Program and the Aptos-Capitola Rotary Club - Vera and Henry Piatt Graduate Fellowship Award.

Chapter I: Introduction

1. Eukaryotic gene expression

The central dogma of molecular biology encompasses the transfer of genetic information to a functional protein. The simplistic view is that DNA is transcribed by RNA polymerase to produce an mRNA which is translated to protein by the ribosome. The steps of eukaryotic gene expression covered by the central dogma need to be expanded to include the pre-mRNA processing steps that must also take place. These steps include the addition of the 5' methyl guanosine cap and 3' polyadenylation, which are essential to stabilize the RNA, protect it from degradation and permit nuclear export (Jurado et al. 2014). Additionally, in most eukaryotic genes, the pre-mRNA sequence is divided into regions called introns and exons, which alternate down the pre-mRNA transcript (Berget et al. 1977; Chow et al. 1977). The information for protein translation is held within the exons and is interrupted by intervening introns. The removal of the introns to generate a continuous protein-coding sequence is crucial for proper gene expression. The mechanism by which introns are identified and correctly removed is called pre-mRNA splicing (Will and Luhrmann 2011; Plaschka et al. 2019).

3. Pre-mRNA splicing

Pre-mRNA splicing uses sequences within the pre-mRNA to identify the regions that need to be removed and the regions that need to be kept (Figure I-1). Many researchers use a model pre-mRNA to study splicing, which consists of a 5' or upstream exon, followed by an intronic intervening sequence, and then a 3' or

downstream exon. There is a 5' splice site at the very 5' end of the intron, and towards the 3' end, there is a sequence known as the branch point sequence that contains the branch point adenosine used for splicing chemistry. The branch point is followed by a tract of pyrimidine bases, known as the polypyrimidine tract and finally, at the end of the intron is the 3' splice site. The identification of branch point sequence determines the 3' splice site, together with the identification of the 5' splice site establishes the boundaries of the intron. The 5' exon and the 3' exon are ligated together, and the intron is removed to generate a mature mRNA (Wahl et al. 2009; Will and Luhrmann 2011).

4. The spliceosome

The macromolecular machine that carries out pre-mRNA splicing is called the spliceosome. The spliceosome is a multi-mega Dalton entity that is composed of ribonucleic acid (RNA) and protein. There are five small nuclear ribonucleoprotein particles (snRNPs) named U1, U2, U4, U5, and U6 snRNPs. Each snRNP is a subcomplex that is composed of one corresponding RNA molecule (U1, U2, U4, U5, and U6 snRNA) and many proteins. For example, the U2 snRNP contains the U2 snRNA and the core U2 associated proteins along with SF3B and SF3A subcomplexes. The snRNP complexes come together *de novo* on each intron to be removed from the pre-mRNA transcript in a stepwise fashion and undergo multiple rearrangements through a process known as the spliceosome cycle (Figure I-2). Understanding how the spliceosome works is still a major question for the field of splicing. There are multiple intermediate steps of assembly that are known and characterized; however, new complexes continue to emerge, giving details and

insight into the intricate rearrangements that must occur throughout the spliceosome cycle. The increasing number of cryo-EM models of the spliceosome through many steps of the spliceosome cycle are crucial for the advancement of the field. However, the structures often create more questions than they answer because they imply large rearrangements that are required from early assembly through the formation of a catalytic spliceosome. Additionally, the limited resolution of the cryo-EM structures leaves gaps in the models. Traditional biochemical approaches are necessary to confirm the structural models and can help to provide insight into rearrangements that have yet to be captured in the structures.

5. Splicing Chemistry

The chemistry of pre-mRNA splicing is well-understood (Figure I-1). Once the spliceosome is fully assembled, an active site forms to catalyze two coordinated SN2 displacement reactions that result in the removal of the intron and the joining of the exons. In the first reaction, known as step 1, the branch point adenosine acts as the nucleophile to transfer the phosphodiester bond at the 5' splice site. The result is a free 5' exon with a 3' hydroxyl group and a lariat intron intermediate still connected to the 3' exon. In the subsequent reaction, called step 2, the 3' hydroxyl on the 5' exon acts as a nucleophile and attacks the phosphodiester bond at the 3' splice site to ligate the 5' and 3' exons and release the lariat intron. We can recapitulate this intricate reaction, which occurs at each intron to be removed, in a test tube. We can visualize both steps of splicing chemistry using an *in vitro* splicing assay with a radiolabeled pre-mRNA and HeLa nuclear extract. The *in vitro* splicing assay allows us to follow pre-mRNA, the first step products, lariat intron – 3' exon, free 5' exon, as

well as second step products, the ligated exons or mature mRNA, and released intron.

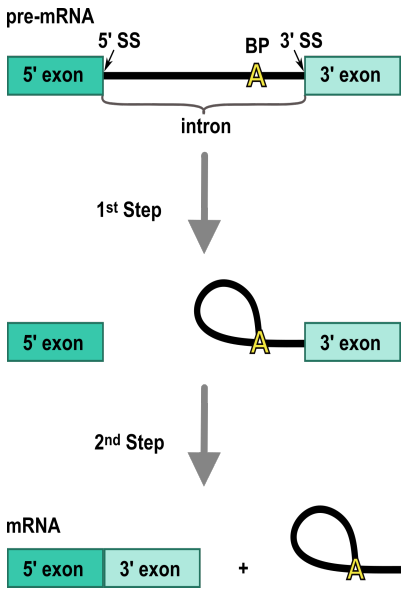


Figure I-1. **Splicing chemistry.** Splicing of a model pre-mRNA, the boxes represent exons, the line represents the intron. The 5' splice site (5' SS), branchpoint adenosine (A), and 3' splice site (3'SS) are labeled. Two transesterification reactions (step 1 and step 2), each represented by a gray arrow, cleave and ligate the exons.

6. Spliceosome assembly

The large-scale rearrangements during spliceosome assembly and the complexity of these rearrangements and interactions have made it incredibly difficult for the field to understand all the steps of spliceosome assembly completely. However, we know when the components join and leave, giving an ordered progression of complexes termed E, A, B, and C (Figure I-2). The first step in

spliceosome assembly is the recognition of the 5’ splice site, which can be found at the 5’ exon-intron boundary. The U1 snRNP is responsible for identifying the correct 5’ splice site. Next, the U2 snRNP helps identify the 3’ end of the intron when the U2 snRNA base pairs with the branch point sequence of the pre-mRNA to form a structure known as the branch helix. Once U1 and U2 snRNP join, A-complex is formed. The U4, U5, and U6 snRNP join the spliceosome as a pre-constructed complex known as the tri-snRNP. Once the tri-snRNP joins the spliceosome, the very 5’ end of the U2 snRNA base pairs with the U6 snRNA, resulting in the formation of B complex (Wu and Manley 1989). Multiple rearrangements take place and U1 and U4 snRNP leave the complex and we arrive at C complex. The position of the RNAs and protein components create the active site within the complex, where both steps of splicing chemistry occur. Finally, the spliced mRNA is released, and then the remaining U2, U5, and U6 snRNPs are disassembled from the lariat intron, which is then degraded (Will and Luhrmann 2011; Wilkinson et al. 2020).

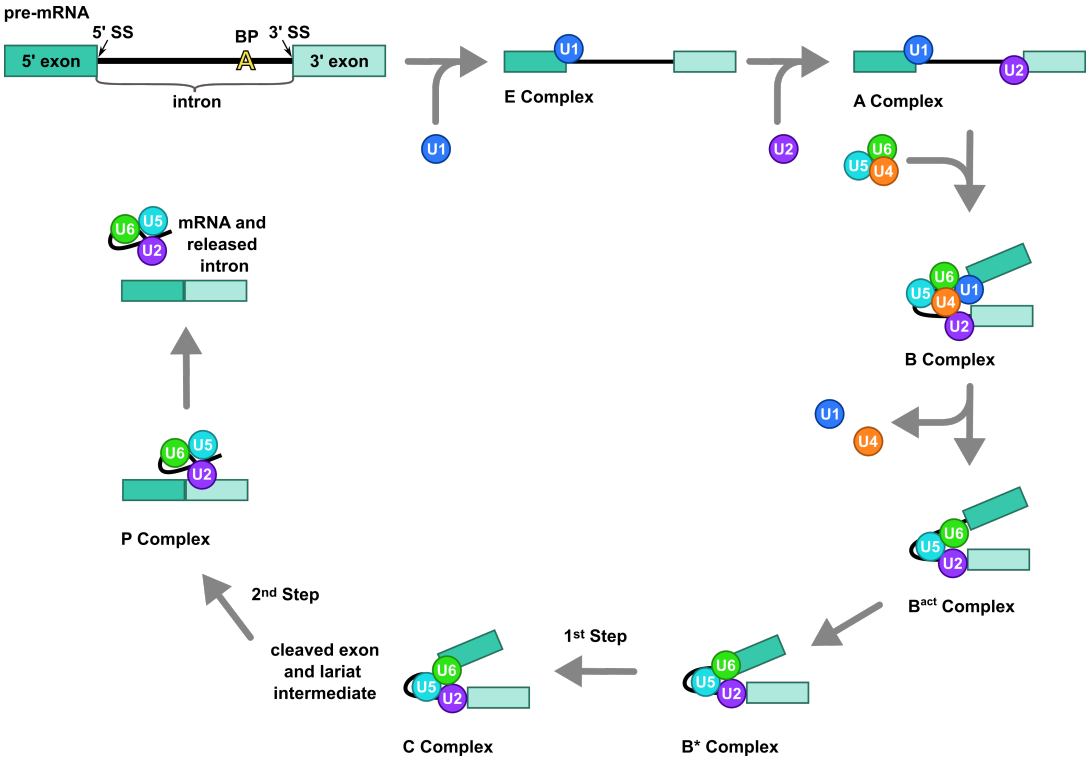


Figure I-2. **Model pre-mRNA and spliceosome assembly.** The model pre-mRNA is shown in the top left corner, the boxes represent exons, the line represents the intron. The 5' splice site (5' SS), branchpoint adenosine (A), and 3' splice site (3'SS) are labeled. The circles represent snRNPs and are color-coded as U1 snRNP, dark blue; U2 snRNP, purple; U4 snRNP, orange; U5 snRNP, light blue; U6 snRNP, green. The stepwise spliceosome assembly intermediates are shown.

7. SF3B inhibitors

A group of natural products including spliceostatin A, pladienolide B, and herboxidiene, were initially discovered as cytotoxic to cancer cells (Nakajima et al. 1996; Sakai et al. 2002; Sakai et al. 2004). These compounds were studied further and discovered to target the protein SF3B1, which is a component of the U2 snRNP, and inhibit splicing (Kaida et al. 2007; Kotake et al. 2007; Hasegawa et al. 2011). In the presence of the compounds, the spliceosome assembly pathway is clearly

interrupted, with an accumulation of A-complex and a loss of later complexes (Roybal and Jurica 2010; Corrionero et al. 2011; Folco et al. 2011; Effenberger et al. 2014). The A-complex formed in the presence of inhibitor is unstable, demonstrated by sensitivity to heparin (Corrionero et al. 2011; Effenberger et al. 2014). Effenberger et al. showed that transition out of A-complex is also hindered by the presence of SF3B inhibitors. The role of SF3B1 may extend past branch point recognition, given that its association with the spliceosome appears to be regulated and continues past the first step of splicing chemistry, which suggests it may function at multiple stages of assembly (Coltri et al. 2011; Lardelli et al. 2010; Ilagan et al. 2013). The open questions regarding SF3B inhibitors in the spliceosome currently are: What happens in an inhibitor stalled spliceosome? Does the inhibitor affect the formation of the branch helix? What is the U2 snRNA secondary structure arrangement in inhibitor stalled spliceosomes?

8. The U2 snRNP

The target of the SF3B inhibitors, SF3B, is a component of the U2 snRNP. The U2 snRNP can be separated into three main parts, a core complex and the two protein subcomplexes SF3A and SF3B. The core contains the U2 snRNA and proteins that associate with its 3' end. The subcomplexes SF3A and SF3B both contain multiple proteins, interact with the 3' and 5' end of the U2 snRNA, respectively. When the three parts are in complex together, they form the 17S U2 snRNP, which refers to the Svedberg coefficient at which the particle sediments in a gradient (Kramer et al. 1999). The interaction between the U2 core complex and SF3B and SF3B subcomplexes can be disrupted in the presence of high salt

concentration. With increasing concentrations of salt SF3B is removed, and a 15S U2 snRNP remains, and when SF3A is removed it becomes the 12S U2 snRNP (Kramer et al. 1999). The early structural studies of the U2 snRNP indicated two globular structures linked by an RNase sensitive region, but now we have a cryo-EM structure of the U2 snRNP which more completely models the 5' and 3' halves of the complex (Behrens et al. 1993; Zhang et al. 2020). The protein components of the U2 snRNP must rearrange extensively relative to each other, within the assembled spliceosomes. The different arrangements of the proteins within the U2 snRNP and splicing complexes lead to the open questions: How are these rearrangements coordinated? How do these rearrangements connect to the structural changes in the U2 snRNA?

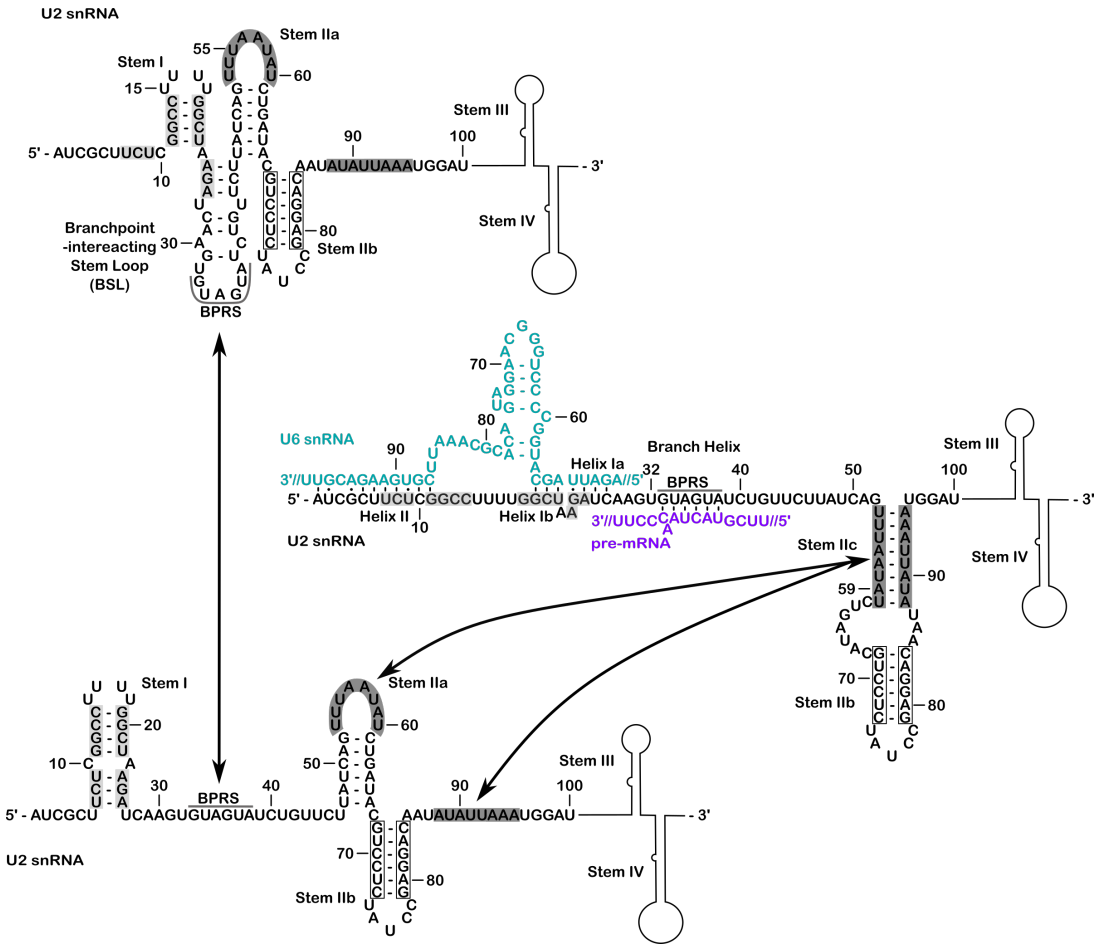


Figure I-3. **U2 snRNA secondary structure dynamics.** The model of the U2 snRNA secondary structure dynamics. The U2 snRNA is shown in black text and the nucleotides are numbered. The key features and different secondary structure elements are labeled. Regions that base pair in a stem are indicated by shaded boxes. The U2 snRNA can base pair with complementary regions to create multiple stem loop structures. The U2 snRNA can adopt multiple conformations. The pre-mRNA is shown in purple. The U6 snRNA is shown in teal. Interactions via base pairing are illustrated with dashes.

9. U2 snRNA dynamics

The RNA within the U2 snRNP, the U2 snRNA, is highly dynamic and must juggle both intramolecular and intermolecular interactions. It is responsible for

engaging the branch point sequence in the intron of the pre-mRNA as well as interacting with the 3' end of the U6 snRNA (Ast et al. 2001). The multiple different interactions make the 5' end of U2 snRNA quite dynamic, while the 3' end is more static. The intramolecular interactions that make the U2 snRNA so dynamic start with Stem I, the bottom half of the stem which shares nucleotides with the Branch point-interacting Stem Loop (BSL) (Figure I-3) (Perriman and Ares 2010). The Branch point-interacting Stem Loop must also unwind to interact with the branch point sequence of the pre-mRNA to form the branch helix. Stem I at the 5' end of the U2 snRNA must completely unwind to interacts with the U6 snRNA. An additional toggle downstream of the Branch point-interacting Stem Loop includes two consecutive stem loops, Stem IIa and Stem IIb. The loop of Stem IIa is complementary to the region downstream of Stem IIb. When Stem IIa is unwound, the loop can base pairs with the region downstream to form the mutually exclusive Stem IIc.

