UNIVERSITY OF CALIFORNIA, SAN DIEGO

Structural Insights Into RNA Splicing

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Chemistry

by

Jessica K. Peters

Committee in charge:

Professor Navtej Toor, Chair
Professor Neal Devaraj
Professor Simpson Joseph
Professor Amy Pasquinelli
Professor F. Akif Tezcan

2017

The dissertation of Jessica K. Peters is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____
Chair

University of California, San Diego

2017

Epigraph

If two people agree on everything, one of them is unnecessary.

- Attributed to Winston Churchill

Table of Contents

List of Abbreviations

APS	Ammonium Persulfate
DI-VI	Domains 1-6
DTT	Dithiothreitol
EBS	Exon Binding Sequence
EDTA	Ethylenediaminetetraacetic Acid
EM	Electron Microscopy
EMSA	Electrophoretic Mobility Shift Assay
IBS	Intron Binding Sequence
IEP	Intron Encoded Protein
LB	Lysogeny Broth
MAD	Multiwavelength Anomalous Dispersion
mRNA	Messenger RNA
NAIM	Nucleotide Analog Interference Mapping
NMR	Nuclear Magnetic Resonance
nt	Nucleotide
NTP	Nucleotide Triphosphate
ORF	Open Reading Frame
PCR	Polymerase Chain Reaction
rRNA	Ribosomal RNA
RT	Reverse Transcriptase
SAXS	Small Angle X-ray Scattering
SDS	Sodium Dodecyl Sulfate

TL-TLR	Tetraloop-Tetraloop Receptor
WC	Watson-Crick
WT	Wild Type

List of Figures

List of Tables

Acknowledgements

First and foremost, I want to acknowledge my mentor for the past five and a half years, Professor Navtej Toor. Thank you for being so relentlessly optimistic. I may not have always seemed appreciative of that persistent optimism in the moment, but it helped to motivate me to set up one more crystal tray or make one more mutation instead of giving up. To my past and present lab mates (Dr. Aaron Robart, Dr. Russell Chan, Tim Wiryaman, Anastassia Gomez, Dan Haack) – thank you for being my comrades in arms through my grad school experience. Aaron – your immense experience working with RNA, and more specifically with group II introns, was so helpful to me. I am incredibly grateful for the years we overlapped in the Toor lab and for the assistance you provided me as a mentor and, I hope, peer through various scientific discussions. I have no doubts you will be very successful running your own research lab. Russell – you are probably one of the kindest people I have ever met. Don't ever change. Seriously, it's not every day you meet someone who is more than happy to spend so much of their time and resources to help someone else, never asking for anything in return. Science needs more people like you. Tim & Ana – thank you for being such wonderful office-mates for my final year in grad school. I was a little worried when Ana brought out the kimchi shortly after moving into our office, but the smells only got better and the conversations more valuable with time.

Although I never did any research in the Figueroa lab (bar one strange experiment that needed to be performed under argonne), I somehow managed to infiltrate the ranks and become an honorary member. My guess is that it had something

to do with Brandon (vide infra), but we may never know. Regardless, thank you to lab members past and present for all the fun times and for being such kick-ass, hard-working scientists.

My time at UCSD would have been far lonelier if not for the many friends I made when I moved to California. Dr. Dan Shiek – thank you for always being game for an evening out, and for the everlasting supply of colorful comments, especially the ones that somehow seem to get Brandon in trouble. Jamie Schiffer and Naomi Handly – thank you for being such fantastic role models. I will always admire how selflessly you both live your lives. Melissa Helm – thank you for including me in Camelot gatherings and being a fellow cat lover. And I guess for peer-pressuring me into downloading Snapcat (also known as Snapchat to those who use it for anything besides sending pictures and videos of cats). Sam – thank you for all the fun roommate dinners/shopping trips/mani-pedis/etc and for being a good listener as I aired frustrations about my day. I'm glad such a brief period of living with me and Juno inspired you to get a cat of your own. Though he is a weirdo, it only makes us love him more. I don't have the space or energy to go on about all the friends I've made in California, but I am forever grateful for the many laughs and thoughtful conversations I've shared with all of you.

I've been so fortunate to have friends from high school and college who are still present and sources of support in my life. Soohee Cho and John Tillotson – thank you for making chemistry classes, and especially labs, so much fun at Georgia Tech. I'm so glad we're able to get together after not having seen each other for many years and pick

up where we left off. Lauren Stollar – thank you for the care packages and gifts throughout the years. It means the world to me to have such a caring friend who, even though may be 2,000 miles away, recognizes when I'm feeling lonely and reaches out. The same goes for you, Rebecca Pearl. I am so grateful your family decided to move in next door to mine nearly twenty years ago. My life has only been better because of it. You have always been like a sister to me, and although we may never live in the same city again, I can't imagine my life without you in it. There are only a handful of people whom I could talk to for hours on the phone, not caring if the conversation would lag and the silent moments become longer. You, my friend, are at the top of that list.

Brandon – you have these moments when you are, without a doubt, the grouchiest, Grinchiest, most curmudgeonly person I know. Thankfully those moments are brief (though others may disagree with my definition of brief). In their absence, you happen to be someone whom I admire very much. Not a day goes by that I don't look up to you and your kick ass work ethic. Spending so much time with you (and the rest of the Figueroa gang) has motivated me to push myself to learn more, work harder, and be a better version of myself every single day. Thank you for being the yin to my yang and bringing out the best in me. I'm so grateful to have you as my best friend and partner in crime. Sorry I'm not sorry.

Lastly, I need to acknowledge the ones who tolerated me and inspired me the longest: my brother and dad. I've been so lucky to have grown up surrounded by such smart, inquisitive, and hard-working people. Though you may drive me crazy after only interacting with you for a few minutes, I wouldn't trade you for anyone else. There

aren't enough words in my vocabulary or enough space in this dissertation to describe to the full extent how much I appreciate everything you've done for me. Thank you for the unconditional love and support through the years. I want to especially acknowledge my dad for all that you've done for Zack and myself. Thank you for supporting whatever it was that we wanted to do. Whether it was driving hours to the middle of nowhere, GA to see your son play soccer or buying Georgia Tech season football tickets to see your daughter on the field at half time, you were there for us every step of the way. You also showed us that it's ok to fail because you'll only find another way to succeed in a greater way. I can say, with absolute certainty, that I would not be where I am today if it wasn't for you. Thank you.

Chapter 3, in part, is currently being prepared for submission to *Nature Structural & Molecular Biology* for publication of the material: Chan, R.T. and Peters, J.K., Robart A.R., Rajashankar, K.R. & Toor, N. Structural basis for the second step of group II intron splicing. The dissertation author is a co-first author of this paper.

Chapter 4, in part, is currently being prepared for submission for publication of the material. Peters, J.K. & Toor, N. A kinetic and thermodynamic analysis of domain VI dynamics in the group II intron *P.li.LSUI2*. The dissertation author is the first author or this paper.

Chapter 5, in full, is a reprint with permission of the material as it appears in *Nature*: Robart, A.R., Chan, R.T., Peters, J.K., Rajashankar, K.R. & Toor, N. Crystal

structure of a eukaryotic group II intron lariat. *Nature* **514**, 193-197 (2014). Copyright 2014, Nature Publishing Group. The dissertation author is the third author of this paper.

Chapter 6 is adapted with permission from Peters, J.K. & Toor, N. Group II intron lariat: structural insights into the spliceosome. *RNA Biology* **95**, 9244-9249 (2015). Copyright 2015, Taylor & Francis, LLC (http://www.tandfonline.com). The dissertation author is the first author of this paper.

This work was supported, in part, by the National Institute of Health Molecular Biophysics Training Program and the ARCS Foundation.

Vita

2011 Bachelor of Science, Georgia Institute of Technology

2013 Master of Science, University of California, San Diego

2017 Doctor of Philosophy, University of California, San Diego

Publications

Robart, A.R., Chan, R.T., Peters, J.K., Rajashankar, K. R. & Toor, N. Crystal structure of a eukaryotic group II intron lariat. *Nature* **514**, 193-197 (2014).

Peters, J. K. & Toor, N. Group II intron lariat: structural insights into the spliceosome. *RNA Biology* **12**, 913-917 (2015).

Chan, R.T., Peters, J.K., Robart, A.R., Rajashankar, K. R. & Toor, N. Structural basis for the second step of group II intron splicing. Manuscript in preparation for submission to *Nature Structural & Molecular Biology*.

Peters, J.K. & Toor, N. A kinetic and thermodynamic analysis of domain VI dynamics in the group II intron *P.li.LSUI2*. Manuscript in preparation for submission.

Abstract of Dissertation

Structural Insights Into RNA Splicing

by

Jessica K. Peters

Doctor of Philosophy in Chemistry

University of California, San Diego, 2017

Professor Navtej Toor, Chair

The formation of branched lariat RNA is an evolutionarily conserved feature of RNA splicing reactions for both group II and spliceosomal introns. In both splicing systems, the lariat is formed in the first of two transesterification reactions as a result of the 2' hydroxyl of a conserved bulged adenosine within the intron attacking the phosphate backbone at the 5' splice site. In the second step, the 3' hydroxyl of the 5' exon attacks the 3' splice site, resulting in ligation of the exons and release of the intron as a lariat. Due to the mechanistic and structural similarities, it is believed that group II

introns and the spliceosome share a common ancestor. Though many structures have been published of both group II introns and the spliceosome, the precise structural requirements for splicing in both systems remain unknown. In the case of group II introns, all of the structures published prior to the work in this dissertation are of a primitive group IIC intron from the bacterium *Oceanobacillus iheyensis* (*O.i.*). This intron splices through a hydrolytic pathway and therefore does not provide a comprehensive understanding of branch formation, which severely limits the conclusions that can be applied to our understanding of the spliceosome. This dissertation describes various experiments that have been performed to better understand how group II introns and the RNA components of the spliceosome active site are able to properly fold and place substrates within the single active site for the two subsequent reactions.

Two new crystal structures of the lariat-forming IIB intron from the brown algae *Pylaiella litoralis*, *P.li.LSUI2*, at different stages of splicing are outlined in this dissertation: the intermediate lariat–3' exon and post-catalytic lariat. Two different arrangements of the catalytic triplex, the residues which create the active site scaffold, are observed in these structures. However, examination of all published *O.i.* crystal structures reveals the same catalytic triplex arrangement throughout splicing. Therefore, the rearrangement that occurs in the more evolved *P.li.LSUI2* intron is likely facilitating the transition between the stages of splicing and relevant to branch formation and exon ligation in the spliceosome. These structures are also the first in which domain VI is visualized. Domain VI is integral to the branching pathway as it contains the first-step

nucleophilic residue. For the first time, domain II can be seen forming two tetraloop-receptor interactions, π-π' and η-η', with domain VI proximal to the bulged adenosine. These tertiary interactions appear to stabilize domain VI. However, disrupting either or both interactions simultaneously results in a stalling of the second step. Additionally, iridium hexammine binds in the major groove of domain VI and promotes exon ligation. Taken together, this suggests that domain VI is dynamic and plays an important role in facilitating both steps of splicing.

Chapter 1: Introduction

1.1 Discovery of group II introns

When RNA research became popular in the mid-1900s, it was thought that RNA played only three major roles within cells: 1) as the translatable message containing our genetic information (mRNA), 2) as the structural framework of the ribosome (rRNA), and 3) as the scaffold linking amino acid building blocks to the genetic code (tRNA). Carl Woese, Francis Crick, and Leslie Orgel originally proposed the idea that RNA could play another important role as catalysts in 1968[1–3], but it wasn't until 1982 that evidence was found for such RNAs. Catalytic RNAs, or ribozymes, were discovered in the early 1980s by Tom Cech[4,5] and Sid Altman[6], an achievement which awarded them the Nobel Prize in chemistry in 1989. Their introduction of the idea that RNAs could catalyze reactions, a macromolecular role previously thought only to be held by proteins, led to the discovery of many more catalytic RNAs. Over the next few decades, RNAs were discovered or created with a broad range of functions including self-cleavage[4,5,7–9], cyclization[10], polymerization[11], branching[12], tRNA processing[6], and many more.

The particular catalytic RNA discovered by Tom Cech was a self-splicing intron RNA from the protozoan *Tetrahymena thermophila*. This class of RNA catalyzes its self-removal from flanking exon sequences through two transesterification reactions utilizing an exogenous guanosine as the nucleophile in the first step (Figure 1.1). As a true ribozyme, this intron does this in the absence of any proteins. Shortly after this discovery, more self-cleaving introns were found interrupting genes in the mitochondria

of yeast and fungi[13]. Although these introns were also capable of splicing in the absence of proteins, they did not share any sequence homology with group I introns. Investigation of *Saccharomyces cerevisiae* mitochondrial RNAs using electron microscopy revealed covalent circular intron RNA products (Figure 1.2) similar to those produced by the spliceosome[14]. This was particularly interesting because mitochondria do not contain spliceosomal introns or the machinery to remove them. This circular product, later referred to as a lariat, provided evidence for a new class of auto-catalytic introns more closely related to spliceosomal introns. Based on this and phylogenetic analyses, these auto-catalytic introns have been classified as group I and group II introns.

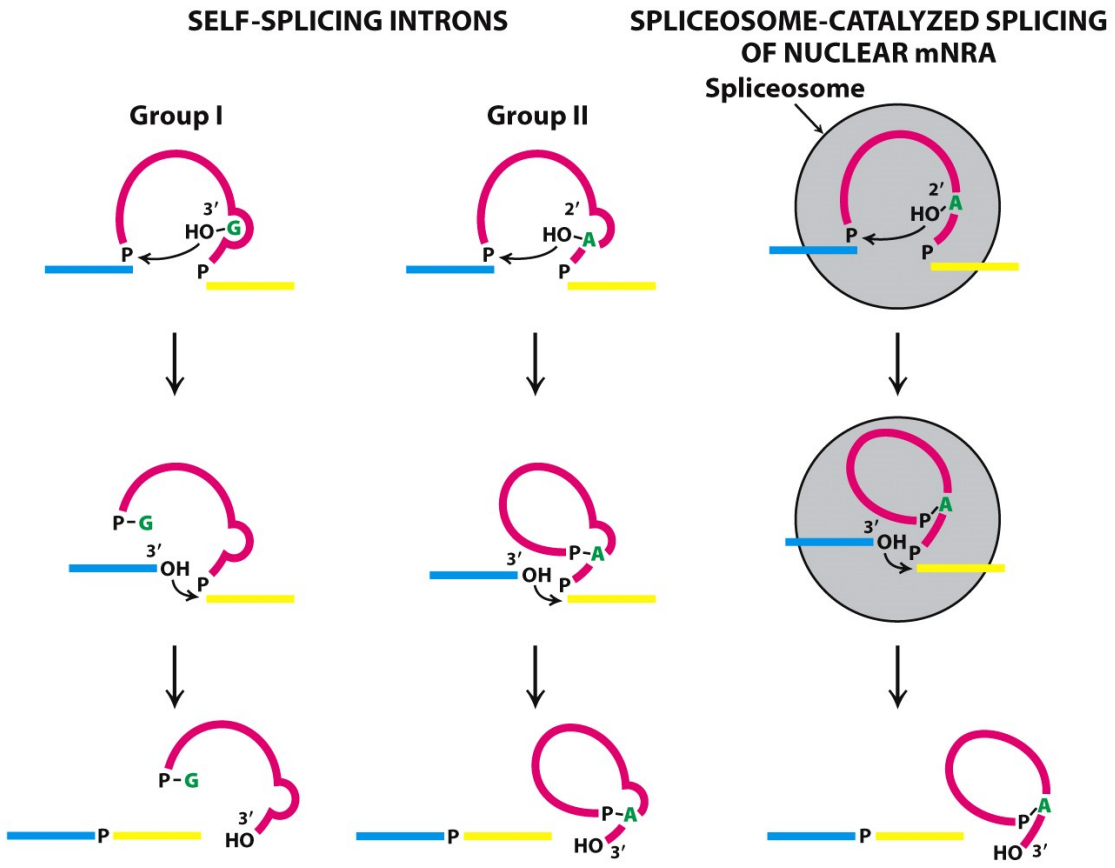

Figure 1.1: **Mechanism of splicing in three splicing systems.** 5′ exon is shown in blue, 3′ exon in yellow, intron in pink, and 1st step nucleophile in green. left) Group I intron splicing proceeds via nucleophilic attack of the 3′ OH of an exogenous guanosine. In the next reaction, the 3′ OH of the 5′ exon attacks the 3’ splice site. middle) Group II intron splicing proceeds via nucleophilic attack of the 2′ OH of a conserved adenosine. The second step proceeds as in group I introns. right) Spliceosomal splicing proceeds identically to group II introns except that the complex is composed of both RNA and proteins. Taken from Biochemistry, 7th Ed[15].

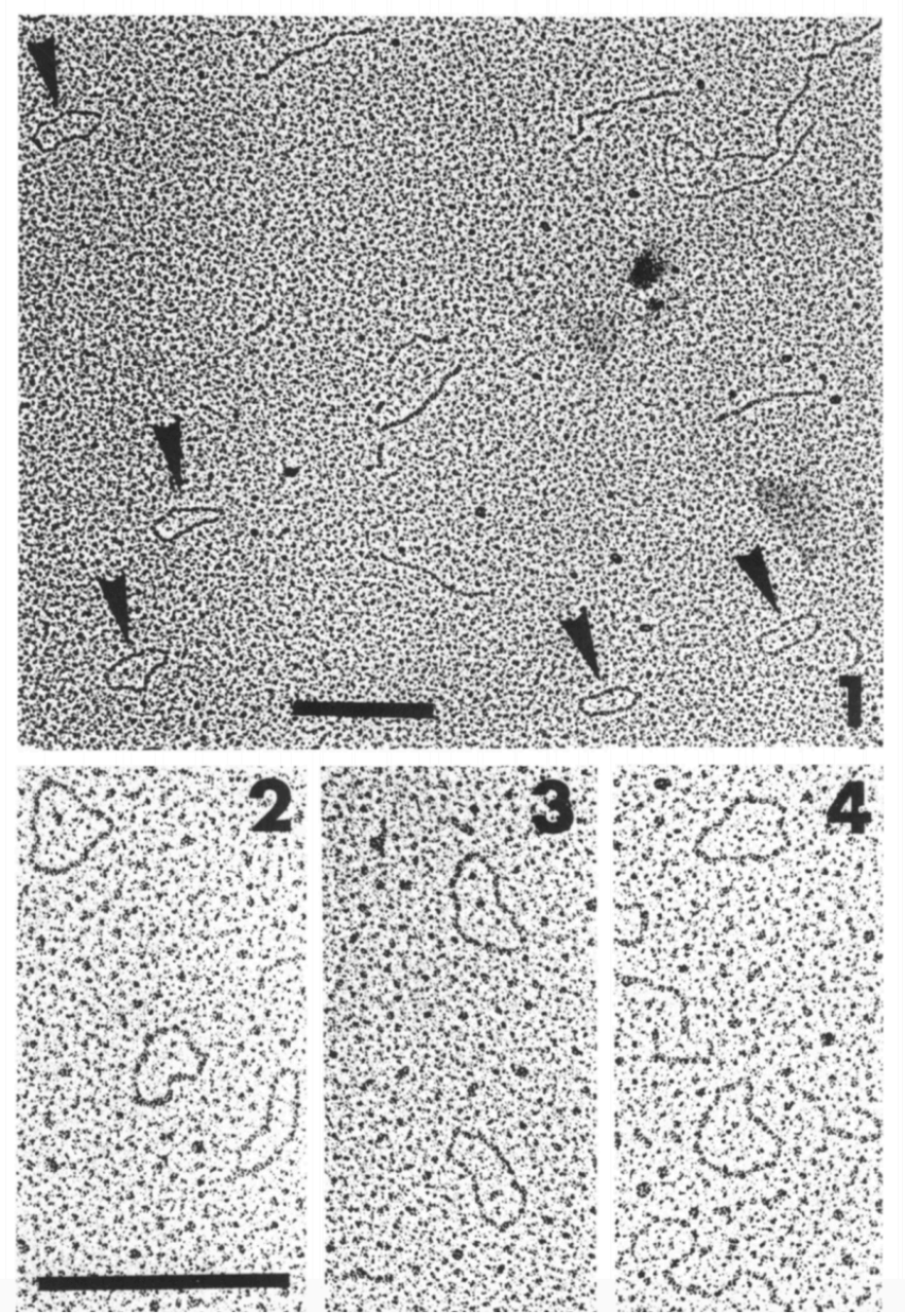

Figure 1.2: **Electron micrographs showing Saccharomyces cerevisiae mitochondrial RNA.** 1) Randomly selected field of ~1Kb RNA contains both circular (indicated with arrows) and linear RNAs. Circular RNAs are assumed to be covalent products because they persist in harsh denaturing conditions (70% formamide at 80°C). 2-4) Fields selected to show circular RNAs. The bars represent 0.25 μm. Taken from Halbreich et al[14].

Though both group I and group II introns splice via two sequential transesterification reactions, the nucleophile for the first step is different (Figure 1.1). In group II introns, the nucleophile for the first step is the 2′ hydroxyl of a conserved adenosine near the 3′ end of the intron, whereas in group I introns it is the 3′ hydroxyl of an exogenous guanosine. Therefore, group I intron splicing results in a linear intron whereas group II intron splicing results in a lariat intron with a unique 2′-5′ bond between the first intron nucleotide and the conserved adenosine[16,17]. The second step proceeds the same in the self-splicing systems: the 3′ hydroxyl of the 5′ exon attacks the 3′ splice site resulting in ligated exons and a free intron. Spliceosome-catalyzed splicing follows the same mechanism as group II introns, however it is both protein- and ATP-dependent. In addition, both transesterification reactions are chemically identical in group II introns and the spliceosome with regard to their scissile phosphate Rp and Sp dependencies[18–21]. The unique lariat bond in these systems also allows for reverse-splicing and, in the case of group II introns, retrotransposition[22–24].

1.2 Phylogenetic classification of group II introns

Group II introns are found in all three domains of life, though are rarely found in archaea[25,26]. Within eukarya, group II introns are limited to the organellar genomes of plants, fungi, protists, and an annelid worm[27–30]. Some group II introns contain and open reading frame (ORF) that codes for an intron-encoded protein (IEP), or maturase[31]. The IEP is a multifunctional protein that plays two roles. First, it binds to the RNA to stabilize the overall structure and promote splicing. Next, it remains bound to the intron

RNA to facilitate invasion of DNA through the reverse splicing activity of the intron and the reverse transcriptase activity of the IEP. As long as the ribozyme remains active and associated with its IEP, it can continuously invade DNA with minimal disruption of gene expression. Nearly all (>99%) of group II introns in bacteria encode an IEP[32]. However, group II introns in the organelles of lower eukaryotes more commonly contain degenerate ribozymes, lack an IEP, or encode degenerate IEPs no longer capable of intron mobility[13,33–36]. Bacterial group II introns are typically inserted intergenically, whereas nearly all organellar group II introns are located within housekeeping genes. In the case of organellar group II introns, their ability to splice is critical to the host's survival. Because these introns are often degenerate or immobile, they have adapted to use host-supplied splicing factors[34,37,38].

Group II introns are subdivided into three classes based on secondary structure and exon recognition: IIA, IIB, and IIC (Figure 1.3). All classes can be found in bacteria, whereas only IIA introns have been found in mitochondrial genomes and IIB introns in chloroplast genomes[39]. Of the three, IIC introns are thought to be the most primitive[25]. They are significantly smaller than IIA/B introns (400 vs. 600-800 nucleotides), and they contain a different catalytic triad sequence than IIA/B introns and the spliceosome (CGC vs. AGC)[40]. IIA and IIB introns typically form large amounts of lariat RNA during *in vitro* self-splicing reactions, whereas IIC introns primarily splice through hydrolysis *in vitro* to form linear intron[41]. Additionally, IIC introns tend to be more promiscuous in their 5′ and 3′ splice site selections *in vitro*, often resulting in internal cleavage products[41]. It is only with the addition of an IEP that the IIC intron

predominantly forms lariat, indicating that IIC introns exhibit a much less stable RNA fold[42].

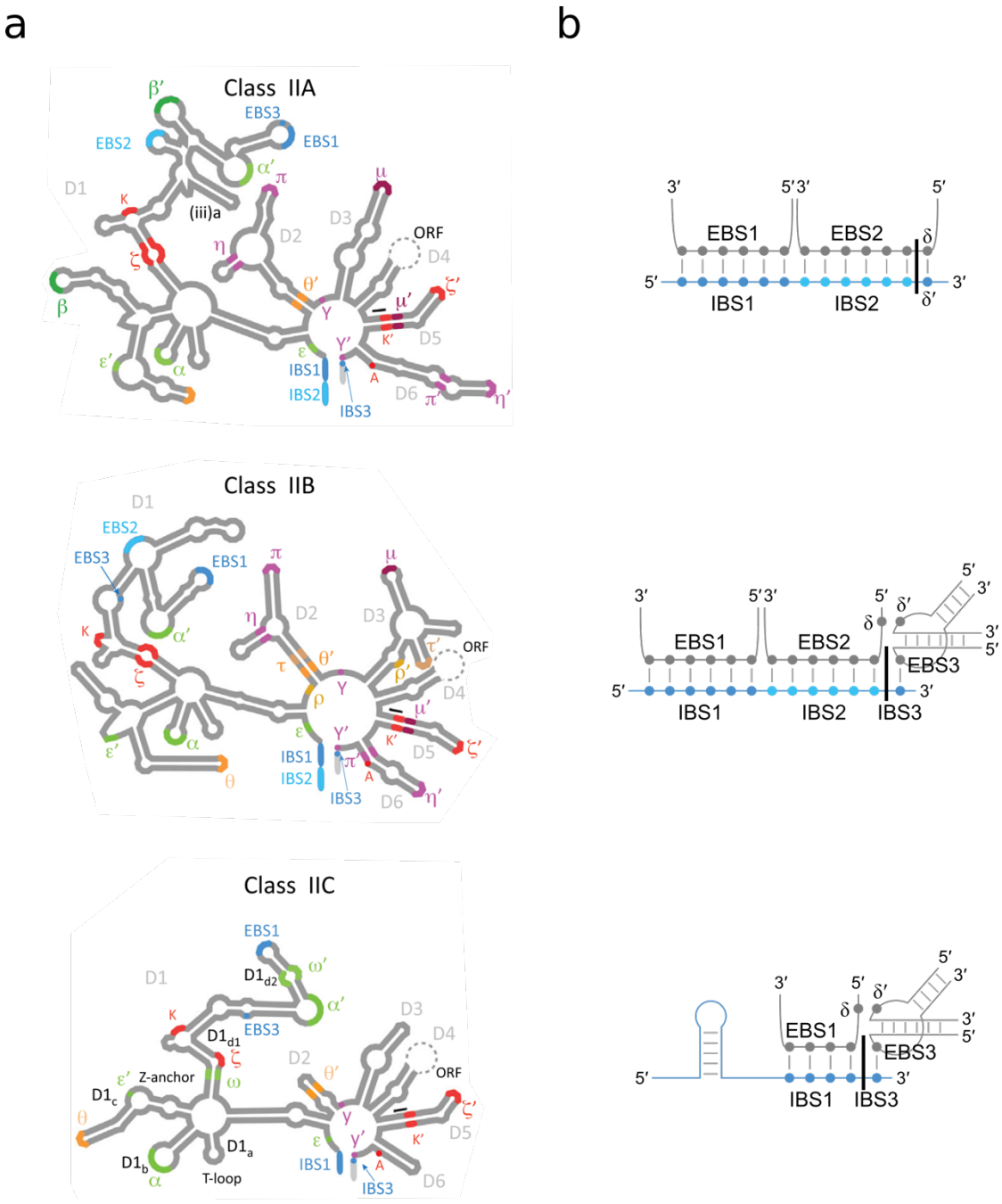

Figure 1.3: **The three subclasses of group II introns.** (a) Secondary structure diagrams of the three subclasses of group II introns. The 6 domains are numbered D1-D6. Long-range tertiary contacts are color-coded and labeled with Greek letters. EBS: exon-binding site; IBS: intron-binding site; ORF: open reading frame. Taken from Zhao and Pyle, 2017[43]. (b) Base-pairing interactions used in exon recognition of each subclass.

1.3 The structure that begets function in group II introns

Group II introns can be easily recognized by their characteristic secondary structure in which six domains (DI-DVI) emanate out from a central hub like spokes on a wheel (Figure 1.4). DI, the largest of the six domains and located at the 5′ end of the intron, acts as a scaffold upon which the remainder or the ribozyme folds. It also contains sequences that base-pair with the 5′ and 3′ exons to correctly position them in the active site for catalysis[44]. DII participates in four tetraloop-receptor interactions within the intron, most important of which are the two with DVI that are implicated to be dynamic throughout splicing[45]. DIII acts as a structural brace in a manner similar to an architectural buttress. DIV encodes the ORF for the IEP. DV contains the highly conserved catalytic triad and two-nucleotide bulge that form the active site and properly position the two catalytic Mg^{2+}, therefore awarding it the title of "catalytic domain." DVI, located at the 3′ end of the intron, contains the conserved bulged adenosine that acts as the nucleophile during the first step of splicing.

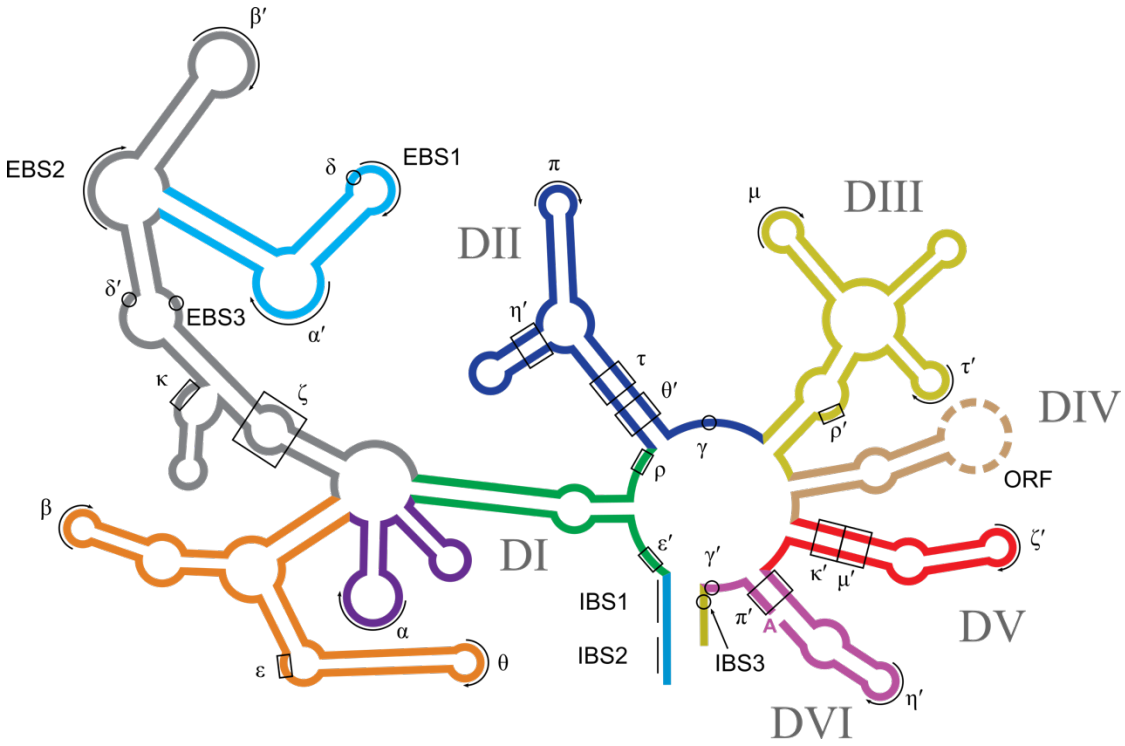

Figure 1.4: **Secondary structure of an unspliced group IIB intron.** The intron is organized into six domains, listed as DI-DVI, radiating out from a central hub. Long-range tertiary interactions are indicated in Greek letters. Coloring is consistent with published group II intron crystal structures. DI is shown as green, purple, orange, grey, and cyan. DII is shown as blue. DIII is shown as yellow. DVI is shown as wheat with the open reading frame as a dashed line. DV is shown as red. DVI is shown as magenta. The 5′ exon is shown as sky blue. The 3′ exon is shown as gold.

Depending on the intron subclass, there are between 7 and 12 known tertiary interactions that stabilize the ribozyme architecture, the vast majority of which involve DI. As mentioned, DI contains sequences, referred to as exon binding sequences (EBS), which pair with both the 5′ (EBS1 and EBS2) and 3′ exon (EBS3) to properly position them within the active site during the different stages of splicing. Additionally, a base-pairing interaction between DI and residues 3 and 4 at the conserved 5′ end of the intron

(GUGYG) also forms to correctly position the 5′ splice site (ε-ε′). Disruption of this interaction results in decreased splicing efficiency that can be recovered with compensatory mutations[46].

Two kissing loop interactions, α-α′[44] and β-β′[47], present within DI are thought to be among the first tertiary contacts to form during ribozyme folding[48]. Though the sequence and length of these interactions is not conserved, evidence of their coevolution suggests an importance to the overall architecture of the ribozyme[40]. In fact, disruption of α-α′ results in defective splicing[49]. It is likely that α-α′ and β-β′ are integral to the proper folding of DI, upon which the remainder of the ribozyme folds. Once the ribozyme has folded to a near-native state, crucial inter-domain contacts form[48]. The two tetralooop-receptor interactions between DI and DV, κ-κ′[50] and ζ-ζ′[33], are positioned at opposite ends of DV to create a ~45° bend in DV at the two-nucleotide bulge[51]. This bend allows the two-nucleotide bugle to properly interact with the catalytic triad and coordinate the two catalytic Mg^{2+} within the active site. Interestingly, DI and DV alone can be combined to form a minimal group II ribozyme that is capable of some form of catalytic activity, though it is much lower than wild type[52,53]. This highlights the crucial role DI plays as a scaffold that is key to correctly positioning active site residues.

DII has a "Y" shaped secondary structure with two inter-domain tertiary contacts in the stem (τ-τ′ and θ-θ′) and one contact in each arm (η-η′ and π-π′) for a total of 4 contacts. In a recent crystal structure of a IIB intron, the two arms of the DII "Y" are coaxially stacking alongside DVI, allowing for both the η-η′[54] and π-π′[45] tetraloop-

receptor interactions to be docked[45]. Deleting DII was shown to have no effect on the first splicing reaction, but resulted in an inhibition of the second reaction[54]. Because of this, it was hypothesized that DVI was undocked for the first step, then underwent a large-scale ratcheting movement to dock into DII for the second step[55]. However, low resolution crystallographic data of the same intron in the pre-catalytic state suggests that η-η′ is engaged prior to branching[45]. Therefore, a new model of DVI rearrangements has been proposed to incorporate evidence that η-η′ is constant and π-π′ is dynamic (see Chapter 5).

DIII size and shape varies, but is conserved within each subclass. The most complex DIII structure is within the group IIB introns. In this class, DIII contains an internal loop followed by a three-way junction of three stem loops. There are three known tertiary interactions in which DIII participates: ρ-ρ′[45] (with DI), τ-τ′[45] (with DII), and μ-μ′[56] (with DV). The deletion or mutagenesis of this domain results in a much lower splicing efficiency which can be restored with the addition of DIII in trans[57,58]. Therefore, it has been referred to as a "catalytic effector" for group II intron splicing as its presence dramatically increases the rate of splicing[57,59,60]. Recent crystal structures have shed light on how DIII is able to have such a strong effect on catalysis when it is not directly interacting with the substrates or active site. We now know that the μ-μ′ interaction is positioned on the opposite side of DV from the active site likely serving to brace DV, and the τ-τ′ interaction with DII appears to stabilize the side of DII opposite from DVI[45].

In the secondary structure, the six domains are connected via single-stranded junction nucleotides. The length of these junctions is variable among the subclasses, with IIC introns containing the shortest sequences. The junction between domains II and III (J2/3) is of particular importance because it forms two crucial tertiary contacts. The first contact made by J2/3 is between a single nucleotide in the middle of J2/3, γ, and the final nucleotide of the intron, γ′. These residues have coevolved to form a Watson-Crick base pair that is required for the second step of splicing[46]. Mutagenesis of this interaction has been shown to severely decrease the rate and fidelity of the second step[61,62]. The γ-γ′ interaction likely works with the IBS3-EBS3 (or δ-δ′ in IIA introns) base-pairing interaction to correctly position the 3′ splice site in the active site with a kinked backbone to facilitate splicing. The second tertiary contact made by J2/3 is from the two conserved nucleotides immediately downstream of γ that participate in base triples with the first two nucleotides of the catalytic triad in DV[51,63,64]. Deletion of these two nucleotides results in a dramatic inhibition of the second step of splicing, likely because their removal alters the placement of the γ nucleotide and makes it unable to pair with γ′.

Previously, a crystal structure of a hydrolytic IIC intron from the bacterium *Oceanobacillus iheyensis* (*O. iheyensis*) revealed that the highly conserved DV forms an active site containing two catalytic magnesium ions[51]. Consecutive base triples consisting of nucleotides from the catalytic triad, DV two-nucleotide bulge, J2/3, and J4/5 form the catalytic triplex, which makes up the scaffold for the active site and catalytic Mg^{2+} binding. This organization was first observed in the crystal structure of a

hydrolytic IIC intron from the bacterium *Oceanobacillus iheyensis* (*O. iheyensis*)[51] and then later in a lariat-forming IIB intron from the brown algae *Pylaiella littoralis* (*P.li.LSUI2*)[45], though the *O. iheyensis* catalytic triplex did not contain J4/5 nucleotides as this intron lacks a junction sequence between domains IV and V.

DVI is the final domain of intron and contains the conserved bulged adenosine required for lariat formation in the first step[65]. It was first visualized in 2014 in the crystal structure of the *P.li.LSUI2* intron[45]. Given that this intron catalyzes lariat formation[66], it has contributed to a new understanding of the structural requirements for splicing in higher eukaryotes. Mutagenesis of the branch-site nucleotide indicates that alterations of the base identity are tolerated, and splicing proceeds with no apparent loss of fidelity but at a much slower rate[67]. The location of the branch-site nucleotide with respect to the base of the stem and the catalytic triad of DV is also strictly enforced, with the ideal arrangement being a 3-nt J5/6 and a 4-bp stem at the base of DVI[68].

1.4 The evolution of group II introns and similarities to the spliceosome

It has been shown that a single group II intron sequence could be fragmented into two or three segments and assembled to trans-splice both *in vitro* and *in vivo*, indicating a modular structural arrangement of these introns[47,53,69–72]. In fact, naturally occurring bipartite and tripartite group II intron systems have been found in plant mitochondria and plant/algal chloroplasts in which separate precursor mRNAs form one intermolecular structure to splice and join exon sequences from the different

transcripts[73]. This provides strong evidence that spliceosomal introns may have evolved from the retrotransposition of group II introns.

Figure 1.5 shows the current hypothesis of how group II introns evolved into the eukaryotic introns that are removed by the spliceosome. It is thought that after symbiosis of single-celled prokaryotes into an archaebacterial host, intron sequences then became a part of the host genome through group II introns retrotransposition and genome transfer following endosymbiont lysis. Gene expression was likely impeded by translation of unspliced transcripts and co-transcriptional translation, which then provided evolutionary pressure to develop a nuclear envelope and a more efficient system to remove introns (i.e., the spliceosome). Fragmentation of group II introns in organelles led to trans-splicing ribozymes, whereas fragmentation in the nucleus led to the spliceosome.

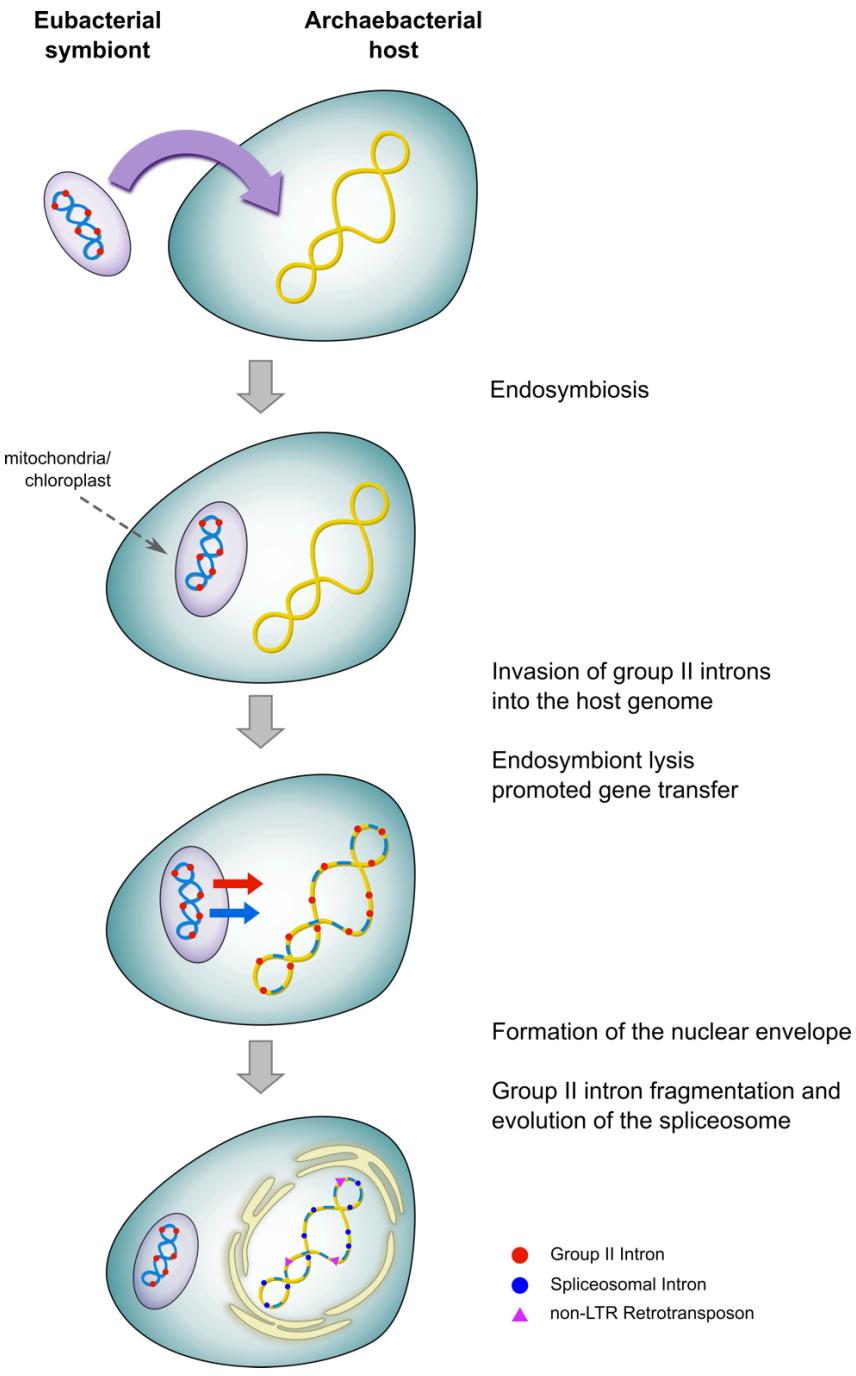

Figure 1.5: **Endosymbiont event and origin of spliceosomal introns.** Eukaryotes are thought to have evolved through the symbiosis of prokaryotes (endosymbiont event. Archaebacterial hosts likely engulfed prokaryotic symbionts that would later evolve into mitochondria and chloroplasts. Horizontal gene transfer and group II intron invasion then resulted in altering of the host genome. Formation of the nuclear envelope then led to the evolution of the spliceosome to remove introns before nuclear export and translation could occur.

The spliceosome is a multi-megadalton complex consisting of 5 snRNAs and up to 150 associated proteins in humans[74,75]. It catalyzes the removal of introns through two sequential transesterification reactions in a mechanism chemically identical to the one occurring in group II introns[20,21,76]. Recent biochemical evidence has also shown that the spliceosome is a ribozyme that uses a two-metal-ion mechanism of catalysis[77]. Two spliceosomal RNAs, U2 and U6, comprise the active site and are capable of performing a splicing-related reaction in the absence of spliceosomal proteins[78,79]. The core RNA architecture of both systems is quite similar (Figure 1.6), and high resolution cryo-EM structures of the spliceosome reveal a nearly identical active site architecture[80–83]. The snRNA U6 forms an internal stem loop that is almost indistinguishable from the fold of DV in group II introns. Lastly, the sequence homology and crystal structure of spliceosomal protein Prp8 indicate that this protein very likely evolved from a group II IEP[84].

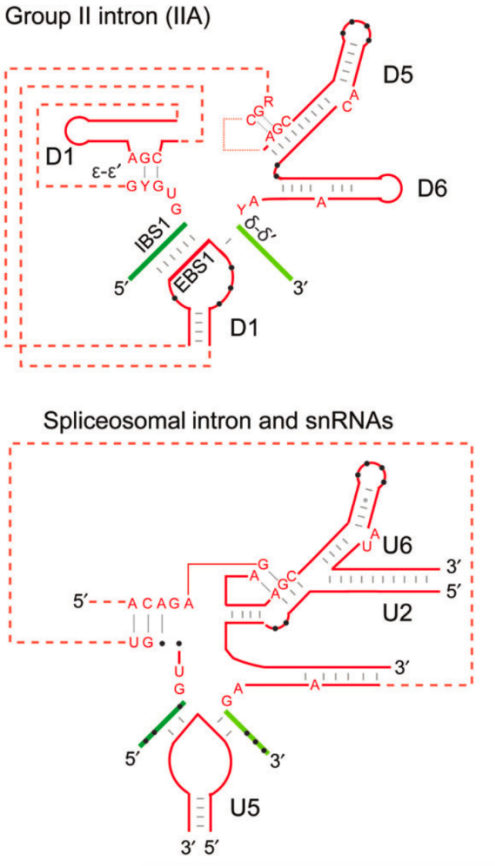

Figure 1.6: **Comparison of the group II intron ribozyme and spliceosomal snRNA structure.** (top) Group IIA intron secondary structure diagram. (bottom) Secondary structure pairing of snRNAs U2/U6 and U5. Non-analogous RNA sequences (dashed red lines) are omitted for clarity. Taken from McNeil et al. (2016)[85].

1.5 Goal of the dissertation

Though previous crystal structures of the group IIC intron from the bacterium *Oceanobacillus iheyensis* provided important structural understanding of group II intron folding and active site architecture, this intron is primitive and does not form lariat in the absence of its associated maturase protein. Therefore, the group II intron structural studies performed prior to this dissertation do not contribute to the understanding of

splicing via the branching pathway. This pathway is more relevant as it is the pathway used by the spliceosome and group II introns in their *in vivo* environments. The following chapters outline work done to elucidate the structural requirements for branching and exon ligation in lariat-forming group II introns and the spliceosome.

1.6 References

1. Woese, C. *The genetic code : the molecular basis for genetic expression.* (Harper & Row, 1967).

2. Crick, F. H. The origin of the genetic code. *J. Mol. Biol.* **38,** 367–79 (1968).

3. Orgel, L. E. Evolution of the genetic apparatus. *J. Mol. Biol.* **38,** 381–93 (1968).

4. Cech, T. R., Zaug, A. J. & Grabowski, P. J. In vitro splicing of the ribosomal RNA precursor of Tetrahymena: involvement of a guanosine nucleotide in the excision of the intervening sequence. *Cell* **27,** 487–96 (1981).

5. Kruger, K., Grabowski, P., Zaug, A., Sands, J., Gottschling, D. E., Cech, T. R. Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell* **31,** 147–57 (1982).

6. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35,** 849–57 (1983).

7. Schmelzer, C. & Schweyen, R. J. Self-splicing of group II introns in vitro: mapping of the branch point and mutational inhibition of lariat formation. *Cell* **46,** 557–65 (1986).

8. Forster, A. C. & Symons, R. H. Self-cleavage of plus and minus RNAs of a virusoid and a structural model for the active sites. *Cell* **49,** 211–20 (1987).

9. Saville, B. J. & Collins, R. A. A site-specific self-cleavage reaction performed by a novel RNA in Neurospora mitochondria. *Cell* **61,** 685–96 (1990).

10. Buzayan, J. M., Hampel, A. & Bruening, G. Nucleotide sequence and newly formed phosphodiester bond of spontaneously ligated satellite tobacco ringspot virus RNA. *Nucleic Acids Res.* **14,** 9729–43 (1986).

11. Johnston, W. K., Unrau, P. J., Lawrence, M. S., Glasner, M. E. & Bartel, D. P. RNA-Catalyzed RNA Polymerization: Accurate and General RNA-Templated Primer Extension. *Science* **292,** 1319–1325 (2001).

12. Johansen, S. & Vogt, V. M. An intron in the nuclear ribosomal DNA of Didymium iridis codes for a group I ribozyme and a novel ribozyme that cooperate in self-splicing. *Cell* **76,** 725–34 (1994).

13. Michel, F., Umesono, K. & Ozeki, H. Comparative and functional anatomy of group II catalytic introns--a review. *Gene* **82,** 5–30 (1989).

14. Halbreich, A., Pajot, P., Foucher, M., Grandchamp, C. & Slonimski, P. A pathway of cytochrome b mRNA processing in yeast mitochondria: specific splicing steps and an intron-derived circular DNA. *Cell* **19,** 321–9 (1980).

15. Berg, J. M., Tymoczko, J. L. & Stryer, L. *Biochemistry*. (W.H. Freeman, 2012).

16. Peebles, C. L., Perlman, P. S., Mecklenburg, K. L., Petrillo, M. L., Tabor, J. H., Jarrell, K. H. A self-splicing RNA excises an intron lariat. *Cell* **44,** 213–23 (1986).

17. van der Veen, R., Arnberg, A. C., van der Horst, G., Bonen, L., Tabak, H. F., Grivell, L. A. Excised group II introns in yeast mitochondria are lariats and can be formed by self-splicing in vitro. *Cell* **44,** 225–34 (1986).

18. Padgett, R. A., Podar, M., Boulanger, S. C. & Perlman, P. S. The stereochemical course of group II intron self-splicing. *Science* **266,** 1685–8 (1994).

19. Podar, M., Perlman, P. S. & Padgett, R. A. The two steps of group II intron self-splicing are mechanistically distinguishable. *RNA* **4,** 890–900 (1998).

20. Moore, M. J. & Sharp, P. A. Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. *Nature* **365,** 364–8 (1993).

21. Maschhoff, K. L. & Padgett, R. A. The stereochemical course of the first step of pre-mRNA splicing. *Nucleic Acids Res.* **21,** 5456–62 (1993).

22. Zimmerly, S., Guo, H., Perlman, P. S. & Lambowitz, A. M. Group II intron mobility occurs by target DNA-primed reverse transcription. *Cell* **82,** 545–54 (1995).

23. Zimmerly, S., Guo, H., Eskes, R., Yang, J., Perlman, P. S., Lambowitz, A. M. A group II intron RNA is a catalytic component of a DNA endonuclease involved in intron mobility. *Cell* **83,** 529–38 (1995).

24. Yang, J., Zimmerly, S., Perlman, P. S. & Lambowitz, A. M. Efficient integration of an intron RNA into double-stranded DNA by reverse splicing. *Nature* **381,** 332–5 (1996).

25. Rest, J. S. & Mindell, D. P. Retroids in Archaea: Phylogeny and Lateral Origins. *Mol. Biol. Evol.* **20,** 1134–1142 (2003).

26. Lambowitz, A. M. & Zimmerly, S. Group II Introns: Mobile Ribozymes that Invade DNA. *Cold Spring Harb. Perspect. Biol.* **3,** a003616 (2011).

27. Belfort, M., Derbyshire, V., Parker, M. M., Cousineau, B. & Lambowitz, A. M. *Mobile Introns : Pathways and Proteins*. *Mobile DNA II* (ASM Press, 2002). doi:10.1128/9781555817954.ch31

28. Lambowitz, A. M. & Zimmerly, S. Mobile group II introns. *Annu. Rev. Genet.* **38,** 1–35 (2004).

29. Toro, N., Jiménez-Zurdo, J. I. & García-Rodríguez, F. M. Bacterial group II introns: not just splicing. *FEMS Microbiol. Rev.* **31,** 342–358 (2007).

30. Vallès, Y., Halanych, K. M. & Boore, J. L. Group II introns break new boundaries: Presence in a bilaterian's genome. *PLoS One* **3,** e1488 (2008).

31. Lambowitz, A. M., Caprara, M. G., Zimmerly, S. & Perlman, P. S. *Group I and group II ribozymes as RNPs: clues to the past and guides to the future*. *The RNA World* (Cold Spring Harbor Laboratory Press, 1999).

32. Candales, M. A., Duong, A., Hood, K. S., Tony, L., Neufeld, R. A. E., Sun, R., McNeil, B. A., Li, W., Jarding, A. M., Zimmerly, S. Database for bacterial group II introns. *Nucleic Acids Res.* **40,** D187-90 (2012).

33. Costa, M. & Michel, F. Frequent use of the same tertiary motif by self-folding RNAs. *EMBO J.* **14,** 1276–85 (1995).

34. Barkan, A. in *Molecular Biology and Biotechnology of Plant Organelles: Chloroplasts and Mitochondria* (eds. Daniell, H. & Chase, C.) 295–322 (Springer Netherlands, 2004). doi:10.1007/978-1-4020-3166-3_11

35. Bonen, L. Cis- and trans-splicing of group II introns in plant mitochondria. *Mitochondrion* **8,** 26–34 (2008).

36. Barkan, A. Genome-wide analysis of RNA-protein interactions in plants. *Methods Mol. Biol.* **553,** 13–37 (2009).

37. Germain, A., Hotto, A. M., Barkan, A. & Stern, D. B. RNA processing and decay in plastids. *Wiley Interdiscip. Rev. RNA* **4,** 295–316 (2013).

38. Brown, G. G., Colas des Francs-Small, C. & Ostersetzer-Biran, O. Group II intron splicing factors in plant mitochondria. *Front. Plant Sci.* **5,** 35 (2014).

39. Toor, N., Hausner, G. & Zimmerly, S. Coevolution of group II intron RNA structures with their intron-encoded reverse transcriptases. *RNA* **7,** 1142–52 (2001).

40. Dai, L., Toor, N., Olson, R., Keeping, A. & Zimmerly, S. Database for mobile group II introns. *Nucleic Acids Res.* **31,** 424–6 (2003).

41. Toor, N., Robart, A. R., Christianson, J. & Zimmerly, S. Self-splicing of a group IIC intron: 5' exon recognition and alternative 5' splicing events implicate the stem-loop motif of a transcriptional terminator. *Nucleic Acids Res.* **34,** 6461–71 (2006).

42. Robart, A. R., Seo, W. & Zimmerly, S. Insertion of group II intron retroelements after intrinsic transcriptional terminators. *Proc. Natl. Acad. Sci.* **104,** 6620–5 (2007).

43. Zhao, C. & Pyle, A. M. Structural Insights into the Mechanism of Group II Intron Splicing. *Trends Biochem. Sci.* (2017). doi:10.1016/j.tibs.2017.03.007

44. Michel, F. & Jacquier, A. Long-range intron-exon and intron-intron pairings involved in self-splicing of class II catalytic introns. *Cold Spring Harb. Symp. Quant. Biol.* **52,** 201–12 (1987).

45. Robart, A. R., Chan, R. T., Peters, J. K., Rajashankar, K. R. & Toor, N. Crystal structure of a eukaryotic group II intron lariat. *Nature* **514,** 193–197 (2014).

46. Jacquier, A. & Michel, F. Base-pairing interactions involving the 5' and 3'-terminal nucleotides of group II self-splicing introns. *J. Mol. Biol.* **213,** 437–47 (1990).

47. Michel, F. & Ferat, J.-L. Structure and Activities of Group II Introns. *Annu. Rev. Biochem.* **64,** 435–461 (1995).

48. Fedorova, O., Waldsich, C. & Pyle, A. M. Group II intron folding under near-physiological conditions: collapsing to the near-native state. *J. Mol. Biol.* **366,** 1099–114 (2007).

49. Harris-Kerr, C. L., Zhang, M. & Peebles, C. L. The phylogenetically predicted base-pairing interaction between alpha and alpha' is required for group II splicing in vitro. *Proc. Natl. Acad. Sci. U. S. A.* **90,** 10658–62 (1993).

50. Konforti, B. B., Liu, Q. & Pyle, A. M. A map of the binding site for catalytic domain 5 in the core of a group II intron ribozyme. *EMBO J.* **17,** 7105–17 (1998).

51. Toor, N., Keating, K. S., Taylor, S. D. & Pyle, A. M. Crystal structure of a self-spliced group II intron. *Science* **320,** 77–82 (2008).

52. Koch, J. L., Boulanger, S. C., Dib-Hajj, S. D., Hebbar, S. K. & Perlman, P. S. Group II introns deleted for multiple substructures retain self-splicing activity. *Mol. Cell. Biol.* **12,** 1950–8 (1992).

53. Michels, W. J. & Pyle, A. M. Conversion of a group II intron into a new multiple-turnover ribozyme that selectively cleaves oligonucleotides: elucidation of reaction mechanism and structure/function relationships. *Biochemistry* **34,** 2965–77 (1995).

54. Chanfreau, G. & Jacquier, A. An RNA conformational change between the two chemical steps of group II self-splicing. *EMBO J.* **15,** 3466–76 (1996).

55. Costa, M., Déme, E., Jacquier, A. & Michel, F. Multiple tertiary interactions involving domain II of group II self-splicing introns. *J. Mol. Biol.* **267,** 520–36 (1997).

56. Fedorova, O. & Pyle, A. M. A conserved element that stabilizes the group II intron active site. *RNA* **14,** 1048–56 (2008).

57. Podar, M., Dib-Hajj, S. & Perlman, P. S. A UV-induced, Mg(2+)-dependent crosslink traps an active form of domain 3 of a self-splicing group II intron. *RNA* **1,** 828–40 (1995).

58. Fedorova, O., Mitros, T. & Pyle, A. M. Domains 2 and 3 interact to form critical elements of the group II intron active site. *J. Mol. Biol.* **330,** 197–209 (2003).

59. Griffin, E. A., Qin, Z., Michels, W. J. & Pyle, A. M. Group II intron ribozymes that cleave DNA and RNA linkages with similar efficiency, and lack contacts with substrate 2'-hydroxyl groups. *Chem. Biol.* **2,** 761–70 (1995).

60. Xiang, Q., Qin, P. Z., Michels, W. J., Freeland, K. & Pyle, A. M. Sequence Specificity of a Group II Intron Ribozyme: Multiple Mechanisms for Promoting Unusually High Discrimination against Mismatched Targets. *Biochemistry* **37,** 3839–3849 (1998).

61. Chanfreau, G. & Jacquier, A. Interaction of intronic boundaries is required for the second splicing step efficiency of a group II intron. *EMBO J.* **12,** 5173–80 (1993).

62. Robart, A. R., Montgomery, N. K., Smith, K. L. & Zimmerly, S. Principles of 3' splice site selection and alternative splicing for an unusual group II intron from Bacillus anthracis. *RNA* **10,** 854–62 (2004).

63. Mikheeva, S., Murray, H. L., Zhou, H., Turczyk, B. M. & Jarrell, K. A. Deletion of a conserved dinucleotide inhibits the second step of group II intron splicing. *RNA* **6,** 1509–15 (2000).

64. de Lencastre, A. & Pyle, A. M. Three essential and conserved regions of the group II intron are proximal to the 5'-splice site. *RNA* **14,** 11–24 (2008).

65. Altura, R., Rymond, B., Seraphin, B. & Rosbash, M. Sequence requirements for branch formation in a group II self-splicing intron. *Nucleic Acids Res.* **17,** 335–54 (1989).

66. Costa, M., Fontaine, J. M., Loiseaux-de Goër, S. & Michel, F. A group II self-splicing intron from the brown alga Pylaiella littoralis is active at unusually low magnesium concentrations and forms populations of molecules with a uniform conformation. *J. Mol. Biol.* **274,** 353–64 (1997).

67. Liu, Q., Green, J. B., Khodadadi, A., Haeberli, P., Beigelman, L., Pyle, A. M.. Branch-site selection in a group II intron mediated by active recognition of the adenine amino group and steric exclusion of non-adenine functionalities. *J. Mol. Biol.* **267,** 163–71 (1997).

68. Chu, V. T., Adamidi, C., Liu, Q., Perlman, P. S. & Pyle, A. M. Control of branch-site choice by a group II intron. *EMBO J.* **20,** 6866–76 (2001).

69. Jacquier, A. & Jacquesson-Breuleux, N. Splice site selection and role of the lariat in a group II intron. *J. Mol. Biol.* **219,** 415–28 (1991).

70. Jarrell, K. A., Dietrich, R. C. & Perlman, P. S. Group II intron domain 5 facilitates a trans-splicing reaction. *Mol. Cell. Biol.* **8,** 2361–6 (1988).

71. Belhocine, K., Mak, A. B. & Cousineau, B. Trans-splicing of the Ll.LtrB group II intron in Lactococcus lactis. *Nucleic Acids Res.* **35,** 2257–2268 (2007).

72. Belhocine, K., Mak, A. B. & Cousineau, B. Trans-splicing versatility of the Ll.LtrB group II intron. *RNA* **14,** 1782–1790 (2008).

73. Bonen, L. Trans-splicing of pre-mRNA in plants, animals, and protists. *FASEB J.* **7,** 40–6 (1993).

74. Wahl, M. C., Will, C. L. & Lührmann, R. The spliceosome: design principles of a dynamic RNP machine. *Cell* **136,** 701–18 (2009).

75. Valadkhan, S. & Jaladat, Y. The spliceosomal proteome: at the heart of the largest cellular ribonucleoprotein machine. *Proteomics* **10,** 4128–41 (2010).

76. Yean, S.-L., Wuenschell, G., Termini, J. & Lin, R.-J. Metal-ion coordination by U6 small nuclear RNA contributes to catalysis in the spliceosome. *Nature* **408,** 881–884 (2000).

77. Fica, S. M., Tuttle, N., Noval, T., Li, N.-S., Lu, J., Koodathingal, P., Dai, Q., Staley, J. P., Piccirilli, J. A. RNA catalyses nuclear pre-mRNA splicing. *Nature* **503,** 229–34 (2013).

78. Valadkhan, S., Mohammadi, A., Wachtel, C. & Manley, J. L. Protein-free spliceosomal snRNAs catalyze a reaction that resembles the first step of splicing. *RNA* **13,** 2300–11 (2007).

79. Valadkhan, S., Mohammadi, A., Jaladat, Y. & Geisler, S. Protein-free small nuclear RNAs catalyze a two-step splicing reaction. *Proc. Natl. Acad. Sci.* **106,** 11901–6 (2009).

80. Keating, K. S., Toor, N., Perlman, P. S. & Pyle, A. M. A structural analysis of the group II intron active site and implications for the spliceosome. *RNA* **16,** 1–9 (2010).

81. Yan, C., Hang, J., Wan, R., Huang, M, Wong, C. L., Shi, Y. Structure of a yeast spliceosome at 3.6-angstrom resolution. *Science* **349,** 1182–91 (2015).

82. Galej, W. P., Wilkinson, M. E., Fica, S. M, Oubridge, C., Newman, A. J., Nagai, K. Cryo-EM structure of the spliceosome immediately after branching. *Nature* **537,** 197–201 (2016).

83. Yan, C., Wan, R., Bai, R., Huang, G. & Shi, Y. Structure of a yeast step II catalytically activated spliceosome. *Science* **355,** 149–155 (2017).

84. Galej, W. P., Oubridge, C., Newman, A. J. & Nagai, K. Crystal structure of Prp8 reveals active site cavity of the spliceosome. *Nature* **493,** 638–43 (2013).

85. McNeil, B. A., Semper, C. & Zimmerly, S. Group II introns: versatile ribozymes and retroelements. *WIREs RNA* **7,** 341–55 (2016).

Chapter 2: Determination of the U2/U6 snRNA crystal structure

2.1 Abstract

The spliceosome is the eukaryotic ribonucleoprotein (RNP) complex responsible for the removal, or splicing, or introns from pre-mRNA to form mature mRNA. Virtually all genes undergo alternative splicing to produce multiple protein isoforms, thus increasing coding capacity. The complexity of the spliceosome (5snRNAs and over 150 associated proteins) makes it difficult to study structurally. U2 and U6, which are highly conserved, are thought to form the active site because they are the only snRNAs required for both steps of splicing, and they are capable of catalysis in the absence of any proteins. The active site of group II introns is composed of conserved RNA sequences equivalent to U2 and U6, and they both share a similar mechanism of splicing. Crystallization experiments of a U2-U6 snRNA complex have resulted in crystals that diffract to a resolution of 1.8–2.9 Å. Heavy metal soaks with $YbCl_3$ reveal two anomalous peaks ~4.3 Å apart, which is indicative of a two-metal ion mechanism of catalysis.

2.2 Introduction

The spliceosome, an incredibly dynamic ribonucleoprotein (RNP), removes introns from flanking exon sequences through two sequential transesterification reactions (branching and exon ligation). This process is exceedingly complex, involving 5 snRNAs and up to 150 associated proteins in humans[1,2]. To make matters more

complex, both branching and exon ligation take place within a single active site, necessitating a rearrangement of products and reactants between the reactions. How the spliceosome is able to sense the progression of splicing and induce the correct structural rearrangements remains unclear. Although proteins play an important role both structurally and in promoting catalysis[3,4], the catalytic core itself is composed of the snRNAs U2 and U6[5–8]. In fact, it has been shown that these two snRNAs are capable of catalyzing both steps of splicing *in vitro* in the absence of spliceosomal proteins[9–11].

Until recently, the complex and dynamic nature of the spliceosome has made it very difficult to study structurally as a whole. The heterogeneity of splicing complexes makes it a bad candidate for crystallography, and its multi-megadalton size puts it well over the upper limit for NMR. Recent advances in cryo-electron microscopy (cryo-EM) technology have resulted in a so-called "Resolution Revolution[12]" where single-particle cryo-EM has broken the barrier into atomic resolution with macromolecules >200 kDa[13]. Due to these advances in detector technology and data processing techniques, there has been a recent outpouring of spliceosome structures determined via cryo-EM[14–23]. These structures have shown that the RNA organization of the active site of the spliceosome is nearly identical to that of group II introns, which are catalytic RNAs that are capable of splicing in the absence of proteins[24]. The mechanism of splicing is also nearly identical in both systems in which the nucleophile for the first step is the 2′ hydroxyl of a conserved adenosine[25] (Figure 2.1). In group II introns, the active site, commonly referred to as the catalytic triplex, is composed of consecutive base triples involving nucleotides from the catalytic triad and the two-nucleotide bulge from domain

V, and the junctions J2/3 and J4/5. The formation of these base triples positions the non-bridging oxygens in the two-nucleotide bulge backbone as ligands for two Mg^{2+} to bind. Metal rescue, crosslinking and biochemistry experiments performed in yeast reveal the presence of base triples in the spliceosome analogous to those found in the active site of group II introns[26]. The formation of these base triples promotes catalytic metal binding, catalyzing both steps of splicing.

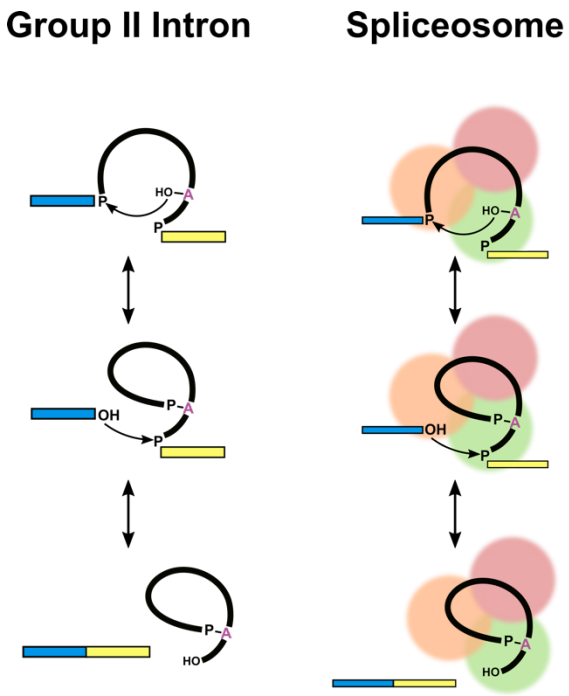

Figure 2.1: **Splicing mechanism of group II introns and the spliceosome.** In both systems, the 2′ hydroxyl of a conserved adenosine within the intron attacks the 5′ splice site to form a branched intron with a unique 2′-5′ covalent bond. In the next reaction, the 3′ hydroxyl of the 5′ exon attacks the 3′ splice site, resulting in ligated exons and a free lariat intron.

The structural and functional similarities between these systems led to the prediction that the U2/U6 active site scaffold also contains base triples that act as a

platform for catalytic Mg^{2+} binding. However, metal ion identity and location in the recent cryo-EM spliceosome structures remains ambiguous, with each structure labelling different pockets of density as the catalytic metals. Improvements in cryo-EM technology have resulted the ability to push the limits of structure determination, but cryo-EM still has its disadvantages. Most notably, metal ion identity cannot currently be determined using electron microscopy. Using x-ray crystallography, however, the precise location and identity of metal ions can be determined from heavy metal anomalous scattering.

To understand the precise mechanism of the two transesterification reactions, we crystallized a U2-U6 snRNA complex. Soaking crystals with $YbCl_3$ revealed two anomalous peaks ~4.3 Å apart. This distance is approximately the distance observed in enzymes and catalytic RNAs utilizing a two metal ion mechanism of catalysis[27,28].

2.3 Crystallization of the U2/U6 snRNA complex

Initial efforts to obtain the structure of the U2/U6 snRNA duplex were based off of small angle x-ray scattering (SAXS) and NMR results published by the laboratory of Sam Butcher[29]. In 2012, they published the solution structure of a truncated U2/U6 duplex using SAXS, NMR, and molecular modeling. The U2/U6 complex contains a three-helix junction and forms an extended "Y" shape. This indicated that their truncated U2/U6 complex was an ideal candidate for structure determination using x-ray crystallography.

The U2/U6 complex used by Burke et al. 2012 was modified to contain a GAAA tetralooop to cap helix II and its high affinity receptor proximal to helix III (Figure 2.2) to facilitate intermolecular interactions. Helix II and III lengths were varied and each RNA was screened against the 48-condition Natrix screen from Hampton Research. Impressively, crystals of one of the first RNA constructs tested (U2/U6 wt) grew in Natrix 13 as a cluster of fine needles (Figure 2.3a) that diffracted to ~3.4 Å. To improve resolution and promote single crystal growth, U2/U6 wt RNA in Natrix 13 was screened against the additive screen from Hampton Research. The addition of 1% PEG 3350 improved resolution to ~1.85 Å. Although the crystals grew in longer, they were still clustered together, obstructing the ability to loop a single crystal (Figure 2.3b). Serial seeding[30,31] produced single crystals of the same resolution (Figure 2.3c). These crystals stood up to radiation damage well and one or more full datasets were able to be collected on a single crystal.

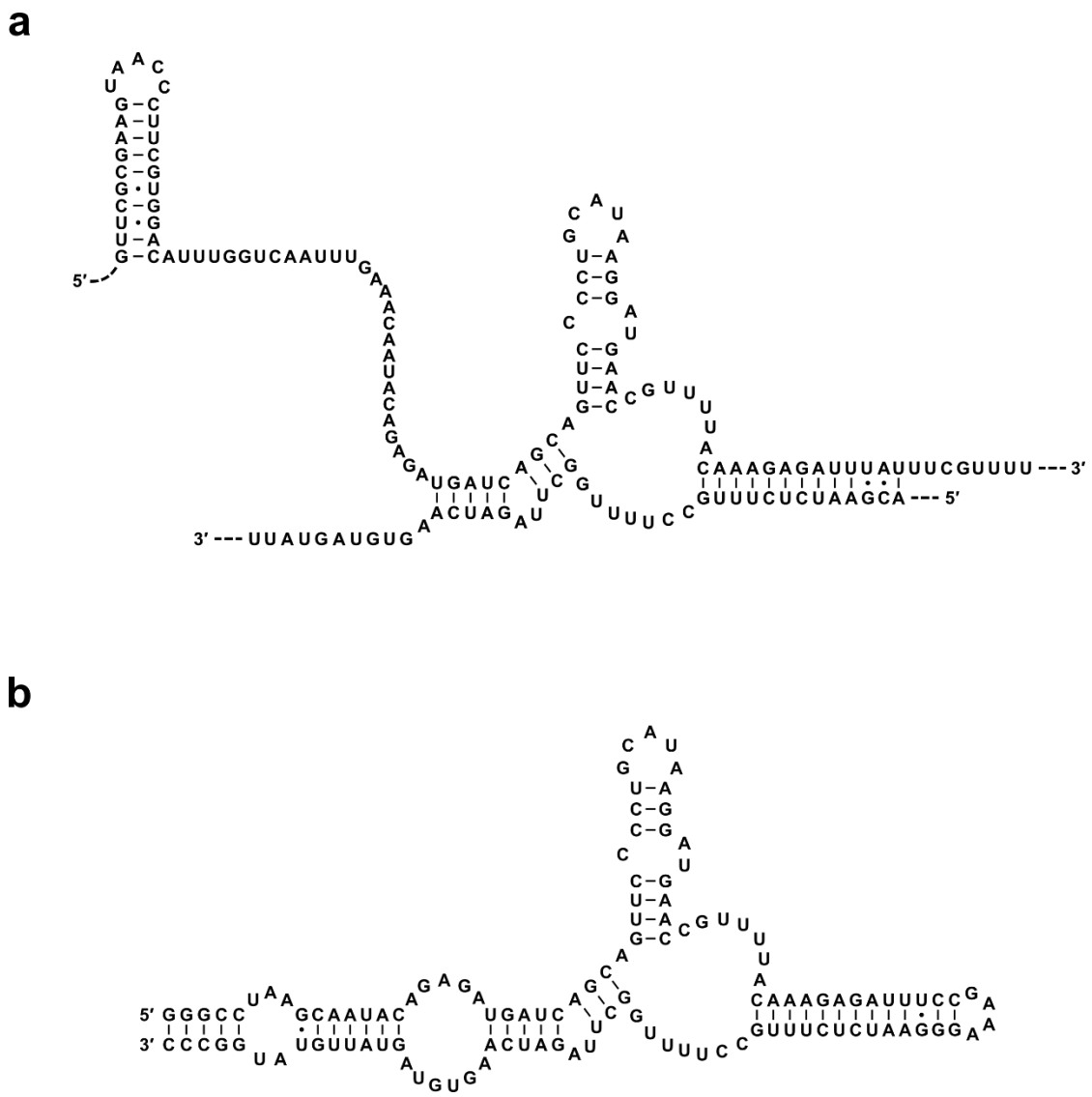

Figure 2.2: **Secondary structure of the U2/U6 snRNA complex from Saccharomyces cerevisiae.** (a) Core U2/U6 duplex sequences. Extraneous non-pairing sequences are omitted for clarity. (b) Truncated U2/U6 complex used for structure determination.