Genetic mutants of U2 snRNA in yeast revealed the switch between these two mutually exclusive structures occurs at different stages of spliceosome assembly (Perriman and Ares 2007; Hilliker et al. 2007). The U2 snRNA is highly conserved which reflects that the base sequence is important for allowing the base pairing interactions as well as the switch between structures. The cryo-EM models alongside the genetic data have created a more complete model for U2 snRNA in the spliceosome. However, there are still regions that are either unknown or unmodeled in the structural data. These different regions may have different functions in spliceosome assembly, but conformational specificity of specific steps and regulation of conformational changes are not understood. These gaps can be filled with the inclusion of chemical probing data.

10. Contributions of this Dissertation

Overall, my studies contribute to the overarching goal of the Jurica lab to create a comprehensive and precise model of the spliceosome throughout the stages of assembly, catalysis, and disassembly. I also aim to identify interactions that are disrupted in the presence of SF3B inhibitors to learn about their activity and the mechanisms of U2 snRNP structural changes in splicing.

In Chapter II, I present the results of chemical probing experiments targeting U2 snRNA in different forms of the U2 snRNP, A-complex spliceosomes, and SF3B inhibitor stalled spliceosomes. I also explored the possibilities and intrinsic capabilities of U2 snRNA to adopt specific secondary structures using a synthetic U2 snRNA. My data support a similar conformation of the U2 snRNA in A-complex and inhibitor stalled complex. Looking at the U2 snRNP and the synthetic RNA, I have evidence of a newly described structural element in the U2 snRNA that is an extension of Stem I. In Chapter III, I employed a technique to temporally separate the two steps of splicing chemistry to investigate the role of SF3B inhibitors after the first chemical step of splicing. In addition to the accumulation of A-complex, and the ability for the spliceosome to move out of A-complex, I demonstrated the inhibitors also interfere with exon ligation. In summary, my biochemical studies help interpret recent cryo-EM structures and fill in some remaining gaps to build a more complete model of the U2 snRNA in the spliceosome.

Chapter II: U2 snRNA from snRNP to inhibitor stalled spliceosomes

1. Abstract

The U2 snRNP is an essential component of the spliceosome. It is responsible for branch point recognition via base-pairing of the U2 snRNA with an intron to form a branch helix in the spliceosome A-complex. Small molecule inhibitors target the SF3B component of the U2 snRNP and interfere with A-complex formation during spliceosome assembly. We previously found that the first inhibitor-stalled-complex (ISC) is less stable than A-complex and hypothesized that SF3B inhibitors interfere with U2 snRNA secondary structure changes required to form the branch helix. Using RNA chemical modifiers, I probed the U2 snRNA conformation in A-complex and ISCs. The reactivity pattern for U2 snRNA in the ISC is indistinguishable from that of A-complex, suggesting that they have the same secondary structure conformation, including the presence of the branch helix. This observation suggests ISC instability does not stem from an alternate RNA conformation and instead points to the inhibitors interfering with protein component interactions that normally stabilize the U2 snRNP's association with an intron. In addition to splicing complexes, I also probed the U2 snRNA in the free U2 snRNP in the presence of inhibitor and again my results pointed to no effect on secondary structure. However, in the absence of SF3A and SF3B, increased protection of nucleotides upstream of Stem I suggest a change of the secondary structure at the very 5' end of U2 snRNA. Probing synthetic U2 snRNA in the absence of proteins results in similar protections and predicts a previously uncharacterized extension of

Stem I. Because this stem must be disrupted for SF3A and SF3B proteins to stably join the snRNP, the structure has the potential to influence snRNP assembly and recycling after spliceosome disassembly.

2. Introduction

The U2 small nuclear ribonucleoprotein particle (snRNP) is a vital component of the human spliceosome. The spliceosome is the macromolecular machine that carries out pre-mRNA splicing, an essential step in eukaryotic gene expression. The spliceosome is composed of five snRNPs which assemble *de novo* on an intron to be removed from a pre-mRNA transcript (Wahl et al. 2009). The U2 snRNP is composed of the U2 snRNA, core proteins, SF3A and SF3B subcomplexes (Brosi et al. 1993; Kramer et al. 1999). During the spliceosome cycle, the U2 snRNP is responsible for branch point recognition via base-pairing interaction with the pre-mRNA to select the branch point adenosine that participates in the first chemical step of splicing (Wu and Manley 1989; Parker et al. 1987). The correct identification of the branch point sequence is important as it designates the 3' end of the intron (Smith et al. 1989). When the U4/U5/U6 tri-snRNP joins, the 5' end of U2 snRNA base pairs with the 3' end of U6 snRNA, which helps form the active site of the spliceosome (Zhang et al. 2017). U2 snRNP's SF3A and SF3B complex proteins hold the branch helix in the catalytic core prior to activation (Bertram et al. 2017). Activation docks the branch helix into the active site resulting in coincident loss of SF3A and SF3B complexes (Yan et al. 2016; Wan et al. 2019). After both steps of splicing occur, the spliceosome must disassemble. The U2 snRNA is released as part of the intron lariat spliceosome (ILS), and the interactions with U6 snRNA and the intron are disrupted

(Wan et al. 2017). Finally, the SF3A and SF3B proteins must rejoin before U2 snRNP can carry out another round of splicing.

The U2 snRNA can be split into two main parts: the dynamic 5' end which is capable of multiple secondary structure rearrangements and, by comparison, a more stable 3' end which has no secondary structure rearrangements. The 3' end of the RNA folds into two large stable stem loop structures that interact with the core proteins U2A' and U2B", along with the Sm ring, which stabilizes the 3' end of the RNA molecule (Price et al. 1998; Zhang et al. 2020). The 5' end contains several stem loop structures that adopt alternate intramolecular and intermolecular conformations. At the 5' end of the U2 snRNA is Stem I, which must unwind to base pair with the U6 snRNA during formation of the spliceosome's active site (Ares and Weiser 1995; Staley and Guthrie 1998; Zhang et al. 2017). The lower portion of Stem I can also exchange to form a different stem known as the Branch point-interacting Stem Loop (BSL). The BSL contains the branch point recognition sequence (BPRS) in its loop and conserved nucleotides forming an imperfect stem (Perriman and Ares 2010). Once the U2 snRNP engages the intron, the BSL must be unwound to fully base pair with the branch point sequence in the branch helix (Perriman and Ares 2010; Liang and Cheng 2015; Zhang et al. 2020). The next region downstream of the BSL forms one of two mutually exclusive secondary structures, Stem IIa and Stem IIc (Perriman and Ares 2007; Hilliker et al. 2007). The loop of Stem IIa is complementary to the 3' half of Stem IIc, which can form when Stem IIa unwinds. Mutational analysis of U2 snRNA in yeast demonstrated a switch between the two stem loops, with Stem IIa promoting spliceosome assembly while Stem IIc promotes splicing chemistry (Zavanelli and Ares 1991; Yan et al. 1998; Perriman and Ares

2007; Hilliker et al. 2007). The multiple secondary structure arrangements that have been observed and the various interactions the U2 snRNA facilitates bring up the questions: Have we observed all the possible secondary structure arrangements of the U2 snRNA? How are the known rearrangements ordered during spliceosome assembly? and Does U2 snRNA adopt other secondary structure conformations after the core U2 snRNP is released from the ILS and before rejoining with the SF3A and SF3B complexes?

Recent cryo-EM structures model some of the U2 snRNA structures before and after different conformational switches and give insight into the way the RNA functions in the spliceosome. The 17S U2 snRNP structure shows density for the top of Stem I, the BSL, and Stem IIa (Zhang et al. 2020). The cryo-EM model of the yeast A-complex showed U2 snRNA interacting with the intron in an extended branch helix and a Stem IIa structure (Figure I-3) (Plaschka et al. 2018). Cryo-EM models after tri-snRNP addition show the interactions between U2 and U6 snRNAs, while structures after activation additionally show Stem IIc after the branch helix is docked in the spliceosome active site (Yan et al. 2015; Galej et al. 2016; Yan et al. 2016; Zhang et al. 2017; Wan et al. 2017). In all the cryo-EM structures, certain nucleotides that map to flexible single stranded regions that connect the stems and some loops cannot be modeled.

The order of important interactions that drive A-complex assembly, including the formation of the branch helix and the unwinding of Stem I, is still an open question. Comparison of the structures of 17S U2 snRNP and the A-complex (Plaschka et al. 2018; Zhang et al. 2020) suggests that several significant

rearrangements must occur in U2 snRNA structure and protein interactions, but how and the order in which the changes are realized are not obvious from the structure. For example, the topology of the branch helix necessitates an undocking of either Stem IIa or the BSL for the intron to thread between U2 snRNA and SF3A/SF3B contacts (Plaschka et al. 2018; Zhang et al. 2018).

At the end of the spliceosome cycle, the structure of U2 snRNA upon release from the intron lariat spliceosome is also not known. The structure of the U2 snRNA after it disengages from the intron lariat spliceosome is likely different than the full U2 snRNP, because interactions with U6, the intron, SF3A, and SF3B are absent (Fourmann et al. 2016; Wan et al. 2017). Before it can participate in another round of splicing, the U2 snRNA must adopt the BSL and Stem IIa structures seen in the 17S U2 snRNP (Zhang et al. 2020).

One tool available to study U2 snRNP function is SF3B inhibitors. These compounds interfere with U2 snRNP's interaction with the intron and arrest spliceosome assembly at an inhibitor-stalled-complex (ISC) that is unstable relative to A-complex (Roybal and Jurica 2010; Effenberger et al. 2016b; Effenberger et al. 2016a). Previous studies suggested that the ISC has a branch helix, but that it may differ from the branch helix in A-complex (Corrionero et al. 2011). How the inhibitors stall the spliceosome is not fully understood.

In this study, I address questions regarding U2 snRNA structure in early spliceosome assembly by chemically probing splicing complexes stalled by either the absence of functional tri-snRNP (A-complex) or in the presence of SF3B inhibitors (ISC). Surprisingly, I detect no differences in the U2 snRNA secondary structure in

the ISC compared to the functional A-complex, and my data are consistent with the presence of the branch helix, Stem IIa, and Stem IIb. Additionally, my results point to the top of Stem I being intact, a region of U2 snRNA that is not modeled in the cryo-EM structure of A-complex. I conclude that the SF3B inhibitors do not interfere with the crucial steps of BSL unwinding and branch helix formation, and that Stem I is unwound after those events occur.

To further understand SF3B inhibitor effects on U2 snRNA structure, I also probed U2 snRNP with and without the SF3A and SF3B proteins present. Again, the inhibitors had no discernable impact on the U2 snRNA structure, but U2 snRNP lacking SF3A/B components revealed unexpected protections in the 5' end of U2 snRNA. Both endogenous and synthetic protein-free U2 snRNA gave rise to a similar pattern of protections that are predicted to extend Stem I. With synthetic U2 snRNA, this structure appears to be quite stable and is not fully disrupted by single nucleotide changes. I postulate that the structure forms upon release of U2 snRNP from the intron lariat complex after splicing and may be important for maintaining an inactive conformation of U2 snRNP until the recruitment of SF3 components to generate a snRNP ready for another round of spliceosome assembly.

3. Results

3.1. U2 snRNA secondary structure is maintained in SF3B inhibited splicing complexes

During spliceosome assembly, the U2 snRNA secondary structure rearranges extensively, adopting multiple conformations. Comparing the recent cryo-

EM structures of a *H. sapiens* 17S U2 snRNP (Zhang et al. 2020) and *S. cerevisiae* A-complex spliceosomes (Plaschka et al. 2018) indicates that the BSL must unwind to allow the branch point recognition sequence (BPRS) to form base pairs with the branch point sequence in a branch helix, while the 3' half of the BSL forms ~10 base pair-like interactions with intron nucleotides upstream of the branch point. The conformation of the first 29 nucleotides of U2 snRNA, including the Stem I region is not modeled in A-complex, presumably due to structural flexibility. Based on the location of the U2 snRNA in the two cryo-EM structures, several rearrangements are necessary to achieve the changes in the topology of RNA/RNA and RNA/protein interactions during A-complex assembly. Since SF3B inhibitors interfere with A-complex function, I hypothesized that the U2 snRNA in ISC exists in an intermediate structure between that in the 17S U2 snRNP and the A-complex.

To test this idea, I assembled spliceosome complexes in HeLa nuclear extract on a pre-mRNA substrate with a branch point sequence 5'-UACUAAC-3' that is perfectly complementary to the U2 snRNA BPRS. The pre-mRNA substrate also contains the hairpin recognition sequence for MS2 (MS2 bacteriophage coat protein), which binds tightly to the fusion protein MS2:MBP (maltose binding protein) for affinity purification (Ilagan and Jurica 2014). To accumulate A-complex, I depleted the nuclear extract of functional tri-snRNP using endogenous RNase H and oligonucleotides complementary to the U4 and U6 snRNA (Effenberger et al. 2016b). To accumulate ISCs, I treated nuclear extract with 1μM spliceostatin A (SSA) prior to spliceosome assembly (Effenberger et al. 2016b). I isolated the spliceosome complexes by size exclusion followed by amylose affinity selection for MS2:MBP. The complexes were then immediately probed with either dimethyl sulfate (DMS) to

modify the Watson-Crick faces of adenosine and cytosine bases or 1-Methyl-7-nitroisatoic anhydride (1m7) to acylate the ribose 2'-OH. Reactive nucleotides are chemically modified and correlate with unpaired bases for DMS and flexibility of the RNA backbone for 1m7, which occurs in unstructured regions. Nucleotides that are protected from modification are unreactive and correlate with stable base pairing in a stem or structured by protein interactions. Modifications of the U2 snRNA were detected by primer extension with an oligonucleotide complementary to nucleotides 97-117. Modifications cause reverse transcriptase to stop at the preceding base, which results in a band at the correlating nucleotide position in a sequencing gel. Because reverse transcriptase can also be stopped by endogenous pseudouridylation and 2' O-methylation of U2 snRNA or by highly stable secondary structure, we compared our results to control reactions lacking either DMS or 1m7 reagents. To more easily interpret the data, I mapped consistent probe-induced stops observed across replicate probing reactions onto different U2 snRNA secondary structure models (Figure II-1D & II-1E).

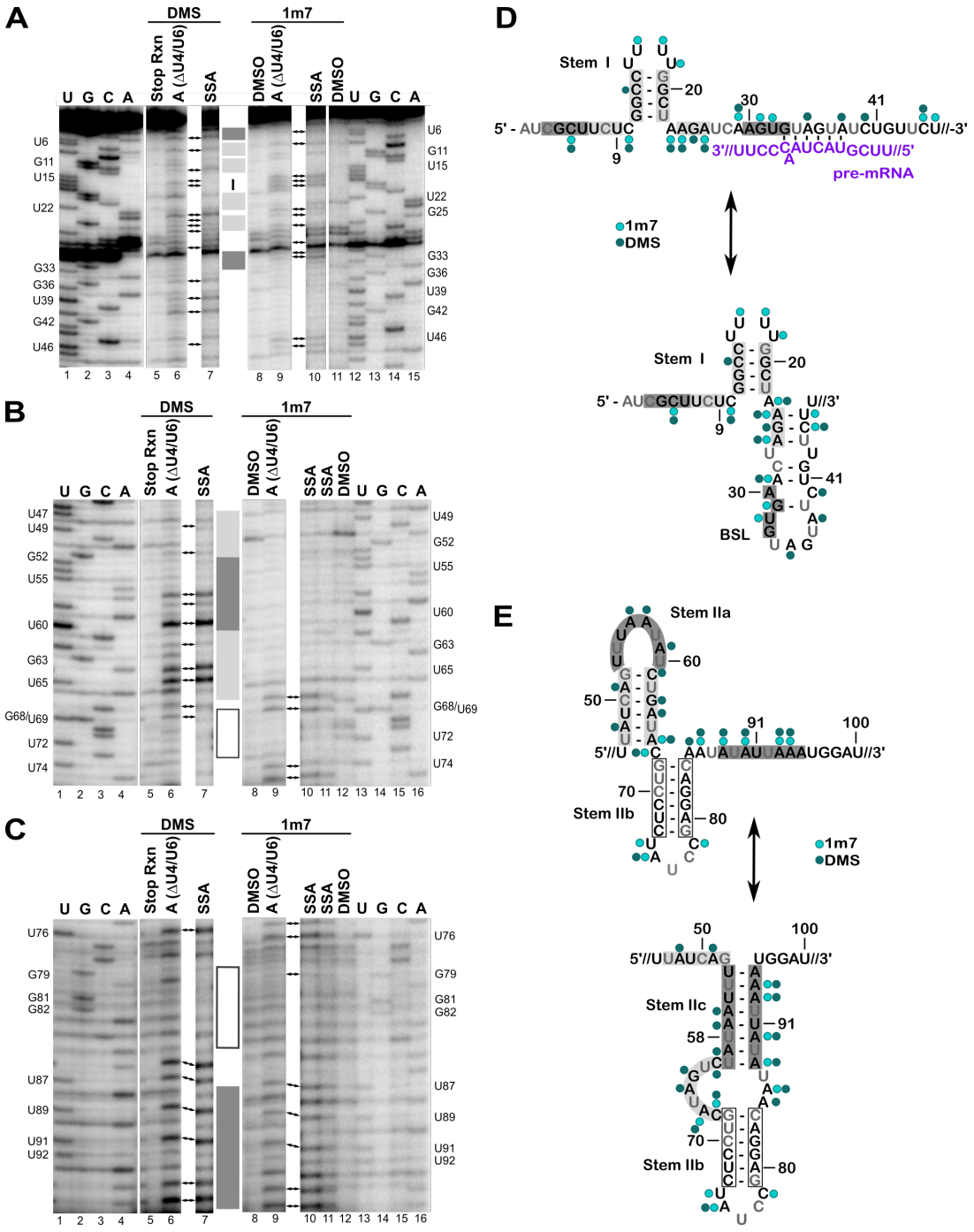


Figure II-1. **Accessibility of the U2 snRNA in A-complex and ISCs.** (A) Chemical modification of A-complex and ISC with DMS and 1m7. Modification of the RNA were detected by primer extension analysis with an oligonucleotide complementary to U2 snRNA nucleotides 97-117. Sequencing lanes are indicated with the corresponding nucleotides, and nucleotide positions are indicated to the left and right of each panel. The structural elements of the U2 snRNA are indicated by shaded boxes at the

center. Arrows point to bands observed consistently across experiments done in triplicate. Nucleotides 1-46 of the U2 snRNA. Lanes 1-4 and 13-16 are ladders. Lanes 5, 8 & 12 are unmodified controls. Lane 6 is A complex, lane 7 is ISC treated with DMS. Lane 8 is A complex, lane 9 & 10 is ISC treated with 1m7. (B) Same as A, nucleotides 47-74 (C) Same as A, nucleotides 75-95 of the u2 snRNA. (D) The accessibility of the U2 snRNA in the two complexes was mapped to two competing structural models of the U2 snRNA, with the top structure matching the reactivity more closely. The U2 snRNA is shown in black text and the nucleotides are numbered. The pre-mRNA is shown in purple. Nucleotides shown in grey, indicate background bands and an inability to map reactivity to the nucleotide. The different secondary structure elements are labeled. Regions that base pair in a stem are indicated by shaded boxes. The circles indicate reactivity at that nucleotide, light green is reactive to 1m7 and dark green is reactive to DMS. Nucleotides 1-46 of the U2 snRNA. (E) Same as D, nucleotides 46-100 of the U2 snRNA.