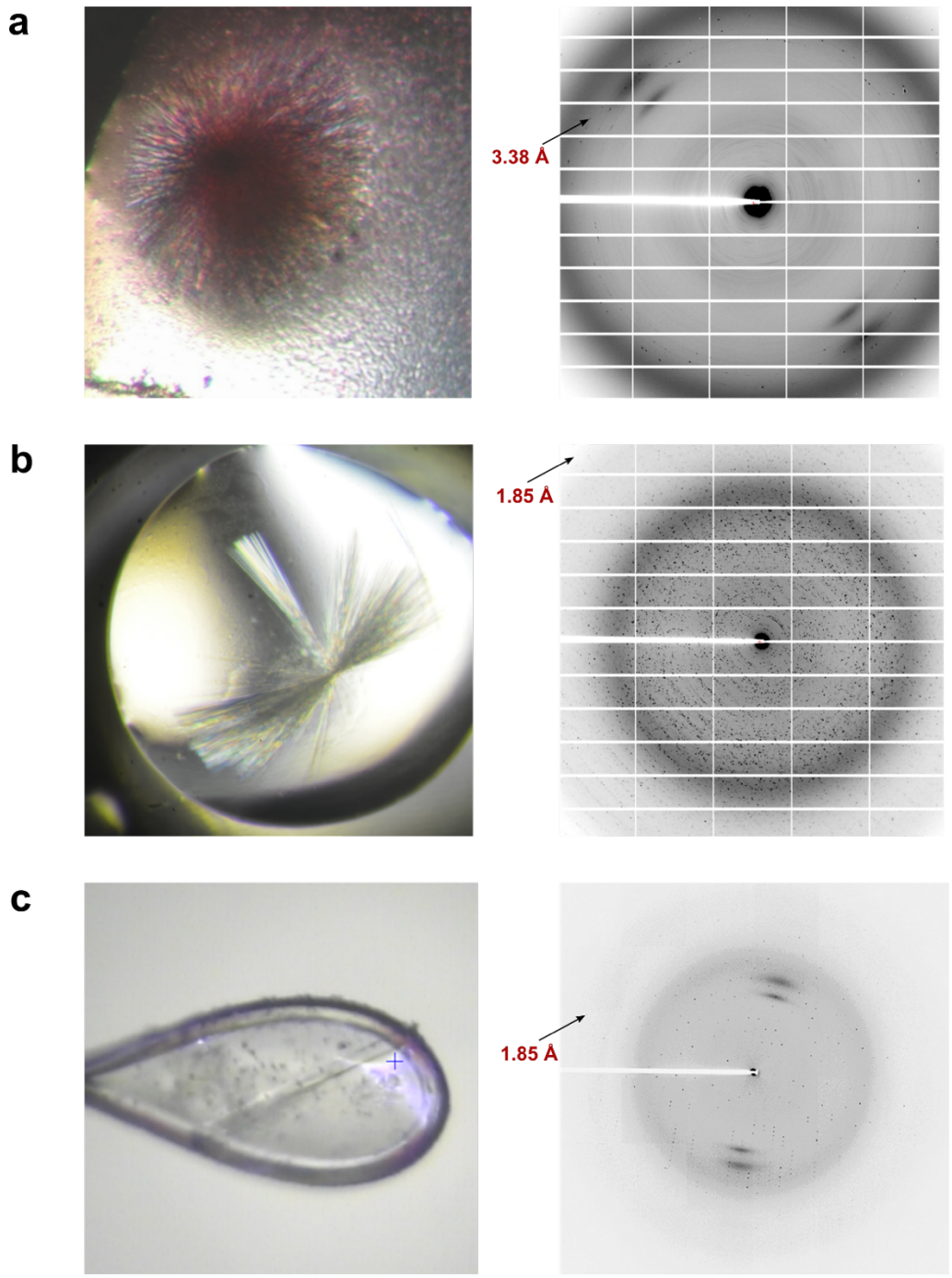

Figure 2.3: **U2/U6 wt crystals and diffraction images.** (a) Preliminary crystals grown in 1.7 M $(NH_4)_2SO_4$. The resulting diffraction image shows rings instead of single spots because cyrstals grew in large clumps of needles. (b) Cyrstals resulting from using 1% PEG 3350 as an additive. Fewer needles formed from a single nucleation point resulting in many overlapping diffraction patterns. (c) Single crystal nucleated via seeding suspended in a nylon loop. The resulting single diffraction pattern extends to 1.85 Å.

Performing molecular replacement using the solution structures of the U2/U6 complex did not yield any promising results, nor did performing molecular replacement with smaller A-form RNA fragments as laid out by William Scott[32]. Therefore, U2/U6 wt + 1% PEG 3350 crystals were soaked with $YbCl_3$ to obtain initial phase information. $YbCl_3$ is commonly used as an anomalous scatterer to experimentally determine initial phases in RNA crystals. It exhibits the same octahedral coordination geometry as Mg^{2+} and preferentially binds to sites containing highly coordinated magnesium ions[33,34]. $YbCl_3$ soaks revealed two anomalous peaks ~4.3 Å apart (Figure 2.4). It is well known that 4 Å is the preferred distance between metals in a two-metal ion mechanism[28]. Hence, two anomalous peaks at such a distance suggests that the U2/U6 complex folds in a functionally relevant way to properly position the catalytic triplex to bind two Mg^{2+}.

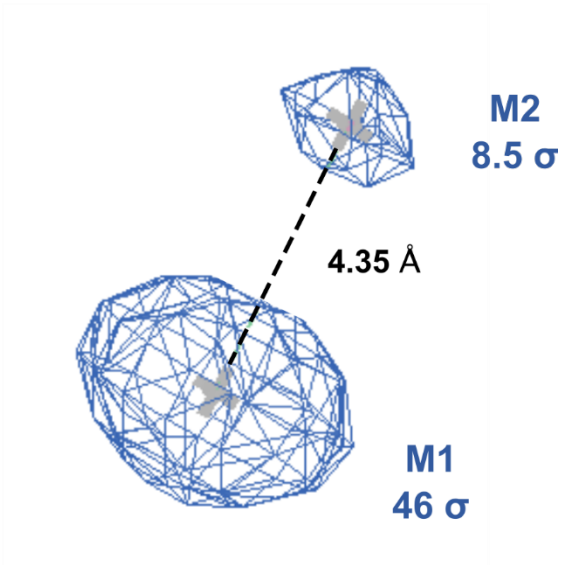

Figure 2.4: **Yb^{3+} anomalous density from U2/U6 crystals.** The Yb^{3+} anomalous map (blue mesh) reveals two large peaks 4.3 Å apart. 4 Å is the preferred distance for a two-metal mechanism[28].

Multiwavelength Anomalous Dispersion (MAD) was used to obtain density for model building. The density appeared to contain sections resembling A-form helices crossing over one another with clear density for base pairs (Figure 2.5). Surprisingly, distance between the density resembling stacked base pairs was too close to accommodate a proper RNA backbone. The defective nature of this density is likely due to reticular twinning[35].

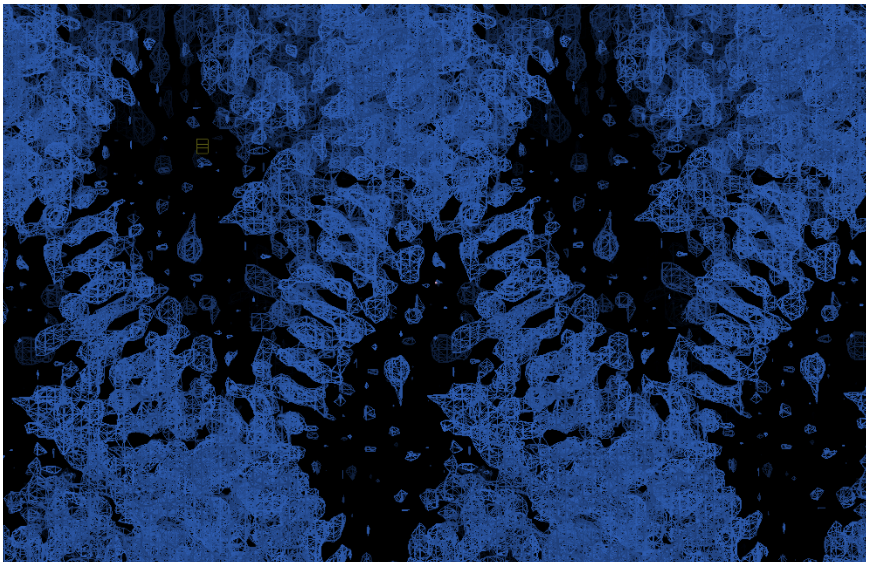

Figure 2.5: **Initial density obtained from multi-wavelength anomalous dispersion (MAD) of U2/U6 crystals soaked in $YbCl_3$.** The preliminary density resembles a crisscross pattern of nucleic acid helices.

2.4 U2/U6 embedded in group II introns

In addition to crystallizing the free U2/U6 RNA complex, I performed crystallization trials of a construct containing the U2/U6 complex embedded in domain III of the group II intron, *Oceanobacillus iheyensis* (*O.i.*). The crystal structure of this

group II intron had been previously determined in 2008 by Dr. Navtej Toor[33]. In that structure, it is evident that domain III contains a stem loop that is not making any intramolecular contacts and is extending out from the intron, making it a good candidate for chimeric engineering for structure determination. I engineered a construct in which the U2/U6 complex is inserted into domain III of *O.i.* (Figure 2.6). Domain II and IV stem lengths were varied and constructs were screened against the 48-condition Natrix screen from Hampton Research. One construct produced crystals that diffracted to ~6 Å, but attempts to improve resolution were unsuccessful.

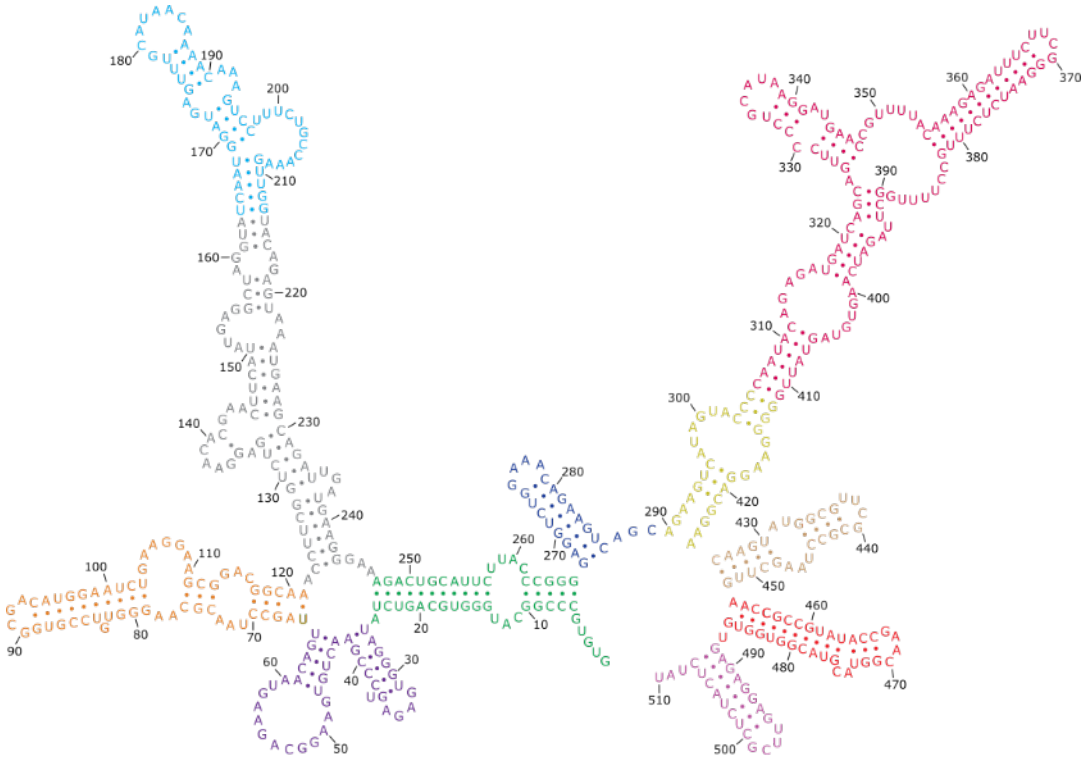

Figure 2.6: **Secondary structure showing U2/U6 embedded in domain III of the *Oceanobacillus iheyensis* (*O.i.*) group II intron.** Domain coloring of *O.i.* is consistent with the coloring of the *P.li.LSUI2* intron shown in Fig. 5.3 except that the U2/U6 complex is shown in pink.

2.5 The use of protein chaperones to assist in crystallization

In an effort to obtain high resolution crystals that are not twinned, we turned to the use of protein chaperones[36,37]. Even stable, conformationally homogeneous RNAs may have difficulty creating well-ordered crystals. This is most likely due to the fact that the surface of RNA molecules is composed of a regular arrangement of negatively charged phosphates. In the absence of well-defined intermolecular RNA-RNA tertiary interactions, it is likely that the RNA molecules may pack in an inconsistent way,

making subtly different intermolecular contacts. Crystallographers have engineered specific RNA-RNA[38] and RNA-protein[39,40] interactions into RNA constructs to differentiate the RNA surface landscape to facilitate more precise crystal contacts.

Several RNA-binding proteins have been used to aid the crystallization of RNA including the human spliceosomal protein U1A[40] and recombinant antibody fragments (Fabs)[41,42]. U1A has been most frequently used as a crystallization module because of its extensive characterization[40,43], high binding affinity (~20 pM)[43], and relatively small RNA-binding sequence (10 nucleotide loop)[44]. We engineered the U1A binding sequence into the end of helix II (Figure 2.7) and bound to recombinantly expressed U1A. An electrophoretic mobility shift assay confirmed RNP assembly (Figure 2.8a). There is no observable difference between a U1A:RNA ratio of 0.5:1 and 1.5:1 suggesting that either U1A binding becomes saturated at ~50% of RNA bound or that the concentration determined for U1A is incorrect. A U1A:RNA ratio of 2:1 was used to prepare RNP for crystallization trials and the excess, unbound U1A was removed by filtration. Initial crystallization trials resulted in crystals that grew in Index 7 (Hampton Research) and diffracted to ~2.8 Å. Unfortunately, these crystals exhibited signs of pseudomerohedral twinning and further attempts to eliminate twinning, as had been previously effective for crystals of the enzyme Dicer[45], were unsuccessful.

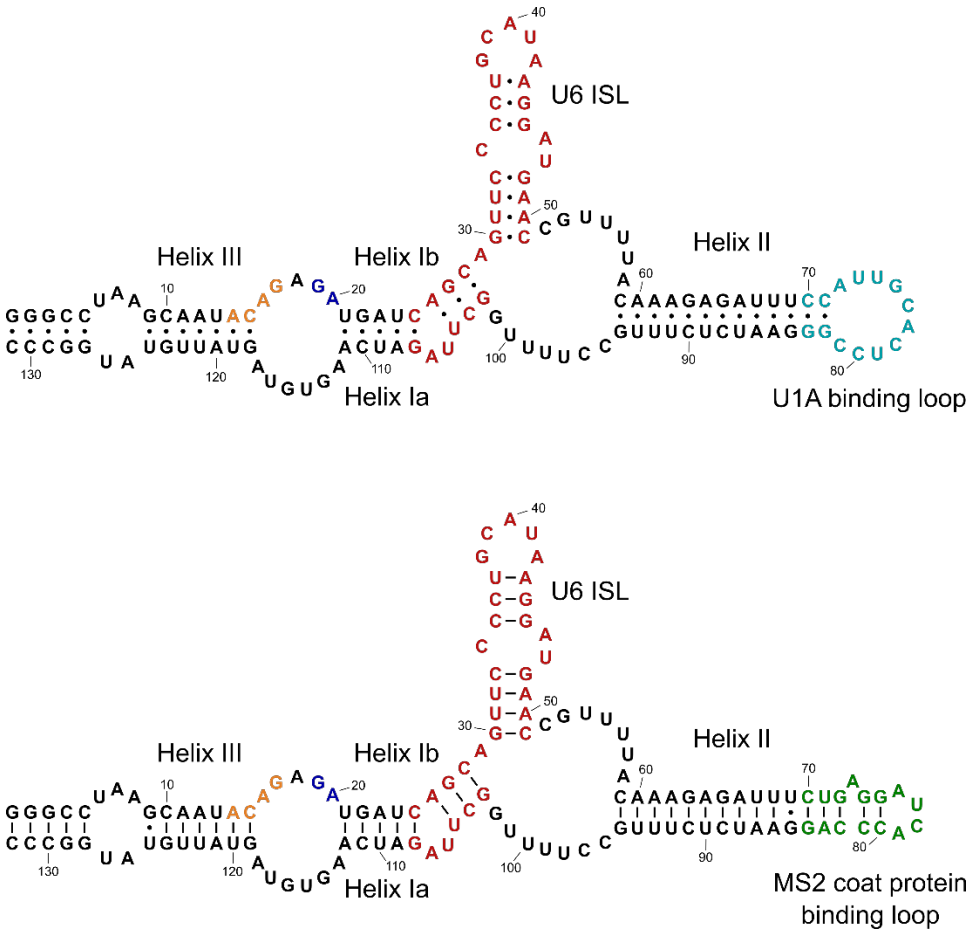

Figure 2.7: **U2/U6 snRNA secondary structures with U1A and MS2 coat protein binding sequences**.

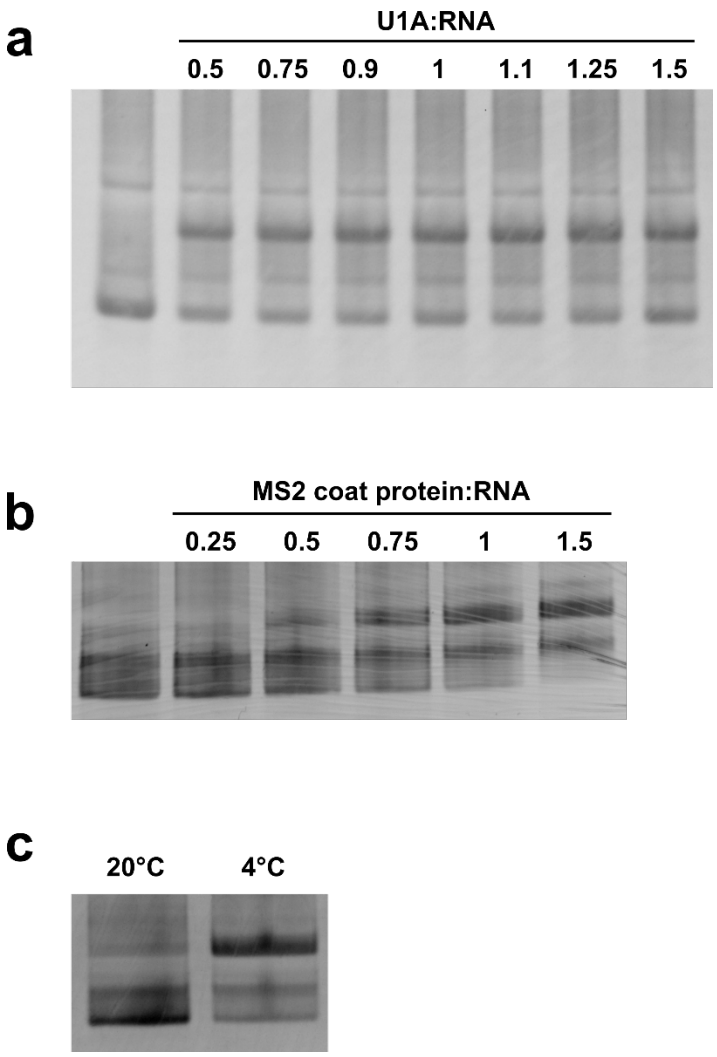

Figure 2.8: **Representative U1A and MS2 EMSA.** (a) Native PAGE showing an electrophoretic mobility shift of the U2/U6 complex upon binding U1A. There is no observable difference between a U1A:RNA ratio of 0.5:1 and 1.5:1. (b) Native PAGE showing an electrophoretic mobility shift of the U2/U6 complex upon binding the MS2 coat protein. Protein:RNA ratio is calculated for the MS2 coat protein dimer. (c) Native PAGE showing U2/U6-MS2 RNP assembly at room temperature vs. 4^{o}C. Each sample was assembled with two-fold molar excess of the MS2 coat protein and incubated at the indicated temperature for 24 hours.

The MS2 coat protein is another potential crystallization module. Each MS2 coat protein is ~14 kDa and only the dimer is capable of binding RNA. The larger size (~28

kDa MS2 coat protein dimer vs. 10.9 kDa U1A) and pM binding affinity to a15-nt stem loop[46,47] make it a promising candidate as a crystallization aide. Currently, the MS2 coat protein is frequently used for the biochemical purification of RNA-protein complexes[48] and the detection of RNA in living cells[49] because of its tight binding to the cognate RNA sequence. I wondered if the MS2 coat protein's tight binding and larger size would make it useful as a crystallization chaperone. The U2/U6 snRNA complex was engineered to contain the high affinity MS2 hairpin in helix II (Figure 2.7). Interestingly, the protein was only found to bind RNA at 4^{o}C. RNP formation was only observed at room temperature for short (< 1 hour) incubation times (Figure 2.8b).

2.6 Methods

In vitro transcription

U2/U6 constructs with variable helix II and III stem lengths and loop sequences were made via overlapping PCR to contain a HindIII restriction site at the 3′ end and cloned into pUC57 using the EvoRV cute site. Plasmids containing the U2/U6 sequence followed by a HindIII restriction site were linearized prior to *in* vitro transcription. Run-off transcription was performed overnight at 37°C using 0.04 mgml^{-1} linearized plasmid in 25 mM $MgCl_2$, 2 mM spermidine, 5 mM DTT, 40 mM Tris-HCl pH 7.5, 0.05% Triton X-100, 2.5 mM of each NTP, T7 RNA polymerase purified in house and thermostable inorganic pyrophosphatase (New England Biolabs). $CaCl_2$ was added to a final concentration of 1.2 mM. The reaction mixture was treated with turbo DNase (Life Technologies) for 45 min followed by proteinase K (Life Technologies) for 1 hr. U2/U6

RNA was washed sufficiently with 10 mM $MgCl_2$, 5 mM sodium cacodylate pH 6.5 using an Amicon Ultra-15 centrifugal filter with a 30 kDa molecular weight cut off and concentrated to 10 $mgml^{-1}$. For U2/U6-U1A and U2/U6-MS2 constructs, RNA was incubated with 2 molar equivalents of either U1A or MS2 coat protein in 10 mM $MgCl_2$, 5 mM sodium cacodylate pH 6.5 at room temperature for 3 hours then concentrated to 10 $mgml^{-1}$.

Crystallization

Initial crystallization trials were performed using an Art Robbins crystallization robot to set sitting drops against the Natrix, Index, and Crystal screens from Hampton Research. Equal volumes of RNA (10 $mgml^{-1}$) were mixed with 0.5 mM spermine and the precipitant solution. In the case of trays set with the additive screen, an equal volume of each additive was added to the RNA/spermine/precipitant mixture. Crystals were exchanged into solutions containing 0.5 mM spermine, 15 mM $Mg(Oac)_2$, 50 mM sodium cacodylate pH 6.0, 1% PEG 3,350, saturated $(NH_4)_2SO_4$ both without and with 15 mM $YbCl_3$ (U2/U6 alone); or 0.5 mM spermine, 100 mM NH_4OAc, 20 mM $MgCl_2$, 50 mM HEPES pH 7.0, 6% PEG 8,000, and 32% MPD (Oi-U2U6); or 0.5 mM spermine, 100 mM citric acid pH 3.5, 3.6 M NaCl (U2U6-U1A).

Seeding

Initial crystals were grown in sitting drops by means of vapor diffusion at 30°C using 10 $mgml^{-1}$ RNA, 0.5 mM spermine, 15 mM $Mg(Oac)_2$, 50 mM sodium cacodylate pH 6.0, 1% PEG 3,350, and 1.7 M $(NH_4)_2SO_4$. Clusters of needles appeared within 16-48 hours and were transferred to a stabilization solution containing 15 mM $Mg(Oac)_2$,

50 mM sodium cacodylate pH 6.0, 1% PEG 3,350, and saturated $(NH_4)_2SO_4$. Crystals were homogenized using ceramic Seed Beads (Hampton Research) and streaked into drops containing 6 $mgml^{-1}$ RNA equilibrated against 15 mM $Mg(Oac)_2$, 50 mM sodium cacodylate pH 6.0, 1% PEG 3,350, and 1.7 M $(NH_4)_2SO_4$ at 30°C. Single needle crystals appeared within 16-48 hours.

Multiwavelength Anomalous Dispersion

Crystals were either grown or soaked (24 hrs) in 5 mM ytterbium (III) chloride. The high salt concentrations in the crystallization conditions necessitated use of a higher ytterbium (III) chloride concentration to permeate the crystals. X-ray data sets were collected at NE-CAT's 24-ID-C beamline at the Advanced Photon Source (Argonne National Laboratory, Argonne, Illinois). Data was processed using HKL-2000[50], heavy metal sites were identified with SHELXD[51] and phasing done using Phenix[52] and SHELXE[53]. RNA nucleotides were modeled using COOT[54] and the RCrane plugin[55]. The phylogenetically predicted secondary structure (Figure 5.1) guided modeling into the electron density. Structure refinement was done using Buster[56], Phenix[52], DEN[57], and Phenix.Erasser[58]. All software was compiled by SBGrid[59].

Electrophoretic mobility shift assays

U2/U6-U1A samples were prepared by incubating 10 μg RNA with 0.1-1.5 molar equivalents of U1A in 15 μL of 10 mM $MgCl_2$, 5 mM sodium cacodylate pH 6.5 at 37°C for 30 min. Samples were diluted to 0.1 $\mu g \mu L^{-1}$ with loading buffer and 0.75 μg was loaded per lane on a 5% 37.5:1 acryalmide:bisacrylamide, 4% glycerol, 0.5X TBE gel. U2/U6-MS2 samples were prepared by incubating 1 μg RNA with 0.25-1.5 molar

equivalents of MS2 coat protein in 10 mM $MgCl_2$, 100 mM NaCl, and 50 mM HEPES pH 7.5 at room temperature for 1hr (Figure 2.8b), room temperature for 5 hrs or 4°C for 32 hours. Samples were run on 5% 37.5:1 acrylamide:bisacrylamide, 4% glycerol, 0.5X TBE gels. All gels were run at a constant 5 watts at 4°C.

2.7 Acknowledgements

We thank Kanagalaghatta Rajashankar and the staff of the NE-CAT beamlines at the Advanced Photon Source (APS) of Argonne National Laboratory for their beamline support and innumerable helpful discussions regarding data collection on small crystals and crystal packing. We thank Aaron Robart for his mentorship and advice regarding crystallization methods. We also thank Juno the cat for her gracious donation of whiskers that were used to seed crystals. J.K.P. was supported by the UCSD Molecular Biophysics Training Program funded by NIH predoctoral training grant 5T32GM008326. This work was supported by NIH grant 5R01GM102216 awarded to N.T. NE-CAT is supported by NIH grant 8P41GM103403-10 and APS is supported by the U.S. DOE under Contract No. DE-AC02-06CH11357.

2.8 References

1. Wahl, M. C., Will, C. L. & Lührmann, R. The spliceosome: design principles of a dynamic RNP machine. *Cell* **136,** 701–18 (2009).

2. Valadkhan, S. & Jaladat, Y. The spliceosomal proteome: at the heart of the largest cellular ribonucleoprotein machine. *Proteomics* **10,** 4128–41 (2010).

3. Rasche, N., Dybkov, O., Schmitzová, J., Akyildiz, B., Fabrizio, P., Lührman, R.. Cwc2 and its human homologue RBM22 promote an active conformation of the spliceosome catalytic centre. *EMBO J.* **31,** 1591–604 (2012).

4. Galej, W. P., Oubridge, C., Newman, A. J. & Nagai, K. Crystal structure of Prp8 reveals active site cavity of the spliceosome. *Nature* **493,** 638–43 (2013).

5. Fica, S. M., Tuttle, N., Novak, T., Li, N.-S., Lu, J., Koodathingal, P., Dai, Q., Staley, J. P., Piccirilli, J. A. RNA catalyses nuclear pre-mRNA splicing. *Nature* **503,** 229–34 (2013).

6. Hausner, T. P., Giglio, L. M. & Weiner, A. M. Evidence for base-pairing between mammalian U2 and U6 small nuclear ribonucleoprotein particles. *Genes Dev.* **4,** 2146–56 (1990).

7. Datta, B. & Weiner, A. M. Genetic evidence for base pairing between U2 and U6 snRNA in mammalian mRNA splicing. *Nature* **352,** 821–824 (1991).

8. Madhani, H. D. & Guthrie, C. A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. *Cell* **71,** 803–17 (1992).

9. Valadkhan, S., Mohammadi, A., Jaladat, Y. & Geisler, S. Protein-free small nuclear RNAs catalyze a two-step splicing reaction. *Proc. Natl. Acad. Sci.* **106,** 11901–6 (2009).

10. Valadkhan, S., Mohammadi, A., Wachtel, C. & Manley, J. L. Protein-free spliceosomal snRNAs catalyze a reaction that resembles the first step of splicing. *RNA* **13,** 2300–11 (2007).

11. Jaladat, Y., Zhang, B., Mohammadi, A. & Valadkhan, S. Splicing of an intervening sequence by protein-free human snRNAs. *RNA Biol.* **8,** 372–7 (2011).

12. Kühlbrandt, W. The Resolution Revolution. *Science* **343,** (2014).

13. Merk, A., Bartesaghi, A., Banjeree, S., Falconieri, V., Rao, P., Davis, M. I., Pragani, R., Boxer, M. B., Earl, L. A., Milne, J. L. S., Subramaniam, S. Breaking Cryo-EM Resolution Barriers to Facilitate Drug Discovery. *Cell* **165,** 1698–707 (2016).

14. Yan, C., Hang, J., Wan, R., Huang, M., Wong, C. L., Shi, Y. Structure of a yeast spliceosome at 3.6-angstrom resolution. *Science* **349,** 1182–91 (2015).

15. Boesler, C., Rigo, N., Agafonov, D. E., Kastner, B., Urlaub, H., Will, C. L., Lührmann, R. Stable tri-snRNP integration is accompanied by a major structural rearrangement of the spliceosome that is dependent on Prp8 interaction with the 5' splice site. *RNA* **21,** 1993–2005 (2015).

16. Galej, W. P., Wilkinson, M. E., Fica, S. M, Oubridge, C., Newman, A. J., Nagai, K. Cryo-EM structure of the spliceosome immediately after branching. *Nature* **537,** 197–201 (2016).

17. Wan, R., Yan, C., Bai, R., Wang, L., Huang, M., Wong, C. L., Shi, Y. The 3.8 Å structure of the U4/U6.U5 tri-snRNP: Insights into spliceosome assembly and catalysis. *Science* **351,** 466–75 (2016).

18. Wan, R., Yan, C., Bai, R., Huang, G. & Shi, Y. Structure of a yeast catalytic step I spliceosome at 3.4 Å resolution. *Science* **353,** 895–904 (2016).

19. Yan, C., Wan, R., Bai, R., Huang, G. & Shi, Y. Structure of a yeast activated spliceosome at 3.5 Å resolution. *Science* **353,** 904–11 (2016).

20. Rauhut, R., Fabrizio, P., Dybkov, O., Hartmuth, K., Pena, V., Chari, A., Kumar, V., Lee, C.-T., Urlaub, H., Kastner, B., Stark, H., Lührmann, R. Molecular architecture of the Saccharomyces cerevisiae activated spliceosome. *Science* **353,** 1399–1405 (2016).

21. Yan, C., Wan, R., Bai, R., Huang, G. & Shi, Y. Structure of a yeast step II catalytically activated spliceosome. *Science* **355,** 149–155 (2017).

22. Fica, S. M., Oubridge, C., Galej, W. P., Wilkinson, M. E., Bai, X.-C., Newman, A. J., Nagai, K. Structure of a spliceosome remodelled for exon ligation. *Nature* **542,** 377–380 (2017).

23. Bertram, K., Agafonov, D. E., Liu, W.-T., Dybkov, O., Will, C. L., Hartmuth, K., Urlaub, H., Kastner, B., Stark, H., Lührmann, R. Cryo-EM structure of a human spliceosome activated for step 2 of splicing. *Nature* **542,** 318–323 (2017).

24. Bonen, L. & Vogel, J. The ins and outs of group II introns. *Trends Genet.* **17,** 322–31 (2001).

25. Cech, T. R. The generality of self-splicing RNA: relationship to nuclear mRNA splicing. *Cell* **44,** 207–10 (1986).

26. Fica, S. M., Mefford, M. A., Piccirilli, J. A. & Staley, J. P. Evidence for a group II intron-like catalytic triplex in the spliceosome. *Nat. Struct. Mol. Biol.* **21,** 464–71 (2014).

27. Yang, W., Lee, J. Y. & Nowotny, M. Making and breaking nucleic acids: two-Mg2+-ion catalysis and substrate specificity. *Mol. Cell* **22,** 5–13 (2006).

28. Steitz, T. A. & Steitz, J. A. A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl. Acad. Sci. U. S. A.* **90,** 6498–502 (1993).

29. Burke, J. E., Sashital, D. G., Zuo, X., Wang, Y.-X. & Butcher, S. E. Structure of the yeast U2/U6 snRNA complex. *RNA* **18,** 673–83 (2012).

30. Bergfors, T. Seeds to crystals. *J. Struct. Biol.* **142,** 66–76 (2003).

31. Zhu, D.-Y., Zhu, Y.-Q., Xiang, Y., Wang, D.-C. Optimizing protein crystal growth through dynamic seeding. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **61,** 772–775 (2005).

32. Robertson, M. P., Chi, Y.-I. & Scott, W. G. Solving novel RNA structures using only secondary structural fragments. *Methods* **52,** 168–72 (2010).

33. Toor, N., Keating, K. S., Taylor, S. D. & Pyle, A. M. Crystal structure of a self-spliced group II intron. *Science* **320,** 77–82 (2008).

34. Adams, P. L., Stahley, M. R., Kosek, A. B., Wang, J. & Strobel, S. A. Crystal structure of a self-splicing group I intron with both exons. *Nature* **430,** 45–50 (2004).

35. Dauter, Z. & Jaskolski, M. Crystal pathologies in macromolecular crystallography. *Postepy Biochem.* **62,** 401–407 (2016).

36. Flores, J. K., Walshe, J. L. & Ataide, S. F. RNA and RNA–Protein Complex Crystallography and its Challenges. *Aust. J. Chem.* **67,** 1741–1750 (2014).

37. Ferré-D'Amaré, A. R. & Doudna, J. A. in *Current Protocols in Nucleic Acid Chemistry* 7.6.1-7.6.13 (John Wiley & Sons, Inc., 2000). doi:10.1002/0471142700.nc0706s00

38. Ferre Â -D, A. R., Â, A., Zhou, K. & Doudna, J. A. A General Module for RNA Crystallization. *J. Mol. Biol.* **279,** 621–31 (1998).