The overall patterns of reactivity in A-complex and ISC are very similar (Figure II-1). All but one U in the loop of Stem I are reactive to 1m7 (Figure II-1A, lane 9 & 10 compared to lane 8, Figure II-1D), while the upper part of the stem is generally protected, although C13 shows some modification with DMS (Figure II-1A, lane 6 & 7 compared to lane 5). Together the pattern of modification indicates that in both complexes, the top of Stem I is not yet unwound (Figure II-1A & Figure II-1D). Nucleotides 24-26, which can form either the lower region of Stem I or the mutually exclusive BSL are reactive to both reagents, meaning that this region is not base paired (Figure II-1A, lane 6 & 7 compared to lane 5, lane 9 & 10 compared to lane 8, Figure II-1D). DMS and 1m7 reactivity of nucleotides A29 and C45 further support the BSL being unwound, and I conclude that SSA does not interfere with BSL unwinding (Figure II-1A & Figure II-1D).

Nucleotides from the branch point recognition sequence through U44 appear protected to 1m7 in both complexes (Figure II-1A, lanes 9 & 10 compared to lane 8, Figure II-1D). Because 1m7 does not react well with the 2'-OH group when the

nucleotide is in an A-form helix (Mortimer and Weeks 2007), this result means that these nucleotides could be in the extended branch helix duplex. This region of the RNA may be further constrained by interactions with SF3A2 and SF3A3, which are modeled next to U2 snRNA homologous nucleotides 39-45 in the A-complex cryo-EM structure (Plaschka et al. 2018). I conclude that the inhibitor does not block branch helix formation.

Nucleotides A35, A38, and C40 in the same region show some reactivity to DMS in A-complex (Figure II-1A, lane 6 & 7 compared to lane 5, Figure II-1D). This result is surprising for A35 and A38, which are expected to base pair with the two uracil residues in the 5'-UACUAAC-3' branch point sequence of the intron in A-complex included in the pre-mRNA substrate used to assemble these complexes (Figure II-1D). The mild reactivity may be due to deviations compared to the perfect A-form helical conformation in the A-complex cryo-EM model that affects base-pairing interactions. DMS reactivity of the same nucleotides is sightly enhanced in the ISC. The SF3B1 protein goes from an open to a closed conformation upon spliceosome activation (Maji et al. 2019). SSA is predicted to interfere with SF3B1 closing that in the A-complex cryo-EM structure appears to stabilize the branch helix in the A-complex cryo-EM model. In the context of an open SF3B1 conformation, the U2 snRNA may be sampling more than one base-pairing configuration, which would briefly expose the base pairing faces of A35 and A38 to DMS modification.

The similarity between A-complex and ISCs continues through the Stem IIa/IIb/IIc regions (Figure II-1B & 1C). Nucleotides C45 and U46 are reactive to 1m7 in the stretch of nucleotides that links the extended branch helix and Stem IIa. The

low reactivity of the next seventeen nucleotides (47-63) to 1m7 is consistent with the presence of Stem IIa as modeled in the *S. cerevisiae* A-complex (Figure II-1B, lane 9 & 10 compared to lane 8, Figure II-1E). The lack of 1m7 reactivity in the loop of Stem IIa is explained by the tetraloop structure that docks onto SF3B proteins. A64 in the stem is somewhat reactive to DMS (Figure II-1B, lane 6 & 7 compared to lane 5), while both A66 and C67, which link Stem IIa and IIb show stronger reactivity to both DMS and 1m7 (Figure II-1B, lane 6 & 7 compared to lane 5, 9 & 10 compared to lane 8, Figure II-1E). DMS data for nucleotides U47-A66 also support the predicted Stem IIa structure, with nucleotides A48, A51, and C61 in the stem showing weak reactivity and A56, A57, and A59 in the U-turn loop showing stronger reactivity (Figure II-1B, lane 6 & 7 compared to lane 5, Figure II-1E). Surprisingly, both G63 and A64 at the base of the stem are highly reactive to DMS (Figure II-1B, lane 6 & 7 compared to lane 5, Figure II-1E). DMS modification of guanosine at N7 is typically inefficient and should not result in a stop in the primer extension reaction. A64 is modeled as base paired with U49, so the unusual DMS reactivity suggests some sort of structural strain may impact the local chemical environment.

Residues in Stem IIb show the predicted protection of residues G68-C73 (Figure II-1B, lane 6 & 7 compared to lane 5, lane 9 & 10 compared to lane 8, Figure II-1E) and their base pairing partners G79-C83 in both complexes (Figure II-1C, lane 6 & 7 compared to lane 5, lane 9 & 10 compared to lane 8, Figure II-1E). In contrast, U74 and A75 in the loop of Stem IIb, are highly reactive (Figure II-1C, lane 6 & 7 compared to lane 5, lane 9 & 10 compared to lane 8, Figure II-1E). The remaining nucleotides in the loop yield strong stops in both experimental and control reactions and cannot be evaluated (Figure II-1C, lane 6 & 7 compared to lane 5, lane 9 & 10

compared to lane 8). Finally, nucleotides leading into' and continuing to the 3' half of Stem IIc (U86-A94) are generally reactive to both reagents, consistent with its mutually exclusive conformation relative to Stem IIa (Figure II-1C, lane 6 & 7 compared to lane 5, lane 9 & 10 compared to lane 8, Figure II-1E).

Together my data indicate that the conformation of U2 snRNA observed in the *S. cerevisiae* cryo-EM A-complex structure is maintained in human A-complex with an extended branch helix, Stem IIa, and Stem IIb present. I also find that branch helix formation is not significantly impacted by the SF3B inhibitor SSA. This result is consistent with previous psoralen crosslinking studies showing U2 snRNA interaction with an intron, albeit differently positioned relative to the branch point, can be detected in the presence of inhibitor (Corrionero et al. 2011). Because chemical reactivity patterns of nucleotides in U2 snRNA are not significantly different between A-complex spliceosomes and ISCs, SF3B, the target of the inhibitor, must mediate other aspects of spliceosome assembly. Beyond the inability to stably recruit tri-snRNP for the next stage of spliceosome assembly, ISCs also readily disassociate when challenged with the polyanion heparin as compared to A-complex (Effenberger et al. 2016b). In that context, my results imply that branch helix formation alone is not sufficient to stabilize U2 snRNP's engagement with the intron, but that SF3B1 closure over the branch helix, which the inhibitors are proposed to block, is also required. That closure may also signal correct branch point adenosine selection for continued spliceosome assembly. Finally, because the upper region of Stem I is intact in A-complexes arrested by either depletion of functional tri-snRNP or by the presence of SSA, we speculate that unwinding of Stem I to enable interactions with U6 snRNA when tri-snRNP is recruited likely follows SF3B1 closure.

3.2. U2 snRNA structure in the snRNP is dependent on SF3A/B association

The recent cryo-EM structure of U2 snRNP revealed a conformation in which the U2 snRNA adopts four consecutive stem loops: the upper region of Stem I, the BSL, Stem IIa, and Stem IIb (Zhang et al. 2020). Protein interactions, especially with SF3B1 and SF3B2, stabilize Stem IIa relative to the mutually exclusive Stem IIc, which forms after SF3B and SF3A proteins leave the spliceosome during catalytic activation (Zhang et al. 2020). The juxtaposition of DDX46, HTATSF1, and SF3A3 around the BSL may stabilize its structure relative to the competing lower region of Stem I (Zhang et al. 2020). Notably, the position of the BSL is incompatible with the closed conformation of SF3B1 observed in activated spliceosomes (Haselbach et al. 2018). Even with the trove of information provided by the cryo-EM model, it is important to emphasize that the U2 snRNP structure is not rigid. There is a high degree of flexibility between the 5' and 3' structural modules, and the cryo-EM model reflects only one confirmation.

In 1993, Behrens et al reported DMS probing of U2 snRNA in the 17S snRNP, which includes SF3A and SF3B, and the 12S U2 snRNP, which lost the same proteins after high salt treatment (Behrens et al. 1993). However, at that time, the BSL and Stem IIc conformations were not yet realized. I revisited the same experiments to test whether SF3B inhibitors, which are proposed to stabilize SF3B1 in an open conformation affect U2 snRNA structure in the snRNP. I also aimed to reinterpret the results in the context of the recent cryo-EM model of 17S U2 snRNP (Zhang et al. 2020). For my analysis, I used HeLa nuclear extract with either 150 mM potassium chloride (KCl) for 17S U2 snRNP or 420 KCl for 12S U2 snRNP lacking

the SF3B and SF3A proteins (Kramer et al. 1999) both in the presence or absence of SSA. Nucleotide reactivity was mapped as described for A complex and the ISC (Figure II-2C & II-2D).

Similar to the results with splicing complexes, I see no distinct differences in DMS reactivity with SSA treatment (Supplementary Figure II-1). For the first 23 nucleotides, only C10, C14, and C21 top of Stem I are protected in the intact U2 snRNP (Figure II-2A, lane 4 compared to lane 3, Figure II-2C). C13 in the stem shows some reactivity, similar to what I saw in A-complex spliceosomes (Figure II-2A, lane 4 compared to lane 3, Figure II-2C). This pattern is consistent with the U2 snRNP cryo-EM model, where the lower region of Stem I is unwound to accommodate the formation of the BSL (Zhang et al. 2020). A24, which is in the linker between Stem I and the BSL is protected, which may reflect an interaction with SF3A3 (Figure II-2A, lane 4 compared to lane 3, Figure II-2C). SF3A3 is notable in that it nestles a short alpha-helix between Stem IIa and the base of the BSL, where it positions a tryptophan to stack on the last base pair of the BSL (Zhang et al. 2020). SF3A3 maintains that same position in A-complex where it stacks the same tryptophan at the end of the branch helix (Plaschka et al. 2018). Limited reactivity for nucleotides A26, C28, A38, C40, and C45 in the BSL stem, along with a reactive A35 in the BSL loop indicates its presence as well (Figure II-2A, lane 4 compared to lane 3, Figure II-2C). Nucleotides in Stem IIa follow the same reactivity pattern as in A-complex, including the unusual reactivity of G63. Stem IIb nucleotides are protected, except for A75 in its loop (Figure II-2B, lane 2 compared to lane 1, Figure II-2D). Finally, nucleotides in the 3’ half of Stem IIc are reactive as expected with the

presence of the mutually exclusive Stem IIa (Figure II-2B, lane 2 compared to lane 1, Figure II-2D).

For the most part, U2 snRNA in the snRNP incubated in high salt conditions exhibits a similar pattern of reactivity as the intact snRNP with two notable exceptions. First, nucleotides in Stem IIa show increased reactivity that indicate its destabilization, which was also noted previously (Behrens et al. 1993). Unexpectedly there is no corresponding decrease in nucleotide reactivity that supports a switch to Stem IIc (Figure II-2B, lane 3 compared to lane 6, Figure II-2D). In contrast, U2 snRNA probed in the absence of all proteins shows strong protections that are consistent with the switch to Stem IIc and correlating increased reactivity of nucleotides in the mutually exclusive Stem IIa (Figure II-2B, lane 4 compared to lane 5, Figure II-2D). This result suggests that SF3B and SF3A components are needed to stabilize Stem IIa and that either the core snRNP or a factor in nuclear extract prevents Stem IIc formation. Such a factor could serve to promote the rejoining of SF3B and SF3A after U2 snRNP is released from the spliceosome.

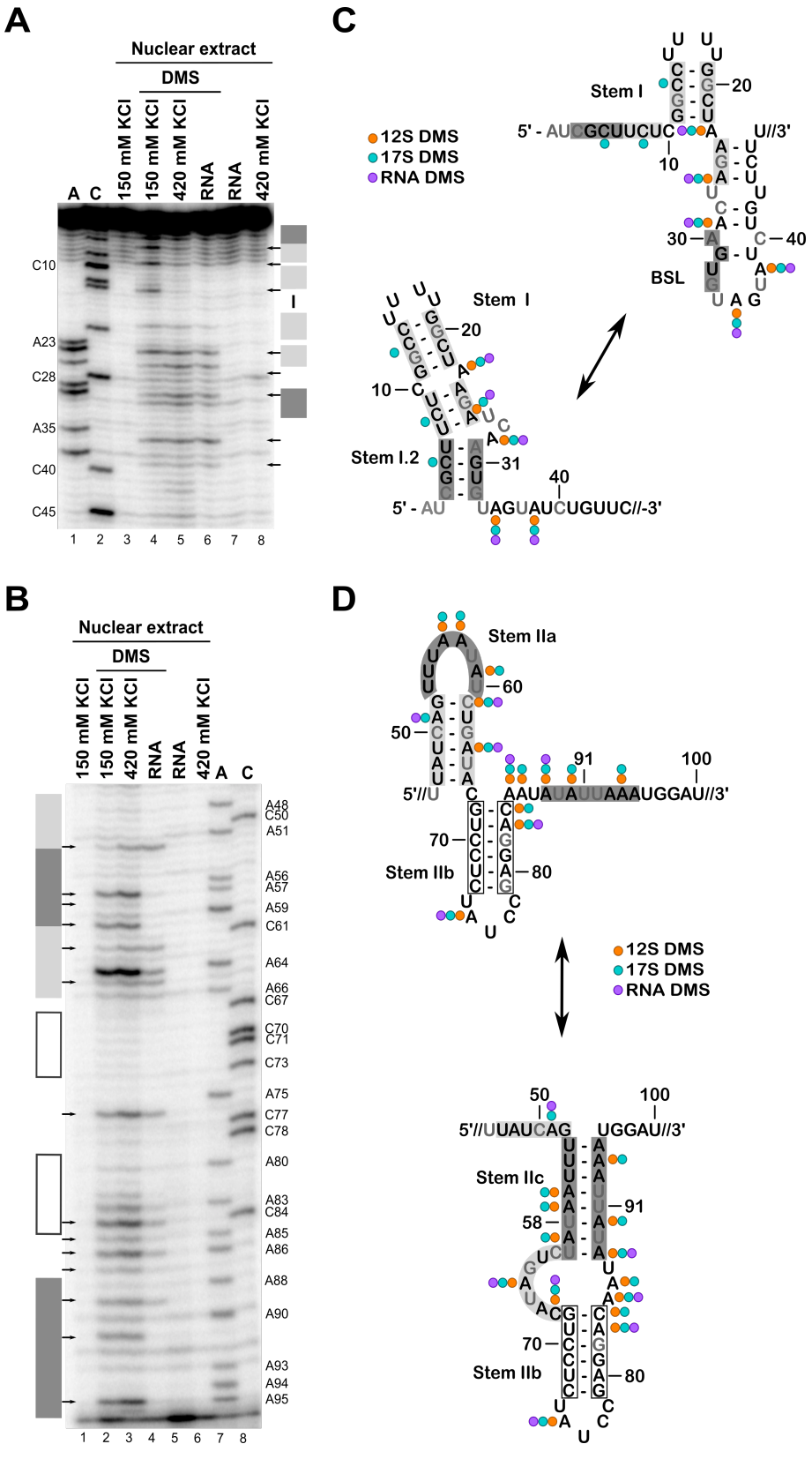

A
Nuclear extract
DMS
150 mM KCl
150 mM KCl
420 mM KCl
RNA
RNA
420 mM KCl
A C
C10
A23
C28
A35
C40
C45
I
1 2 3 4 5 6 7 8
B
Nuclear extract
DMS
150 mM KCl
150 mM KCl
420 mM KCl
RNA
RNA
420 mM KCl
A C
A48
C50
A51
A56
A57
A59
C61
A64
A66
C67
C70
C71
C73
A75
C77
C78
A80
A83
C84
A85
A86
A88
A90
A93
A94
A95
1 2 3 4 5 6 7 8
C
12S DMS
17S DMS
RNA DMS
Stem I
5' - AU GCUUCUC A U//3'
10
20
30
40
BSL
Stem I
Stem I.2
5' - AU UAGUAUCUGUUC//-3'
31
D
Stem IIa
60
50
91
100
5'//U C AAUAUAUUAAAUGGAU//3'
70
Stem IIb
80
12S DMS
17S DMS
RNA DMS
50
100
5'//UUAUCAG UGGAU//3'
Stem IIc
58
91
70
Stem IIb
80

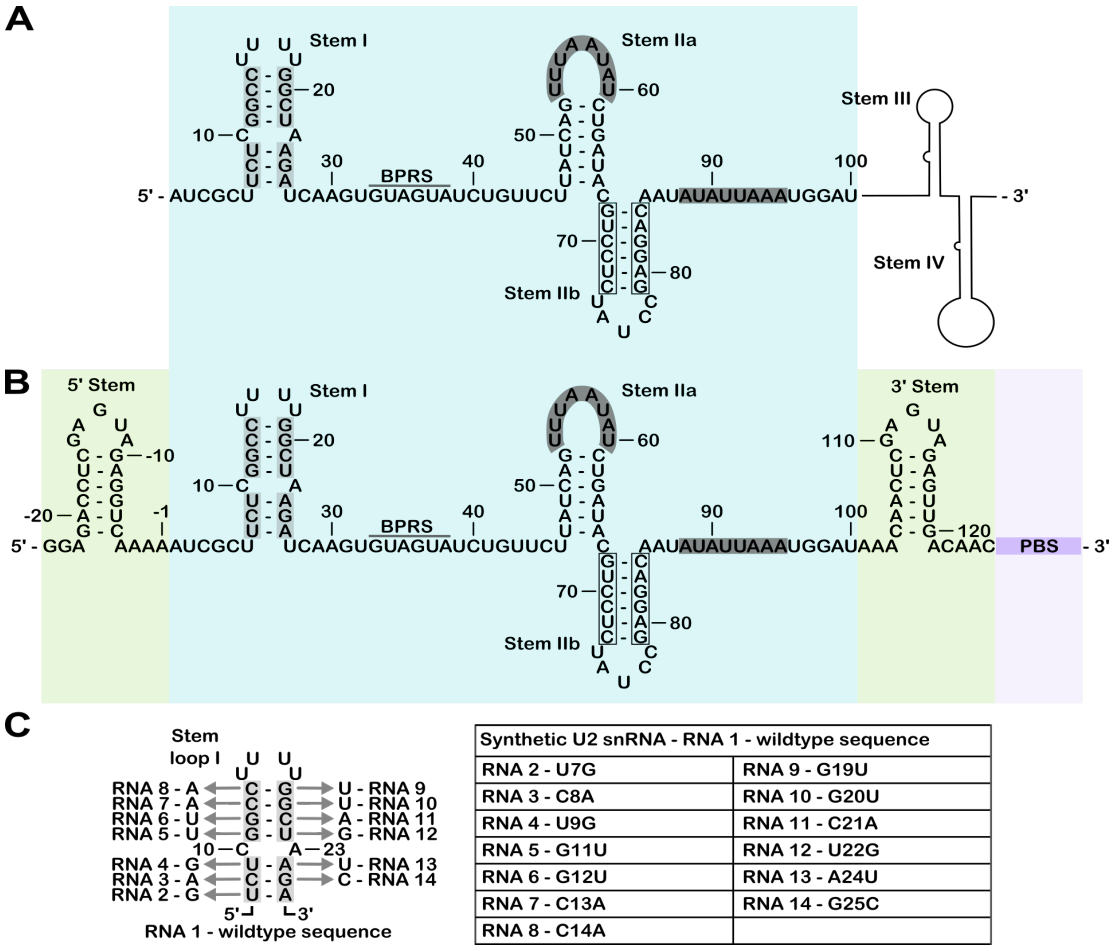
Figure II-2. **Accessibility of the U2 snRNA in U2 snRNP and RNA.** (A) Chemical modification of U2 snRNP and U2 snRNP exposed to high salt and RNA with DMS. Modifications of the RNA were detected by primer extension analysis with an oligonucleotide complementary to U2 snRNA nucleotides 97-117. Sequencing lanes are indicated with the corresponding nucleotides, and nucleotide positions are indicated to the right/left of each gel. The structural elements of the U2 snRNA are indicated by shaded boxes at the left/right. Arrows point to bands observed consistently across experiments done in triplicate. Nucleotides 1-46 of the U2 snRNA. Lanes 1 & 2 are ladders. Lanes 3, 7 & 8 are unmodified controls. Lane 4 is 17S U2 snRNP, lane 5 is 12S U2 snRNP and lane 6 is RNA treated with DMS. (B) Same as A, nucleotides 47-95 of the U2 snRNA. Lanes 7 & 8 are ladders. Lanes 1, 5 & 6 are unmodified controls. Lane 2 is 17S U2 snRNP, lane 3 is 12S U2 snRNP and lane 4 is RNA treated with DMS. (C) The accessibility of the U2 snRNA in the three conditions were mapped to two competing structural models of the U2 snRNA. The U2 snRNA is shown in black text and the nucleotides are numbered. Nucleotides shown in grey, indicate background bands and an inability to map reactivity to the nucleotide. The different secondary structure elements are labeled. Regions that base pair in a stem are indicated by shaded boxes. Arrows point to bands observed consistently across experiments done in triplicate. The circles indicated reactivity at that nucleotide, green is 17S U2 snRNP, orange is 12S U2 snRNP and purple is RNA. Nucleotides 1-46 of the U2 snRNA. (D) Same as C, nucleotides 46-100 of the U2 snRNA.

The second difference between U2 snRNP in the lower and higher salt conditions is in the 5' end. Nucleotides C5, C8, C10, and C14 become protected when SF3A and SF3B proteins are destabilized with higher salt (Figure II-2A, lane 5 compared to lane 8, Figure II-2C). C8, C10 and C14 protections were also noted in the early study by Behrens et al. (Behrens et al. 1993). In addition, I see a similar increase in protection with protein-free U2 snRNA (Figure II-2A, lane 6 compared to lane 7, Figure II-2C). This pattern would be consistent with an extended Stem I, except that structure is mutually exclusive with the BSL, and there is not a correlating increase in reactivity in BSL nucleotides A38, C40, and C45 (Figure II-2B, lane 5 compared to lane 8, Figure II-2C). How can protections between three competing strands be explained? The expectation is that one strand would have reactive

nucleotides because the other two strands are paired. Importantly, the same protection pattern is also observed in protein-free U2 snRNA, ruling out the possibility that a protein is responsible (Figure II-2B, lane 6 compared to lane 7, Figure II-2C). Furthermore, it suggests that the interactions responsible for the protections are intrinsic to the RNA and might not involve canonical base pairing.

3.3. Chemical mapping of synthetic U2 snRNA predicts an extended Stem I

To further investigate the unexpected patterns of protection that we observe in U2 snRNA in the absence of SF3B and SF3A proteins, I turned to a synthetic U2 snRNA. I generated templates to transcribe the first 100 nucleotides of U2 snRNA flanked by two stable stem loops to use as probing standards and a 3' tail for primer extensions. Notably, the synthetic RNA lacks all of the endogenous modifications found in U2 snRNA isolated from nuclear extract. I also generated a series of U2 snRNAs with mutations in the Stem I region to examine the importance of individual base pairs on the structure. I treated the synthetic U2 snRNAs with 1m7, DMS, and CMCT (1-cyclohexyl-(2-morpholinoethyl) carbodiimide metho-p-toluene sulfonate) and mapped nucleotide reactivity by primer extension followed by capillary electrophoresis, and reactivities were quantified using HiTrace (Yoon et al. 2011). Different modification types were separately used as restraints for structure prediction with the RNAstructure Fold algorithm (Kladwang et al. 2011; Tian et al. 2014).



Synthetic U2 snRNA - RNA 1 - wildtype sequence	
RNA 2 - U7G	RNA 9 - G19U
RNA 3 - C8A	RNA 10 - G20U
RNA 4 - U9G	RNA 11 - C21A
RNA 5 - G11U	RNA 12 - U22G
RNA 6 - G12U	RNA 13 - A24U
RNA 7 - C13A	RNA 14 - G25C
RNA 8 - C14A	

Figure II-3. **Synthetic U2 snRNA and mutations to disrupt Stem I.** (A) The model of the U2 snRNA and the corresponding position in the synthetic construct. The U2 snRNA is shown in black text and the nucleotides are numbered. The key features and different secondary structure elements are labeled. Regions that base pair in a stem are indicated by shaded boxes. (B) The synthetic construct contains, two additional stems (5' Stem and 3' Stem) as well as a primer binding sequence (PBS) for primer extension. (C) The mutations are diagrammed on the secondary structure model and listed in the table.

With 1m7, nucleotides in the loop of Stem I, Stem IIb and the 3' half of Stem IIa (nt 61-67) of the synthetic U2 snRNA construct are highly reactive like those in the loop of the reference stems (Figure II-4A). Nucleotides in the loop (U34-A35) of

the BSL also show higher reactivity (Figure II-4A). Nucleotides in the stem of the BSL generally show low reactivity, although more than the highly protected nucleotides in Stem IIb, Stem IIc, and the reference stems (Figure II-4A). Notably, nucleotides at the 5' end of the U2 sequence (C6-C10) and the 5' half of Stem IIa are also protected (Figure II-4A). Secondary structures predicted based on the 1m7 reactivity uniformly contained Stem I, Stem IIb, Stem IIc, and the two reference stems (Figure II-4C). Nucleotides in the 5' half of Stem IIa were predicted to base pair with a synthetic buffer sequence situated between the end of the U2 snRNA and the last reference stem (Figure II-4C). The software also predicted four continuous base pairs between C3-U6 and A30-G33, which extends Stem I (Figure II-4C). This proposed extension of Stem I is mutually exclusive with the BSL, and while it explains the protections observed in the 5' end of U2 snRNA, the low reactivity of nucleotides in the 3' half of the BSL (nt 39-44) remains unaccounted for (Figure II-4A). One issue with the structure prediction algorithm I used is the assumption of a single RNA conformation. I tried a second algorithm that considers the probability of multiple structures (Lorenz et al. 2011), but the top-ranking predictions all strongly resembled the previous prediction. Notably, the reactivity pattern for nucleotides 1-42 strongly resembles my observations for U2 snRNP in the absence of SF3A/SF3B complexes and protein-free U2 snRNA isolated from nuclear extracts.

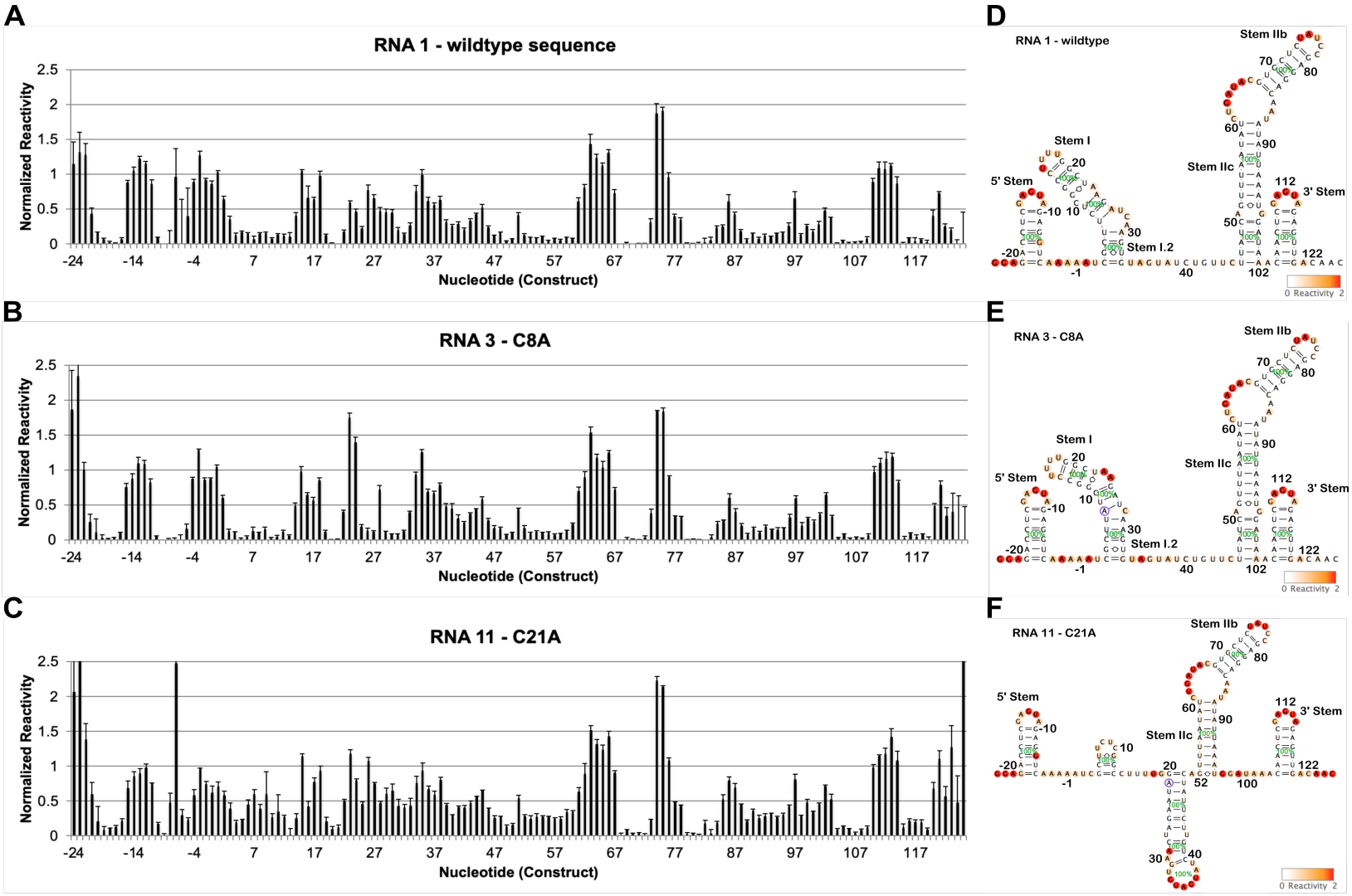

A
RNA 1 - wildtype sequence
Normalized Reactivity
Nucleotide (Construct)
B
RNA 3 - C8A
Normalized Reactivity
Nucleotide (Construct)
C
RNA 11 - C21A
Normalized Reactivity
Nucleotide (Construct)
D
RNA 1 - wildtype
Stem IIb
Stem I
5' Stem
Stem IIc
3' Stem
Stem I.2
0 Reactivity 2
E
RNA 3 - C8A
Stem IIb
Stem I
5' Stem
Stem IIc
3' Stem
Stem I.2
0 Reactivity 2
F
RNA 11 - C21A
Stem IIb
5' Stem
Stem IIc
3' Stem
0 Reactivity 2

Figure II-4. **Chemical mapping of synthetic U2 snRNA predicts an extended Stem I.** The synthetic RNAs were modified with 1m7. Modification of the RNAs were detected by primer extension analysis with an oligonucleotide complementary to the primer binding sequence of the RNA. (A) RNA 1 (wildtype), (B) RNA 3 (C8A), (C) RNA 11 (C21A). Within each panel, (left) the normalized reactivities are plotted across the nucleotides, each bar plotted represents experiments conducted in triplicate or greater with the respective standard deviation as error bars. (right) Color-coded schematic of the normalized reactivity is shown on the RNAstructure predicted secondary structure model. The U2 snRNA wildtype or mutant is shown in black text and the nucleotides are numbered. Regions that are modeled to base pair are labeled with a green percentage, which indicated the probability of each structural element. The mutation of each RNA construct is indicated by a purple circle.

When I examined the effects of single nucleotide mutations for base-paired nucleotides in Stem I, I did not always see the expected increase in reactivity (Figure II-4A, II-4B, II-4C, Supplementary Figure II-2, Supplementary Figure II-3, Supplementary Figure II-4, & Supplementary Figure II-5). Predicted structures based on the sequence and reactivity data often showed a shift in base pairing partners, but not a disruption of the stem. For example, mutating C8 in the lower half of Stem I to an A did not increase its reactivity, and results in a predicted partner swap of G25 for U27, along with the base-pairing shifts for neighbors U9 and G10 base pairs (Figure II-4B & II-4E). The most dramatic change in reactivity patterns occurred when C21 in the upper region of Stem I was changed to A, which results in the loss of the stem in the correlating predicted structures along with a slightly modified BSL structure that includes a four-nucleotide stem extension (Figure II-4C & II-4F). Combining the chemical mapping data and structure prediction models, I can conclude for almost all of the mutants, they likely include Stem I, Stem IIa, and Stem IIc (Figure II-4D, II-4E, II-4F, Supplementary Figure II-2, Supplementary Figure II-3, Supplementary Figure II-4, & Supplementary Figure II-5).