39. Ferre Â -D, A. R., Â, A. & Doudna, J. A. Crystallization and Structure Determination of a Hepatitis Delta Virus Ribozyme: Use of the RNA-Binding Protein U1A as a Crystallization Module. *J. Mol. Biol.* **295,** 541–56 (2000).

40. Use of the spliceosomal protein U1A to facilitate crystallization and structure determination of complex RNAs. *Methods* **52,** 159–167 (2010).

41. Shechner, D. M., Grant, R. A., Bagby, S. C., Koldobskaya, Y., Piccirilli, J. A., Bartel, D. P. Crystal structure of the catalytic core of an RNA-polymerase ribozyme. *Science* **326,** 1271–5 (2009).

42. Koldobskaya, Y., Duguid, E. M., Shechner, D. M., Suslov, N. B., Ye, J., Sidhu, S. S., Bartel, D. P., Koide, S., Kossiakoff, A. A., Piccirilli, J. A. A portable RNA sequence whose recognition by a synthetic antibody facilitates structural determination. *Nat. Struct. Mol. Biol.* **18,** 100–6 (2011).

43. Hall, K. B. & Stump, W. T. Interaction of N-terminal domain of U1A protein with an RNA stem/loop. *Nucleic Acids Res.* **20,** 4283–90 (1992).

44. Oubridge, C., Ito, N., Evans, P. R., Teo, C. H. & Nagai, K. Crystal structure at 1.92 A resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature* **372,** 432–8 (1994).

45. Macrae, I. J. & Doudna, J. A. An unusual case of pseudo-merohedral twinning in orthorhombic crystals of Dicer. *Acta Crystallogr. D. Biol. Crystallogr.* **63,** 993–999 (2007).

46. LeCuyer, K. A., Behlen, L. S. & Uhlenbeck, O. C. Mutants of the bacteriophage MS2 coat protein that alter its cooperative binding to RNA. *Biochemistry* **34,** 10600–6 (1995).

47. Johansson, H. E., Dertinger, D., LeCuyer, K. A., Behlen, L. S., Greef, C. H., Uhlenbeck, O. C. A thermodynamic analysis of the sequence-specific binding of RNA by bacteriophage MS2 coat protein. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 9244–9 (1998).

48. Yoon, J.-H., Srikantan, S. & Gorospe, M. MS2-TRAP (MS2-tagged RNA affinity purification): tagging RNA to identify associated miRNAs. *Methods* **58,** 81–7 (2012).

49. Bertrand, E., Chartrand, P., Schaefer, M., Shenoy, S. M., Singer, R. H., Long, R. M. Localization of ASH1 mRNA particles in living yeast. *Mol. Cell* **2,** 437–45 (1998).

50. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–26 (1997).

51. Schneider, T. R. & Sheldrick, G. M. Substructure solution with SHELXD. *Acta Crystallogr. D. Biol. Crystallogr.* **58,** 1772–9 (2002).

52. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, R. C., Zwart, P. H. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr.* **66,** 213–21 (2010).

53. Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr*. A. **64,** 112–22 (2008).

54. Emsley, P. & Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr*. **60,** 2126–2132 (2004).

55. Keating, K. S. & Pyle, A. M. RCrane: semi-automated RNA model building. *Acta Crystallogr. D. Biol. Crystallogr.* **68,** 985–95 (2012).

56. Blanc, E., Roversi, P., Vonrhein, C., Flensburg, C., Lea, S. M., Bricogne, G. Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr. D. Biol. Crystallogr.* **60,** 2210–21 (2004).

57. Schröder, G. F., Levitt, M. & Brunger, A. T. Super-resolution biomolecular crystallography with low-resolution data. *Nature* **464,** 1218–22 (2010).

58. Chou, F.-C., Sripakdeevong, P., Dibrov, S. M., Hermann, T. & Das, R. Correcting pervasive errors in RNA crystallography through enumerative structure prediction. *Nat. Methods* **10,** 74–6 (2013).

59. Morin, A., Eisenbraun, B., Key, J., Sahschagrin, P. C., Timothy, M. A., Ottaviano, M. Sliz, P. Collaboration gets the most out of software. *Elife* **2,** e01456 (2013).

Chapter 3: Structural basis for the second step of group II intron splicing

3.1 Abstract

The group II intron and the spliceosome share a common active site architecture containing multiple base triples that are essential for catalysis and splice via the same lariat-forming mechanism. Due to these similarities, they are thought to be evolutionarily related. Here we report the 3.7 Å crystal structure of a eukaryotic group II intron in the lariat-3′ exon form, immediately preceding the second step of splicing. The intact 3′ splice site is positioned in the active site adjacent to the catalytic metal ions and the 3′-OH nucleophile of the 5′ exon. This represents the first structure of either the group II intron or the spliceosome with an intact 3′ splice site. Comparisons with the post-catalytic structure reveal extensive rearrangements occurring within the catalytic triplex. As splicing progresses, the conserved linker sequences J2/3, J4/5 and J5/6 engage in a "shuffling" action within the catalytic triad of domain V to facilitate 3′ splice site capture for the transition towards the second step of splicing and exon ligation. Mutagenesis of this triplex results in first- or second-step splicing defects consistent with its role in the spatiotemporal positioning of substrates within the active site during the splicing reaction. Therefore, we hypothesize that each stage of splicing is associated with a specific, unique arrangement of the catalytic triplex.

3.2 Introduction

Group II introns are self-splicing ribozymes that catalyze two transesterification reactions to excise themselves from nascent mRNA. In the first step of splicing, the 2′-OH of a bulged adenosine residue is used as the nucleophile to attack the 5′ splice site. This is followed by a second step in which the free 3′-OH of the 5′ exon attacks the 3′ splice site to form ligated exons. The chemistry of this RNA splicing reaction is identical to that found in the splicing of nuclear introns by the spliceosome in eukaryotes. Recently, cryo-EM structures of the spliceosome were determined at high-resolution and revealed that its active site architecture is highly conserved with that of group II introns, with both utilizing a two-metal-ion mechanism to catalyze RNA splicing[1–10]. This structural homology suggests an evolutionary link between group II introns and the spliceosome.

Structures are available of both the group II intron and the spliceosome at different stages of catalysis; however, the intact 3′ splice site has not yet been visualized in either splicing system. In fact, no density for the 3′ exon has been seen in any of the spliceosome structures. The precise location of the 3′ splice site would provide insight into the mechanism of the second step of splicing. Here we report the 3.6 Å crystal structure of a eukaryotic group IIB intron from the brown algae *Pylaiella littoralis* (*P.li.LSUI2*) in the lariat-3′ exon form with an intact 3′ splice site docked into the catalytic core. This is the first structure of a group II intron immediately preceding the second step of splicing. The intact 3′ splice site is aligned in the active site adjacent to the two catalytic Mg^{2+} ions and the 3′-OH nucleophile of the 5′ exon. We also observe

conformational rearrangements within the active site in the transition from the lariat-3′ exon form to ligated exon product.

3.3 Structure of the lariat-3′ exon intermediate

The structure of lariat-3′ exon intermediate in the pre-second step (pre-2S) state was determined at 3.6 Å resolution (Rfree = 25.4%) (Figure 3.1 and Table 3.1). The intact 3′ splice site was biochemically confirmed with a sequence analysis of the crystallized RNA (Figure 3.2). In addition, the refinement of the previous post-catalytic structure[11] (iridium hexammine derivative) was further improved to yield an Rfree = 25.5% (Table 3.1). The presence of iridium hexammine favors the post-catalytic state (Figure 3.3), whereas its absence stalls splicing *in crystallo* prior to the second step.

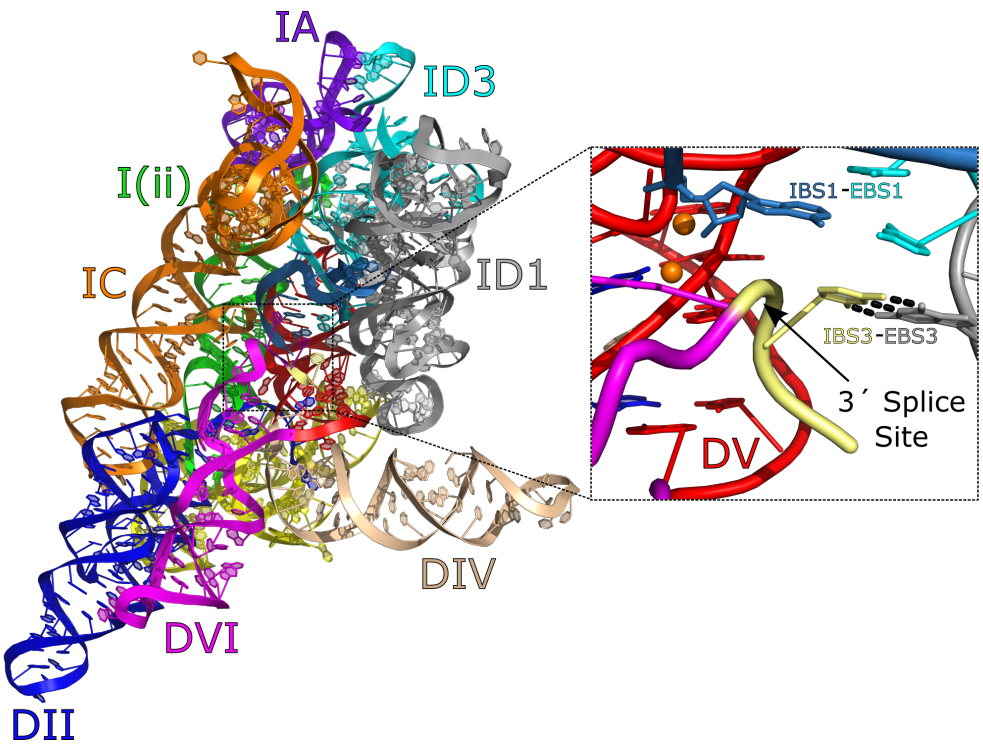

Figure 3.1: **Structure of the lariat-3′ exon intermediate.** Inset image shows the intact 3′ splice site positioned in close proximity to the catalytic two-metal-ion center and the 3′-OH nucleophile of 5′ exon. The IBS3-EBS3 interactions positions serves to properly position the 3′ splice site.

Table 3.1: Data collection and refinement statistics

	P.li.LSUI2 Pre-2S	*P.li.LSUI2* Post-Catalytic Rerefine
Data collection		
Space group	$C222_1$	$C222_1$
Cell		
a, *b*, *c* (Å)	165.01, 256.89, 137.23	163.7, 255.4, 136.8
α, β, γ (°)	90, 90, 90	90, 90, 90
Resolution (Å)	150.00-3.70 (3.83-3.70)	150.0-3.68 (3.74-3.68)
R_{sym} or R_{merge}	11.4 (58.6)	14.9 (>100)
I / ?I	15.1 (1.21)	6.4 (0.6)
Completeness (%)	97.1 (92.9)	99.9 (99.9)
Redundancy	3.5 (3.2)	6.8 (3.8)
Refinement		
Resolution (Å)	50.08-3.70 (3.82-3.70)	81.83-3.68(3.77-3.68)
No. reflections	31200	31107
R_{work} / R_{free}	21.63/25.40(32.12/34.09)	20.99/25.48(36.93/42.48)
No. atoms	13881	14107
RNA	13541	13479
Ligand/ion	104	382
Water	236	246
B-factors		
RNA	200.4	192.7
Ligand/ion	157.9	179.5
Water	172.0	159.0
R.m.s. deviations		
Bond lengths	0.002	0.009
Bond angles (°)	0.516	1.080

* Values in parentheses are for highest-resolution shell.

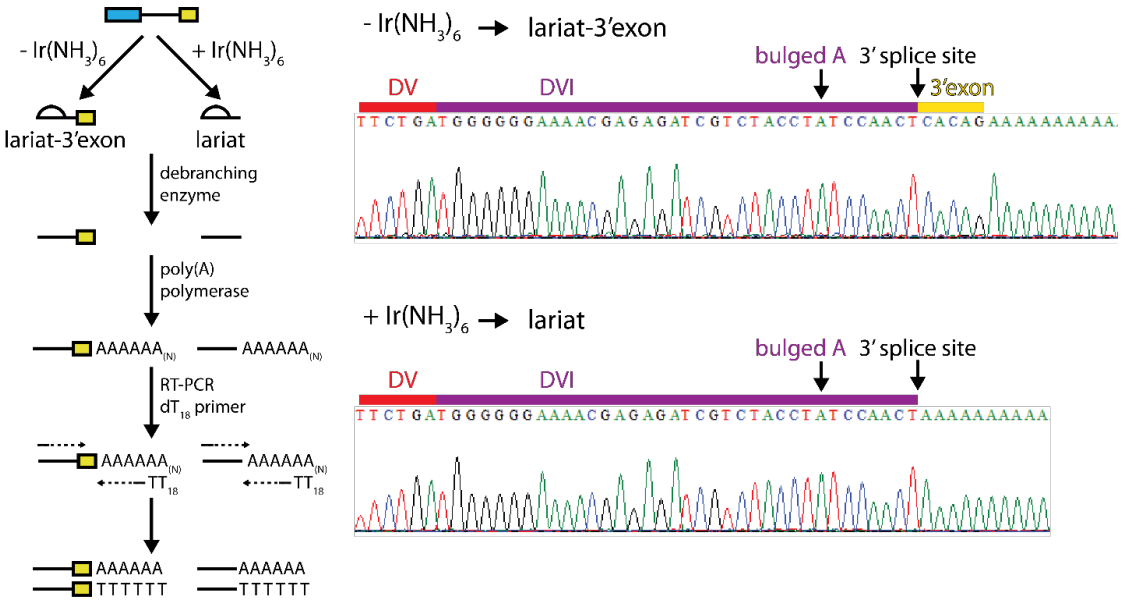

Figure 3.2: **Sequence analysis of the 3′ end of crystallized RNA from the pre-2s and post-catalytic states.** (left) Crystallized RNA was first incubated with debranching enzyme to digest the 2′-5′ phosphodiester of the lariat. This allows 3′ end mapping without steric clashes with the reverse transcriptase and the lariat bond during primer extension. A poly(A) tail was added to the 3′ end using poly(A) polymerase. Primer extension was using an oligo dT primer and reverse transcriptase. (right) Fluorescent sequencing traces show an intact 3′ splice site in the absence of iridium hexammine, whereas the iridium hexammine-containing sample does not contain 3′ exon. This supports the x-ray data, which shows density for the 3′ splice site in the absence of iridium hexammine.

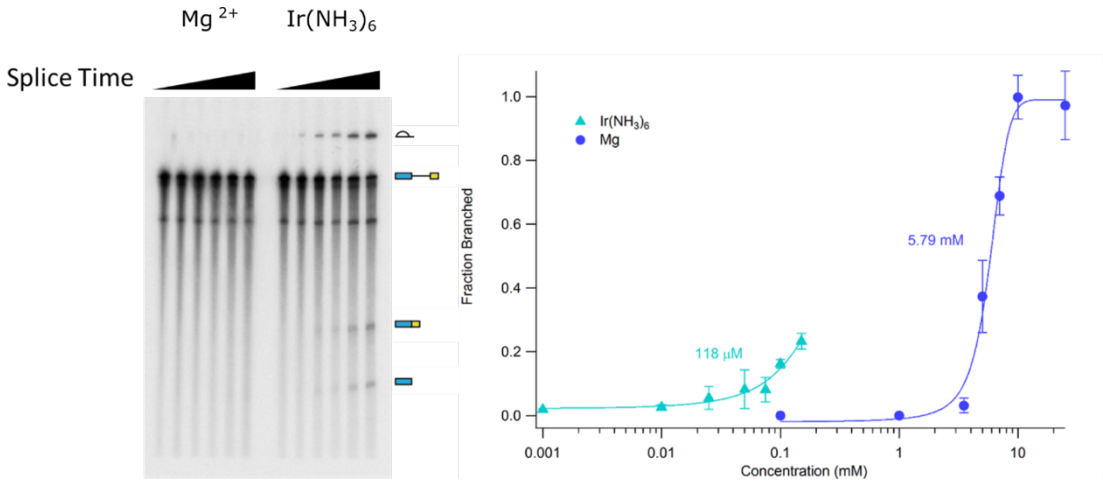

Figure 3.3: **Iridium hexammine stimulates the forward splicing reaction.** (left) At a limiting magnesium concentration of 3mM, iridium hexammine strongly favors the post-catalytic state with no visible pre-2s state during the splicing time course. In the absence of iridum hexammine, the pre-2s state is easily visible during the time course. (right) Iridium hexammine stimulates splicing at micromolar concentrations.

In order to avoid model bias, Fo-Fc omit maps were calculated by deleting nucleotides spanning the 3′ splice site (Figure 3.4) for both the pre-2s and post-catalytic structures. This reveals density for an intact 3′ splice site in the pre-2s crystals and confirms that this structure represents the state directly preceding the second step. A comparison of both structures reveals significant differences within the catalytic domain V (DV), the branch site helix domain VI (DVI), and the 3′ end of the intron. Examination of the intron core reveals the 3′-OH of the 5′ exon poised to engage in nucleophilic attack at the intact 3′ splice site (Figure 3.1). The scissile phosphate junction at the 3′ splice site is highly distorted and kinked as predicted by Chan *et al.* (2012)[12]. This distortion probably aids in presenting the correct scissile phosphate to the catalytic metal ions for cleavage in the second step in addition to applying additional strain on this bond to assist cleavage. The 3′ splice site is positioned in the core via DVI and the IBS3-

EBS3 interaction. IBS3-EBS3 consists of a single Watson-Crick pair between the first nucleotide of the 3′ exon (IBS3) and an internal intron base (EBS3), while domain 6 is covalently attached immediately upstream of the 3′ splice site. The 3′ splice site is also positioned by a single Watson-Crick pair between the 3′ terminal nucleotide of the intron and a residue from J2/3 (γ-γ′ interaction). This interaction is critical for placing the scissile phosphate of the 3′ splice site near the catalytic metal ions. γ-γ′ is not engaged in the observed pre-2s state, thus accounting for the 10 Å distance between the 3′ splice site and the catalytic metal ions M1 and M2. The γ nucleotide from J2/3 is flipped 180° away from the active site and therefore cannot form the pairing with the 3′ terminal nucleotide (γ′) of the intron. The lack of this interaction accounts for the fact that the 3′ splice site is intact in this structure. This also suggests that γ-γ′ is a transient and dynamic contact.

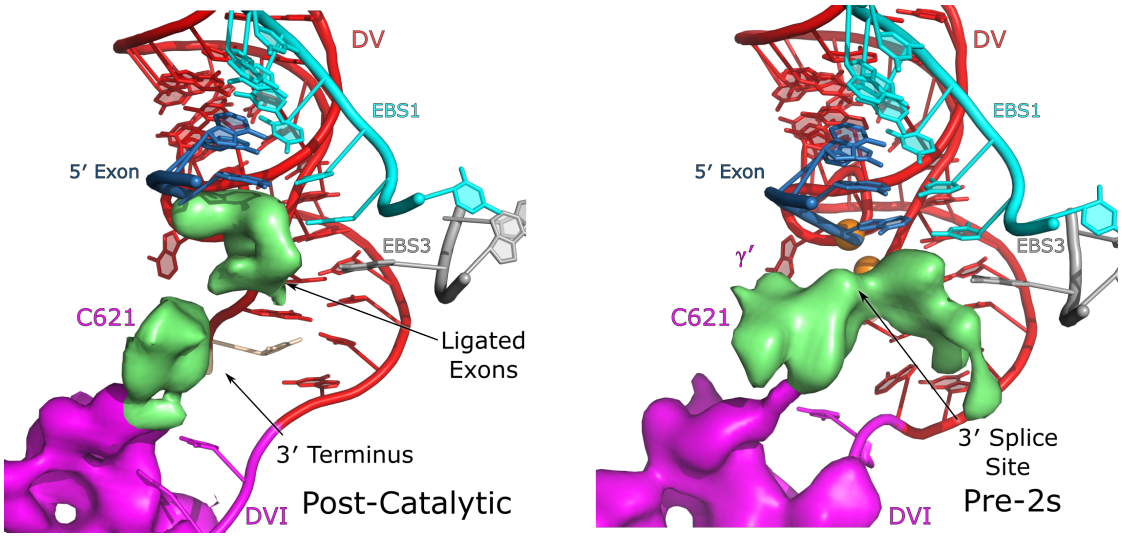

Figure 3.4: **Density for the 3′ splice site in the lariat 3′-exon intermediate.** Fo-Fc omit maps (contoured at 3 σ) were calculated for the region spanning the 3′ splice site in both the post-catalytic and pre-2s states. The cleaved 3′ end is clearly visible in the post-catalytic state versus the intact 3′ splice site in the pre-2s structure.

A comparison of this pre-2s structure with the post-catalytic state suggests that the pathway between the first and second steps of splicing consists of multiple conformational rearrangements that set the stage for 3′ splice site capture and cleavage (Figure 3.5). A comparison of the pre-2s and post-catalytic structures reveals that the basal stem of DVI (containing the bulged adenosine) undergoes a “compression” in the transition to the second step. Figure 3.6 shows a superposition of the DVI region from both structures. In particular, the backbone of the bulged adenosine residue is distorted once exon ligation is completed in the post-catalytic state. Given that the scissile phosphate of the 3′ splice site is not in a position to be cleaved in the pre-2s structure, we hypothesize that this compression of the DVI helix assists in positioning the 3′ splice

site into close proximity to the catalytic metals ions and the 3′-OH nucleophile of the 5′ exon for exon ligation.

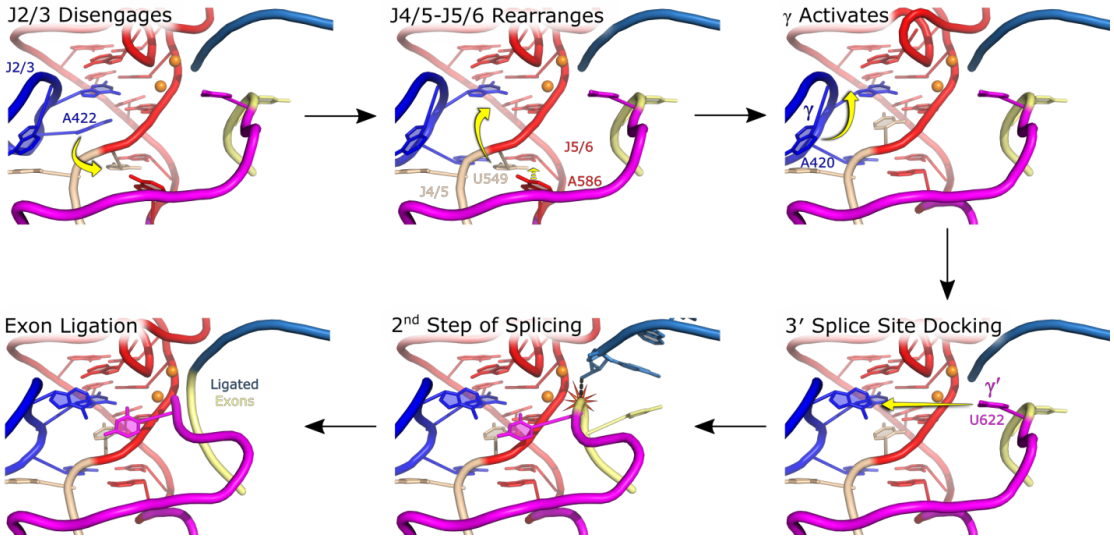

Figure 3.5: **Model for the second step of group II intron splicing.** This model illustrates events occurring after the first step and the exit of the 2′-5′ lariat phosphodiester bond from the active site. The catalytic triplex prepares for the second step with J2/3 residue A422 disengaging from A551. U549 replaces the position previously occupied by A422. EBS3-IBS3 then forms to anchor the 3′ end of the intron and causes A586 of J5/6 to displace U549 at the base of DV to activate the catalytic triplex for the second step. Finally, A420 (γ) enters its active conformation by stacking under A573. The formation of γ-γ′ draws the 3′ splice site into the active site and promotes proper exon ligation.

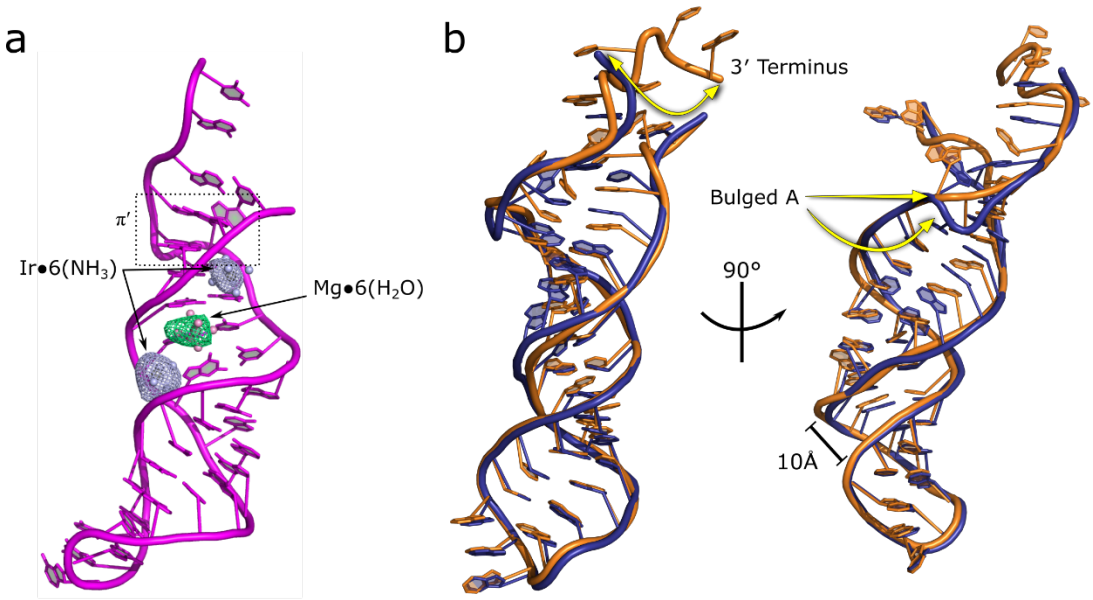

Figure 3.6: **Superposition of DVI in the post-catalytic and pre-2s states.** (a) Two $Ir(NH_3)_6^{3+}$ and one $Mg(H_2O)_6^{2+}$ are bound to the major groove of DVI in the post-catalytic structure. Ir^{3+} anomalous maps (light blue) are contoured to 2.0σ, $Mg(H_2O)_6^{2+}$ Fo-Fc map (green) is contoured to 3.6σ. (b) The pre-2s (orange) and post-catalytic (blue) DVI conformations are superimposed. This reveals conformational rearrangements at the basal region of the DVI stem. The backbone of the bulged adenosine residue in particular undergoes repositioning with less distortion in the pre-2s state.

3.4 Novel catalytic triplex arrangement in the lariat-3′ exon intermediate

The catalytic triplex in group II introns consists of consecutive base triples involving nucleotides from the catalytic triad, the two-nucleotide bulge, and a linker sequence between domains II and III (J2/3). Marcia and Pyle (2012) were the first to discover the existence of conformational rearrangements in the catalytic triplex of the hydrolytic group IIC intron[13]. However, these rearrangements were only observed through growth of the crystals in an ionic condition not supporting splicing. This work revealed that J2/3 is disengaged from the catalytic triad in the pre-catalytic state and is

engaged in the active form. Prior to these structural studies, Jarrell (2000) had observed that mutagenesis or deletion of J2/3 resulted in strong inhibition of splicing in the *aI5γ* intron[14]. Therefore, the cumulative evidence suggests that the residues involved in the catalytic triplex are playing an essential role in the transition between the two steps of splicing.

The structure of the post-catalytic eukaryotic IIB intron revealed the first evidence for additional junction nucleotides (other than J2/3) participating in base triples within the catalytic triplex. In comparison with the IIC intron, we observed in the IIB intron that A422 of J2/3 disengages with A551 of the catalytic triad to form a base triple with U549 of J4/5 and U584 of J5/6 at the base of DV[11] (Figure 3.7). Surprisingly, the pre-2s structure presented in the current study reveals yet another configuration of the catalytic triplex (Figure 3.7). In comparison with the post-catalytic structure, A586 of J5/6 has taken the place of U549 in the base triple below DV, which promotes U549 to form a base triple with A551 and U583 of the triad. We hypothesize that this catalytic triplex represents an active conformation that corresponds to the second step of splicing. In light of these two additional triplex configurations, it is probable that the triplex observed in the bacterial group IIC intron corresponds to the first step of splicing. IIC introns are known to engage in random hydrolytic cleavage at the 3′ end, which removes J5/6, domain VI and the 3′ splice site[15,16]. Therefore, in the absence of essential nucleotides required for the second step, the IIC intron becomes locked into the first step configuration.

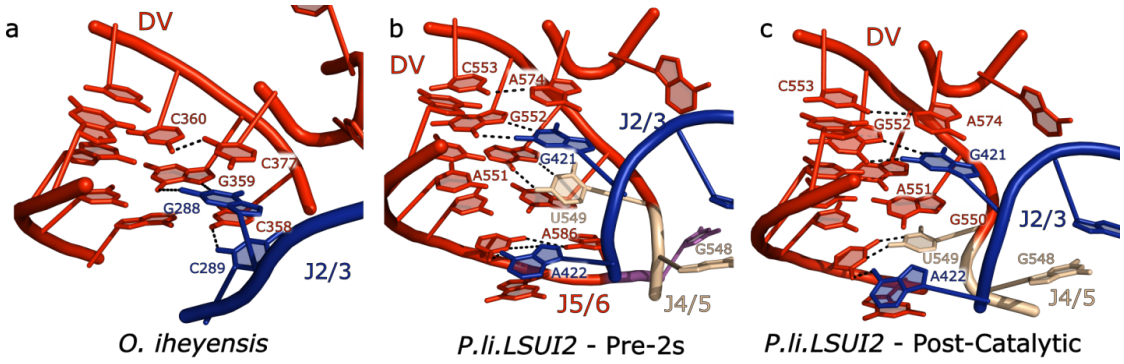

Figure 3.7: **Different configurations of the catalytic triplex.** This shows the three known configurations of the catalytic triplex in group II introns. The configuration in *O. iheyensis* is thought to represent the configuration required for the first step.

3.5 Catalytic triplex nucleotides modulate RNA splicing

To test the function of the junction nucleotides in the *P.li.LSUI2* intron, we altered the sequence and length of the J2/3, J4/5 and J5/6 linkers (Figure 3.8). Mutations within J2/3 had significant effects upon both steps of splicing depending on the specific alteration. A J2/3 ΔGA mutant resulted in a severe catalytic defect with almost no splicing. However, a J2/3 G241C mutant resulted in a large inhibition of the second step with a 4-fold greater accumulation of lariat-3′ exon intermediate compared to WT. In contrast, mutagenesis of the J4/5 linker largely results in an inhibition of the first step with the +GC mutant resulting in a 4-fold increase in unspliced, precursor RNA relative to WT. The J5/6 linker resulted in the most diverse effects upon splicing with the largest magnitude inhibition of either the first or second step with various mutations. It was previously seen in the aI5γ intron that mutations in the J5/6 linker result in alternative 3′ splice site selection. This is consistent with a role for J5/6 and the catalytic triplex in

ensuring the fidelity of splice site selection. Taken together, the cumulative data are consistent with a model in which dynamic rearrangements of the catalytic triplex are responsible for the positioning and removal of exon substrates from the active site in the transition between the two steps of splicing (Figure 3.5). The catalytic triplex also exists in the spliceosome, however in this case, protein-mediated conformational rearrangements may have taken over the role played by the dynamic triplex in group II introns (Figure 3.9).

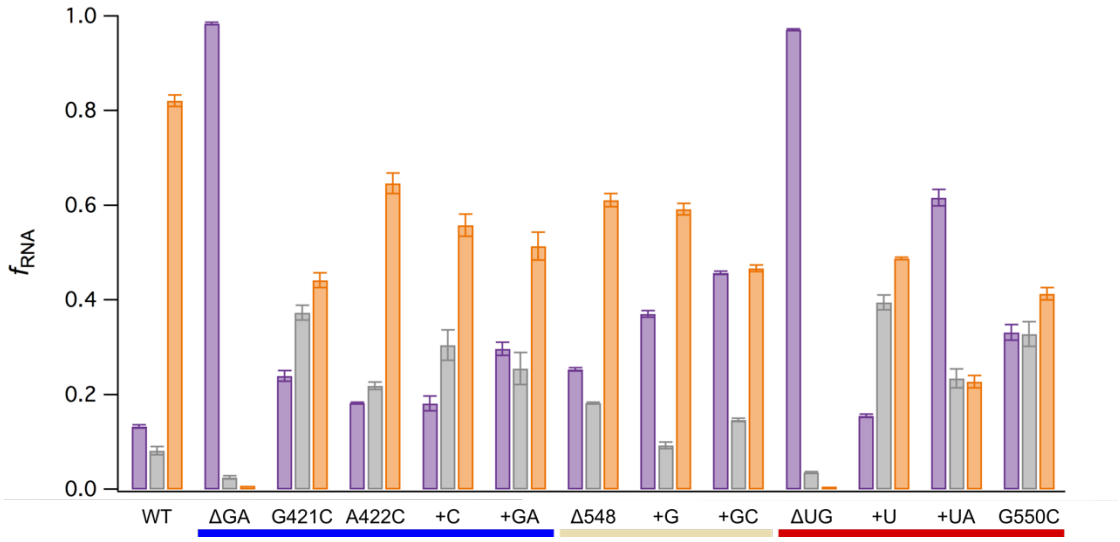

Figure 3.8: **Fraction of splicing components for junction mutants**. Fraction of precursor (purple), linear intron-3′ exon with lariat-3′ exon (grey), and linear intron with lariat (orange) are shown for wild type and several mutations of key residues in J2/3 (blue underline), J4/5 (wheat underline), or J5/6 (red underline). The ΔUG mutation of J5/6 was the only mutation observed to alter the preference of the first step nucleophile as this mutation results in a complete loss of branched product.

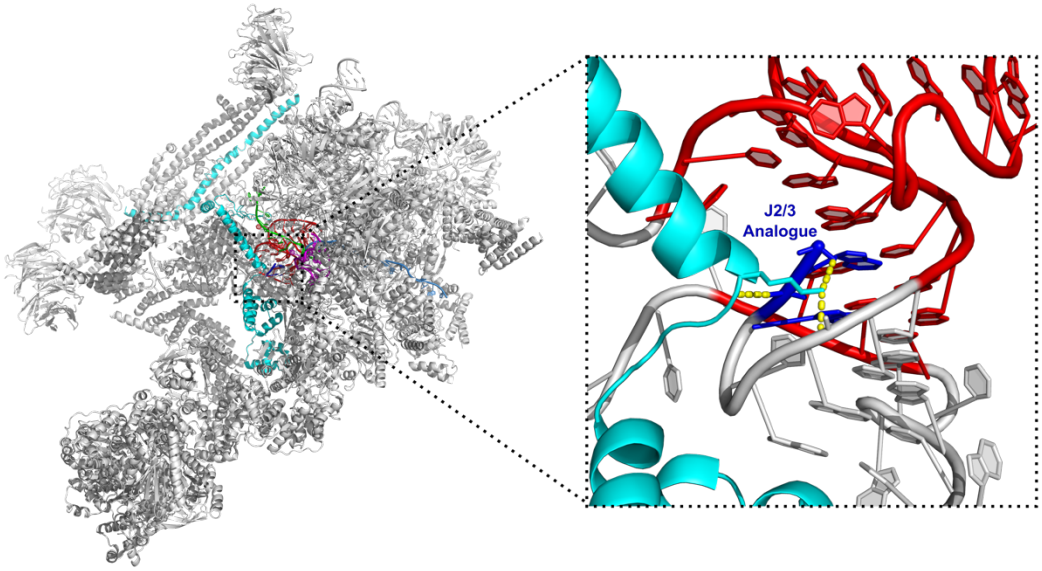

Figure 3.9: **The core architecture of the spliceosome.** The spliceosome structure contains analogous junction nucleotides to the group II intron (shown in blue and wheat, Figure 3.7). Cef1 (cyan), spliceosomal protein and component of the Prp19 complex, spans a large distance into the overall structure and makes contacts with the backbone of analogous J2/3 and J4/5 residues (PDB 3JB9). Mutations of 2 Cef1 residues are strong suppressors of second step splicing defects, indicating that this protein plays a significant role in communicating the fidelity of second step substrates[17].