4. Discussion

The unchanged chemical probing reactivities and protections within the different regions of the U2 snRNA show that the SF3B inhibitors have little impact on U2 snRNA structure in both the spliceosome and in the snRNP. Both functional A-complex and ISC spliceosomes have reactivity patterns that indicate the presence of the upper region of Stem I, the branch helix and Stem IIa and IIb. This means that the BSL has unwound and the branch point recognition sequence engaged the intron in the branch helix before the SF3B inhibitor impacted assembly. In the A-complex structure, the branch helix is held in place by SF3B1 in a closed conformation, in which the branch point adenosine is sequestered in a pocket formed between two SF3B1 HEAT repeats. SF3B inhibitors have been shown to occupy the same pocket and block SF3B1 closing. The presence of the branch helix in the ISC suggests SF3B closing is not required for its formation. Instead, SF3B closing could signal that a proper branch helix is present to promote the next stage of spliceosome assembly. It would be interesting to chemically probe the ISC assembled on pre-mRNAs with branch point sequences that do not perfectly match the consensus sequence UACUAAC to see if strong base pairing alone stabilizes the branch helix in ISC. It would also be good to determine whether the same intron nucleotides are involved in base pairing interactions in the two complexes. I also found that the top of Stem I, a region not observed in the cryo-EM structure of A-complex, remains intact in both complexes. This result means that the branch helix formation and SF3B1 closure precedes Stem I unwinding. Together, my results indicate that SF3B inhibitors do not affect BSL unwinding or branch helix formation but that SF3B1 closure may be required for Stem I to be opened to engage U6 snRNA.

In the full U2 snRNP with SF3A and SF3B, the pattern of DMS modification of the U2 snRNA is consistent with the cryo-EM structure of the 17S U2 snRNP. When reactivity is mapped to competing secondary structure models, the pattern of reactivity for the 17S U2 snRNP is more consistent with the structure containing the top of Stem I, the BSL, Stem IIa and Stem IIc (Figure II-2C & II-2D). Loss of SF3A and SF3B complexes in high salt conditions to yield a 12S U2 snRNP results in the loss of Stem IIa, but surprisingly not the formation of Stem IIc. Because Stem IIc is present in the RNA alone, I conclude that either the remaining core U2 snRNP proteins or another factor in nuclear extract may interfere with Stem IIc formation. Such a factor could function in recycling of the U2 snRNP after spliceosome disassembly.

In both the 12S U2 snRNP and free RNA, there are protections consistent with the BSL and the top of stem I, but also new protections at the very 5' end are not easy to fit into current secondary structure models (especially C5). These protections also extend to synthetic U2 snRNA constructs probed with 1m7. When used as restraints for RNA structure prediction, the RNAStructure algorithm favors an extended Stem I structure and leaves the slightly less protected nucleotides in the 3' half of the BSL as single stranded. The algorithm likely models the extended Stem I due to the proximity and potential to form canonical base pairing interactions (AU, GC, and GU) (Reuter and Mathews 2010). Notably, the potential to form an extended Stem I is conserved in U2 snRNA across nearly all species. In the future, it will be important to probe U2 snRNA from sequences with divergent 5' ends to see if the protections are conserved, although it will be challenging to rule out the possibility that the conservation is enforced by the requirement to base pair with U6 snRNA in

the spliceosome. Comparably, the BSL was also overlooked for many years because similar conservation was attributed to U6 snRNA base pairing needs.

If U2 snRNA takes on an extended Stem I, the expectation is that nucleotides in the competing structure would be unpaired and reactive to chemical probes. For the extended Stem I, the competing structure is the BSL, which the RNAStructure program models as single stranded. However, the relatively low reactivity of nucleotides in the BSL region indicates that the story is not as simple. Importantly, the structure of U2 snRNA's 5' end in the absence of SF3B and SF3A proteins has the potential to impact both the biogenesis of and recycling of U2 snRNP, where interactions with SF3A and SF3B must be established. Additional studies, such as determining a cryo-EM structure of the 12S U2 snRNP will be important for seeing what the nature of this structure really is.

The flexibility of the U2 snRNA throughout the spliceosome cycle is a marvel. The RNA swaps intramolecular and intermolecular base pairing partners in a dynamic exchange. My chemical probing studies both fill in gaps in the cryo-EM models of spliceosome complexes and serve as a starting point for exploring even more possible conformations of the remarkable U2 snRNA.

5. Methods

5.1. HeLa nuclear extract

Nuclear extract was prepared from HeLa cells grown in DMEM/F-12 1:1 and 5% (v/v) newborn calf serum as previously described (Dignam et al. 1983) with the

exception that HEPES replaced Tris in all buffers. For high salt conditions, nuclear extract was dialyzed for four hours at 4˚C in a 10 kD cut-off Slide-A-Lyzer™ MINI Dialysis Device (Thermo Scientific™) in buffer containing 20 mM HEPES pH 7.9, 420 mM potassium chloride, 0.2 mM EDTA, 20% glycerol, 0.5 mM dithiothreitol. Nuclear extract for A-complex assembly was depleted of U4 and U6 snRNAs using oligonucleotides complementary to the specified snRNAs and endogenous RNase H, as previously described in (Effenberger et al. 2016b).

5.2. Spliceosome complex purification

Spliceosome complexes were purified as previously described in Ilagan et al. 2015 with the exceptions described below. A-complex and SF3B-inhibited spliceosome were assembled *in vitro* for 15 minutes and purified by size exclusion chromatography followed by amylose affinity selection. Purified splicing complexes were eluted in 150 mM potassium chloride, 5 mM EDTA, 20 mM HEPES pH 7.9, 1 mM dithiothreitol, 10 mM maltose.

5.3. Chemical probing by DMS

U2 snRNP in nuclear extract prepared at 150 mM potassium chloride or 420 mM potassium chloride supplemented with additional HEPES (80 mM), magnesium acetate (2 mM) and was chemically modified with dimethyl sulfate (DMS, Sigma Aldrich) (0.7% v/v) for 5 minutes at room temperature. Reactions were quenched with 37.5% betamercaptoethanol (BME), 0.375 M sodium acetate.

Purified spliceosomes in 90 mM potassium chloride, 3 mM EDTA, 33 mM HEPES pH 7.9, 8mM magnesium acetate, .6 mM dithiothreitol, 6 mM maltose were

chemically modified with DMS (0.33%) at room temp for 10 minutes. Reactions were quenched with 30% BME, 0.3M sodium acetate.

5.4. Chemical probing by SHAPE

Purified spliceosomes in 90 mM potassium chloride, 3 mM EDTA, 33 mM HEPES pH 7.9, 8mM magnesium acetate, .6 mM dithiothreitol, 6 mM maltose were chemically modified with 10 mM 1-Methyl-7-nitroisatoic anhydride (1m7, a gift from M. Ares) at room temperature for 10 minutes and then supplemented to 0.3 M sodium acetate.

5.5. Primer extension by reverse transcription

In preparation for reverse transcription, 10 picomoles of DNA primer (complementary to nucleotides 97-119 of the U2 snRNA sequence) was labeled with γ-^{32}P ATP and purified by sephadex G-25 (Sigma Aldrich) column. Following chemical modification, RNA was purified by phenol:chloroform:iso-amyl alcohol (25:24:1) extraction, chloroform:iso-amyl alcohol (24:1) extraction and ethanol precipitation and resuspended in distilled H2O. The RNA and labeled primer were annealed by incubation at 95˚C for 2 min, 53˚C for 5 min, and on ice for 5 min and then added to reverse transcription reaction. Final reaction concentrations were 50 mM Tris-chloride pH 8.0, 75 mM potassium chloride, 7 mM dithiothreitol, 1 mM dNTPs, 3 mM magnesium chloride and 80U of SSIII (Invitrogen) at 53˚C for 30 minutes. DNA was isolated by the addition of 0.3 M NaAc pH 5.2, 0.5 mM EDTA, 0.05% SDS followed by ethanol precipitation. Sequencing ladders were generated using total RNA isolated from nuclear extract with the addition of a single ddNTP.

DNA species were separated on a 9.6% (v/v) denaturing polyacrylamide gel that was dried onto Whatman paper and visualized by phosphorimaging.

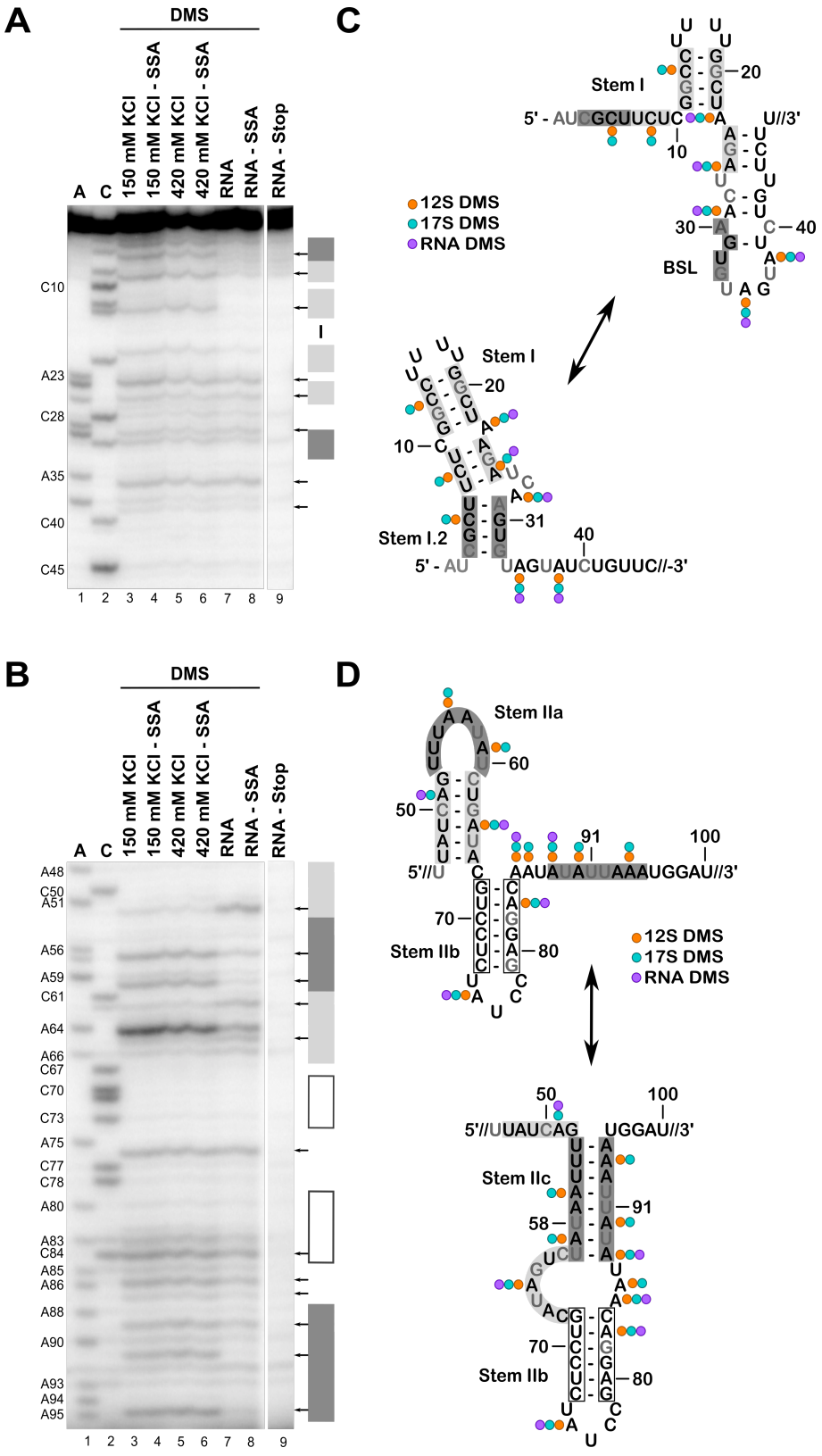
5.6. Synthetic U2 snRNA constructs

Synthetic U2 snRNA constructs were designed generated by PCR. The first 100 nucleotides for the human U2 snRNA sequence flanked by two hairpins. RNA was generated by T7 run-off transcription. Reactions were separated on a 5% polyacrylamide gel and bands identified by UV shadowing. The bands were excised, and RNA was extracted. Each RNA was diluted to a concentration of 6 pmol in 10 µL, heated to 95 ˚C for 5 minutes, and incubated at room temperature for 15 minutes to allow the RNA to fold/adopt a secondary structure.

5.7. Chemical probing and analysis of synthetic RNAs

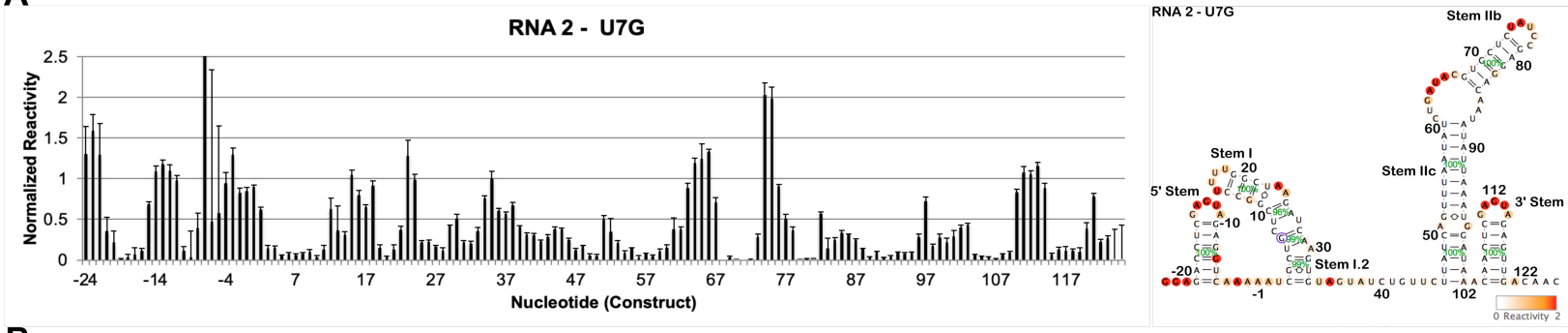
The chemical probing including DMS, 1m7 and CMCT (1-cyclohexyl-(2-morpholinoethyl) carbodiimide metho-p-toluene sulfonate - Sigma Aldrich) as well as analysis of data for synthetic RNA was carried out essentially as described previously in (Palka et al. 2020).

6. Supplemental Figures

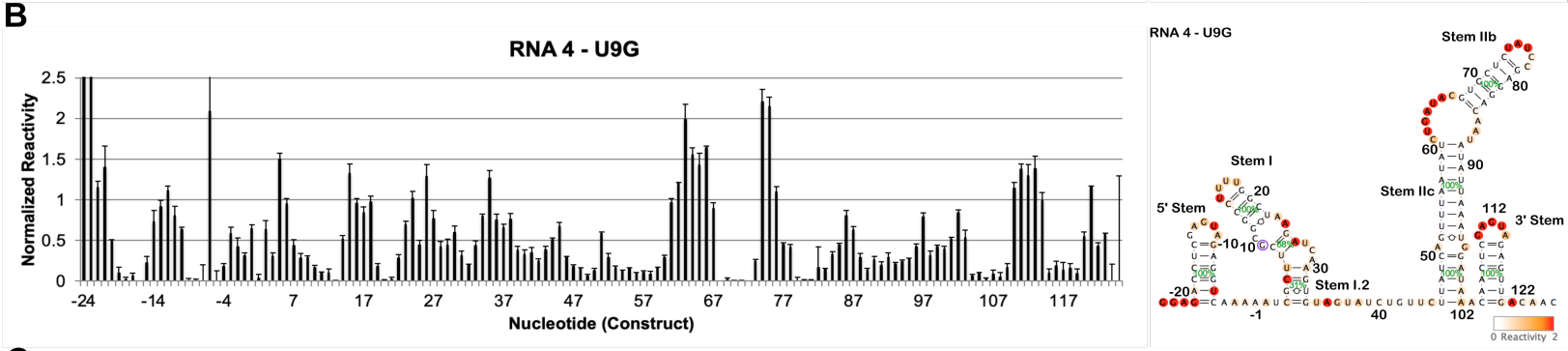

A
DMS
150 mM KCl
150 mM KCl - SSA
420 mM KCl
420 mM KCl - SSA
RNA
RNA - SSA
RNA - Stop
A C
C10
A23
C28
A35
C40
C45
I
1 2 3 4 5 6 7 8 9
B
DMS
150 mM KCl
150 mM KCl - SSA
420 mM KCl
420 mM KCl - SSA
RNA
RNA - SSA
RNA - Stop
A C
A48
C50
A51
A56
A59
C61
A64
A66
C67
C70
C73
A75
C77
C78
A80
A83
C84
A85
A86
A88
A90
A93
A94
A95
1 2 3 4 5 6 7 8 9
C
12S DMS
17S DMS
RNA DMS
Stem I
5' - AU GCUUCUC A U//3'
10
20
30
40
BSL
Stem I
Stem I.2
31
5' - AU UAGUAUCUGUUC//-3'
D
Stem IIa
60
50
91
100
5'//U C AAUAUAUUAAAUGGAU//3'
70
Stem IIb
80
12S DMS
17S DMS
RNA DMS
50
100
5'//UUAUCAG UGGAU//3'
Stem IIc
91
58
70
Stem IIb
80

Supplementary Figure II-1. **Accessibility of the U2 snRNA in U2 snRNP and RNA in the presence of SSA.** (A) Chemical modification of U2 snRNP and U2 snRNP exposed to high salt and RNA in the presence of SSA with DMS. Modification of the RNA were detected by primer extension analysis with an oligonucleotide complementary to U2 snRNA nucleotides 97-117. Sequencing lanes are indicated with the corresponding nucleotides, and nucleotide position are indicated to the right/left of each gel. The structural elements of the U2 snRNA are indicated by shaded boxes at the left/right. Arrows point to bands observed consistently across experiments done in triplicate. The circles indicated reactivity at that nucleotide, green is 17S U2 snRNP, orange is 12S U2 snRNP and purple is RNA. Arrows point to bands observed consistently across experiments done in triplicate. Nucleotides 1-46 of the U2 snRNA. Lanes 1 & 2 are ladders. Lanes 9 is an unmodified control. Lane 3 is 17S U2 snRNP, Lane 4 is 17S U2 snRNP with SSA, lane 5 is 12S U2 snRNP Lane 6 is 12S U2 snRNP with SSA, Lane 7 is RNA, and lane 8 is RNA with SSA treated with DMS. (B) Same as A, nucleotides 47-95 of the u2 snRNA. (C) The accessibility of the U2 snRNA in the three conditions were mapped to two competing structural models of the U2 snRNA. The U2 snRNA is shown in black text and the nucleotides are numbered. Nucleotides shown in grey, indicate background bands and an inability to map reactivity to the nucleotide. The different secondary structure elements are labeled. Regions that base pair in a stem are indicated by shaded boxes. Arrows point to bands observed consistently across experiments done in triplicate. The circles indicated reactivity at that nucleotide, green is 17S U2 snRNP, orange is 12S U2 snRNP and purple is RNA. Nucleotides 1-46 of the U2 snRNA. (D) Same as C, nucleotides 46-100 of the U2 snRNA.