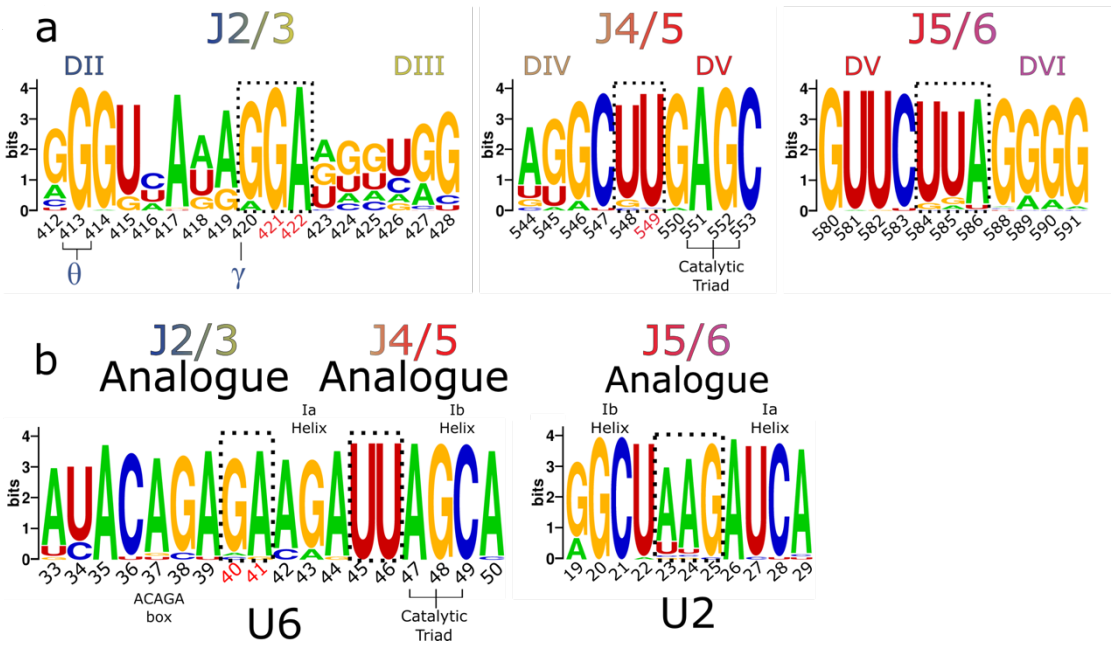

Figure 3.10: **Sequence conservation of junction nucleotides.** (a) The junction nucleotides of the group II intron are highly conserved, and correlate to their observed function in the pre-2s structure. (b) The Spliceosome junction nucleotides identified by either their proximity to the catalytic triad (J4/5 and J5/6 analogues) or their previous biochemical data (J2/3 analogue). Dashed boxes denote the junction nucleotides. Red numbers indicate bases identified to participate in the catalytic triplex. Nucleotide numbers for the group II intron and spliceosome sequences correspond to *P.li.LSUI2* and *S. pombe*.

3.6 Supplementary Discussion: Iridium hexammine favors the post-catalytic state

Previously, we solved the structure of the *P.li.LSUI2* intron in the post-catalytic state using crystals which were soaked in iridium hexammine prior to data collection[9]. Inspection of the major groove of DVI revealed two iridium hexammine molecules and a magnesium hexahydrate near the proposed conformational switch known as the π-π′ interaction. This interaction engages after the first step of splicing to remove the newly formed lariat out of the active site to allow entry of the 3′ splice site, which is crucial

for the transition to the second step of splicing. We had postulated that the region of DVI surrounding the π-π′ interaction is flexible and dynamic to allow the observed 20 Å displacement of the lariat bond out of the active site after the first step. Data was also collected on the native dataset in the absence of iridium hexammine. Rigid body refinement of the iridium hexammine derived structure into the native dataset revealed the presence of multiple positive and negative Fo-Fc peaks within the catalytic triplex of DV and at the 3′ end of the intron (not including the absence of iridium hexammine and minor metal ions). Further analysis showed strong density corresponding to an intact 3′ splice site. In contrast, the iridium hexammine derivative had density for ligated exons in the active site. Therefore, we hypothesized that the native dataset represents the lariat-3′ exon form of the intron immediately preceding the second step of splicing. Further evidence for this hypothesis was provided by an analysis of the 3′ end of the RNA present in both the native crystals and the iridium hexammine derivative. End mapping on this crystallized RNA was done to determine the cleavage pattern at the 3′ splice site (Figure 3.2). The RNA from the crystals was first treated with debranching enzyme to digest the 2′-5′ phosphodiester bond of the lariat. This allows primer extension at the 3′ end without interference from this branched linkage. A polyA tail was then added using polyA polymerase followed by RT-PCR and cloning. This confirmed the presence of an intact 3′ splice site in the native crystals and spliced, post-catalytic intron in the iridium hexammine derivative.

Given that the self-splicing of group II introns is reversible, we hypothesized that iridium hexammine inhibits the backward reaction and favors the equilibrium

towards the post-catalytic-state in the crystals. This is supported by splicing reactions of *P.li.LSUI2* in the presence or absence of iridium hexammine in a limiting Mg^{2+} concentration of 4 mM (Figure 3.3). Conditions lacking iridium hexammine are unable to support splicing. In contrast, supplementation with iridium hexammine at concentrations as low as 50 μM (~80-fold lower than [Mg^{2+}]) is able to promote splicing. The low concentration of iridium hexammine required to initiate splicing suggests that it may be playing a more complex role beyond simple charge neutralization of the phosphate backbone. In addition, analysis of the splicing products reveals that there is no visible lariat-3′ exon intermediate (representing the pre-second step state) for reactions containing iridium hexammine. This is highly unusual in that most group II introns form a significant amount of this intermediate during time courses of *in vitro* splicing reactions. This is in contrast to the splicing reaction at 5.4 mM Mg^{2+} (in the absence of iridium hexammine), which results in the significant accumulation of lariat-3′ exon intermediate. The fact that this intermediate is not visible in the presence of iridium hexammine suggests that this metal complex has a stimulatory effect on the second step to favor the post-catalytic state.

3.7 Methods

Crystallization of *P.li.LSUI2* RNA

The crystallization construct was *in vitro* transcribed and purified as previously described[11]. The same conditions were used to grow the rod-like cyrstals, which were then gradually exchanged into 21% MPD, 100 mM magnesium acetate tetrahydrate, 50

mM MES monohydrate (pH 5.6), 0.5 mM spermine, and 100 mM NaCl; followed by flash freezing in liquid nitrogen. It is important to note that no iridium hexammine was present at any step of crystallization or crystal freezing.

Structure Determination

X-ray data sets were collected at NE-CAT's 24-ID-C beamline at the Advanced Photon Source (Argonne National Laboratory, Argonne, Illinois). Data was processed using HKL-2000[18]. The post-catalytic structure of *P.li.LSUI2* (PDBID 4R0D) was used as a search model for molecular replacement using PHENIX[19]. The last 10 and 5 nucleotides of the intron and 5′ exon, respectively, were deleted to avoid model bias. RNA nucleotides were modeled using COOT[20] and the RCrane plugin[21]. Structure refinement was done using Buster[22], PHENIX[19], DEN[23], and Phenix.Erasser[24]. SBGrid compiled all the software used[25]. Sequence logos were prepared using WebLogos[26].

***In vitro* self-splicing assays**

Constructs used for *in vitro* self-splicing assays were modifications of the wt *P.li.LSUI2* sequence with the DIV ORF removed and a 250 nt 5′ exon and 75 nt 3′ exon. This construct was cloned into pUC57 using the EcoRV cut site. Plasmids were linearized using BamHI to be used as templates for *in vitro* transcription. Radiolabeled transcripts were prepared using T7 RNA polymerase, 5 mM $MgCl_2$, 40 mM Tris-HCl pH 7.5, 0.05% Triton X-100, 10 μCi [α-^{32}P]UTP (3,000 Ci mmol^{-1}), 0.5 mM UTP, and 1 mM other NTPs. RNA was transcribed for 1 hr at 37°C then precursor RNA was gel purified on a denaturing 4% (19:1), 8 M urea, 1X TBE gel. To refold, radiolabeled precursor RNA and an internal control RNA in 40 mM Tris-HCl pH 7.5 was heated at

90°C for 1 min followed by incubation in 40 mM Tris-HCl H 7.5, 10 mM $MgCl_2$ for 15 min. To initiate self-splicing, RNA was mixed with an equal volume of 2X splicing buffer (2 M NH_4Cl, 40 mM Tris-HCl pH 7.5). Reactions were quenched by mixing with an equal volume of 80% formamide, 100 mM EDTA. Splicing products were resolved using denaturing 4% (19:1), 8 M urea, 1X TBE gels. All splicing assays were done in triplicate.

Quantitation of self-splicing products

Splicing gels were exposed to storage phosphor screens and splicing products were quantitated using Quantity One 1-D Analysis Software. Unequal loading was accounted for by normalizing to an internal control RNA. Band intensity was determined by dividing the background-subtracted intensity of each band by the number of Us in the RNA sequence corresponding to the band. All band intensities were then normalized to an unspliced control to give fractional values of input RNA.

3′ end mapping

Native and $Ir(NH_3)_6^{3+}$-soaked crystals were rinsed with excess crystallization solution to remove uncrystallized RNA then dissolved in 50 mM HEPES pH 7, 100 mM NaCl, 1 mM DTT, 0.1 μM Dbr1 and incubated at 37°C for 30 min to hydrolyze the lariat bond. RNA was phenol-chloroform extracted, ethanol precipitated and resuspended in nuclease free water. 1 μg debranched RNA was incubated with 1 mM ATP and 5 units of *E. coli* poly(A) polymerase from NEB at 37°C for 30 min to add a poly (A) tail to the RNA. The RNA was phenol-chloroform extracted, ethanol precipitated and resuspended in nuclease free water. 200 ng oligo$(dT)_{18}$ and 1 μL 10 mM dNTP mix were added to

the RNA and the volume was brought to 13 μL with nuclease free water. To anneal the primer, the mixture was heated at 65°C for 5 min and incubated on ice for 1 min. The RNA was reverse transcribed at 50°C for 30 min using 200 units SuperScript^{TM} III RT, 1X first-strand buffer, and 5 mM DTT in a final volume of 20 μL, and the RT was inactivated by heating at 70°C for 15 min. 2 μL of cDNA was used as a template for PCR using oligo$(dT)_{18}$ as a reverse primer and a forward primer that anneals within the intron. PCR products containing a short poly (A) tail were gel extracted, cloned into pUC57 using the EcoRV cut site, and transformed into *E. coli* to sequence single colonies.

3.8 Acknowledgements

We thank Surajit Banerjee and the staff of the NE-CAT beamlines at the Advanced Photon Source (APS) of Argonne National Laboratory. We thank Rhiju Das and Fang-Chieh Chou for assistance with phenix.erasser for structure refinement and Gerard Bricogne for advice on Buster refinement. R.T.C. was supported by the Cell, Molecular, and Genetics Training Program funded by NIH predoctoral training grant 5T32GM007240. J.K.P. was supported by the UCSD Molecular Biophysics Training Program funded by NIH predoctoral training grant 5T32GM008326. NE-CAT is supported by NIH grant 8P41GM103403-10 and APS is supported by the U.S. DOE under Contract No. DE-AC02-06CH11357. This work was supported by a Hellman Foundation Fellowship and NIH grant 5R01GM102216 awarded to N.T.

Chapter 3, in part, is currently being prepared for submission to *Nature Structural & Molecular Biology* for publication of the material: Chan, R.T. and Peters, J.K., Robart A.R., Rajashankar, K.R. & Toor, N. Structural basis for the second step of group II intron splicing. The dissertation author is a co-first author of this paper.

3.9 References

1. Yan, C., Hang, J., Wan, R., Huang, M, Wong, C. L., Shi, Y. Structure of a yeast spliceosome at 3.6-angstrom resolution. *Science* **349,** 1182–91 (2015).

2. Yan, C., Wan, R., Bai, R., Huang, G. & Shi, Y. Structure of a yeast activated spliceosome at 3.5 Å resolution. *Science* **353,** 904–11 (2016).

3. Wan, R., Yan, C., Bai, R., Huang, G. & Shi, Y. Structure of a yeast catalytic step I spliceosome at 3.4 Å resolution. *Science* **353,** 895–904 (2016).

4. Galej, W. P., Wilkinson, M. E., Fica, S. M, Oubridge, C., Newman, A. J., Nagai, K. Cryo-EM structure of the spliceosome immediately after branching. *Nature* **537,** 197–201 (2016).

5. Fica, S. M., Oubridge, C., Galej, W. P., Wilkinson, M. E., Bai, X.-C., Newman, A. J., Nagai, K. Structure of a spliceosome remodelled for exon ligation. *Nature* **542,** 377–380 (2017).

6. Yan, C., Wan, R., Bai, R., Huang, G. & Shi, Y. Structure of a yeast step II catalytically activated spliceosome. *Science* **355,** 149–155 (2017).

7. Bertram, K., Agafonov, D. E., Liu, W.-T., Dybkov, O., Will, C. L., Hartmuth, K., Urlaub, H., Kastner, B., Stark, H., Lührmann, R. Cryo-EM structure of a human spliceosome activated for step 2 of splicing. *Nature* **542,** 318–323 (2017).

8. Zhang, X., Yan, C., Hang, J., Finci, L. I., Lei, J., Shi, Y. An Atomic Structure of the Human Spliceosome. *Cell* **169,** 918–929.e14 (2017).

9. Plaschka, C., Lin, P.-C. & Nagai, K. Structure of a pre-catalytic spliceosome. *Nature* (2017). doi:10.1038/nature22799

10. Fica, S. M., Mefford, M. A., Piccirilli, J. A. & Staley, J. P. Evidence for a group II intron-like catalytic triplex in the spliceosome. *Nat. Struct. Mol. Biol.* **21,** 464–71 (2014).

11. Robart, A. R., Chan, R. T., Peters, J. K., Rajashankar, K. R. & Toor, N. Crystal structure of a eukaryotic group II intron lariat. *Nature* **514,** 193–197 (2014).

12. Chan, R. T., Robart, A. R., Rajashankar, K. R., Pyle, A. M. & Toor, N. Crystal structure of a group II intron in the pre-catalytic state. *Nat. Struct. Mol. Biol.* **19,** 555–7 (2012).

13. Marcia, M. & Pyle, A. M. Visualizing group II intron catalysis through the stages of splicing. *Cell* **151,** 497–507 (2012).

14. Mikheeva, S., Murray, H. L., Zhou, H., Turczyk, B. M. & Jarrell, K. A. Deletion of a conserved dinucleotide inhibits the second step of group II intron splicing. *RNA* **6,** 1509–15 (2000).

15. Toor, N., Robart, A. R., Christianson, J. & Zimmerly, S. Self-splicing of a group IIC intron: 5' exon recognition and alternative 5' splicing events implicate the stem-loop motif of a transcriptional terminator. *Nucleic Acids Res.* **34,** 6461–71 (2006).

16. Toor, N., Keating, K. S., Fedorova, O., Rajashankar, K., Wang, J., Pyle, A. M. Tertiary architecture of the Oceanobacillus iheyensis group II intron. *RNA* **16,** 57–69 (2010).

17. Query, C. C. & Konarska, M. M. CEF1/CDC5 alleles modulate transitions between catalytic conformations of the spliceosome. *RNA* **18,** 1001–13 (2012).

18. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–26 (1997).

19. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, H. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, R. C., Zwart, P. H. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr.* **66,** 213–21 (2010).

20. Emsley, P. & Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60,** 2126–2132 (2004).

21. Keating, K. S. & Pyle, A. M. RCrane: semi-automated RNA model building. *Acta Crystallogr. D. Biol. Crystallogr.* **68,** 985–95 (2012).

22. Blanc, E., Roversi, P., Vonrhein, C., Flensburg, C., Lean, S. M., Bricogne, G. Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr. D. Biol. Crystallogr.* **60,** 2210–21 (2004).

23. Schröder, G. F., Levitt, M. & Brunger, A. T. Super-resolution biomolecular crystallography with low-resolution data. *Nature* **464,** 1218–22 (2010).

24. Chou, F.-C., Sripakdeevong, P., Dibrov, S. M., Hermann, T. & Das, R. Correcting pervasive errors in RNA crystallography through enumerative structure prediction. *Nat. Methods* **10,** 74–6 (2013).

25. Morin, A., Eisenbraun, B., Key, J., Sanschagrin, P. C., Timony, M. A., Ottaviano, M., Sliz, P. Collaboration gets the most out of software. *Elife* **2,** e01456 (2013).

26. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: A Sequence Logo Generator. *Genome Res.* **14,** 1188–1190 (2004).

Chapter 4: A kinetic and thermodynamic analysis of domain VI dynamics in the group II intron *P.li.LSUI2*

4.1 Introduction

RNA has evolved to fold into many complex tertiary structures to carry out a range of functions varying from peptide bond formation to cleavage of phosphodiester bonds. In order to fold into such compact shapes, distant regions in the primary sequence come together to form long-range tertiary interactions. Among the most common tertiary contacts found in highly structured RNAs is the tetraloop-tetraloop receptor (TL-TLR) interaction. In this interaction, a 4-nucleotide loop (tetraloop) folds into a conserved shape and binds to a dsRNA receptor helix. Natural tetraloop sequences have been found to fall into unique classes with distinct folds. The three most common tetraloop classes are UNCG, CUUG and GNRA, where N is any nucleotide and R is a purine[1]. These are likely the most common because of their higher stability due to the base-pairing between nucleotides 1 and 4, base stacking, and 2′-OH hydrogen bonding[2]. Among rRNAs, the UNCG and GRNA tetraloops account for >70% of all tetraloops[3]. The UNCG tetraloop forms a particularly stable structure, but is unlikely to form interactions with RNA duplexes[4]. Conversely, GNRA tetraloops are commonly found binding to dsRNA receptors to stabilize highly ordered RNAs[5,6].

GNRA tetraloops have been found to bind to at least three receptor motifs: 11-nt, IC3, and minor groove[7]. The 11-nt receptor was first discovered in the *Tetrahymena* group I intron, and because of the very high affinity between this receptor and the GAAA tetraloop, this TL-TLR has been used since as a tool to facilitate intermolecular

self-assembly in crystal packing and RNA tectonics[8–12]. It consists of a 5-nt internal loop with two consecutive adenosines forming an adenosine platform[13] (Figure 4.1b). The IC3 receptor, named after its origin in group IC3 introns, is only slightly different from the 11-nt receptor and binds GNRA tetraloops more promiscuously and with a lower affinity[14,15]. The minor groove receptor is the simplest of the three as it only consists of the minor groove surface of two adjacent base pairs. In this interaction, the third and fourth tetraloop nucleotides form two base triples with the receptor base pairs. The simplicity of this interaction has likely contributed to its abundance in structured RNAs, with >60% of loop-helix interactions from x-ray crystal structures falling into this class[7].

TL-TLR affinities can vary substantially, depending on the sequences involved. The *trans*-splicing group I intron has been used to study relative strengths of different TL-TLR sequences by equating the L2-P8 TL-TLR stability with group I intron catalytic efficiency[6,16]. These results showed variations of the L2-P8 TL-TLR that result in k_{cat}/K_m values spanning three orders of magnitude. So far, most TL-TLR studies have centered around the classical GAAA/11-nt receptor interaction. The incredibly high stability of this interaction makes it an excellent candidate for engineering self-assembling modular RNAs. However, not all naturally occurring TL-TLRs are very stable. For instance, the group IIB intron from the brown algae *Pylaiella litoralis* (*P.li.LSUI2*) contains one GAAA/11-nt receptor interaction (ζ-ζ′) and four weaker TL-TLR interactions[17]. Highly structured RNAs have evolved to contain weak interactions

for two possible reasons: 1) the interaction is not necessary for function or 2) the interaction is dynamic.

The recent crystal structure of *P.li.LSUI2* in the post-catalytic lariat conformation was the first structure of a group II intron that showed domain VI (DVI)[17]. This domain is crucial to lariat formation as it contains the bulged adenosine that acts as the nucleophile in the first step of splicing. This structure revealed that the two arms of DII coaxially stack and form two TL-TLR interactions with DVI, π-π′ and η-η′. The π-π′ interaction, discovered from this crystal structure, is of particular interest because the receptor is directly adjacent to the bulged adenosine. Group II introns use the same active site for both steps of splicing[18], so the bulged adenosine must be positioned in the active site for the first step then displaced prior to the second step to allow room for second step reactants. Therefore, it was surprising to see a TL-TLR interaction so close to the bulged adenosine. The η-η′ interaction had previously been discovered from DEPC modification interference analysis of the ai5γ group II intron, and it was proposed that η-η′ forms after the first step of splicing [19]. However, a low resolution pre-catalytic crystal structure of the *P.li.LSUI2* intron shows that η-η′ is engaged prior to the first step[17]. To probe the possible dynamic nature of the π-π′ and η-η′ interactions, we created a screen of these TL-TLRs with varied stability and performed kinetic studies. These studies show that the π-π′ interaction in *P.li.LSUI2* has evolved to be of moderate stability in order to be dynamic and facilitate both steps of splicing. Though catalysis is less dependent on the strength of η-η′, with <4 kcal/mol adjustments exhibiting only minor affects, ablating this interaction causes a significant second step defect.

4.2 Use of P4-P6 to determine relative binding affinities

In order to knowingly vary the stabilities of the π-π′ and η-η′ TL-TLRs, we first had to determine which TL-TLR combinations would give a range of ΔΔG°′ relative to the *P.li.LSUI2* wt sequences. We chose to use the P4-P6 folding assay established by Silverman & Cech in 1999 to screen TL-TLR combinations for varied stability[20]. The assay was developed to investigate the thermodynamic and cooperative contribution of specific hydrogen bonds to folding of the *Tetrahymena* group I intron P4-P6 domain and has since been used to study the specificity of RNA-RNA helix recognition[21,22]. In this assay, modified P4-P6 constructs are run on non-denaturing polyacrylamide gels at various [$MgCl_2$]. Their free energy of folding (ΔG°′) can then be determined from Hill fits of relative electrophoretic mobility vs. [$MgCl_2$]. Figure 4.1a-b shows the P4-P6 constructs used in this assay. Five different tetraloop sequences were tested against 4 receptor motifs, resulting in a range of electrophoretic mobilities at different [$MgCl_2$] (Figure 4.1c).

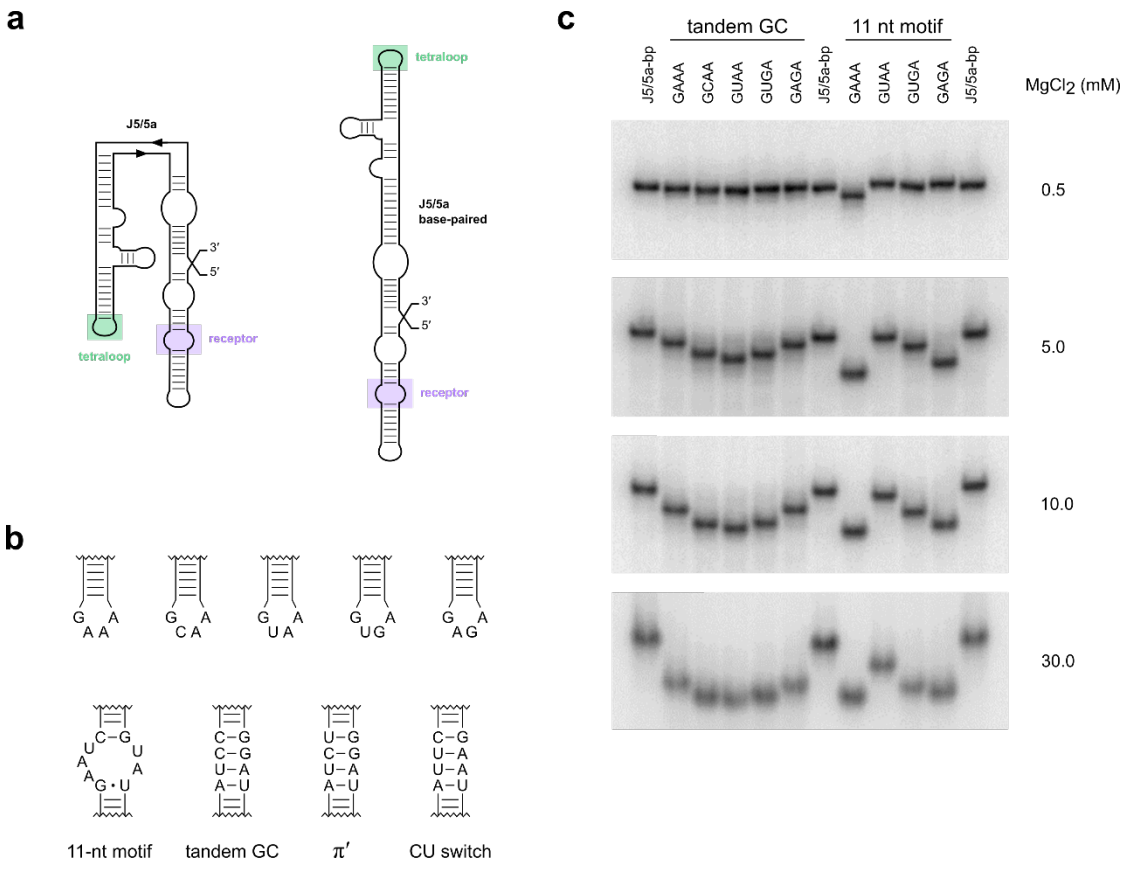

Figure 4.1: **P4-P6 folding assay and TL-TLR modifications.** (a) Secondary structures of P4-P6 constructs used in the folding assay. The modified regions are highlighted in green (tetraloop) and purple (tetraloop-receptor). The construct shown on the right has been modified such that the J5/5a hinge sequence forms WC base-pairs, precluding tetraloop-receptor docking at any magnesium concentration (Silverman and Cech, 1999). (b) Tetraloop (top) and receptor (bottom) modifications tested in the assay. (c) Representative non-denaturing gels of a subset of P4-P6 modifications. J5/5a base-pairing locks the RNA in a rod-like shape, retarding migration on a polyacrylamide gel. TL-TLR modified P4-P6 constructs fold in a [Mg^{2+}]-dependent manner, which is visible by increased mobility at higher [Mg^{2+}].

Plots of electrophoretic mobility relative to an unfolded control (J5/5a-bp) vs. [Mg^{2+}] demonstrate the variable magnesium-dependencies ([$Mg_{1/2}$]) and cooperativities (n) of the different constructs (Figure 4.2 and Table 4.1). Interestingly, the wt P4-P6 construct (GAAA/11-nt motif) shows the highest cooperativity of folding (n). Hill

coefficients are up to 4-fold lower for TL-TLR modifications, though cooperativity does not appear to be correlated with the relative thermodynamic stability.

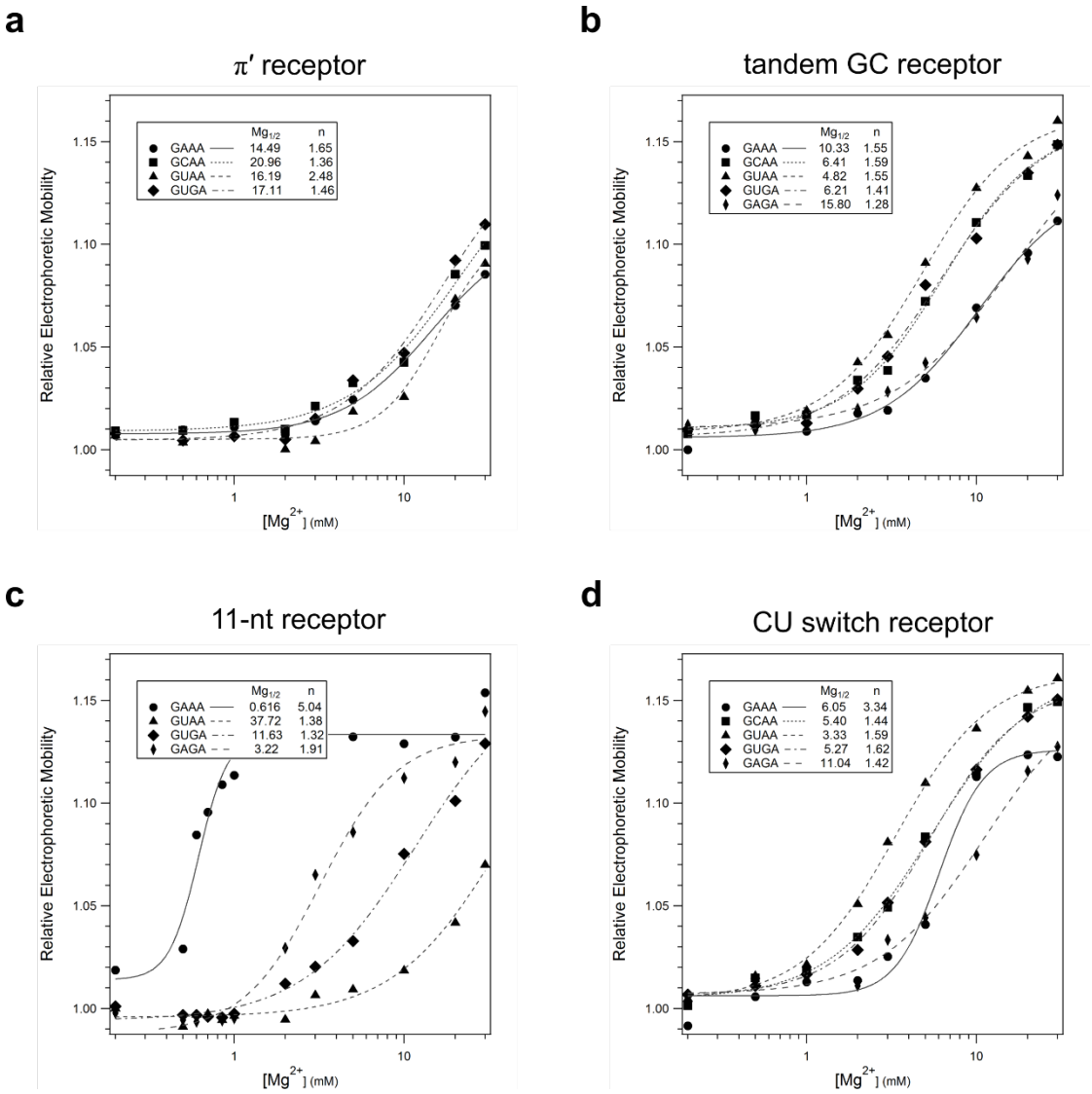

Figure 4.2: **Plots of relative electrophoretic mobility of P4-P6 TL-TLR variants vs. magnesium concentration.** Electrophoretic mobilities of P4-P6 constructs containing a GAAA, GCAA, GUAA, GUGA, or GAGA tetraloop and a (a) π′ receptor (b) tandem GC receptor (c) 11-nt receptor or (d) CU switch receptor (shown in Figure 4.1b) on non-denaturing gels were measured relative to the unfolded J5/5a-bp control and plotted against $[Mg^{2+}]$. The fit values of the Hill equation (see Methods), $[Mg^{2+}]_{1/2}$ and Hill coefficient *n*, are listed for each construct.

Table 4.1: Relative thermodynamic stability (ΔΔG°′) of P4-P6 TL-TLR modifications[†]

tetraloop/receptor sequence	$[Mg^{2+}]_{1/2}$ (mM)	Hill coeffiecient, *n*	ΔΔG°′ (kcal/mol) π-π′	ΔΔG°′ (kcal/mol) η-η′
GUAA/π'	16.187	2.481	1.708	2.070
GAAA/CU switch [a, b]	6.048	3.344	1.160	1.523
GUAA/11-nt motif [b]	37.715	1.375	0.529	0.891
GAAA/π'	14.493	1.652	0.174	0.536
GUGA/π'	17.108	1.464	0.014	0.376
GCAA/π'	20.958	1.359	(0)	0.363
GAAA/tandem GC [a, b]	10.328	1.553	-0.313	0.050
GAGA/tandem GC [b]	15.799	1.284	-0.363	(0)
GAGA/CU switch [b]	11.035	1.424	-0.440	-0.077
GUGA/11-nt motif [b]	11.625	1.319	-0.553	-0.190
GCAA/tandem GC [a]	6.407	1.585	-0.733	-0.370
GUGA/CU switch [a, b]	5.272	1.620	-0.888	-0.525
GUGA/tandem GC [a, b]	6.210	1.414	-0.956	-0.593
GUAA/tandem GC [a, b]	4.822	1.554	-1.040	-0.677
GCAA/CU switch [a]	5.400	1.437	-1.054	-0.691
GAGA/11-nt motif [b]	3.222	1.913	-1.168	-0.805
GUAA/CU switch [a, b]	3.326	1.589	-1.370	-1.007
GAAA/11-nt motif [b]	0.616	5.044	-4.050	-3.687

[†] ΔΔG°′ are shown relative to the *P.li.LSUI2* wt (shown in bold) π-π′ interaction (GCAA/π′) or η-η′ interaction (GAGA/tandem GC).
[a] TL-TLR combinations used in *P.li.LSUI2* π-π′ constructs.
[b] TL-TLR combinations used in *P.li.LSUI2* η-η′ constructs.

4.3 Splicing sensitivity to $[Mg^{2+}]$ and temperature

Though previous studies of the P4-P6 GAAA/11-nt receptor examined the thermodynamic and kinetic contributions of this interaction, our goal was to investigate the kinetics of group II intron splicing, not folding. Therefore, we verified that the ribozyme was fully folded via native-PAGE prior to the initiation of splicing (Figures 4.3 and 4.4). Because we also know that there are no large-scale conformation changes

between the initial and final stages of splicing, and the construct used in this assay contains very short exons (18 and 5 nt), we expect that fully folded unspliced and spliced RNAs will migrate the same on a non-denaturing gel. A fraction of the RNA molecules fold to a compact shape as they are transcribed, but the presence of multiple bands running at a retarded mobility (Figure 4.3, lane 1) indicates that a small population of RNA molecules is stuck in partially unfolded intermediates. Increased salt and temperature initiates splicing and folding of partially unfolded intermediates (lane2). Refolding the intron in the presence of magnesium results in compact bands with high mobility (lanes 5 and 6). Figure 4.4 (first lane) verifies that precursor RNA does not splice under refolding conditions.

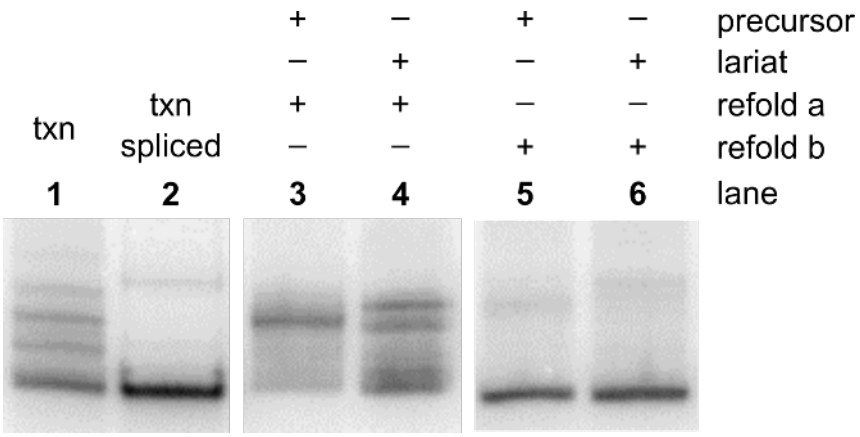

Figure 4.3: **Native-PAGE of *P.li.LSU12* intron folded under different conditions**. Intron RNA was either taken immediately after 3-hr transcription (lane 1), after splicing the transcription (lane2), or after gel extraction and refolding (lanes 3-6) prior to running on a non-denaturing PAGE. To prepare spliced sample, an equal volume of 20 mM $MgCl_2$, 2 M NH_4Cl, 80 mM Tris pH 7.5 was added to the transcription and incubated at 45°C for 30 min. Refolding protocol a: RNA in TE was heated at 95°C for 2 min then placed at room temperature for 15 min. Refolding protocol b: 4V RNA in 40 mM Tris pH 7.5 was heated at 95°C for 1 min then mixed with 1V 5X refolding buffer (50 mM $MgCl_2$, 40 mM Tris pH 7.5) and incubated at 42°C for 15 min. Samples were run on a 5% 37.5:1 acrylamide:bisacrylamide, 4% glycerol, 0.5X TBM gel at 4°C at 6W.

Group II introns have been shown to require much higher salt concentrations and temperature to splice *in vitro* compared to *in vivo*[23]. When splicing *in vivo*, the intron encoded protein (IEP), binds to the ribozyme to assist in folding and therefore promote catalytic activity. When splicing *in vitro*, however, the RNA requires a higher concentration of both monovalent and divalent cations to bring highly negatively charged phosphate backbones in close proximity in order to properly fold 3D RNA structures. Increased temperature can also overcome the splicing deficiency in the absence of the IEP due to increased kinetic energy, though this usually requires higher salt concentrations to counteract increased thermal motion (Figure 4.4).