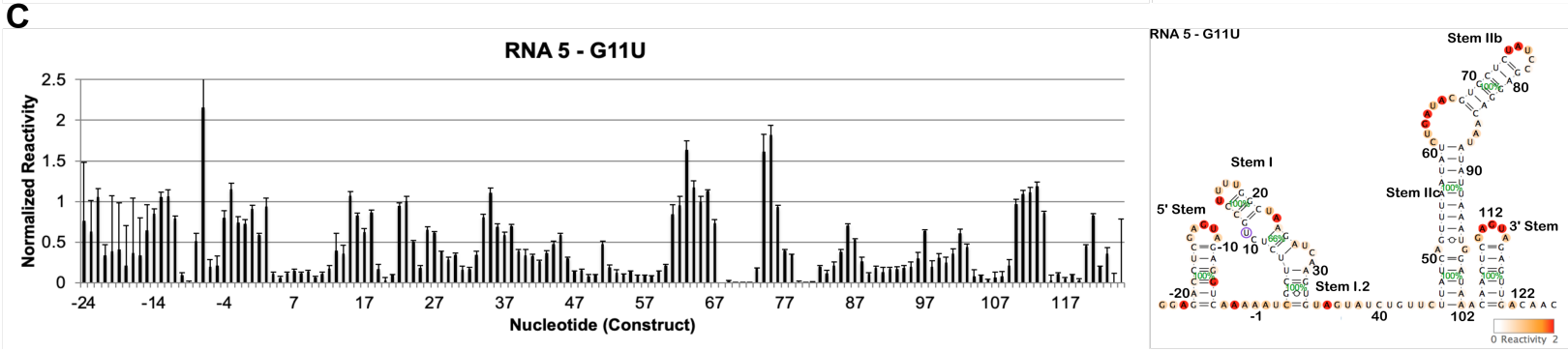
A


RNA 2 - U7G
Normalized Reactivity
Nucleotide (Construct)
RNA 2 - U7G
Stem IIb
Stem I
5' Stem
Stem IIc
3' Stem
Stem I.2
0 Reactivity 2

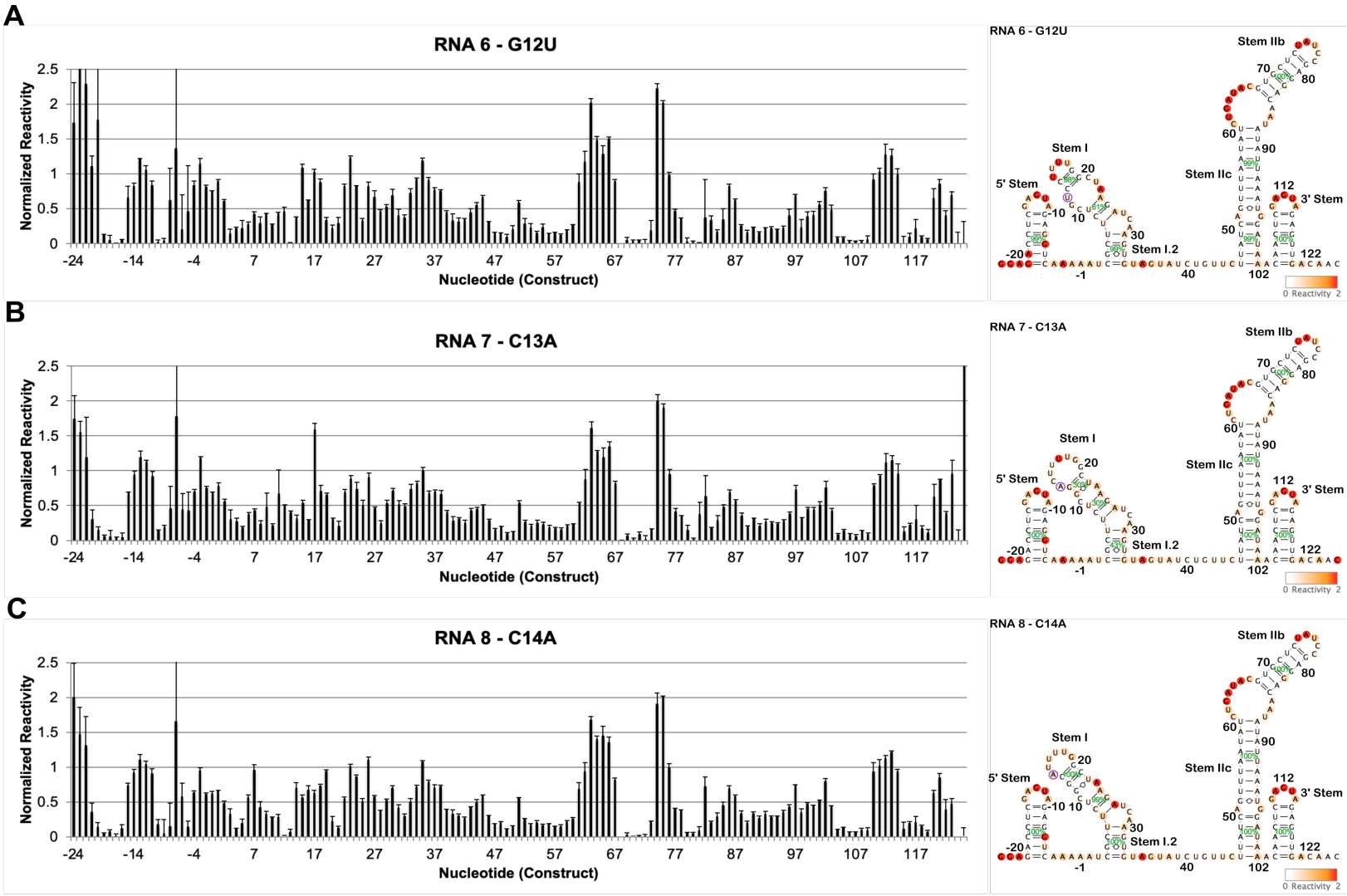
B


RNA 4 - U9G
Normalized Reactivity
Nucleotide (Construct)
RNA 4 - U9G
Stem IIb
Stem I
5' Stem
Stem IIc
3' Stem
Stem I.2
0 Reactivity 2

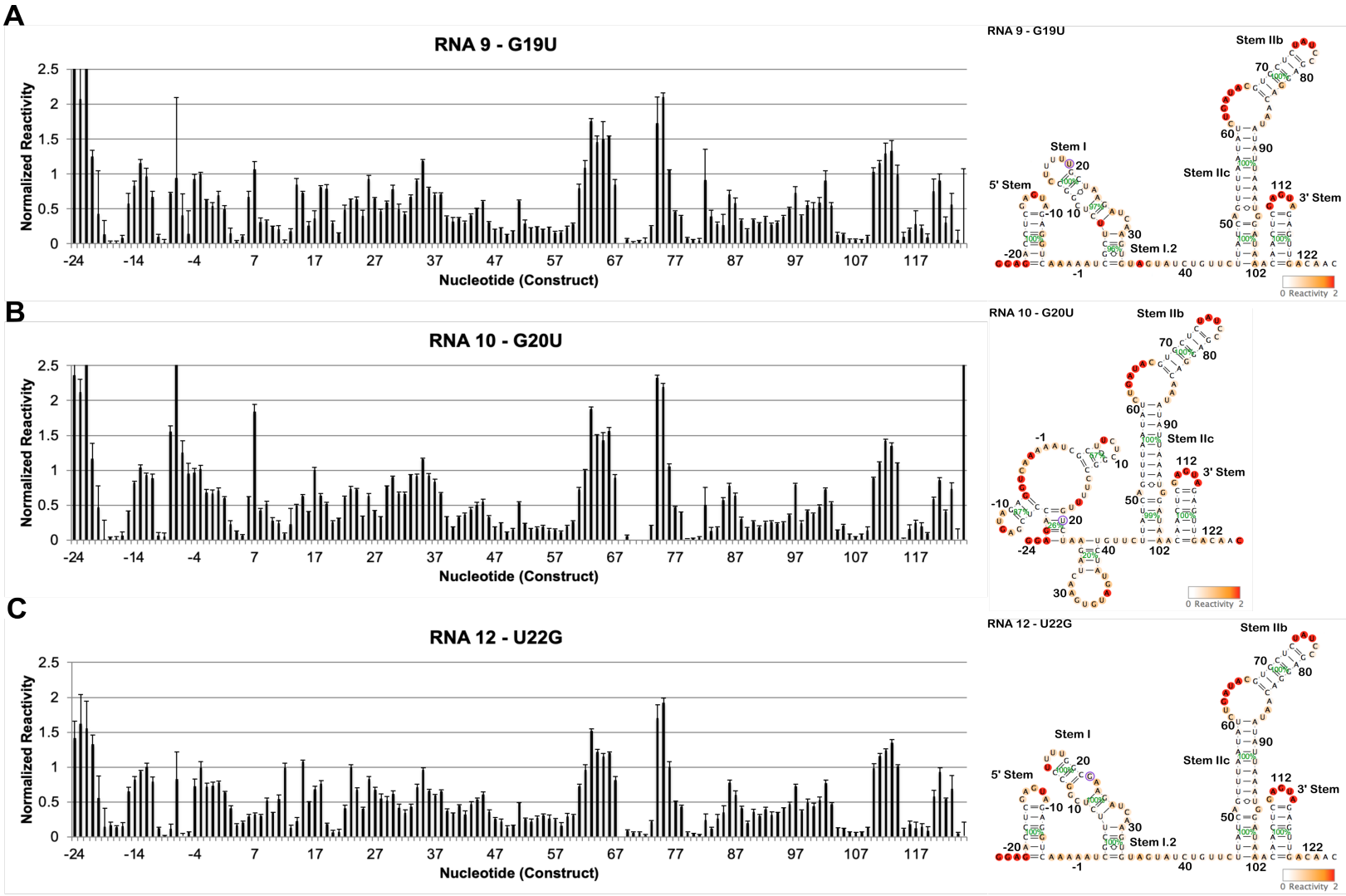
C


RNA 5 - G11U
Normalized Reactivity
Nucleotide (Construct)
RNA 5 - G11U
Stem IIb
Stem I
5' Stem
Stem IIc
3' Stem
Stem I.2
0 Reactivity 2

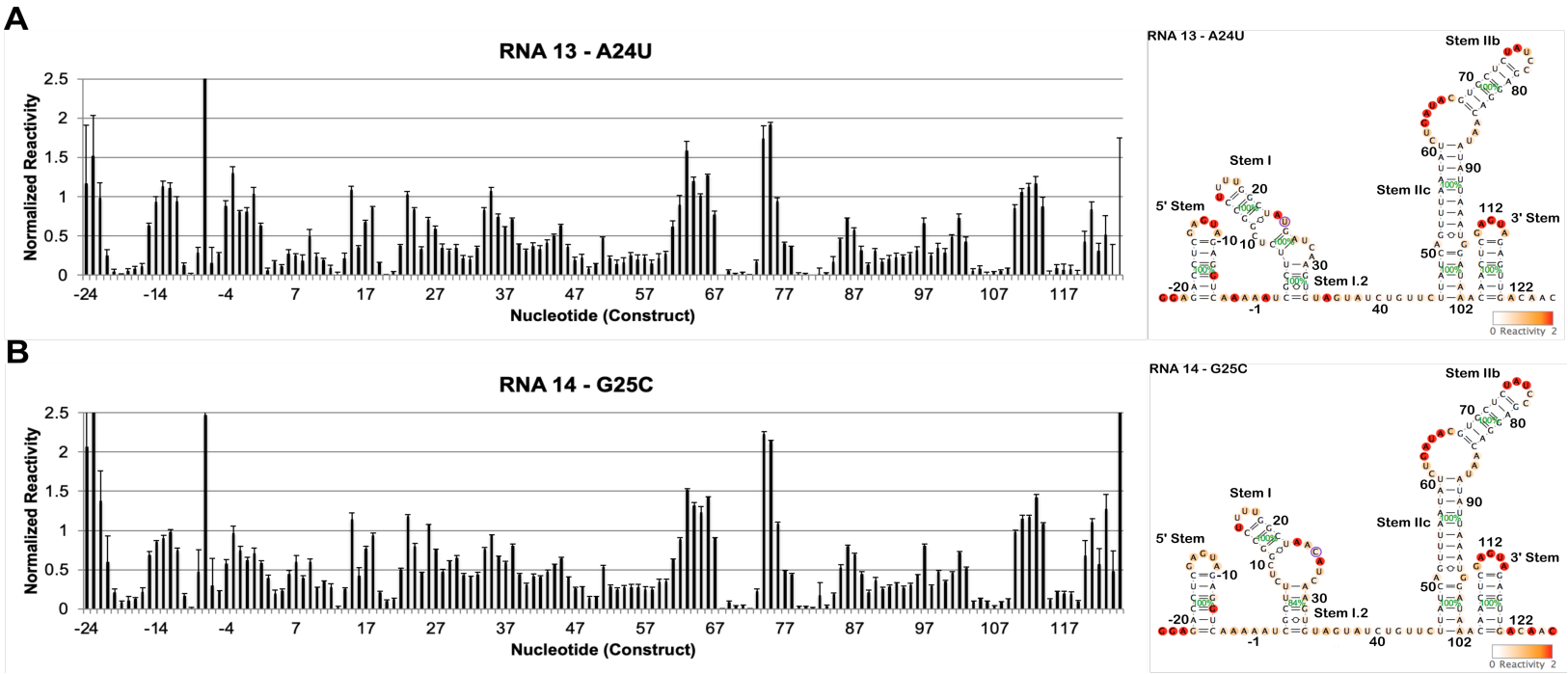
Supplementary Figure II-2. **Chemical mapping of synthetic U2 snRNA mutants, RNA 2, RNA 4, and RNA 5.** The synthetic RNAs were modified with 1m7. Modification of the RNAs were detected by primer extension analysis with an oligonucleotide complementary to the primer binding sequence of the RNA. (A) RNA 2 (U7G), (B) RNA 4 (U9G), (C) RNA 5 (G11U). Within each panel, (left) the normalized reactivities are plotted across the nucleotides, each bar plotted represents experiments conducted in triplicate or greater with the respective standard deviation as error bars. (right) A color-coded schematic of the normalized reactivity is shown on the RNAstructure predicted secondary structure model. The U2 snRNA wildtype or mutant is shown in black text and the nucleotides are numbered. Regions that are modeled to base pair are labeled with a green percentage, which indicated the probability of each structural element. The mutation of each RNA construct is indicated by a purple circle.


A
RNA 6 - G12U
Normalized Reactivity
Nucleotide (Construct)
RNA 6 - G12U
Stem IIb
Stem I
5' Stem
Stem IIc
3' Stem
Stem I.2
0 Reactivity 2
B
RNA 7 - C13A
Normalized Reactivity
Nucleotide (Construct)
RNA 7 - C13A
Stem IIb
Stem I
5' Stem
Stem IIc
3' Stem
Stem I.2
0 Reactivity 2
C
RNA 8 - C14A
Normalized Reactivity
Nucleotide (Construct)
RNA 8 - C14A
Stem IIb
Stem I
5' Stem
Stem IIc
3' Stem
Stem I.2
0 Reactivity 2

Supplementary Figure II-3. **Chemical mapping of synthetic U2 snRNA mutants, RNA 6, RNA 7, and RNA 8.** The synthetic RNAs were modified with 1m7. Modification of the RNAs were detected by primer extension analysis with an oligonucleotide complementary to the primer binding sequence of the RNA. (A) RNA 6 (G12U), (B) RNA 7 (C13A), (C) RNA 8 (C14A). Within each panel, (left) the normalized reactivities are plotted across the nucleotides, each bar plotted represents experiments conducted in triplicate or greater with the respective standard deviation as error bars. (right) A color-coded schematic of the normalized reactivity is shown on the RNAstructure predicted secondary structure model. The U2 snRNA wildtype or mutant is shown in black text and the nucleotides are numbered. Regions that are modeled to base pair are labeled with a green percentage, which indicated the probability of each structural element. The mutation of each RNA construct is indicated by a purple circle.


A
RNA 9 - G19U
Normalized Reactivity
Nucleotide (Construct)
RNA 9 - G19U
Stem IIb
Stem I
5' Stem
Stem IIc
3' Stem
Stem I.2
0 Reactivity 2
B
RNA 10 - G20U
Normalized Reactivity
Nucleotide (Construct)
RNA 10 - G20U
Stem IIb
Stem IIc
3' Stem
0 Reactivity 2
C
RNA 12 - U22G
Normalized Reactivity
Nucleotide (Construct)
RNA 12 - U22G
Stem IIb
Stem I
5' Stem
Stem IIc
3' Stem
Stem I.2
0 Reactivity 2

Supplementary Figure II-4. **Chemical mapping of synthetic U2 snRNA mutants, RNA 9, RNA 10, and RNA 12.** The synthetic RNAs were modified with 1m7. Modification of the RNAs were detected by primer extension analysis with an oligonucleotide complementary to the primer binding sequence of the RNA. (A) RNA 9 (G19U), (B) RNA 10 (G20U), (C) RNA 12 (U22G). Within each panel, (left) the normalized reactivities are plotted across the nucleotides, each bar plotted represents experiments conducted in triplicate or greater with the respective standard deviation as error bars. (right) Color-coded schematic of the normalized reactivity is shown on the RNAstructure predicted secondary structure model. The U2 snRNA wildtype or mutant is shown in black text and the nucleotides are numbered. Regions that are modeled to base pair are labeled with a green percentage, which indicated the probability of each structural element. The mutation of each RNA construct is indicated by a purple circle.


A
RNA 13 - A24U
Normalized Reactivity
Nucleotide (Construct)
RNA 13 - A24U
Stem IIb
Stem I
5' Stem
Stem IIc
3' Stem
Stem I.2
0 Reactivity 2
B
RNA 14 - G25C
Normalized Reactivity
Nucleotide (Construct)
RNA 14 - G25C
Stem IIb
Stem I
5' Stem
Stem IIc
3' Stem
Stem I.2
0 Reactivity 2

Supplementary Figure II-5. **Chemical mapping of synthetic U2 snRNA mutants, RNA 13 and RNA 14.** The synthetic RNAs were modified with 1m7. Modification of the RNAs were detected by primer extension analysis with an oligonucleotide complementary to the primer binding sequence of the RNA. (A) RNA 13 (A24U), (B) RNA 14 (G25C). Within each panel, (left) the normalized reactivities are plotted across the nucleotides, each bar plotted represents experiments conducted in triplicate or greater with the respective standard deviation as error bars. (right) A color-coded schematic of the normalized reactivity is shown on the RNAstructure predicted secondary structure model. The U2 snRNA wildtype or mutant is shown in black text and the nucleotides are numbered. Regions that are modeled to base pair are labeled with a green percentage, which indicated the probability of each structural element. The mutation of each RNA construct is indicated by a purple circle.

Chapter III: SF3B1 inhibitors block exon ligation

1. Introduction

The human spliceosome is responsible for removing intervening sequences from pre-mRNA transcripts. The spliceosome assembles in a stepwise manner, composed of five snRNP sub-complexes. There are multiple rearrangements that must occur as components join and leave the macromolecular complex, interestingly the U2 snRNP joins the spliceosome early in assembly and remains with the spliceosome through both steps of splicing chemistry. The U2 snRNP is responsible for branch point recognition and the U2 snRNA interacts with the branch point sequence of the intron to form the branch helix, which exposes the branch point adenosine. Based on the selection of the branch point, this directs the selection of the correct 3' splice site. During spliceosome activation, Prp2 removes SF3B which coincides with the branch point adenosine being docked into the active site of the spliceosome (Will and Luhrmann 2011; Yan et al. 2019).

The U2 snRNP is composed of a single snRNA, core proteins, SF3A and SF3B subcomplexes. The SF3B subunit is the target of multiple potent splicing inhibitors including Pladienolide B and Spliceostatin A (Kaida et al. 2007; Kotake et al. 2007; Hasegawa et al. 2011). The addition of the inhibitors to an *in vitro* nuclear extract splicing system inhibits splicing at A-complex formation (Roybal and Jurica 2010; Effenberger et al. 2014). Spliceosomes can be assembled in nuclear extract depleted of tri-snRNP components which are necessary to continue past A-complex, accumulating A-complex spliceosomes. These spliceosomes can then be supplemented with additional nuclear extract depleted of U2 snRNP and continue

through spliceosome assembly. However, if splicing inhibitor is added with the U2 snRNP depleted nuclear extract spliceosomes do not progress through assembly steps indicating that the transition from A-complex to B complex is also affected (Effenberger et al. 2016b).

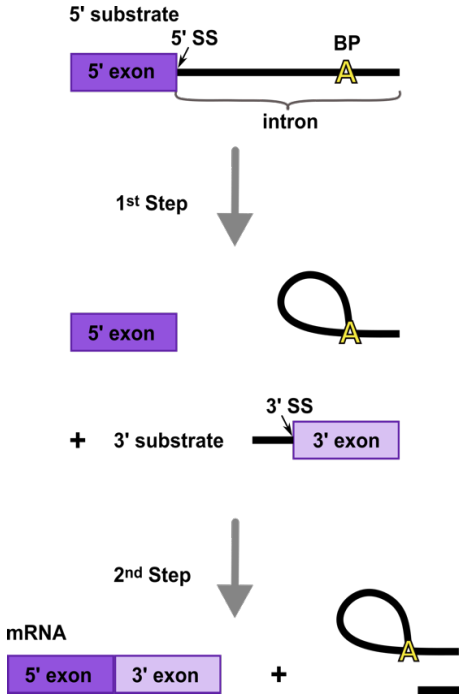


FIGURE III-1**. Bimolecular exon ligation reaction** (A) Schematic of the bimolecular exon ligation reaction. The 5' splice site (5' SS), branchpoint adenosine (A), and 3' splice site (3'SS) are labeled. Two transesterification reactions, each represented by a gray arrow. The 5′ substrate consists of a 5′ exon and intron containing the branch point and PYT and is capable of 1st step splicing chemistry to produce splicing intermediates. The 3′ substrate consists of a 3′ splice site and 3′ exon and is capable of 2nd step chemistry to produce mRNA when added in trans. To test for inhibition, SF3B inhibitor is added with the 3′ substrate.

Prior to the spliceosome structures, mass spec data suggested that SF3B remained with the spliceosome through P complex (Coltri et al. 2011; Ilagan et al. 2013). I asked the question, why would SF3B stay associated with the spliceosome once its task is complete? Could SF3B have a further function in addition to its early role? Is SF3B required for catalysis? To determine possible later roles for SF3B, I combined the splicing inhibitors with other tools to manipulate the *in vitro* splicing system to tease apart the interaction and role of SF3B after branch point recognition. Because the addition of SF3B1 inhibitors to an *in vitro* splicing system at the beginning of the reaction will result in an accumulation of early A-like spliceosome complexes, I needed a strategy to investigate the later steps in spliceosome assembly. The spliceosome moves through multiple assembled complexes to carry out two chemical steps of splicing. In the first step, the branch point adenosine carries out a nucleophilic attack on the 5'SS, creating the free 5' exon and a lariat intron-3'exon intermediate. The second step is exon ligation, the 5' and 3' exon are joined to form an mRNA. I used a biomolecular exon ligation assay to bypass the first point of inhibition, which is the A-complex block (Konforti and Konarska 1995; Anderson and Moore 1997). The bimolecular exon ligation assay temporally isolates the second chemical step of splicing from the first. The bimolecular exon ligation reaction separates the two steps and allows for manipulation of the *in vitro* splicing system between the two steps. This strategy allowed me to add SF3B1 inhibitors after the first step of splicing chemistry had already occurred to investigate if the inhibitors act at the transition between first and second step of splicing chemistry.

2. Results

Given that SF3B1 is present throughout the spliceosome cycle (Agafonov et al. 2011; Ilagan et al. 2013), and the ability of the inhibitors to block the transition out of A-complex in addition to the first block at SF3B's known role in intron recognition (Brosi et al. 1993; Roybal and Jurica 2010; Corrionero et al. 2011; Folco et al. 2011; Effenberger et al. 2014). I hypothesized that in addition to its likely roles in transitioning into and out of A-complex, it also functions at later stages of spliceosome assembly, specifically between the two steps of splicing chemistry. To test this hypothesis, I used the bimolecular exon ligation reaction described above (Figure III-1; (Konforti and Konarska 1995; Anderson and Moore 1997), also see method section). If SF3B1 also has a role in the spliceosome after 1st step chemistry, then exon ligation will be affected by SF3B1 inhibitors. Because bimolecular exon ligation is an inefficient assay, I used a combination of labeled and unlabeled substrates to get a detectable signal. To bypass the early spliceosome assembly blocks induced by the splicing inhibitor Pladienolide B, I incubated an unlabeled pre-mRNA 5′ substrate that contained a 5′ exon and an intron in the presence of nuclear extract. Spliceosomes assemble on this substrate and complete 1st step chemistry to produce a free 5′ exon and lariat intron intermediate that is held in a stalled complex (Konforti and Konarska 1995; Anderson and Moore 1997). I then added a labeled 3′ substrate consisting of a 3′ exon preceded by a 3′ splice site, either in the presence or absence of SF3B inhibitor. In the assay, with just DMSO control added, I detected a labeled mRNA band that is a result of exon ligation (Figure III-2, lane 3). However, the mRNA band nearly disappeared with increasing concentration of Pladienolide B (Figure III-2, lanes 4–9). I repeated the assay with Spliceostatin A, another splicing

inhibitor, with the same result (Figure III-3, lanes 4-9). The inhibition of exon ligation indicates that Pladienolide B and Spliceostatin A interact with their target, SF3B1, in the catalytic spliceosome. It is also consistent with a functional requirement for SF3B1 after 1st step chemistry. There are no data linking SF3B1 directly to splicing catalysis, so it is likely that the protein plays a structural role such as positioning RNA and/or protein, the conformation of which is modulated by the inhibitor. However, I cannot rule out the possibility that the binding of the inhibitors to SF3B1 interferes with the activity of another component of the spliceosome required for exon ligation and that SF3B1 simply is a spectator. I note that relative to the earliest point of inhibition, a higher concentration of the drug is required to block exon ligation. While I have not determined the basis for this difference, although one possibility is that the interaction site in SF3B1 for inhibitors may be partially occluded in the context of the catalytic spliceosome.

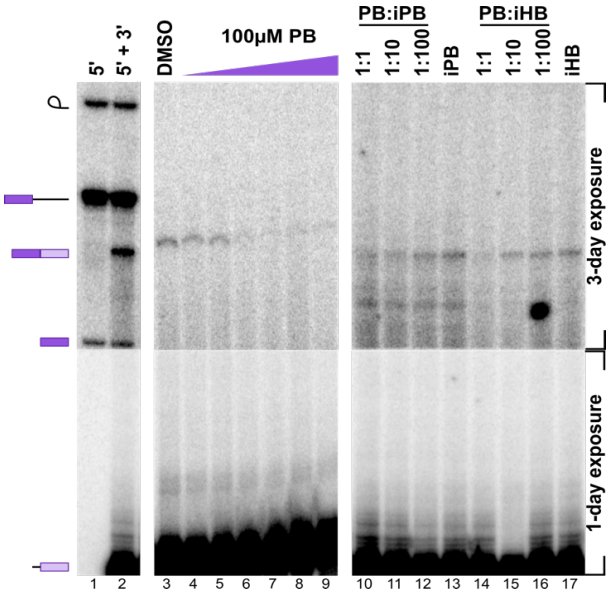


FIGURE III-2. **Bimolecular assay shows that Pladienolide B (PB) inhibits exon ligation.** Denaturing gel analysis of *in vitro* bimolecular exon ligation reactions.

Identities of bands are schematized to the left as (from top to bottom) lariat intermediate, 5′ substrate, mRNA, 5′ exon, and 3′ exon. A 3-d exposure is shown for the top region of the gel, while the bottom region is a 1-d exposure. In lanes 1 and 2, both 5′ and 3′ substrates are radiolabeled. In lanes 3–17, only the 3′ substrate is radiolabeled. In lanes 4–9, increasing concentrations of Pladienolide B (PB) (0.1, 1, 2, 5, 10, 100 μM) were included with the 3′ substrate. In lanes 10–12 and 14–16, active compound was added at 2 μM and increasing concentrations of inactive compounds were added at the indicated ratio. In lanes 13 and 17, inactive compound is at 200 μM.

Three SF3B inhibitors, Spliceostatin A, Pladienolide B, and Herboxidiene, can all inhibit spliceosome assembly and splicing chemistry. While their inactive counterparts do not affect splicing but can compete interchangeably with active inhibitors to restore splicing (Effenberger et al. 2016b; Gamboa Lopez et al. 2021). The competition during the early assembly led me to also test whether inactive analogs could also compete with Pladienolide B and Spliceostatin A inhibition in the catalytic spliceosome, after 1st step of splicing chemistry. Indeed, increasing concentrations of both inactive analogs of Pladienolide B and Herboxidiene restored mRNA production (Figure III-2, lanes 10-13 & 14-16; Figure III-3, lanes 10-12). The successful competition of inactive inhibitors indicates that the inactive competitors likely only work through SF3B1, and the inhibition of exon ligation is also dependent on SF3B1.

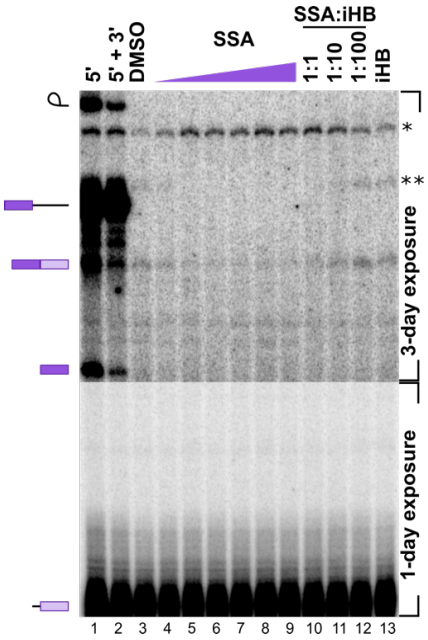


Figure III-3. **Bimolecular assay shows that Spliceostatin A (SSA) inhibits exon ligation.** Denaturing gel analysis of *in vitro* bimolecular exon ligation reactions. Identities of bands are schematized to the left as (from top to bottom) lariat intermediate, 5' substrate, mRNA, 5' exon and 3' exon. A three-day exposure is shown for the top region of the gel, while the bottom region is a one-day exposure. In lanes 1 and 2 both 5' and 3' substrates are radiolabeled. In lanes 3-17 only the 3' substrate is radiolabeled. In lanes 4-9 increasing concentration of Spliceostatin A (SSA) (0.1, 1, 2, 5, 10, 100 μM) were included with the 3' substrate. In lanes 10-12, active Spliceostatin A (SSA) was added at 2 μM and increasing concentration of inactive Herboxidiene (HB) was added at the indicated ratio. In lane 13 inactive Herboxidiene (HB) is at 200 μM. *represents a spiked in loading control of a full-length pre-mRNA. **is the spliced mRNA product of *.

3. Discussion

The observation that SF3B1 inhibitors affect the spliceosome cycle through splicing catalysis represents a new and important addition to our understanding of SF3B1 function in the spliceosome. Importantly, inhibition of SF3B1 at exon ligation is rescued by inactive analogs, suggesting the same activity of SF3B1 demonstrated early in the spliceosome cycle is at work in exon ligation (Effenberger et al. 2016b).

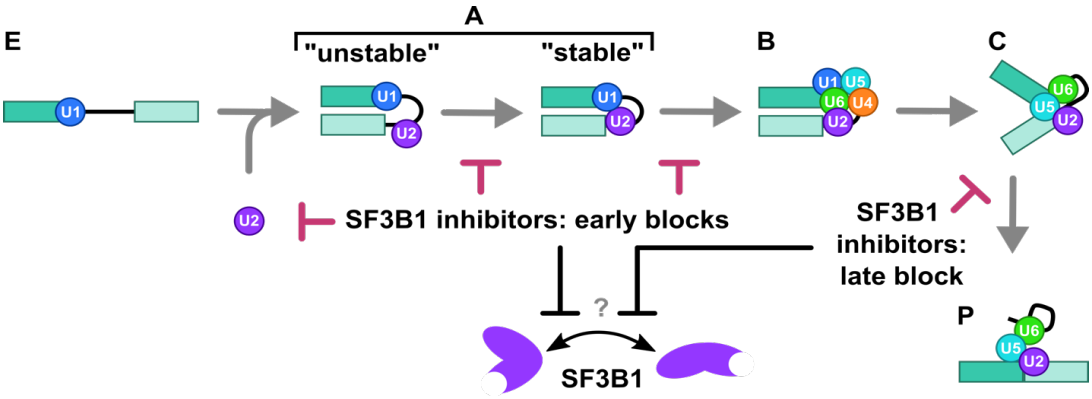


FIGURE III-4. **Model of SF3B1 function in the spliceosome cycle.** Spliceosomes assemble on a pre-mRNA substrate through a series of intermediate splicing complexes (E → A → B → C → P). I propose that SF3B1 inhibitors affect assembly at multiple steps by interfering with a conformational change in SF3B1.