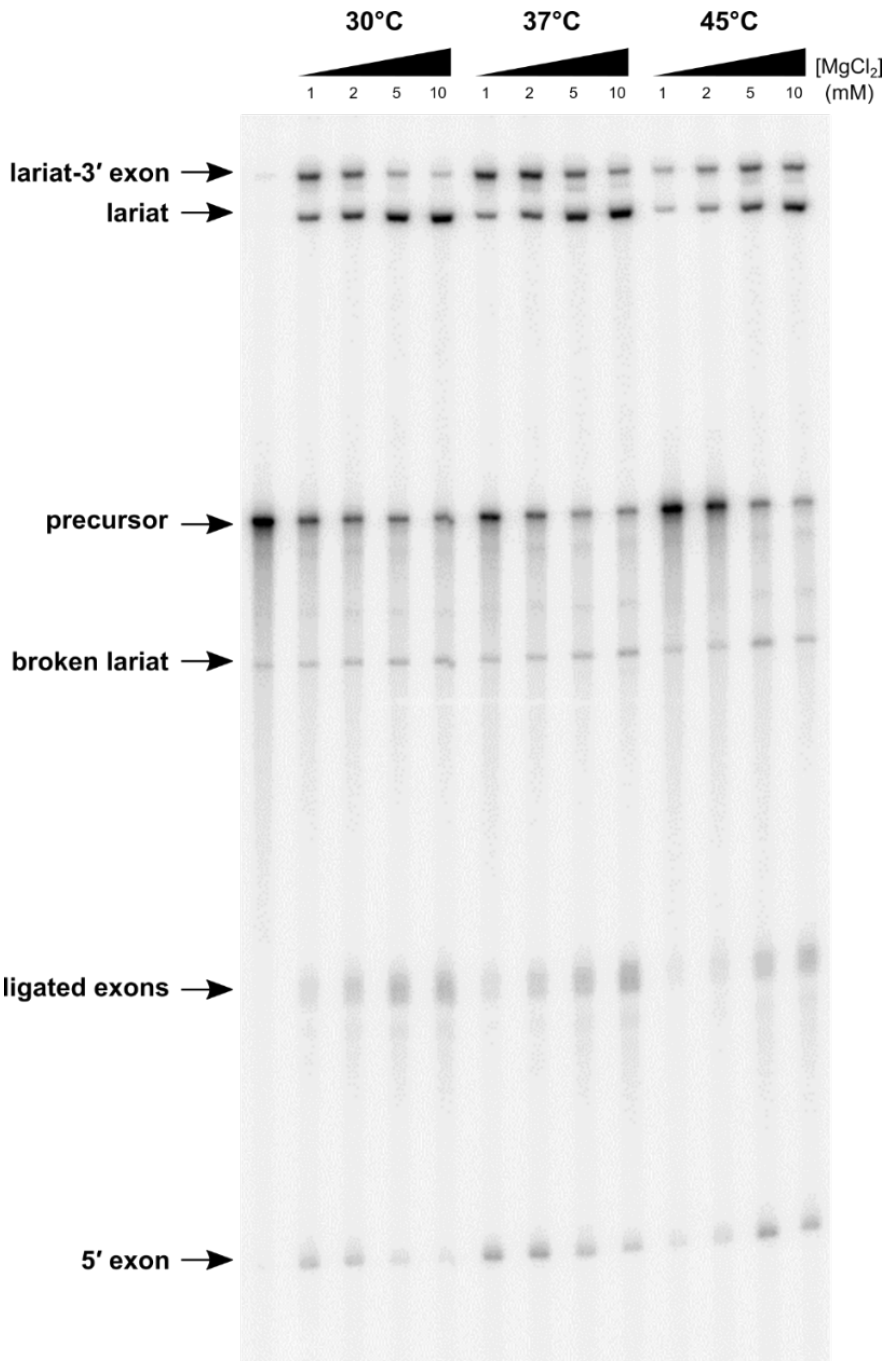

Figure 4.4: **Representative denaturing gel demonstrating effects of [Mg^{2+}] and temperature variations on self-splicing products.** Gel extracted precursor intron (*P.li.LSUI2* GAAA/11-nt receptor η-η′ modification) was refolded in the presence of 10 mM $MgCl_2$ (left lane) and incubated in splicing buffer containing 1 mM NH4Cl, 40 mM Tris-HCl pH 7.5 and 1, 2, 5 or 10 mM $MgCl_2$ at 30, 37, or 45°C for 1 hour. Splicing reactions were ethanol precipitated, resuspended in 50% TE, 50% formamide, and ran on a 6% 29:1 acrylamide:bisacrylamide, 7 M urea PAGE.

4.4 Kinetic analysis of π-π′ and η-η′ variations

It was previously shown that eliminating either the π-π′ or the η-η′ interaction in the *P.li.LSUI2* intron by mutating the tetraloop to an inert UUCG results in an accumulation of intermediate lariat-3′ exon, and eliminating π-π′ and η-η′ simultaneously almost completely blocks exon ligation[17]. Interestingly, disrupting π-π′ or η-η′ increases the rate of the first step of splicing ~2-fold (Table 4.2). A similar result was observed when η-η′ was disrupted in the ai5γ intron[19]. These results imply that both TL-TLR interactions are necessary for exon ligation, but their formation at least to some extent inhibits branching. In the *P.li.LSUI2* lariat crystal structure, π-π′ is engaged and the bulged adenosine is ~20 Å away from the active site. It was hypothesized that π-π′ is disengaged for the first step of splicing, allowing the base of DVI and the bulged adenosine the flexibility to be positioned in the active site. After branch formation, π-π′ can engage to pull the bulged adenosine and the 5′ end of the intron ~20 Å out of the active site.

Table 4.2: First step rate constants for π-π′ and η-η′ knock outs[†]

Construct	K_{slow} (min^{-1})	K_{fast} (min^{-1})	Plateau
wt	0.030	0.53	0.69
Δπ	0.036	0.92	0.65
Δη′	0.011	1.2	0.47
ΔπΔη′	0.014	0.42	0.33

[†] Knock outs were created by mutating the indicated tetraloop sequence to UUCG.

In order to test this proposed splicing transition, two screens were made in which the π-π′ (*P.li.LSUI2*-π) or η-η′ (*P.li.LSUI2*-η) interactions were mutated to different TL-TLR combinations of varied strength (Table 4.1[a,b]). The architecture of the *P.li.LSUI2* intron somewhat limited the possible receptor mutations that could be made. For instance, the distance of the bugled adenosine adenosine from the catalytic triad in DV has been shown to be very important, with the ideal architecture consisting of a 3-nt linker between domains V and VI and a 4-bp stem at the base of DVI[24]. Because of the location of π′ at the base of DVI and its proximity to the bulged adenosine (Figure 4.5), it was decided that only minor groove receptor variations would be used for π′ (tandem GC and CU switch). In contrast, the location of η (DIIa) is more tolerable to insertions because it extends out from the core of the intron (Figure 4.5a,b), so η was changed to the 11-nt and CU switch receptors. Additionally, because P4-P6 constructs containing the π′ receptor have such high $[Mg^{2+}]_{1/2}$ and therefor incomplete electrophoretic mobility curves, accurate $\Delta G^{o\prime}$ values cannot be determined for these constructs and π′ receptor constructs have been omitted from the *P.li.LSUI2*-π screen.

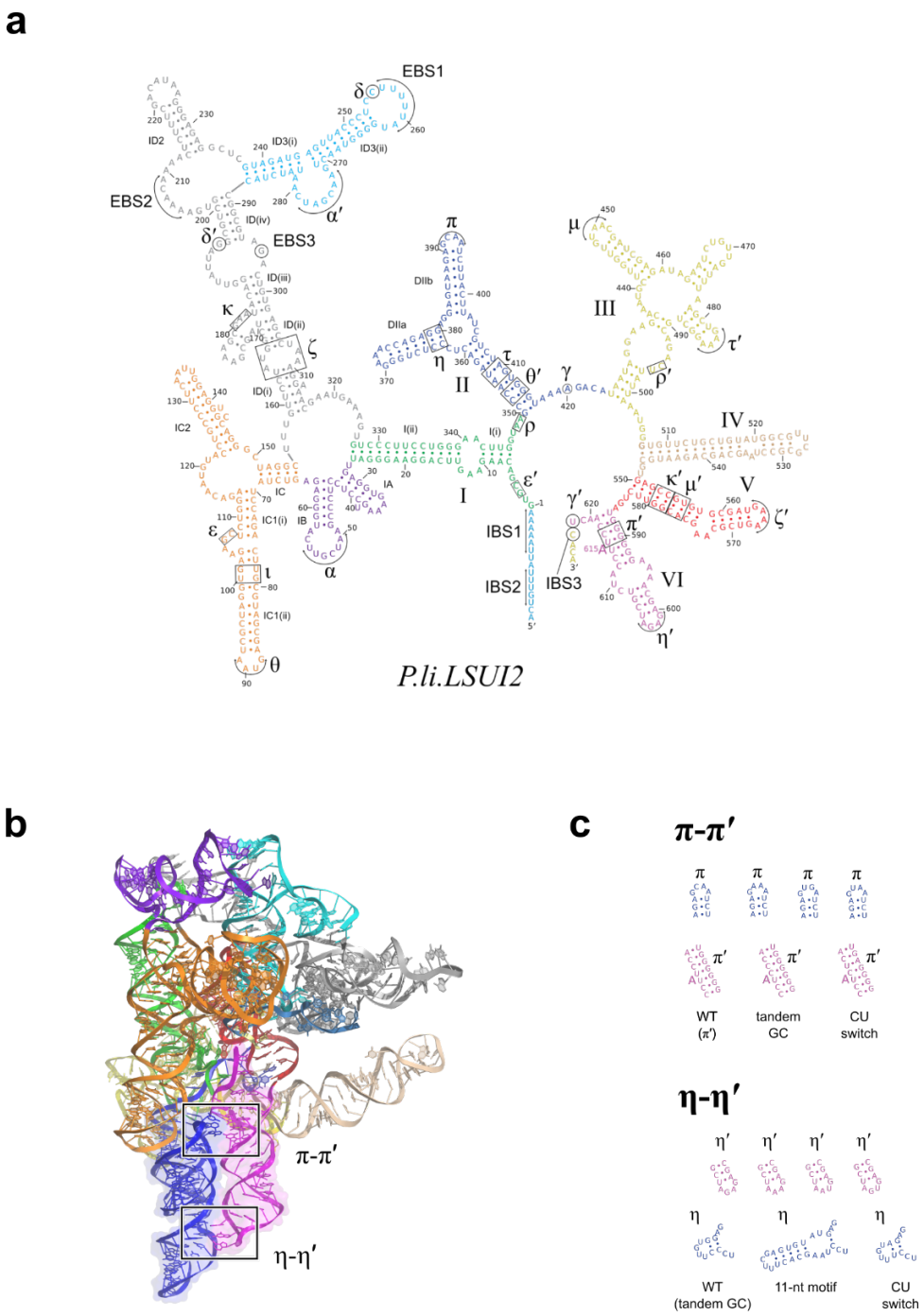

Figure 4.5: **TL-TLR modifications in *P.li.LSUI2*.** (a) Secondary structure of the *P.li.LSUI2* intron. The sequence shown was used for crystallization and contains modifications from the wt sequence in DII and DIV. (b) Crystal structure of *P.li.LSUI2* (4R0D). DII and DVI are highlighted with a surface rendering. DII is shown in blue and DVI in magenta. (c) DII and DVI sequences used for screening varied π-π′ and η-η′ strengths in *P.li.LSUI2*.

Not surprisingly, second step catalysis appears to be correlated with π-π′ stability. As π-π′ stability is decreased by ~2 kcal/mol, f_{lariat} decreases by ~32% (Figure 4.6b) and $f_{lariat\text{-}3'\ exon}$ increases by ~28 % (Figure 4.6a). Because decreasing the strength of π-π′ inhibits second step catalysis, the first step equilibrium is perturbed, which results in an increased fraction of precursor (Figure 4.6c). Although *P.li.LSUI2*-π constructs predominantly splices via the branching pathway, the hydrolytic pathway is used to a small extent, especially when π-π′ is destabilized (Figure 4.6d). These observations are recorded at equilibrium, however, and do not provide any insight into the rate of splicing.

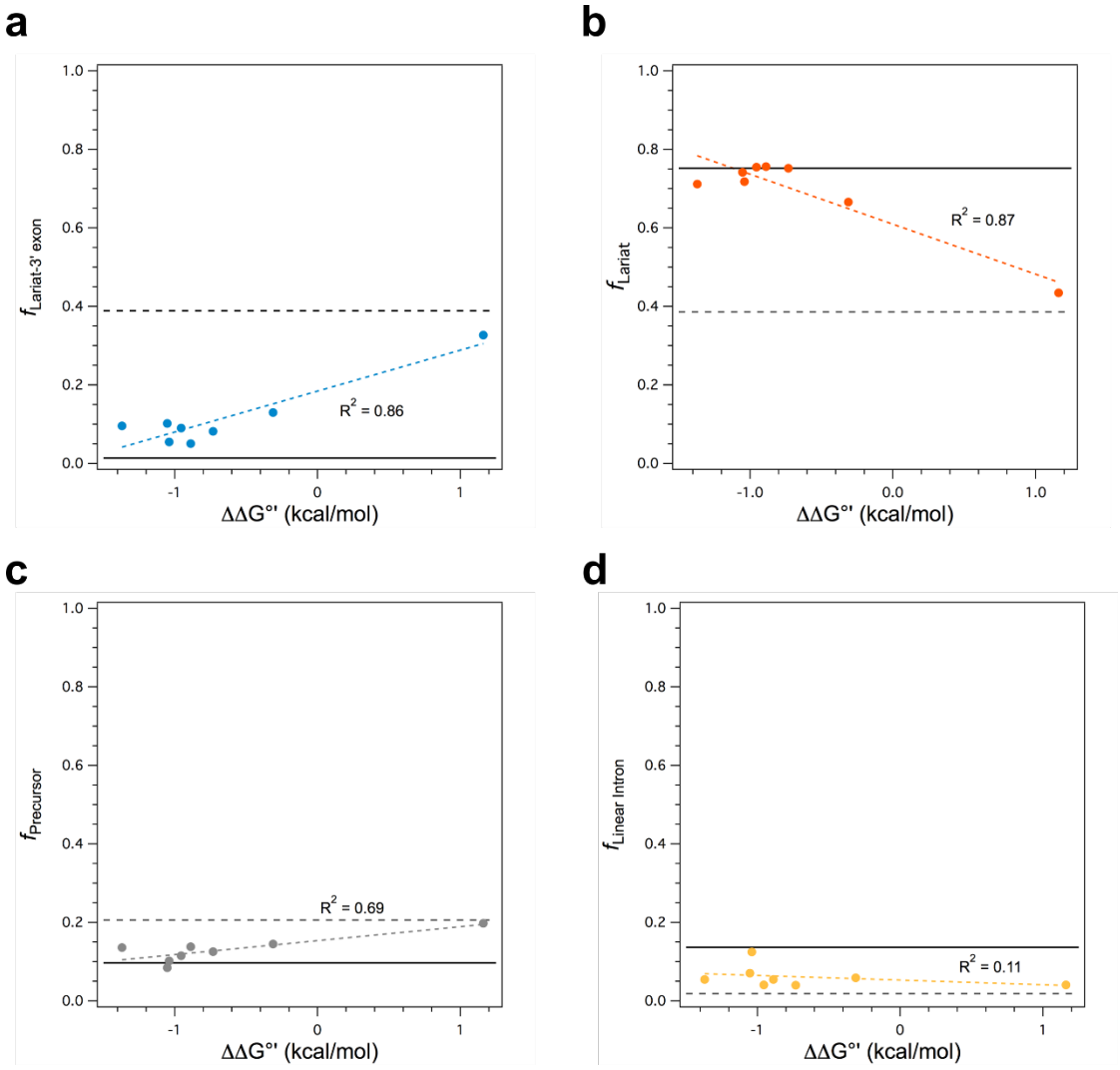

Figure 4.6: **Scatter plots showing fraction of splicing reactants and products vs. ΔΔG°′ for *P.li.LSUI2*-π constructs.** *P.li.LSUI2*-π constructs were made by modifying the π-π′ sequences to those indicated in Table 4.1[a]. The constructs were spliced for 3 hrs and run on a denaturing gel to quantify the fraction of (a) lariat-3′ exon, (b) lariat, (c) precursor (unspliced intron), and (d) linear intron, which were plotted against the ΔΔG°′ values listed in Table 4.1. Solid horizontal lines indicate wt values and dashed horizontal lines indicate Δπ values.

To investigate the effect the strength of π-π′ has on the rates of both steps of splicing, we performed kinetic studies on the pool of *P.li.LSUI2*-π constructs. The rate

of depletion of unspliced RNA was used to monitor the first step, and the rate of lariat + linear intron accumulation was used to monitor the second step (Figure 4.7). Both steps of *P.li.LSUI2* wt splicing are best fit as biphasic kinetic reactions. This is most likely due to the presence of two populations of the intron in solution that splice at different rates (K_{slow} and K_{fast}). However, when spliced under the same conditions as the equilibrium study (Figure 4.6), about 2/3 of *P.li.LSUI2*-π constructs exhibit monophasic kinetics where the rate is on the same order of magnitude as the wt-K_{slow}. Increasing the [$MgCl_2$] to 5 mM reintroduces the fast-splicing population, making the splicing kinetics biphasic and more easily comparable within the *P.li.LSUI2*-π screen.

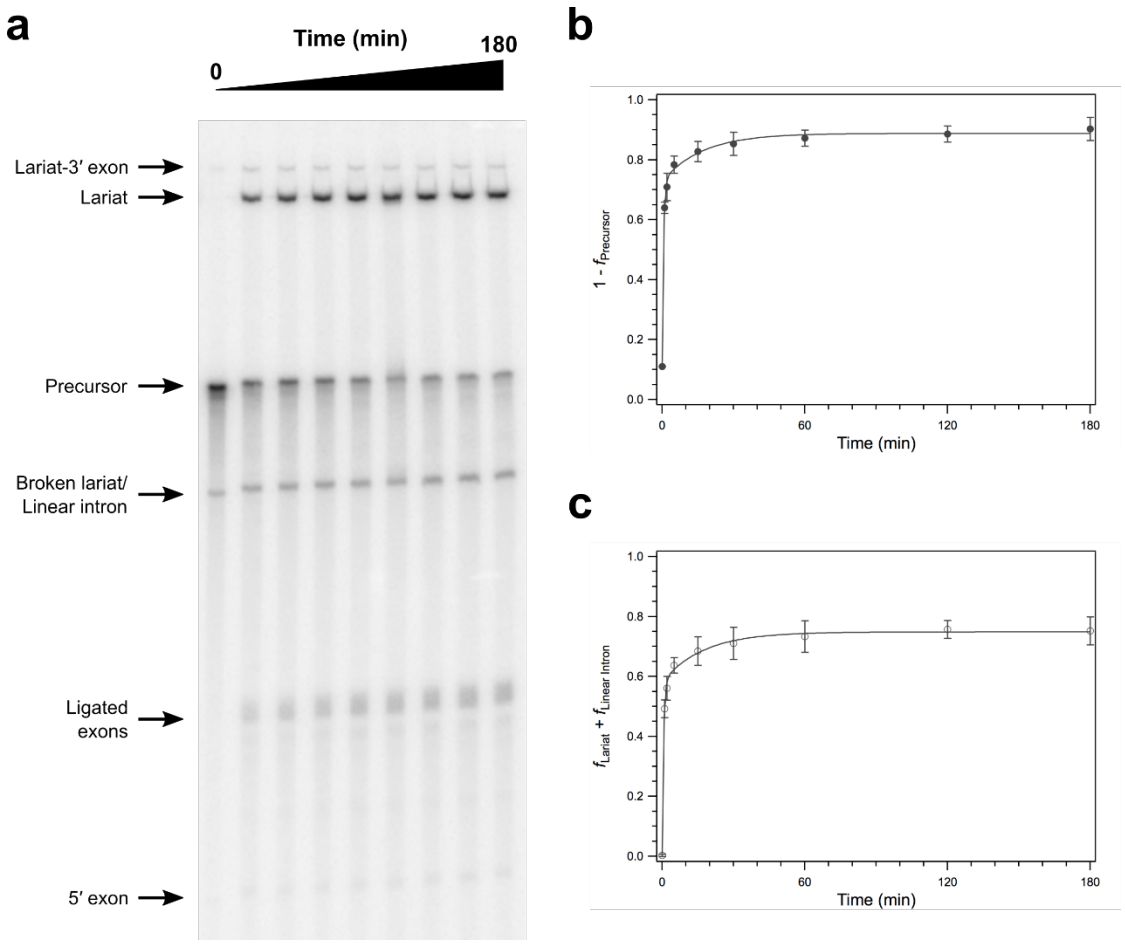

Figure 4.7: **Representative time-course for self-splicing of *P.li.LSUI2* and the kinetic analysis.** (a) Time-course for self-splicing of *P.li.LSUI2* wt splicing in 5 mM $MgCl_2$, 1 M NH_4Cl, and 40 mM Tris pH 7.5 at 30°C. (b) Plot of 1 - $f_{Precursor}$ vs. time (min). (c) Plot of $f_{Lariat} + f_{Linear\ Intron}$ vs. time (min).

The rates of the slower splicing population show an inverse correlation with π-π′ $\Delta\Delta G^{o\prime}$ (Figure 4.8a), though the fit is very poor indicating that it is unlikely K_{slow} follows a trend. The faster splicing population exhibits no change in rate as π-π′ stability is perturbed (Figure 4.8b). These results contrast the effect on K_{fast} when the π-π′ interaction is eliminated by mutating the TL to UUCG (Table 4.2). However, this increased K_{fast} for the first step is observed at 1 mM $MgCl_2$ but not 5 mM $MgCl_2$ (data

not shown), suggesting that the higher [$MgCl_2$] is able to compensate for ribozyme instability and conceal the effect π-π′ has on the rate of the first step.

The amplitude represents the percentage of the RNA population contributing to each rate. Examination of the first step of splicing shows no change in population distribution, however examination of the second step of splicing shows a significant decrease in the fast splicing population (Figure 4.8c,d). This indicates that the slower splicing population is independent of π-π′ formation.

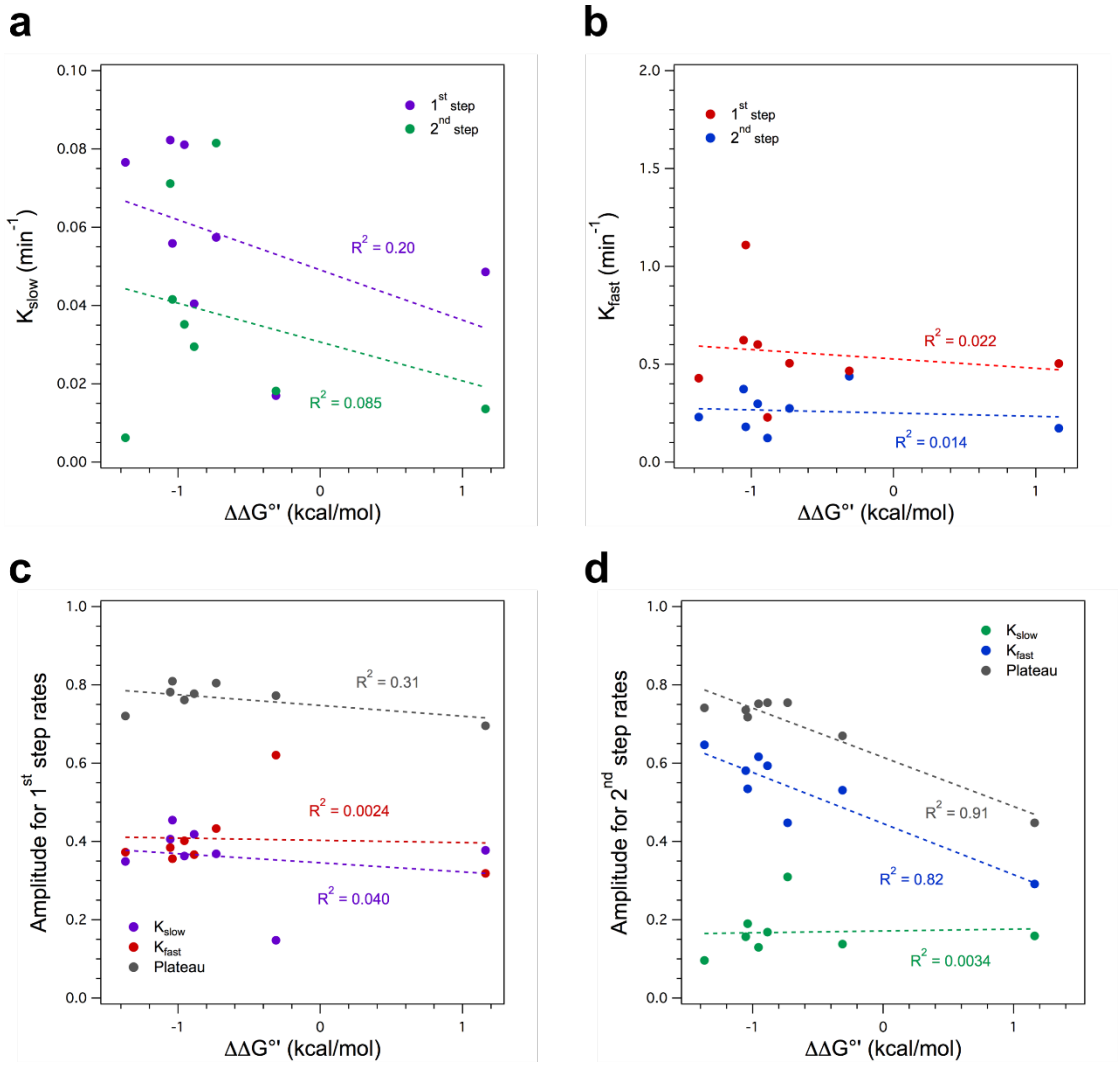

Figure 4.8: **Scatter plots showing splicing rates and amplitudes vs. ΔΔG°′ for *P.li.LSU1*-π constructs.** (a) K_{slow} vs. ΔΔG°′ for 1st step (purple) and 2nd step (green). (b) K_{fast} vs. ΔΔG°′ for 1st step (red) and 2nd step (blue). (c) Amplitude of 1st step K_{slow} (purple) and K_{fast} (red). The plateau of precursor depletion is shown in grey. (d) Amplitude of 2nd step K_{slow} (green) and K_{fast} (blue). The plateau of lariat and linear intron accumulation is shown in grey.

Altogether, this supports the current model that π-π′ is disengaged for the first step and engages prior to the second step to remove the bulged adenosine from the active site. Constructs containing a more stable π-π′ interaction efficiently engage after the

first step of splicing to sequester the bulged adenosine and allow exon ligation. As expected, *P.li.LSUI2* splices via hydrolysis less frequently with π-π′ destabilization because π-π′ is more frequently disengaged, allowing the first step nucleophile access to the active site where branching then takes place.

The *P.li.LSUI2*-η screen shows similar trends as the *P.li.LSUI2*-π screen for $f_{\text{lariat-3′ exon}}$ and $f_{\text{linear intron}}$ (Figure 4.9a,d), however η-η′ ΔΔG°′ variability induces smaller effects on splicing products and the data does not fit a linear trend as well. In fact, the constructs capable of forming the most lariat have ΔΔG°′ that are clustered between -1 and 0 kcal/mol (Figure 4.9b), indicating that the wt η-η′ sequence is among the best sequences for catalysis, and mutations that either increase or decrease stability are less efficient. η-η′ seems to have evolved from a "Goldilocks" scenario in which both weaker and tighter TL-TLR strengths are less efficient. This can also be observed in the K_{fast} amplitudes and rates for the second step of splicing (Figure 4.10b,d).

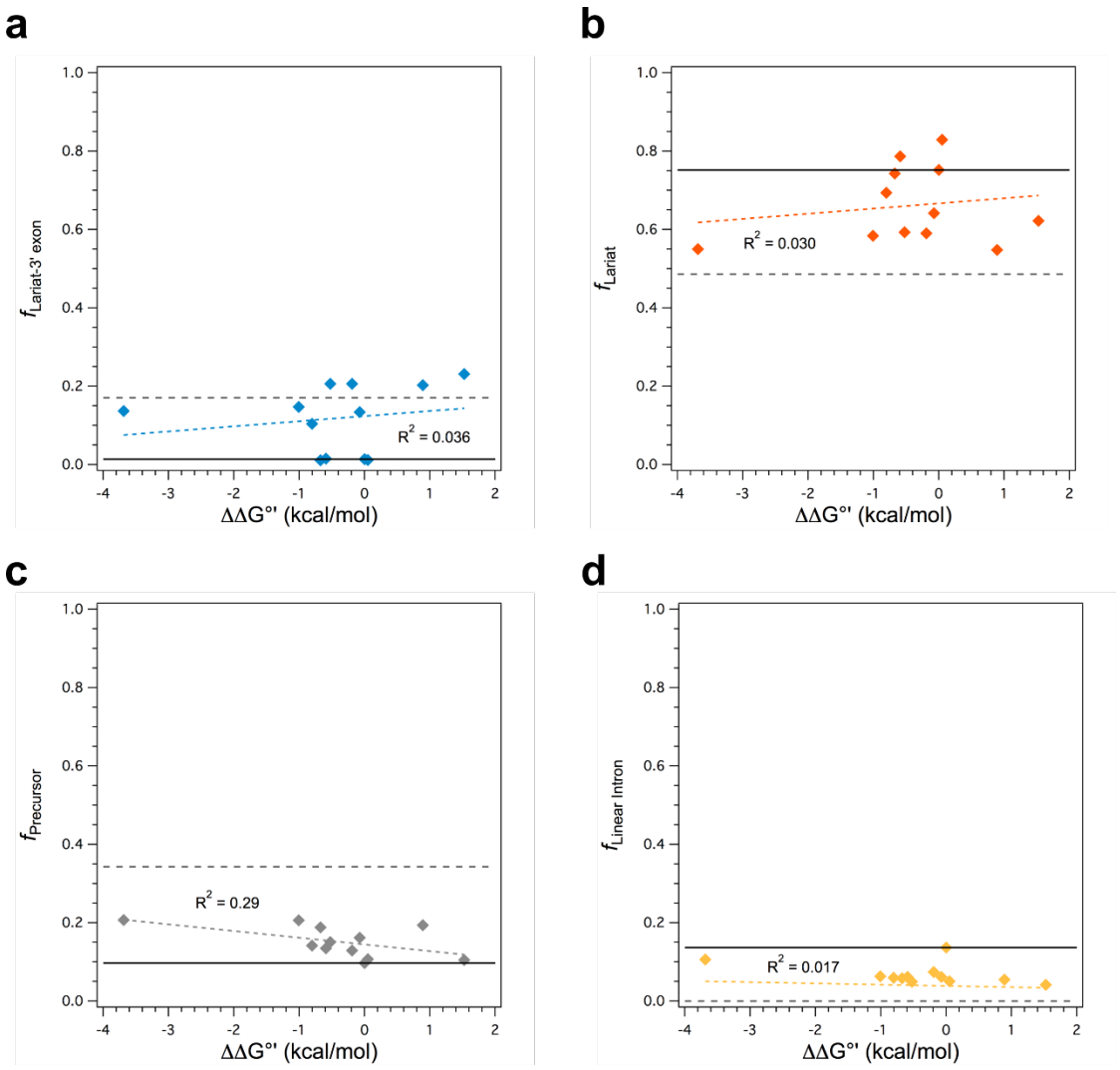

Figure 4.9: **Scatter plots showing fraction of splicing reactants and products vs. ΔΔG°′ for *P.li.LSUI2*-η constructs.** *P.li.LSUI2*-η constructs were made by modifying the η-η′ sequences to those indicated in Table 4.1[b]. The constructs were spliced for 3 hrs and run on a denaturing gel to quantify the fraction of (a) lariat-3′ exon, (b) lariat, (e) precursor (unspliced intron), and (f) linear intron, which were plotted against the ΔΔG°′ values listed in Table 4.1. (c) and (d) show the same data as (a) and (b) except without the outliers highlighted in (a) and (b). Solid horizontal lines indicate wt values and dashed horizontal lines indicate Δη values.

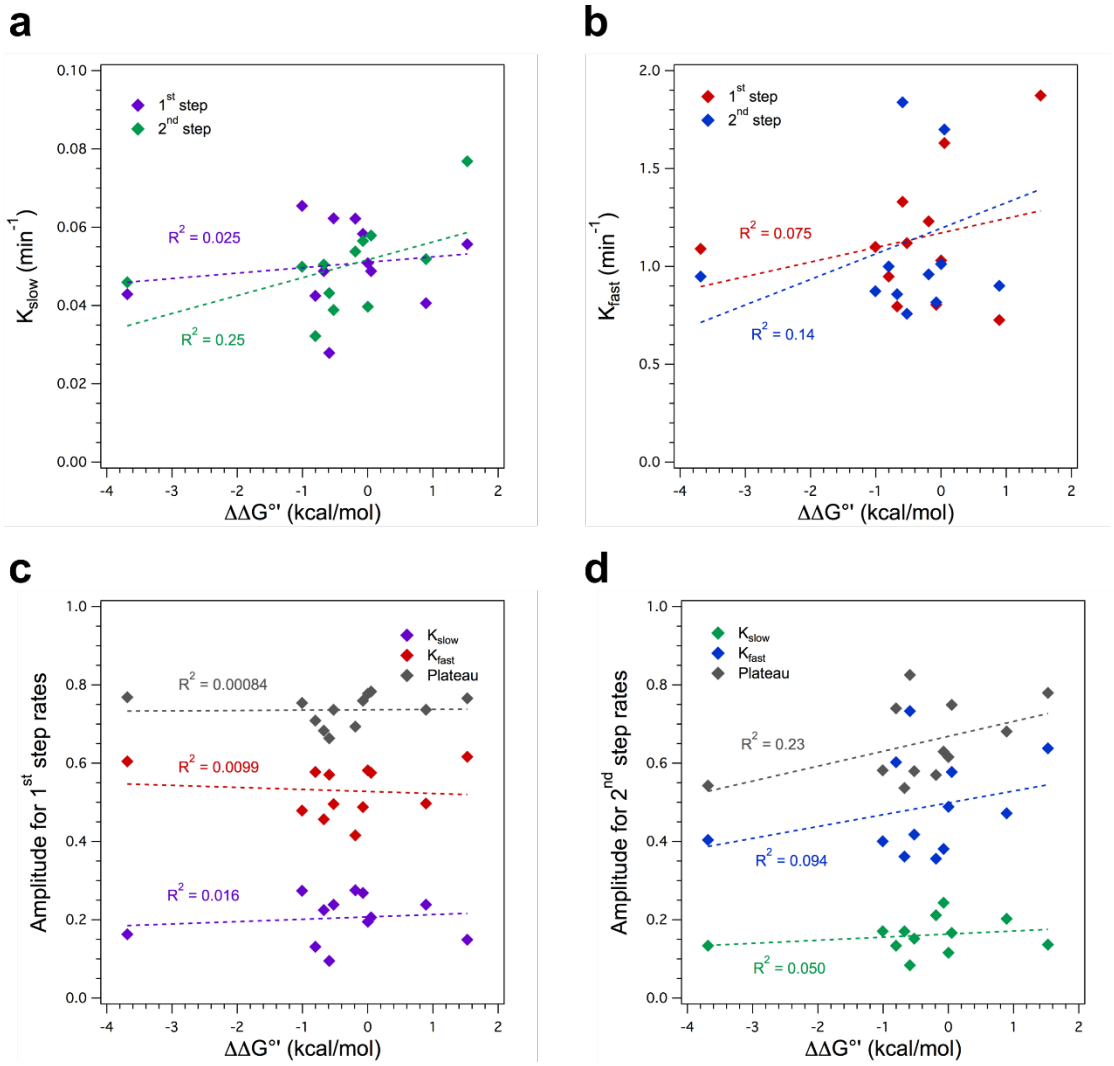

Figure 4.10: **Scatter plots showing splicing rates and amplitudes vs. ΔΔG°′ for *P.li.LSU1*-η constructs.** (a) K_{slow} vs. ΔΔG°′ for 1st step (purple) and 2nd step (green). (b) K_{fast} vs. ΔΔG°′ for 1st step (red) and 2nd step (blue). (c) Amplitude of 1st step K_{slow} (purple) and K_{fast} (red). The plateau of precursor depletion is shown in grey. (d) Amplitude of 2nd step K_{slow} (green) and K_{fast} (blue). The plateau of lariat and linear intron accumulation is shown in grey.

4.5 Domain VI is dynamic

Group IIC introns primarily splice via hydrolysis *in vitro* to form linear intron, nullifying the role of DVI in the first step[25]. However, recent work with the *Oceanobacillus*

iheyensis (*O.i.*) IIC intron has shown that replacing DIc and DVI of this intron with their counterparts from the *Azotobacter vinelandii* (*A.v.*) intron I2 results in *in vitro* lariat formation[26]. Crystal structures of this chimeric intron lariat, Oc19, exhibit different DVI conformations dependent on the presence of an unreactive 5′ exon[27]. The authors suggest that the bulged adenosine exists as a 1-nt bulge for the first step, and a 2-nt bulge for the second step of splicing. It had been originally proposed that the purine rich strand of DVI could pair with different opposite-strand pyrimidines to accommodate π-π′ conformation changes throughout splicing, though only the 1-nt bulge conformation was observed in the *P.li.LSUI2* crystal structure[17].

To test whether *P.li.LSUI2* exhibits similar DVI bulge dependencies, residues were mutated to stabilize either the 1-nt or 2-nt bulge conformations (Figure 4.11a). Construct G592C:C613G, which favors the formation of a 1-nt bulge, undergoes branch formation but cannot ligate exons even when incubated for two hours (Figure 4.11b). In comparison, wt intron splicing is nearly complete after only two minutes. These results are in contrast with those observed for the 1-nt bulge Oc19 chimera which forms wt amounts of lariat but does not properly ligate exons. Construct G591C:C613G, which favors the formation of a 2-nt bulge, does not splice at all via the branching pathway, but instead forms a very small amount of linear product (Figure 4.11b). In contrast, the analogous Oc19 construct primarily splices via hydrolysis but is still capable of lariat formation and forms ligated exons at the same rate as wt. Therefore, *P.li.LSUI2* does not exhibit similar DVI bulge dependencies. This is most likely due to the fact that the *P.li.LSUI2* intron is more evolved than either the *O.i.* or *A.v.* introns and the π-π′ and η-

η′ interactions have evolved to perform the same role as strand slippage in less evolved IIC introns.

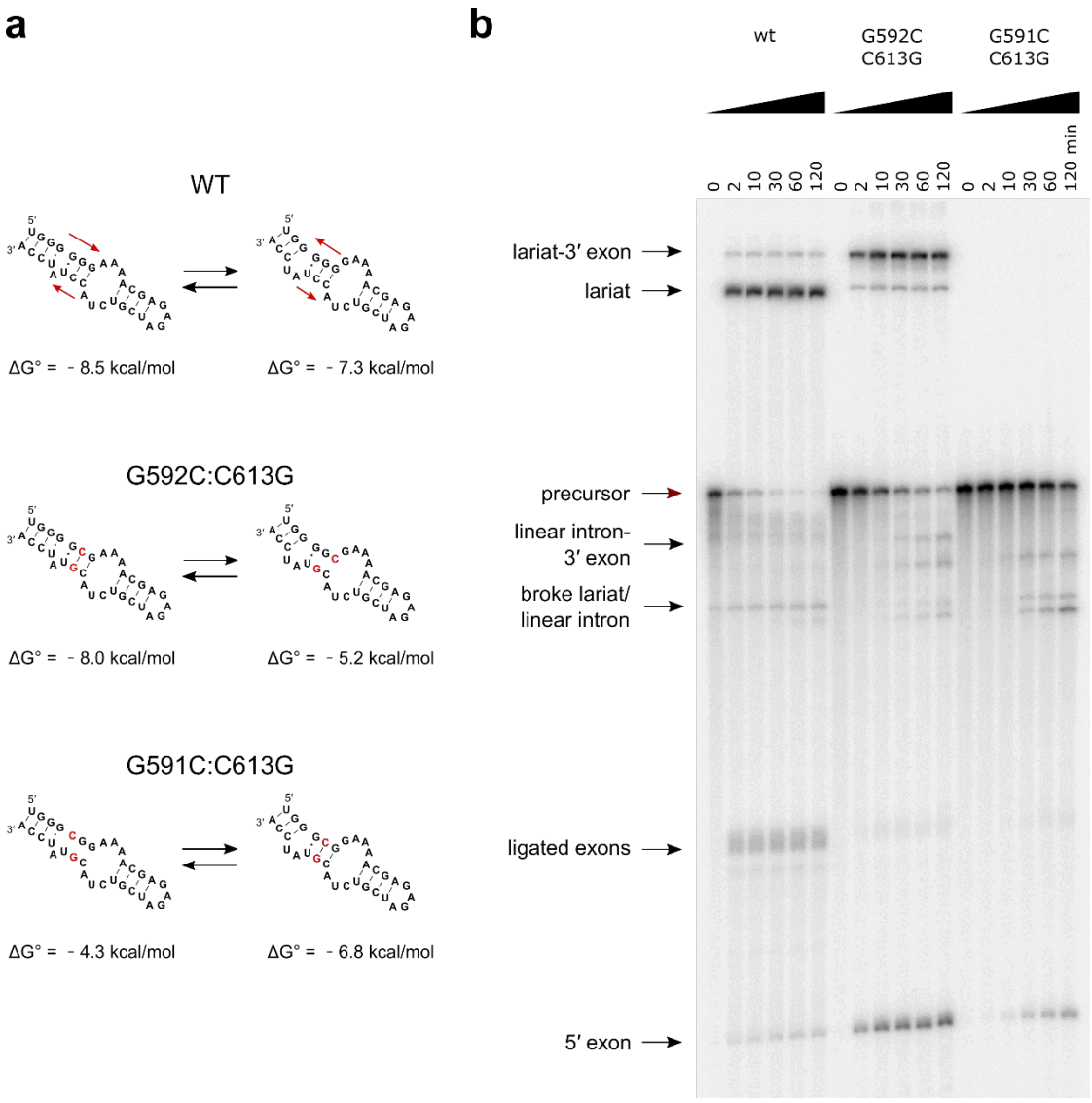

Figure 4.11: **Time-course for self-splicing of DVI bulge constructs**. (a) Equilibrium favored secondary structures of DVI mutants. (b) Time-course for self-splicing of constructs in (a).

4.6 Evolutionary constraints of TL-TLR interactions

P.li.LSUI2 is a chloroplast-like mitochondrial group IIB intron. The introns in this class contain highly conserved long-range tertiary interactions to properly fold the ribozyme[28]. A closer look at the frequency of π-π′ and η-η′ motifs reveals a preference for both interactions (Table 4.3). The most frequent tetraloops found in both interactions are GCAA and GUYA. In ~66% of the constructs in this class, the π′ receptor is tandem GC base pairs (CC:GG). The most frequently observed η receptor sequence is the GCC:GGC helix, found in ~47% of the constructs, with the CC:GG helix being the next most frequent motif in ~33% of constructs. The η sequence was determined by the first two nucleotides of the DIIa stem (Figure 4.5a). Interestingly, the CC:GG motif is found within the GCC:GGC motif and is shifted one base pair into the stem. It is likely that in the constructs containing a GCC:GGC receptor helix, the receptor is the tandem GC motif located one base pair in from the beginning of the stem. Assuming this to be the case, the thermodynamic stabilities of these most common TL-TLR combinations span ~0.3 kcal/mol (Table 4.1). This is a very narrow range compared to the ~2.5-5 kcal/mol tested in the *P.li.LSUI2*-π and -η screens.

Table 4.3: Sequences of π-π′ and η-η′ in group IIB introns

receptor helix*	π tetraloop**				η′ tetraloop**			
	GCAA	GUYA	GVGA	Other	GCAA	GUYA	GVGA	Other
CC:GG	22	21	3	1	5	16	1	2
GCC:GGC	0	0	0	0	11	19	4	1
UC:GA	0	1	3	1	0	2	0	0
CU:AG	2	5	0	2	0	1	2	0
Other	2	1	0	1	0	0	0	0

* Receptor nucleotides are shown in bold. The π′ receptor was determined to be the two base pairs adjacent to the bulged adenosine. The η receptor was determined to be the first two base pairs of the DIIa stem.
** Y represents pyrimidines, V represents any base except U

4.7 Methods

Preparation of RNA

P4-P6 and *P.li.LSUI2* derivatives were cloned into pUC57. To prepare templates for transcription, P4-P6 sequences were amplified using PCR and *P.li.LSUI2* plasmid derivatives were linearized with BamHI. Radiolabeled transcripts were prepared using T7 RNA polymerase, 25 mM $MgCl_2$, 40 mM Tris-HCl pH 7.5, 0.05% Triton X-100, 10 μCi [α-^{32}P]UTP (3,000 Ci mmol^{-1}), 0.5 mM UTP, and 1 mM other NTPs. RNA was transcribed for 1 hr at 37°C then gel purified on a denaturing 4% (19:1), 8 M urea, 1X TBE gel.

P4-P6 folding assay

P4-P6 folding assays were performed as previously described[20] with the following changes. Samples were heated at 55°C for 7 min then placed at 37°C for 15 min in 1X TB (89 mM Tris, 89 mM boric acid, pH ~8.5) with the desired concentration of $MgCl_2$. Nondenaturing 8% (29:1), 1X TB gels were prepared with $MgCl_2$ at the desired final concentration. Gels were run in the same buffer on a Hoeffer SE 600 gel

apparatus at 15 W at 37°C, dried, and exposed to a storage phosphor screen. Igor Pro software was used to analyze the results. Plots of relative electrophoretic mobility vs. $[Mg^{2+}]$ were fit to a Hill equation:

$$y = (\text{base} + (\text{max} - \text{base}))/(1 + \left(\frac{[Mg^{2+}]_{1/2}}{[Mg]}\right)^{n})$$

The Hill coefficient (n) and the concetration of Mg^{2+} needed to fold half of the molecules ($[Mg^{2+}]_{1/2}$ were then used to calculate $\Delta G°'$:

$$\Delta G°' = n\text{RTln}[Mg^{2+}]_{1/2}$$

$\Delta\Delta G°'$ values were obtained by subtracting the $\Delta G°'$ of the wt from the $\Delta G°'$ of the mutant:

$$\Delta\Delta G°' = \Delta G°'(\text{mutant}) - \Delta G°'(\text{wt})$$

***In vitro* self-splicing assays**

Constructs used for *in vitro* self-splicing assays were modifications of the wt *P.li.LSUI2* sequence with the DIV ORF removed and a 250 nt 5′ exon and 75 nt 3′ exon. RNA was prepared as above except that 5 mM $MgCl_2$ was used in the transcription buffer to prevent co-transcriptional splicing. To refold, radiolabeled precursor RNA and an internal control RNA in 40 mM Tris-HCl pH 7.5 was heated at 90°C for 1 min followed by incubation in 40 mM Tris-HCl H 7.5, 10 mM $MgCl_2$ for 15 min. To initiate self-splicing, RNA was mixed with an equal volume of 2X splicing buffer (2 M NH_4Cl, 40 mM Tris-HCl pH 7.5). Reactions were quenched by mixing with an equal volume of 80% formamide, 100 mM EDTA. Splicing products were resolved using denaturing 4% (19:1), 8 M urea, 1X TBE gels, exposed to a storage phosphor screen, and quantitated

using Quantity One 1-D Analysis software. All splicing assays were done in triplicate. Unequal loading was accounted for by normalizing to an internal control RNA. Band intensity was determined by dividing the background-subtracted intensity of each band by the number of Us in the RNA sequence corresponding to the band. All band intensities were then normalized to an unspliced control to give fractional values of input RNA. Rate constants for 5′ splice site cleavage were obtained by fitting plots of precursor depletion vs. time to a biphasic exponential equation. Rate constants for the second step were obtained by fitting plots of lariat and linear intron accumulation vs. time to a biphasic exponential equation. The biphasic nature of the two reactions are most likely from two differently folded populations in solution.

4.8 Acknowledgements

J.K.P. was supported by the UCSD Molecular Biophysics Training Program funded by NIH predoctoral training grant 5T32GM008326. This work was supported by NIH grant 5R01GM102216 awarded to N.T.

Chapter 4, in part, is currently being prepared for submission for publication of the material. Peters, J.K. & Toor, N. A kinetic and thermodynamic analysis of domain VI dynamics in the group II intron *P.li.LSUI2*. The dissertation author is the first author or this paper.

4.9 References

1. Woese, C. R., Winker, S. & Gutell, R. R. Architecture of ribosomal RNA: constraints on the sequence of tetra-loops. *Proc. Natl. Acad. Sci. U. S. A.* **87,** 8467–71 (1990).

2. Varani, G. Exceptionally Stable Nucleic Acid Hairpins. *Annu. Rev. Biophys. Biomol. Struct.* **24,** 379–404 (1995).

3. Woese, C. R., Gutell, R., Gupta, R. & Noller, H. F. Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acids. *Microbiol. Rev.* **47,** 621–69 (1983).

4. Molinaro, M. & Tinoco, I. Use of ultra stable UNCG tetraloop hairpins to fold RNA structures: thermodynamic and spectroscopic applications. *Nucleic Acids Res.* **23,** 3056–63 (1995).

5. Pley, H. W., Flaherty, K. M. & McKay, D. B. Model for an RNA tertiary interaction from the structure of an intermolecular complex between a GAAA tetraloop and an RNA helix. *Nature* **372,** 111–3 (1994).

6. Costa, M. & Michel, F. Frequent use of the same tertiary motif by self-folding RNAs. *EMBO J.* **14,** 1276–85 (1995).

7. Wu, L., Chai, D., Fraser, M. E. & Zimmerly, S. Structural variation and uniformity among tetraloop-receptor interactions and other loop-helix interactions in RNA crystal structures. *PLoS One* **7,** e49225 (2012).

8. Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Szewczak, A. A., Kundrot, C. E., Cech, T. R., Doudna, J. A. RNA tertiary structure mediation by adenosine platforms. *Science* **273,** 1696–9 (1996).

9. Ferre Â -D, A. R., Â, A., Zhou, K. & Doudna, J. A. A General Module for RNA Crystallization. *J. Mol. Biol.* **279,** 621–31 (1998).

10. Jaeger, L., Westhof, E. & Leontis, N. B. TectoRNA: modular assembly units for the construction of RNA nano-objects. *Nucleic Acids Res.* **29,** 455–63 (2001).

11. Geary, C., Baudrey, S. & Jaeger, L. Comprehensive features of natural and in vitro selected GNRA tetraloop-binding receptors. *Nucleic Acids Res.* **36,** 1138–52 (2008).

12. Reiter, N. J., Osterman, A., Torres-Larios, A., Swinger, K. K., Pan, T., Mondragón, A. Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA. *Nature* **468,** 784–789 (2010).

13. Fritsch, V. & Westhof, E. in *The Chemical Biology of Nucleic Acids* (ed. Mayer, G.) 141–174 (John Wiley & Sons, Ltd, 2010). doi:10.1002/9780470664001.ch7

14. Ikawa, Y., Naito, D., Aono, N., Shiraishi, H. & Inoue, T. A conserved motif in group IC3 introns is a new class of GNRA receptor. *Nucleic Acids Res.* **27,** (1999).

15. Ikawa, Y., Nohmi, K., Atsumi, S., Shiraishi, H. & Inoue, T. A comparative study on two GNRA-tetraloop receptors: 11-nt and IC3 motifs. *J. Biochem.* **130,** 251–5 (2001).

16. Costa, M. & Michel, F. Rules for RNA recognition of GNRA tetraloops deduced by in vitro selection: comparison with in vivo evolution. *EMBO J.* **16,** 3289–302 (1997).

17. Robart, A. R., Chan, R. T., Peters, J. K., Rajashankar, K. R. & Toor, N. Crystal structure of a eukaryotic group II intron lariat. *Nature* **514,** 193–197 (2014).

18. Marcia, M. & Pyle, A. M. Visualizing group II intron catalysis through the stages of splicing. *Cell* **151,** 497–507 (2012).

19. Chanfreau, G. & Jacquier, A. An RNA conformational change between the two chemical steps of group II self-splicing. *EMBO J.* **15,** 3466–76 (1996).

20. Silverman, S. K. & Cech, T. R. Energetics and cooperativity of tertiary hydrogen bonds in RNA structure. *Biochemistry* **38,** 8691–702 (1999).

21. Young, B. T. & Silverman, S. K. The GAAA tetraloop−receptor interaction contributes differentially to folding thermodynamics and kinetics for the P4−P6 RNA domain. *Biochemistry* **41,** 12271–12276 (2002).

22. Battle, D. J. & Doudna, J. A. Specificity of RNA-RNA helix recognition. *Proc. Natl. Acad. Sci.* **99,** 11676–81 (2002).

23. Jarrell, K. A., Peebles, C. L., Dietrich, R. C., Romiti, S. L. & Perlman, P. S. Group II intron self-splicing. Alternative reaction conditions yield novel products. *J. Biol. Chem.* **263,** 3432–9 (1988).

24. Chu, V. T., Adamidi, C., Liu, Q., Perlman, P. S. & Pyle, A. M. Control of branch-site choice by a group II intron. *EMBO J.* **20,** 6866–76 (2001).

25. Toor, N., Robart, A. R., Christianson, J. & Zimmerly, S. Self-splicing of a group IIC intron: 5' exon recognition and alternative 5' splicing events implicate the stem-loop motif of a transcriptional terminator. *Nucleic Acids Res.* **34,** 6461–71 (2006).

26. Monachello, D., Michel, F. & Costa, M. Activating the branch-forming splicing pathway by reengineering the ribozyme component of a natural group II intron. *RNA* **22,** 443–55 (2016).

27. Costa, M., Walbott, H., Monachello, D., Westhof, E. & Michel, F. Crystal structures of a group II intron lariat primed for reverse splicing. *Science (80-.).* **354,** aaf9258 (2016).

28. Dai, L., Toor, N., Olson, R., Keeping, A. & Zimmerly, S. Database for mobile group II introns. *Nucleic Acids Res.* **31,** 424–6 (2003).

Chapter 5: Crystal structure of a eukaryotic group II intron lariat

5.1 Abstract

The formation of branched lariat RNA is an evolutionarily conserved feature of splicing reactions for both group II and spliceosomal introns. The lariat is important for the fidelity of 5′ splice-site selection and consists of a 2′-5′ phosphodiester bond between a bulged adenosine and the 5′ end of the intron. To gain insight into this ubiquitous intramolecular linkage, we determined the crystal structure of a eukaryotic group IIB intron in the lariat form at 3.7 Å. This revealed that two tandem tetraloop-receptor interactions, η–η′ and π–π′, place domain VI in the core to position the lariat bond in the post-catalytic state. On the basis of structural and biochemical data, we propose that π–π′ is a dynamic interaction that mediates the transition between the two steps of splicing, with η–η′ serving an ancillary role. The structure also reveals a four-magnesium-ion cluster involved in both catalysis and positioning of the 5′ end. Given the evolutionary relationship between group II and nuclear introns, it is likely that this active site configuration exists in the spliceosome as well.

5.2 Introduction

Splicing of nuclear introns results in the formation of circular RNAs[1] having a branched lariat structure containing an unusual 2′-5′ phosphodiester bond[2,3]. This branched RNA product has also been found in group II introns[4,5], which are self-splicing ribozymes. Defects in lariat formation result in aberrant splicing and human disease[6]. In

higher eukaryotes, splicing of nuclear introns is catalyzed by a large ribonucleoprotein complex called the spliceosome, which is thought to share a common ancestor with group II introns[7,8].

Group II introns are catalytic RNAs with six structural domains (Figure 5.1) that splice via two trans-esterification reactions. In the first step of splicing, the 2′-OH of a bulged adenosine residue is the nucleophile that attacks the 5′ splice site to generate lariat RNA[4,5]. In the second step, the 3′-OH of the 5′ exon attacks the 3′ splice site to form ligated exons and excised intron lariat. The highly conserved domain V (DV) forms the group II intron active site by binding catalytic metal ions[9], and domain VI (DVI) contains the bulged adenosine used as the nucleophile in the first step of splicing[10].

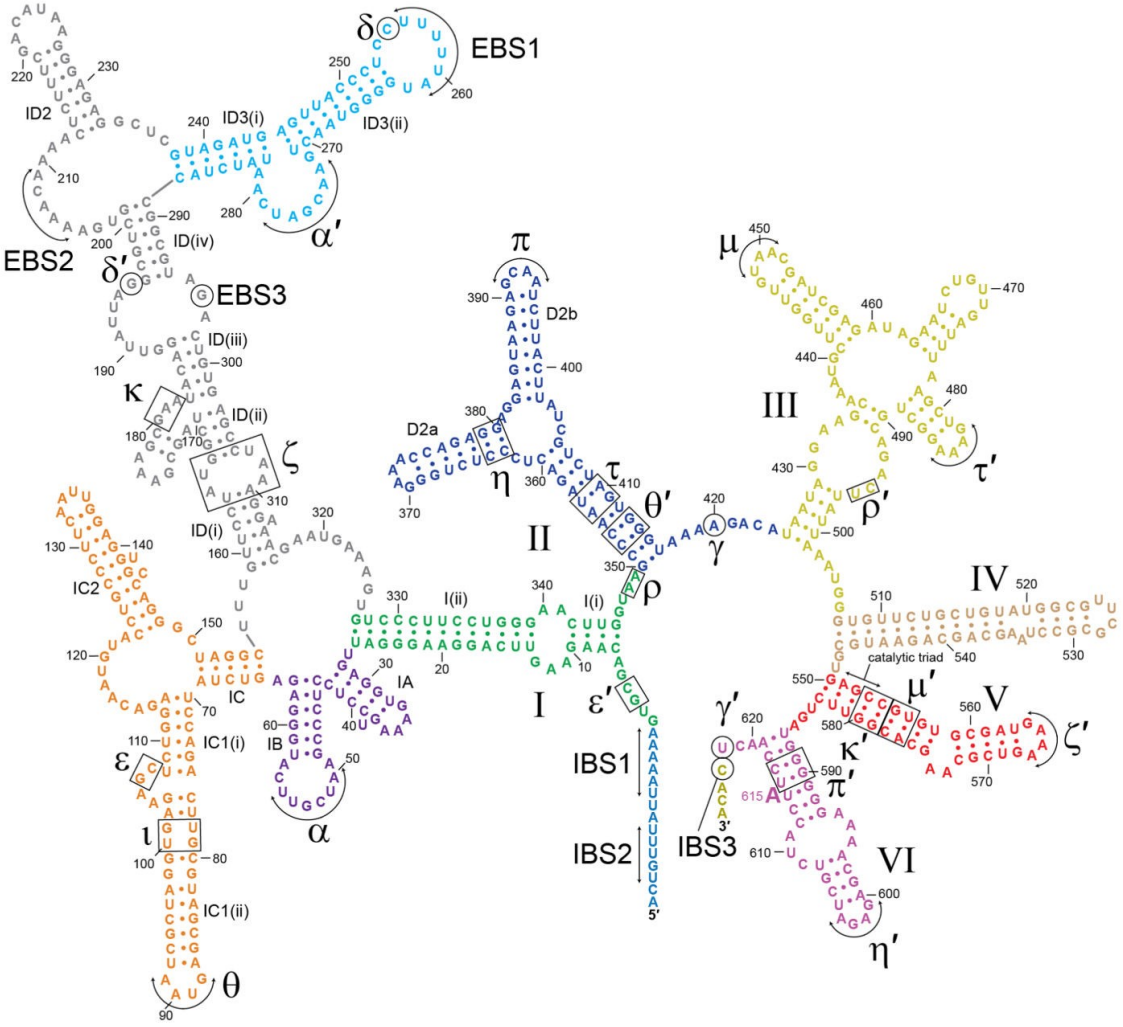

Figure 5.1: **Secondary structure of *P.li.LSUI2* intron crystallization construct.** Tertiary interactions are indicated with Greek letters and domains are labeled with Roman numerals. Colouring of the individual domains is consistent with the overall view of the tertiary structure shown in Fig. 5.3.

Group II introns are divided into three structural classes: IIA, IIB and IIC[11,12]. Historically, the two model systems used to study group II intron structure and function have been 'canonical' eukaryotic IIB introns: *P.li.LSUI2* from the brown algae *Pylaiella littoralis*[13] and aI5γ from the yeast Saccharomyces cerevisiae[4]. However, the only available crystal structure is of a IIC representative from the bacterium *Oceanobacillus*

iheyensis[9]. This idiosyncratic IIC intron class is the most primitive[14] and splices through hydrolysis to form linear intron[15]. In contrast, eukaryotic IIA and IIB introns form lariat, are evolutionarily later branching[14], and therefore more closely related to the spliceosome. We targeted the *P.li.LSUI2* intron for structure determination since it contains a functional DVI that forms large amounts of lariat during splicing[13].

5.3 Overall structure

Here we present the structure of the *P.li.LSUI2* intron in the post-catalytic lariat form with ligated exon product at 3.7 Å resolution (Table 5.1) solved using a Yb^{3+} derivative (Figure 5.2). This represents the first crystal structure of a 2′-5′ branched RNA molecule.

Table 5.1: Data collection and refinement statistics

	P.li.LSUI2 native	DV Triad Mutant	G79A Mutant	Yb^{3+}		Tl^{+}
Data collection				Peak	Inflection	
Space group	$C222_1$	$C222_1$	C222	$C222_1$	$C222_1$	$C222_1$
Cell						
a, *b*, *c* (Å)	163.7, 255.4, 136.8	161.9, 264.5, 137.5	211.7, 457.2, 179.5	164.6, 257.7, 138.0		161.2, 257.7, 136.3
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Wavelength (eV)	11218.5	12662	8949.5	8949.5	8946.5	12657.5
Resolution (Å)	150.0-3.68 (3.74-3.68)	50.0-7.25 (7.37-7.25)	50.0-9.75 (9.92-9.75)	50.0-4.65 (4.73-4.65)	50.0-4.49 (4.57-4.49)	50.0-4.48 (4.56-4.48)
R_{sym}	14.9 (>100)	9.5 (77.9)	8.6 (97.0)	8.8 (73.9)	9.2 (72.8)	8.8 (74.9)
$I/\sigma I$	6.4 (0.6)	13.2 (1.7)	14.3 (1.8)	16.1 (2.1)	9.2 (1.2)	21.8 (2.2)
Completeness	99.9 (99.9)	94.6 (82.5)	96.3 (96.9)	99.9 (99.6)	97.9 (88.7)	99.3 (99.6)
Redundancy	6.8 (3.8)	5.7 (4.7)	6.5 (6.0)	4.4 (3.7)	3.6 (2.8)	7.8 (6.4)
CC*	(0.743)					
Refinement						
Resolution (Å)	81.8-3.68					
No. reflections	31107					
R_{work}/ R_{free}	23.9/27.4					
No. atoms	13979					
RNA	13471					
Ligand/ion	393					
Water	115					
B-factors						
RNA	201.6					
Ligand/ion	181.5					
Water	177.0					
R.m.s.						
Bond lengths	0.017					
Bond angles	1.483					

*Highest resolution shell is shown in parenthesis.

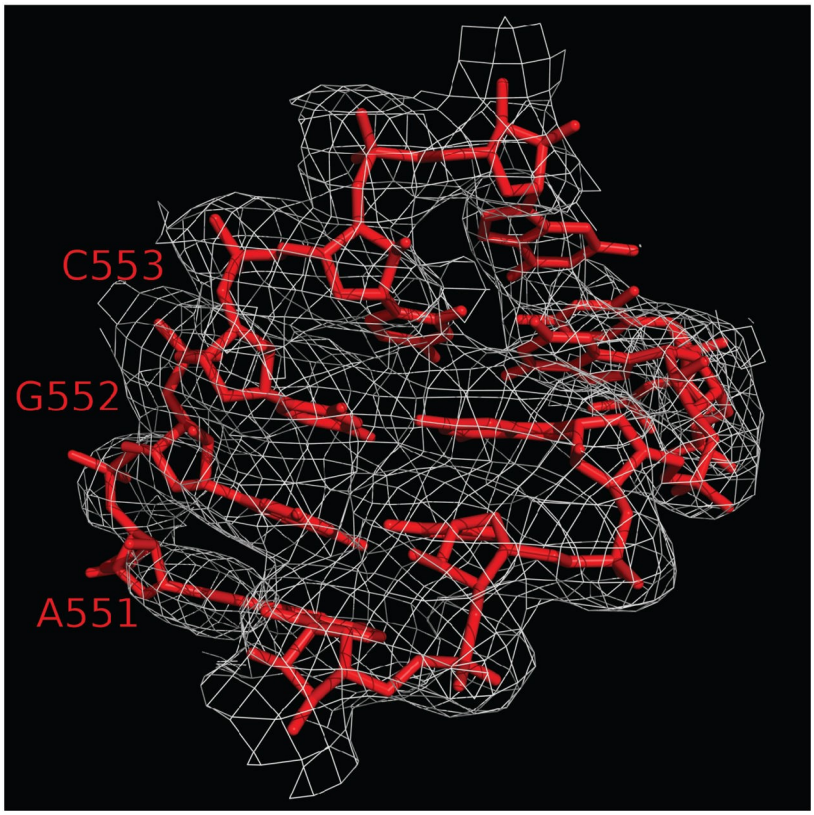

Figure 5.2: **The Yb-MAD experimental density modified map of the portion of DV containing the catalytic triad contoured at 1.8σ**.

Reflecting the ability of IIB introns to form lariat, there are a multitude of unique tertiary interactions in the *P.li.LSUI2* intron compared to the *O. iheyensis* structure (Figure 5.3). These newly visualized contacts include EBS2–IBS2 (Figure 5.4), μ–μ′, ε-ε′ and the canonical form of κ-κ′. Unlike the *O. iheyensis* structure, domains II and III interact with multiple domains through long-range interactions to stabilize the overall fold of the *P.li.LSUI2* intron. We can now visualize the location of DVI within the intron structure (Figure 5.3 and Figure 5.5), as well as the 2′-5′ lariat linkage between the first residue and the bulged adenosine.

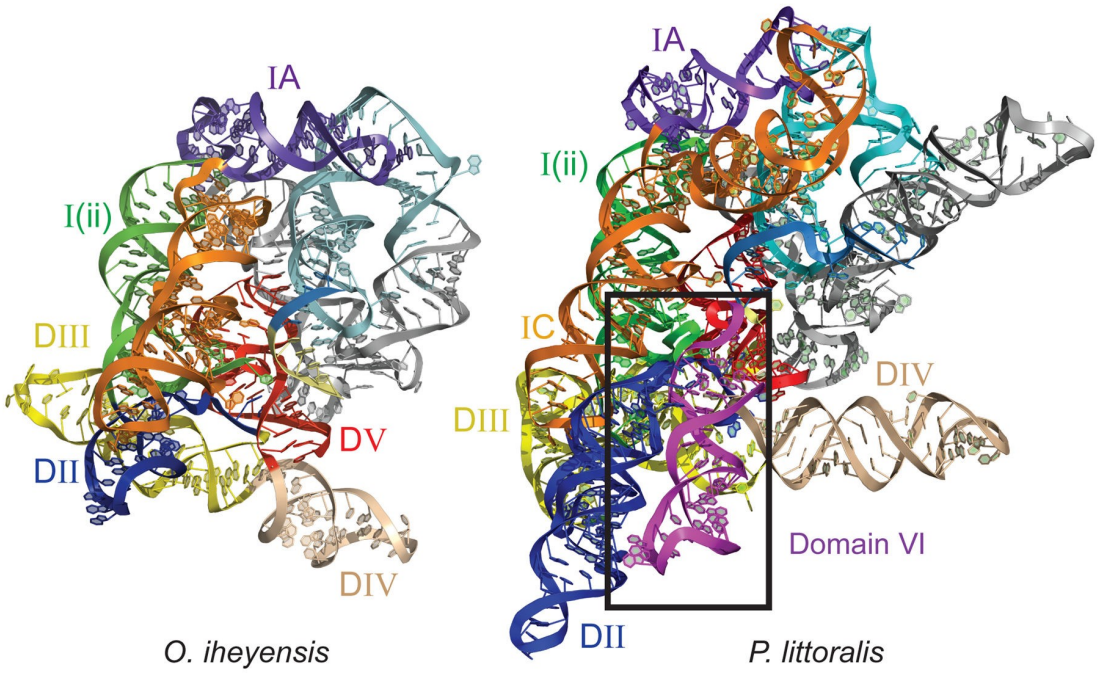

Figure 5.3: **A comparison of the tertiary structures of *O. iheyensis* and *P.li.LSUI2* group II introns.** *P.li.LSUI2* has a significantly larger and more complex structure with a correspondingly greater number of unique RNA tertiary contacts. Domain VI of *P.li.LSUI2* is highlighted within the box.

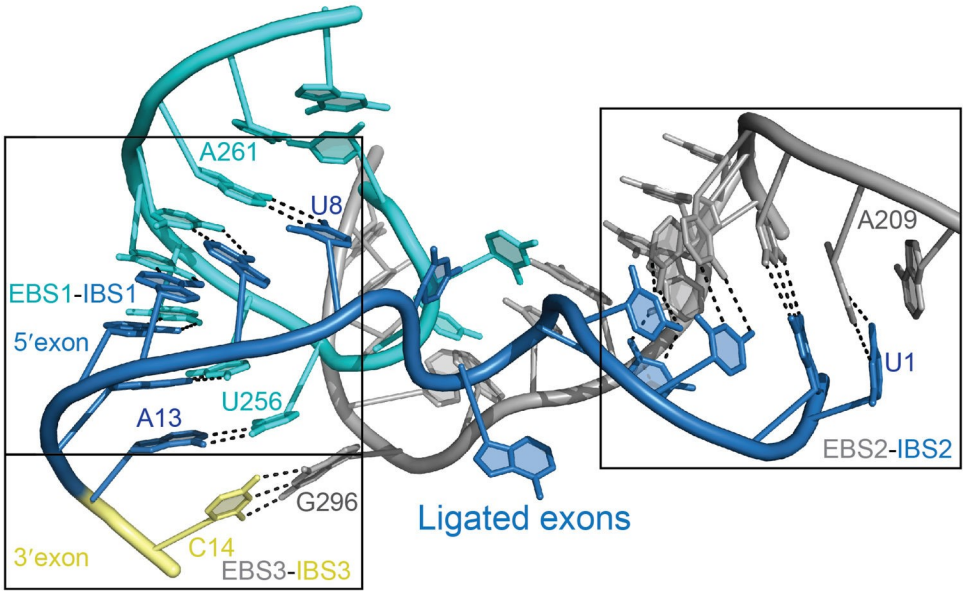

Figure 5.4: **The path of the 5′ exon through the intron structure.** The EBS1-IBS1 and EBS2-IBS2 interactions position the 5′ exon. They do not form a continuous binding interface with the presence of a highly distorted backbone at the junction between these two motifs. As a result, the helical axes of the EBS1-IBS1 and EBS2-IBS2 pairings are positioned ~90° relative to each other. The EBS3-IBS3 interaction places the 3′ exon in the active site.

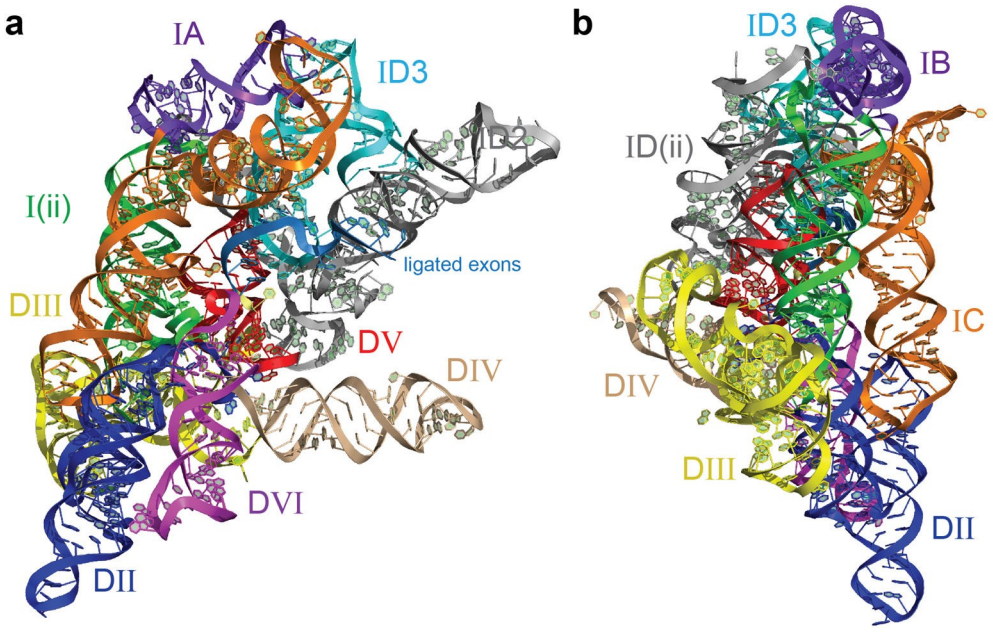

Figure 5.5: **Overall tertiary structure of the *P.li.LSUI2* intron.** Individual domains and subdomains are depicted in different colors. Domain names are labeled with Roman numerals. (a) and (b) show different rotations of the intron structure.