Current cryo-EM structures of SF3B1 demonstrate multiple conformations which are both likely relevant to its function (Cretu et al. 2018). SF3B1 inhibitors are proposed to prop SF3B1 in an “open” conformation and render it unable to clamp down the helix formed between U2 snRNA and the branch sequence of the intron. Initially, I proposed that changes in SF3B1 conformation could be linked to the

rearrangements in U2 snRNA structure that also toggle during spliceosome assembly and which clearly have roles in branch point recognition, A-complex stabilization, and catalytic activation of the spliceosome. Upon further investigation, I observed the change in U2 snRNA structure is not directly affected by the addition of SF3B1 inhibitors at A-complex (See Chapter II). These rearrangements are instead linked to the activity of the RNA-dependent ATPases Prp5 and Prp16 (Perriman et al. 2003; Hilliker et al. 2007; Perriman and Ares 2007; Perriman and Ares 2010). As noted above, SF3B1 interacts with the branch helix during early spliceosome assembly, and subsequently, the branch helix is released during spliceosome activation (Plaschka et al. 2018; Wilkinson et al. 2021; Zhan et al. 2018). Interestingly, SF3B remains associated with the intron after the first step of splicing chemistry (Coltri et al. 2011). Since SF3B is associated with the intron, this opens the question of what is SF3B doing after it releases the branch helix? One possibility is that SF3B can interact with the intron by clamping down and holding it out of the way for second step chemistry to occur and the inhibitor interferes with that function. Alternatively, SF3B could simply be a scaffold to support the rearrangement from 1st to 2nd step chemistry, and the inhibitor weakens the scaffold or disrupts the structure.

4. Methods

4.1. Bimolecular exon ligation

Templates for 5′ and 3′ substrates were generated by PCR of the pre-mRNA substrate derived from the adenovirus major late (AdML) transcript using primers that added a 5′ T7 promoter sequence. ^{32}P-UTP body-labeled G(5′)ppp(5′)G-capped 5 substrate and ^{32}P-UTP body-labeled GMP-capped 3′ substrate were generated by T7 run-off transcription followed by gel purification. Unlabeled G(5′)ppp(5′)G-capped 5′ substrate was generated by T7 run-off transcription followed by size-exclusion chromatography. For bimolecular exon ligation reactions, 5 nM 5′ substrate was incubated with 60 mM potassium glutamate, 2 mM magnesium acetate, 2 mM ATP, 5 mM creatine phosphate, 0.05 mg mL−1 tRNA, and 50% (v/v) HeLa NE at 30°C for 30 min. After 30 min, 3′ substrate with and without SF3B1 inhibitors was added to the reaction and incubated for an additional 60 min.

4.2. Denaturing gel analysis

RNA was extracted from *in vitro* splicing or bimolecular ligation reactions and separated on a 15% (v/v) denaturing polyacrylamide gel. ^{32}P-labeled RNA species were visualized by phosphorimaging and quantified with ImageQuant software (Molecular Dynamics).

References

Agafonov DE, Deckert J, Wolf E, Odenwalder P, Bessonov S, Will CL, Urlaub H, and Luhrmann R. 2011. Semiquantitative proteomic analysis of the human spliceosome via a novel two-dimensional gel electrophoresis method. Mol Cell Biol ***31***: 2667-2682.

Anderson K, and Moore MJ. 1997. Bimolecular exon ligation by the human spliceosome. Science ***276***: 1712-1716.

Ares M, and Weiser B. 1995. Rearrangement of snRNA structure during assembly and function of the spliceosome. Prog Nucleic Acid Res Mol Biol ***50***: 131-159.

Ast G, Pavelitz T, and Weiner AM. 2001. Sequences upstream of the branch site are required to form helix II between U2 and U6 snRNA in a trans-splicing reaction. Nucleic Acids Res ***29***: 1741-1749.

Behrens SE, Tyc K, Kastner B, Reichelt J, and Luhrmann R. 1993. Small nuclear ribonucleoprotein (RNP) U2 contains numerous additional proteins and has a bipartite RNP structure under splicing conditions. Mol Cell Biol ***13***: 307-319.

Berget SM, Moore C, and Sharp PA. 1977. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. Proc Natl Acad Sci U S A ***74***: 3171-3175.

Bertram K, Agafonov DE, Dybkov O, Haselbach D, Leelaram MN, Will CL, Urlaub H, Kastner B, Luhrmann R, and Stark H. 2017. Cryo-EM Structure of a Pre-catalytic Human Spliceosome Primed for Activation. Cell ***170***: 701-713 e11.

Brosi R, Hauri HP, and Kramer A. 1993. Separation of splicing factor SF3 into two components and purification of SF3a activity. J Biol Chem ***268***: 17640-17646.

Chow LT, Gelinas RE, Broker TR, and Roberts RJ. 1977. An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. Cell ***12***: 1-8.

Coltri P, Effenberger K, Chalkley RJ, Burlingame AL, and Jurica MS. 2011. Breaking up the C complex spliceosome shows stable association of proteins with the lariat intron intermediate. PLoS One ***6***: e19061.

Corrionero A, Minana B, and Valcarcel J. 2011. Reduced fidelity of branch point recognition and alternative splicing induced by the anti-tumor drug spliceostatin A. Genes Dev ***25***: 445-459.

Cretu C, Agrawal AA, Cook A, Will CL, Fekkes P, Smith PG, Luhrmann R, Larsen N, Buonamici S, and Pena V. 2018. Structural Basis of Splicing Modulation by Antitumor Macrolide Compounds. Mol Cell ***70***: 265-273 e8.

Dignam JD, Lebovitz RM, and Roeder RD. 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res. ***11***: 1475-1489.

Effenberger KA, Anderson DD, Bray WM, Prichard BE, Ma N, Adams MS, Ghosh AK, and Jurica MS. 2014. Coherence between cellular responses and in vitro splicing inhibition for the anti-tumor drug pladienolide B and its analogs. J Biol Chem ***289***: 1938-1947.

Effenberger KA, Urabe VK, and Jurica MS. 2016a. Modulating splicing with small molecular inhibitors of the spliceosome. Wiley Interdiscip Rev RNA

Effenberger KA, Urabe VK, Prichard BE, Ghosh AK, and Jurica MS. 2016b. Interchangeable SF3B1 inhibitors interfere with pre-mRNA splicing at multiple stages. RNA ***22***: 350-359.

Folco EG, Coil KE, and Reed R. 2011. The anti-tumor drug E7107 reveals an essential role for SF3b in remodeling U2 snRNP to expose the branch point-binding region. Genes Dev ***25***: 440-444.

Fourmann JB, Dybkov O, Agafonov DE, Tauchert MJ, Urlaub H, Ficner R, Fabrizio P, and Lührmann R. 2016. The target of the DEAH-box NTP triphosphatase Prp43 in Saccharomyces cerevisiae spliceosomes is the U2 snRNP-intron interaction. Elife ***5***:

Galej WP, Wilkinson ME, Fica SM, Oubridge C, Newman AJ, and Nagai K. 2016. Cryo-EM structure of the spliceosome immediately after branching. Nature ***537***: 197-201.

Gamboa Lopez A, Allu SR, Mendez P, Chandrashekar Reddy G, Maul-Newby HM, Ghosh AK, and Jurica MS. 2021. Herboxidiene Features That Mediate Conformation-Dependent SF3B1 Interactions to Inhibit Splicing. ACS Chem Biol ***16***: 520-528.

Hasegawa M, Miura T, Kuzuya K, Inoue A, Won Ki S, Horinouchi S, Yoshida T, Kunoh T, Koseki K, Mino K, Sasaki R, Yoshida M, and Mizukami T. 2011.

Identification of SAP155 as the target of GEX1A (Herboxidiene), an antitumor natural product. ACS Chem Biol ***6***: 229-233.

Haselbach D, Komarov I, Agafonov D, Hartmuth K, Graf B, Kastner B, Luehrmann R, and Stark H. 2018. human Bact spliceosome core structure.

Hilliker AK, Mefford MA, and Staley JP. 2007. U2 toggles iteratively between the stem IIa and stem IIc conformations to promote pre-mRNA splicing. Genes Dev ***21***: 821-834.

Ilagan JO, Chalkley RJ, Burlingame AL, and Jurica MS. 2013. Rearrangements within human spliceosomes captured after exon ligation. RNA ***19***: 400-412.

Ilagan JO, and Jurica MS. 2014. Isolation and accumulation of spliceosomal assembly intermediates. Methods Mol Biol ***1126***: 179-192.

Jurado AR, Tan D, Jiao X, Kiledjian M, and Tong L. 2014. Structure and function of pre-mRNA 5'-end capping quality control and 3'-end processing. Biochemistry ***53***: 1882-1898.

Kaida D, Motoyoshi H, Tashiro E, Nojima T, Hagiwara M, Ishigami K, Watanabe H, Kitahara T, Yoshida T, Nakajima H, Tani T, Horinouchi S, and Yoshida M. 2007. Spliceostatin A targets SF3b and inhibits both splicing and nuclear retention of pre-mRNA. Nat Chem Biol ***3***: 576-583.

Kladwang W, VanLang CC, Cordero P, and Das R. 2011. Understanding the errors of SHAPE-directed RNA structure modeling. Biochemistry ***50***: 8049-8056.

Konforti BB, and Konarska MM. 1995. A short 5' splice site RNA oligo can participate in both steps of splicing in mammalian extracts. RNA ***1***: 815-827.

Kotake Y, Sagane K, Owa T, Mimori-Kiyosue Y, Shimizu H, Uesugi M, Ishihama Y, Iwata M, and Mizui Y. 2007. Splicing factor SF3b as a target of the antitumor natural product pladienolide. Nat Chem Biol ***3***: 570-575.

Kramer A, Gruter P, Groning K, and Kastner B. 1999. Combined biochemical and electron microscopic analyses reveal the architecture of the mammalian U2 snRNP. J Cell Biol ***145***: 1355-1368.

Lardelli RM, Thompson JX, Yates JR, 3rd, and Stevens SW. 2010. Release of SF3 from the intron branchpoint activates the first step of pre-mRNA splicing. RNA ***16***: 516-528.

Liang WW, and Cheng SC. 2015. A novel mechanism for Prp5 function in prespliceosome formation and proofreading the branch site sequence. Genes Dev ***29***: 81-93.

Lorenz R, Bernhart SH, Höner Zu Siederdissen C, Tafer H, Flamm C, Stadler PF, and Hofacker IL. 2011. ViennaRNA Package 2.0. Algorithms Mol Biol ***6***: 26.

Maji D, Grossfield A, and Kielkopf CL. 2019. Structures of SF3b1 reveal a dynamic Achilles heel of spliceosome assembly: Implications for cancer-associated abnormalities and drug discovery. Biochim Biophys Acta Gene Regul Mech ***1862***: 194440.

Mortimer SA, and Weeks KM. 2007. A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. J Am Chem Soc ***129***: 4144-4145.

Nakajima H, SATO B, FUJITA T, TAKASE S, TERANO H, and MOKUHARA. 1996. New Antitumor Substances, FR901463, FR901464 and FR901465. I. Taxonomy, Fermentation, Isolation, Physico-chemical Properties and Biological Activities. Journal of Antibiotics ***49***: 1196-1203.

Palka C, Forino NM, Hentschel J, Das R, and Stone MD. 2020. Folding heterogeneity in the essential human telomerase RNA three-way junction. RNA ***26***: 1787-1800.

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

Perriman R, and Ares M, Jr. 2010. Invariant U2 snRNA nucleotides form a stem loop to recognize the intron early in splicing. Mol Cell ***38***: 416-427.

Perriman R, Barta I, Voeltz GK, Abelson J, and Ares M, Jr. 2003. ATP requirement for Prp5p function is determined by Cus2p and the structure of U2 small nuclear RNA. Proc Natl Acad Sci USA ***100***: 13857-13862.

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

Plaschka C, Newman AJ, and Nagai K. 2019. Structural Basis of Nuclear pre-mRNA Splicing: Lessons from Yeast. Cold Spring Harb Perspect Biol ***11***: 1-19.

Plaschka C, Lin PC, Charenton C, and Nagai K. 2018. Prespliceosome structure provides insights into spliceosome assembly and regulation. Nature ***559***: 419-422.

Price SR, Evans PR, and Nagai K. 1998. Crystal structure of the spliceosomal U2B"-U2A' protein complex bound to a fragment of U2 small nuclear RNA. Nature ***394***: 645-650.

Reuter JS, and Mathews DH. 2010. RNAstructure: software for RNA secondary structure prediction and analysis. BMC Bioinformatics ***11***: 129.

Roybal GA, and Jurica MS. 2010. Spliceostatin A inhibits spliceosome assembly subsequent to prespliceosome formation. Nucleic Acids Res ***38***: 6664-6672.

Sakai T, Sameshima T, Matsufuji M, Kawamura N, Dobashi K, and Mizui Y. 2004. Pladienolides, new substances from culture of Streptomyces platensis Mer-11107. I. Taxonomy, fermentation, isolation and screening. J Antibiot (Tokyo) ***57***: 173-179.

Sakai Y, Tsujita T, Akiyama T, Yoshida T, Mizukami T, Akinaga S, Horinouchi S, Yoshida M, and Yoshida T. 2002. GEX1 compounds, novel antitumor antibiotics related to herboxidiene, produced by Streptomyces sp. II. The effects on cell cycle progression and gene expression. J Antibiot (Tokyo) ***55***: 863-872.

Smith CW, Porro EB, Patton JG, and Nadal-Ginard B. 1989. Scanning from an independently specified branch point defines the 3' splice site of mammalian introns. Nature ***342***: 243-247.

Staley JP, and Guthrie C. 1998. Mechanical devices of the spliceosome: motors, clocks, springs, and things. Cell ***92***: 315-326.

Tian S, Cordero P, Kladwang W, and Das R. 2014. High-throughput mutate-map-rescue evaluates SHAPE-directed RNA structure and uncovers excited states. RNA ***20***: 1815-1826.

Wahl MC, Will CL, and Luhrmann R. 2009. The spliceosome: design principles of a dynamic RNP machine. Cell ***136***: 701-718.

Wan R, Bai R, Yan C, Lei J, and Shi Y. 2019. Structures of the Catalytically Activated Yeast Spliceosome Reveal the Mechanism of Branching. Cell ***177***: 339-351.e13.

Wan R, Yan C, Bai R, Lei J, and Shi Y. 2017. Structure of an Intron Lariat Spliceosome from Saccharomyces cerevisiae. Cell ***171***: 120-132.e12.

Wilkinson ME, Charenton C, and Nagai K. 2020. RNA Splicing by the Spliceosome. Annu Rev Biochem ***89***: 359-388.

Wilkinson ME, Fica SM, Galej WP, and Nagai K. 2021. Structural basis for conformational equilibrium of the catalytic spliceosome. Mol Cell ***81***: 1439-1452.e9.

Will CL, and Luhrmann R. 2011. Spliceosome structure and function. Cold Spring Harb Perspect Biol ***3***: a003707.

Wu J, and Manley JL. 1989. Mammalian pre-mRNA branch site selection by U2 snRNP involves base pairing. Genes Dev ***3***: 1553-1561.

Yan C, Wan R, and Shi Y. 2019. Molecular Mechanisms of pre-mRNA Splicing through Structural Biology of the Spliceosome. Cold Spring Harb Perspect Biol ***11***: a032409.

Yan C, Hang J, Wan R, Huang M, Wong CC, and Shi Y. 2015. Structure of a yeast spliceosome at 3.6-angstrom resolution. Science ***349***: 1182-1191.

Yan C, Wan R, Bai R, Huang G, and Shi Y. 2016. Structure of a yeast activated spliceosome at 3.5 A resolution. Science ***353***: 904-911.

Yan D, Perriman R, Igel H, Howe KJ, Neville M, and Ares M, Jr. 1998. CUS2, a yeast homolog of human Tat-SF1, rescues function of misfolded U2 through an unusual RNA recognition motif. Mol Cell Biol ***18***: 5000-5009.

Yoon S, Kim J, Hum J, Kim H, Park S, Kladwang W, and Das R. 2011. HiTRACE: high-throughput robust analysis for capillary electrophoresis. Bioinformatics ***27***: 1798-1805.

Zavanelli MI, and Ares M. 1991. Efficient association of U2 snRNPs with pre-mRNA requires an essential U2 RNA structural element. Genes Dev ***5***: 2521-2533.

Zhan X, Yan C, Zhang X, Lei J, and Shi Y. 2018. Structure of a human catalytic step I spliceosome. Science ***359***: 537-545.

Zhang X, Yan C, Hang J, Finci LI, Lei J, and Shi Y. 2017. An Atomic Structure of the Human Spliceosome. Cell ***169***: 918-929 e14.

Zhang X, Yan C, Zhan X, Li L, Lei J, and Shi Y. 2018. Structure of the human activated spliceosome in three conformational states. Cell Res ***28***: 307-322.

Zhang Z, Will CL, Bertram K, Dybkov O, Hartmuth K, Agafonov DE, Hofele R, Urlaub H, Kastner B, Lührmann R, and Stark H. 2020. Molecular architecture of the human 17S U2 snRNP. Nature ***583***: 310-313.