5.4 Newly visualized tertiary interactions

One of the most highly conserved tertiary contacts in group II introns is the κ–κ′ interaction between the base of the catalytic DV stem and domain I (DI)[16]. The conserved κ sequence GAA, nucleotide A171 from near the κ region, and residues from a GUAAC pentaloop in DIII converge to form a pentuple adenosine base stack (underlined residues) that inserts into the minor groove at the base of DV, rigidly placing the active site into the DI scaffold (Figure 5.6a and 5.7a).

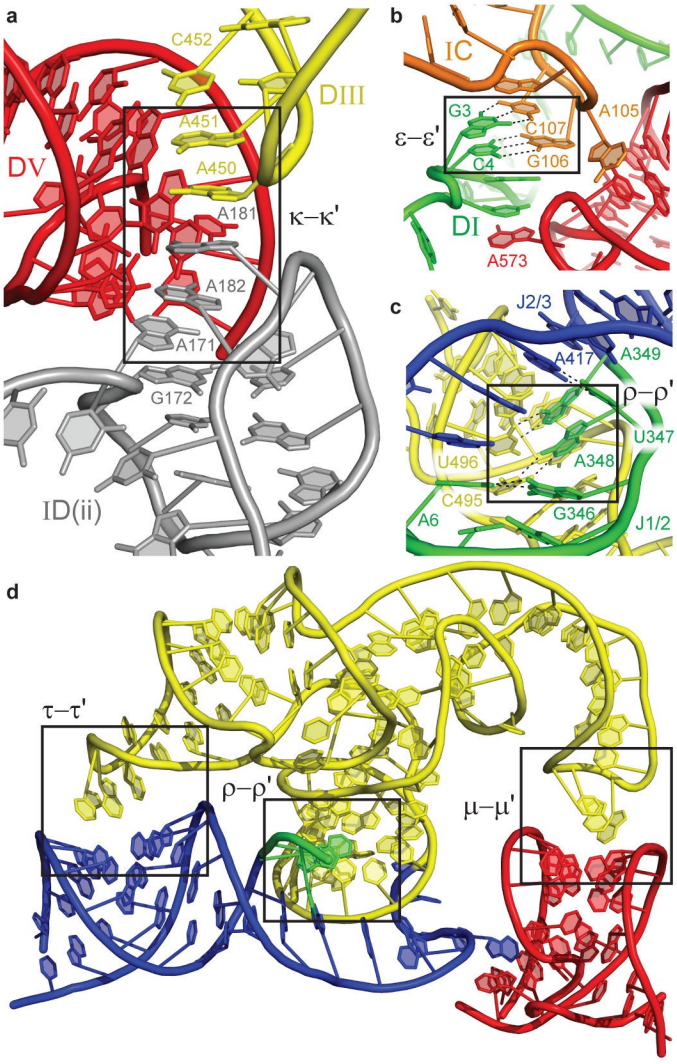

Figure 5.6: **Tertiary interactions in a IIB intron.** (a) The κ loop (grey) and DIII (yellow) converge to form an extended base stack involving five adenosine residues inserting into the minor groove of DV (red). (b) G106 and C107 form Watson-Crick pairs with C4 and G3 from the 5′ end to form the ε-ε′ interaction that positions the 5′ end in the active site. (c) J1/2 (green) interacts with the 5′ end (A6 residue in green), DIII (yellow), and J2/3 (blue). (d) DIII (yellow) acts as a brace on the surface of the intron and forms three tertiary interactions (boxed) with DI (green), DII (blue) and DV (red).

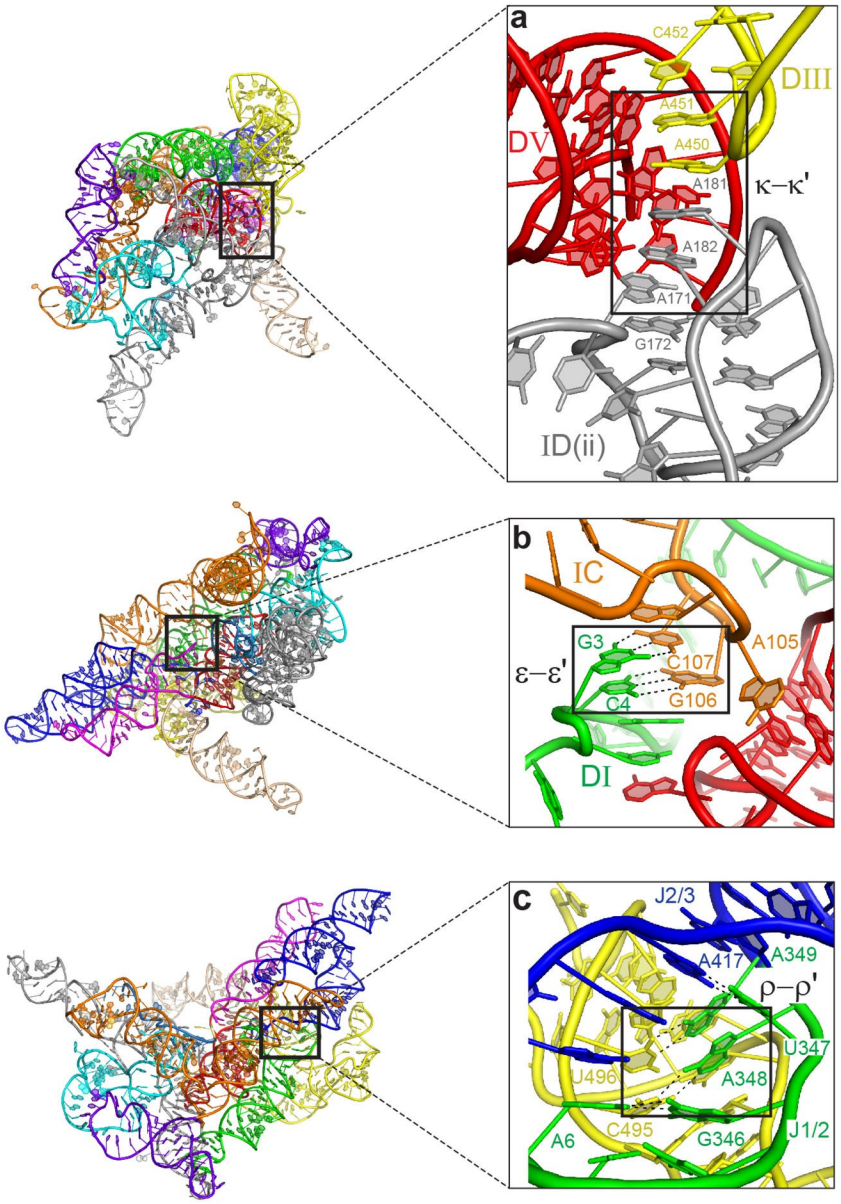

Figure 5.7: **Companion to Fig. 5.6 showing the location of the individual tertiary interactions relative to the overall structure**. (a) κ-κ'. (b) ε-ε'. (c) ρ-ρ'.

The ε–ε′ interaction is critical for catalysis, with disruption through mutagenesis resulting in complete loss of splicing activity[17]. This interaction consists of nucleotides G106 and C107 pairing with C4 and G3 from the 5′ end of the intron (Figure 5.6b and

Figure 5.7b). The end result of these contacts is the formation of five conserved bases stacking in the following order (from bottom to top): A573, U2, G5, C4 and G3. This serves to structure the 5′ end of the intron.

The conserved GUAA linker connecting domains I and II (J1/2) adopts an unusual backbone configuration that interacts with the 5′ end, the junction between domains II and III (J2/3), and DIII (Figure 5.6c and Figure 5.7c). The 5′ end and J2/3 directly interact with the active site through J1/2 positioning these regions to stabilize the core. This new long-range contact (designated as ρ–ρ′) consists of two adenosines from J1/2 docking into the basal stem of DIII (Figure 5.6c).

We can now correlate the function of DIII as a catalytic effector in group II introns. DIII interacts with the intron core through the aforementioned GUAAC pentaloop, which docks into the base of DV (Figure 5.6d). This μ–μ′ interactionf[18] serves to buttress the opposite side of DV from where catalysis takes place. Furthermore, a GAAA tetraloop from DIII interacts with the base of the DII stem (designated as τ–τ′) to provide additional reinforcement. Therefore, DIII functions as an external brace located on the outside surface to stabilize the entirety of the structure. This important role is consistent with deletion or mutagenesis of this domain rendering the intron unstable and not competent for efficient catalysis[19].

5.5 DII positions DVI in the active site

In the *O. iheyensis* structure, DII was markedly truncated to a small stem loop structure, and in many previous biochemical studies of the aI5γ intron, DII was similarly

shortened to study the first step of splicing. We can now visualize the intact DII substructure and find that it serves as a central hub for four different tetraloop receptor interactions (Figure 5.8a). DII makes contacts with domains I, III and VI to organize a large portion of the intron structure.

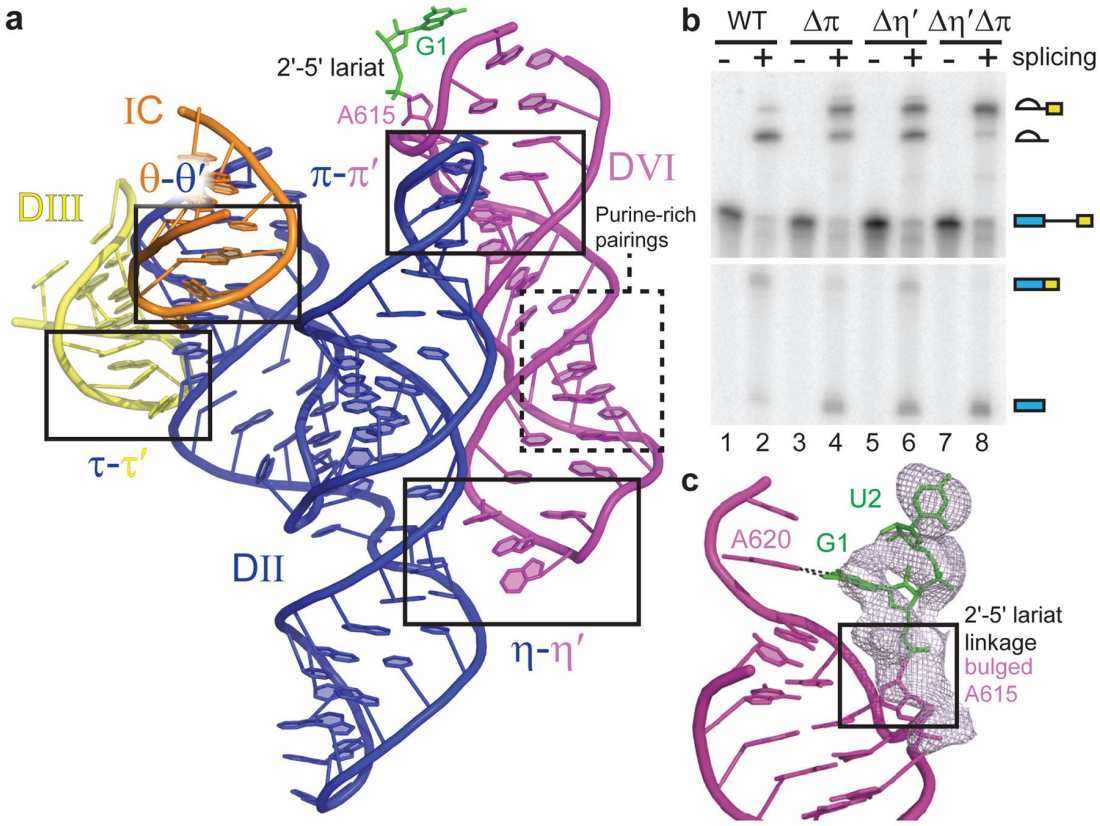

Figure 5.8: **The position of DVI within the intron structure.** (a) DII forms an inverted Y-shaped structure that engages in four distinct tetraloop-receptor interactions with the IC stem (orange), DIII (yellow), and DVI (purple). DVI interacts with DII via two tandem tetraloop-receptor interactions, η-η′ and π-π′. (b) *In vitro* self-splicing assays of the *P.li.LSUI2* intron. The wild-type (WT) intron efficiently catalyzes both steps of splicing and forms intron lariat and ligated exons. Mutagenesis of either η-η′ (Δη′)[20] or π-π′ (Δπ) inhibits the second step, resulting in the accumulation of lariat-3′ exon and 5′ exon. A combination of both mutations (Δη′Δπ) nearly blocks the second step of splicing with predominantly lariat-3′ exon present. (c) Fo–Fc density for the 2′-5′ lariat phosphodiester bond contoured at 3σ. The nucleobase of A615 is disordered and not visualized. This map was calculated using a model deleted for A615, G1, and U2 (shown in stick format) to avoid model bias. The 5′ (G1) and 3′ (A620) ends form a non-canonical base pair.

DII has a 'Y-shaped' RNA secondary structure with two stems, D2a and D2b (Figure 5.1), coaxially stacking on top of each other such that a tetraloop receptor from D2a and a GCAA tetraloop from D2b are facing the same side (Figure 5.8a). This

combination provides a binding interface for DVI, which contains both a GAGA tetraloop and a tetraloop receptor. Therefore, DVI is tightly placed in the core of the intron via two tandem tetraloop receptor interactions with DII. The interaction between D2a and DVI is known as the η-η′ contact[20] and we are designating the newly discovered interaction between D2b and DVI as π–π′. The π–π′ interaction is especially interesting due to its proximity to the bulged adenosine residue A615, which is the nucleophile for the first step of splicing. The π tetraloop interacts with nucleotides directly adjacent to the bulged adenosine (Figure 5.8a) and therefore probably has important effects upon the positioning of this nucleotide within the active site.

Mutagenesis of the GNRA tetraloops to UUCG was done to test the effects of these two interactions on splicing. Disrupting either contact significantly inhibited the second step of splicing, leading to an accumulation of lariat 3′ exon and 5′ exon (Figure 5.8b and Figure 5.9). However, disrupting both interactions simultaneously resulted in a near complete block of the second step. This indicates that π–π′ and η-η′ are synergistic interactions essential for the transition to the second step of splicing and probably function through moving the 3′ splice site (which is attached to DVI) into the active site.

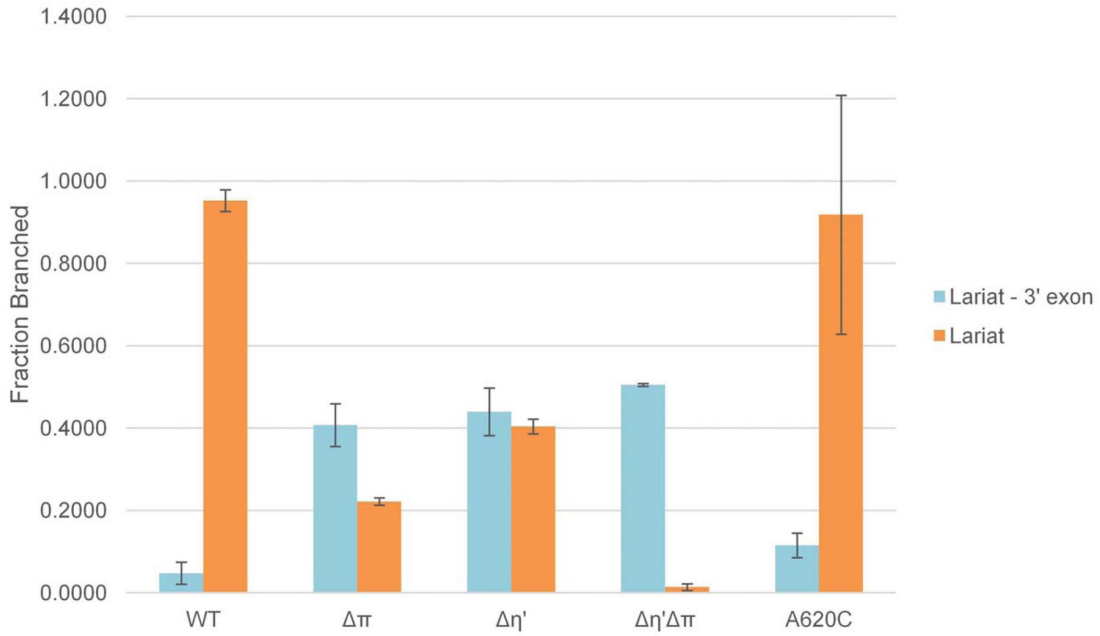

Figure 5.9: **Splicing assays for the DVI mutants showing the proportion of branched product.** Blue and orange bars correspond to lariat-3′ exon and lariat, respectively. The Δη′ and Δπ mutants accumulate large amounts of lariat-3′ exon, thus indicating a second step splicing defect. The Δη′Δπ double mutant is almost completely blocked before the second step. The A620C mutant shows 2.4-fold greater accumulation of lariat-3′ exon compared to the WT intron indicating that the interaction between G1 and A620 is important for the second step. In the yeast aI5γ intron, G1 instead interacts with the penultimate residue[21], indicating a certain degree of flexibility for this pairing. There is evidence for a similar interaction between the termini of nuclear introns[22] involving nearby (but not exactly equivalent) residues, which also has a significant effect upon the second step of splicing. Therefore, the 5′ and 3′ ends of nuclear introns may have a similar arrangement within the spliceosome.

We can now visualize the overall fold of DVI embedded within a catalytically active intron (Figure 5.8a). DVI contains a purine-rich internal loop, which forms non-canonical pairings that induce a slight bend in the helix, allowing DVI to simultaneously form both the π–π′ and η-η′ contacts with DII. These interactions position the ribose sugar of the bulged A615 directly under the 5′ end of the intron. The adenosine in this 2′-5′ lariat linkage is highly constrained, since it is connected to phosphates on three

different sides of the nucleotide (Figure 5.10a). The nucleobase component of the A615 residue is disordered (Figure 5.8c and Figure 5.10b), which is consistent with it not having a role in the later stages of splicing. In the current post-catalytic state, the lariat phosphodiester bond is located ~20 Å from the active site. Therefore, it has undergone a large-scale movement away from the catalytic core after the first step of splicing. In close proximity to the lariat bond, the 5′ and 3′ ends of the intron interact with each other through G1 forming a non-canonical base pair with A620 (Figure 5.8c), which is important for the second step (Figure 5.9).

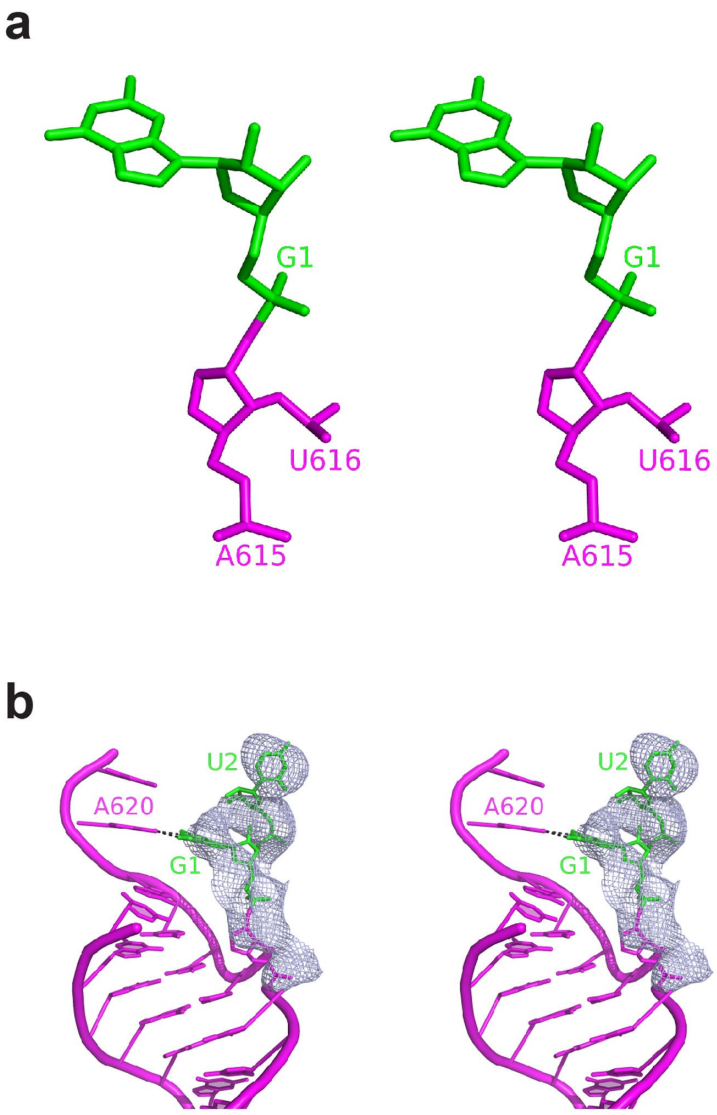

Figure 5.10: **Stereo depiction of the lariat bond.** (a) The lariat 2′-5′ phosphodiester bond in wall-eyed stereo format. (b) Stereo version of Fig. 3c. See Fig. 3c legend for details. Fo-Fc density for the lariat bond contoured at 3σ.

5.6 Active site metal ion configuration

Mg^{2+} ions are an absolute requirement for the catalysis of RNA splicing. To identify active site metal ions, soaks were performed using the anomalous scatterer Yb^{3+}, which exhibits the same octahedral coordination geometry as Mg^{2+} and

preferentially binds to sites containing highly coordinated magnesium ions[9,23]. This revealed four large anomalous peaks in the ribozyme core (Figure 5.11a). Two of these peaks (M1 and M2) are also found in *O. iheyensis*[9] and are embedded within DV to coordinate to the junction phosphate between the ligated exons, while the other two peaks (M3 and M4) are coordinated to the 5′ end of the intron.

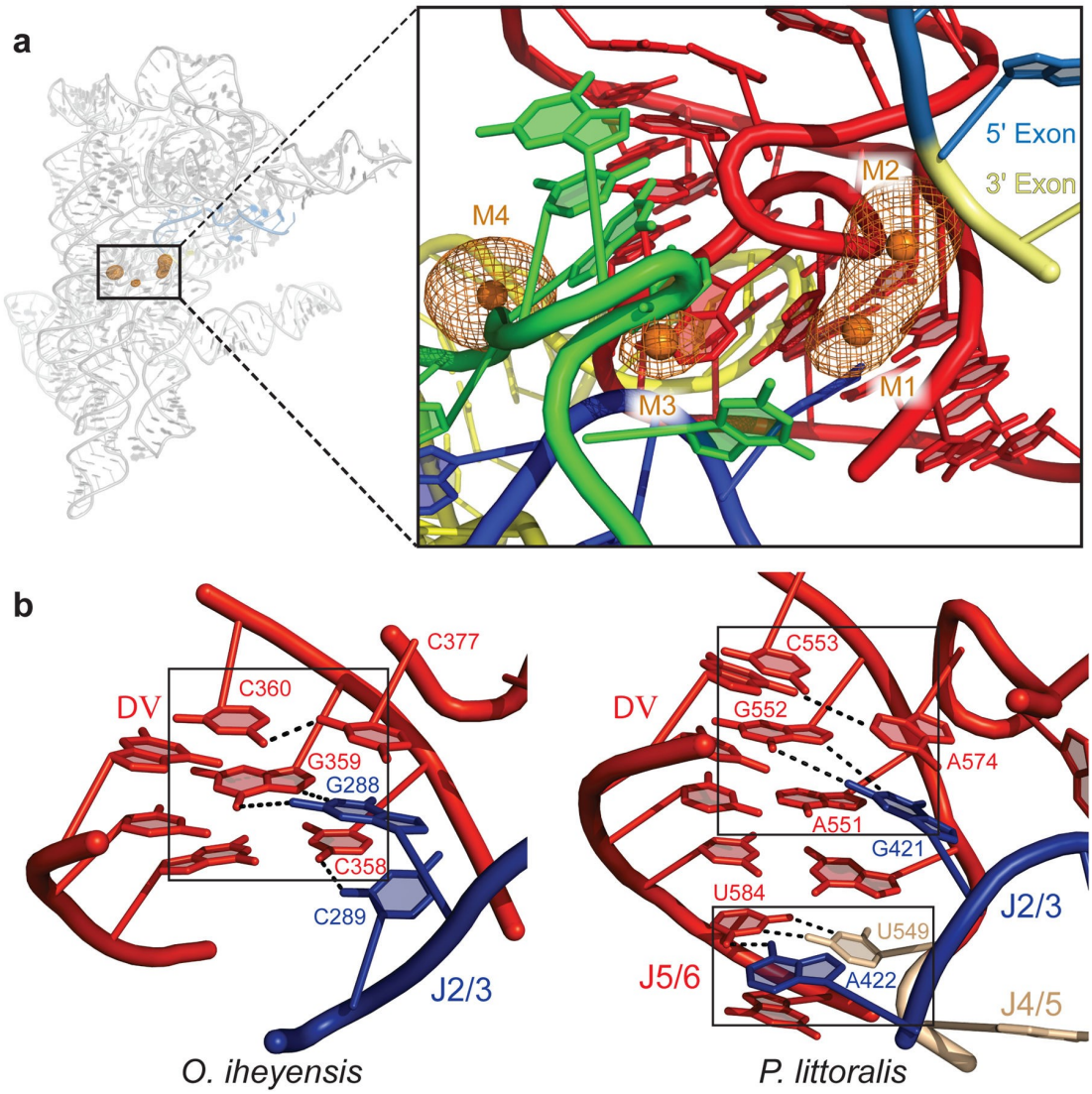

Figure 5.11: **The core of the of *P.li.LSUI2* intron.** (a) Yb^{3+} anomalous map (orange mesh) contoured at 12σ reveals four large peaks, which correspond to highly-coordinated magnesium ions (orange spheres). (b) Comparison of the catalytic triplexes found in the *O. iheyensis* and *P.li.LSUI2* introns. *O. iheyensis* contains a continuous triple helix spanning between the DV catalytic triad (CGC) and J2/3. In *P.li.LSUI2*, the J2/3 residue A422 is disengaged from the catalytic triad (AGC) and forms a base triple with J4/5 and J5/6 nucleotides (U549 and U584, respectively).

M3 is located in a binding pocket formed by the highly conserved 5′ end (Figure 5.12a), which has the sequence GUGCG. Nucleotides in this region exhibit a highly contorted backbone configuration that wraps around all sides of M3 due to the ε–ε′ interaction involving residues G3 and C4 (Figure 5.6). Therefore, we postulate that the primary function of ε-ε′ is to order a crucial metal-binding platform which structures the 5′ end of the intron.

M4 coordinates to conserved IIB intron residues A6, C7 and A341 to stabilize the 5′ end further (Figure 5.12b). These residues are in close spatial proximity to the recently proposed ι motif (Figure 5.1) that is known to promote lariat formation and is predicted to serve as a receptor for positioning DVI (ref. [24]). An ι G79A point mutant was crystallized, revealing strong signals for M1 and M2; however, M3 and M4 were no longer visible (Figure 5.12c, d). Splicing assays of this mutant also show a negative effect on the first step of splicing (Table 5.2). Furthermore, the η–η′ interaction, which forms the metal-binding platform for M3 and M4, persists throughout group II intron catalysis[17]. Taken together, the data suggest that M3 and M4 participate in the first step of splicing by positioning the 5′ splice site in the active site to present the scissile phosphate to M1, M2 and the bulged adenosine. However, it is possible that M3 exists only in the post-catalytic state to stabilize the repositioned lariat bond. Given the rarity of highly coordinated magnesium sites in large RNAs[25], the existence of four such metal ions in close proximity in the active site of *P.li.LSUI2* is striking. In addition, we observe a conserved monovalent ion near the M1/M2 catalytic center (Figure 5.12e).

Table 5.2: Kinetic analysis of *P.li.LSUI2* RNA splicing

	kfast (min−1)	kslow (min−1)
WT	9.51 ± 1.09	0.136 ± 0.0246
G79A	2.97 ± 1.57	0.0767 ± 0.00942

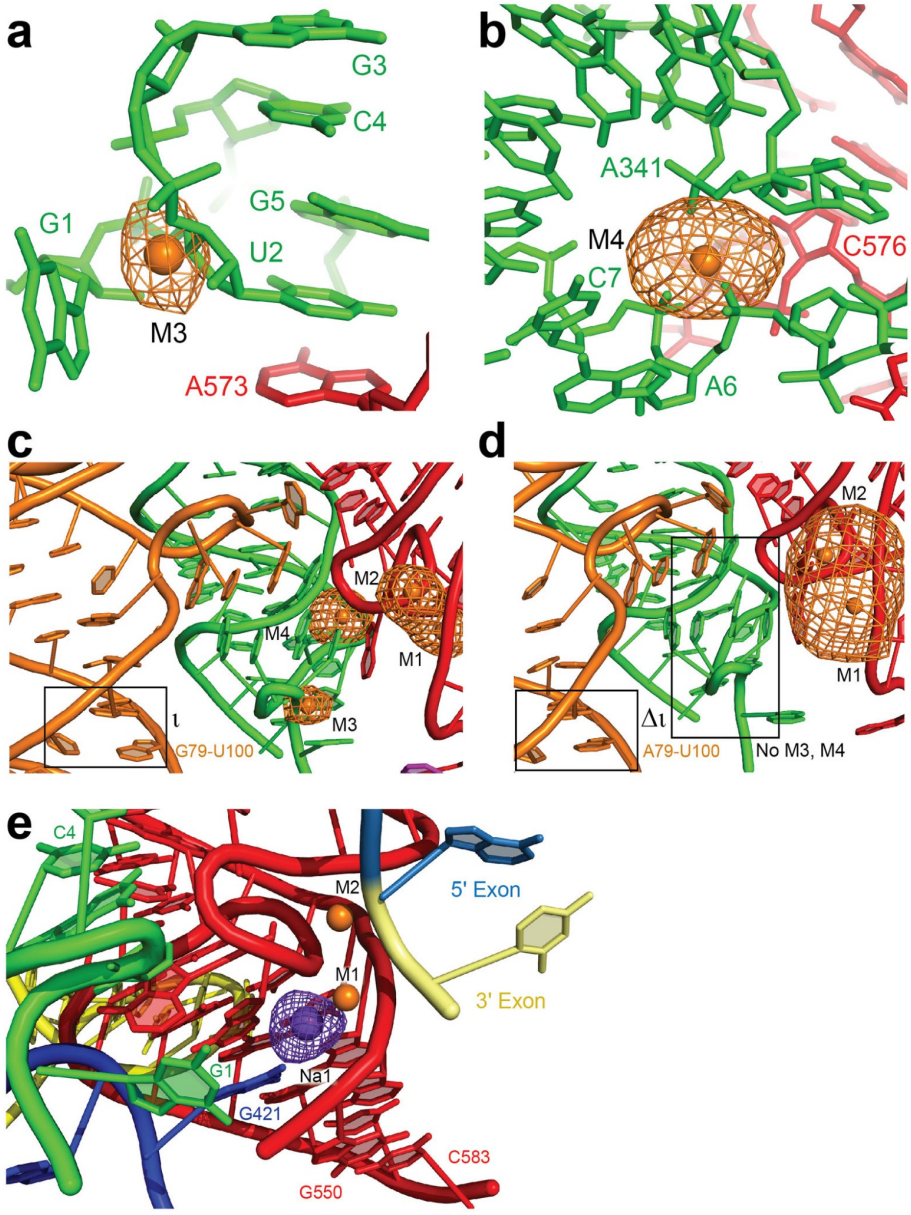

Figure 5.12: **Anomalous peaks in the core of the intron.** (a) and (b) depict the RNA ligands surrounding metals M3 and M4, respectively. Yb^{3+} anomalous map contoured at 9σ. (c) Yb^{3+} anomalous map for WT contoured at 9σ. (d) Compared with the WT intron, the Yb^{3+} anomalous map for the ι G79A mutant (contoured at 4σ) is lacking the peaks corresponding to M3 and M4, even at a lower contour level. (e) Tl^{+} was used as a probe for monovalent ions in the RNA structure[26]. The Tl^{+} anomalous map (purple mesh contoured at 5.5σ) revealed a strong peak located 3.8 Å from M1 that coordinates to the nucleobase of J2/3 residue G421 and the backbone of DV nucleotide G550. This sodium ion Na1 (purple sphere) is significantly closer to M1 than the equivalent K^{+} ion found in *O. iheyensis*[25]. Otherwise, this monovalent ion binding site is relatively conserved between these two introns.

5.7 Catalytic triplex rearrangement

In the *O. iheyensis* structure, J2/3 and residues from DV form a catalytic triplex structure essential for the catalysis of splicing and that is also found in the spliceosome[27]. In the *O. iheyensis* triplex, J2/3 residues G288 and C289 (analogous to G421 and A422 in *P.li.LSUI2*) form base triples with the first two nucleotides of the catalytic triad. Unexpectedly, we observe in the *P.li.LSUI2* structure that one of the J2/3 residues is completely disengaged from the catalytic triad. Specifically, A422 has moved away from the triad and stacks directly underneath the base of the DV helix to form a base triple with residues from the J4/5 and J5/6 linkers (Figure 5.11b). The possibility of J2/3 participating in conformational changes has been previously postulated[28]. However, we can now see the specific nature of this 'switch' with the disengagement of J2/3 from the catalytic triad into an alternate configuration. The linkers between domains are highly conserved for each subclass of group II introns, and based on the *P.li.LSUI2* structure we hypothesize that they are dynamic and modulate splicing.

5.8 π-π' is a dynamic interaction

DVI is proposed to engage in large-scale conformational changes between the two steps of splicing[20,24]. To test this model, a catalytic triad mutant (AGC→GAU) inactive for splicing was crystallized and solved at 7 Å (Figure 5.13). Strong electron density was observed for the η–η′ interaction between DII and DVI, indicating that this contact persists throughout both steps of splicing and that there is no large-scale change

in DVI position. Therefore, a new model is required to explain how DVI mediates the transition between the two steps of splicing.

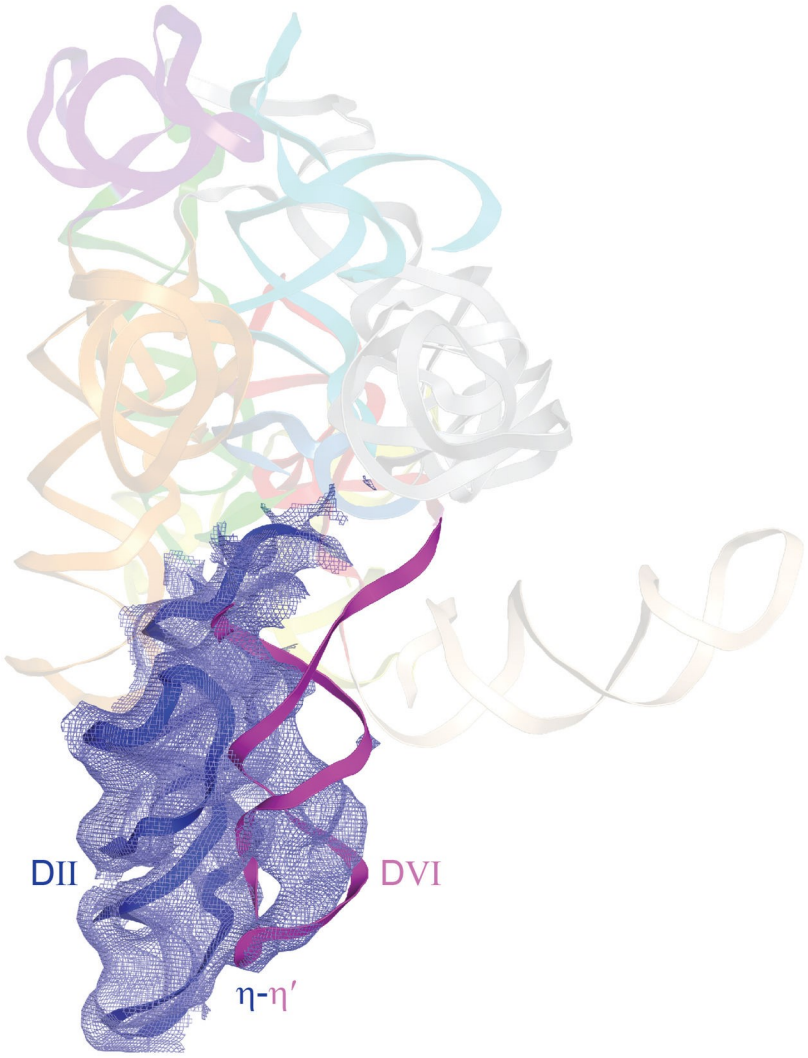

Figure 5.13: **2Fo-Fc density for DVI in the pre-catalytic structure contoured at 1σ.** The η-η' interaction persists throughout the splicing reaction and is visible in the pre-catalytic state. The weaker density for the central region of DVI suggests a partially disordered, dynamic region with possible helical remodeling in the conserved internal loop during splicing. The general pattern of side-by-side packing of domains II and VI persists between the two steps. Catalytic triad mutation consisted of an AGC→GAU substitution.

During the first step, the bulged adenosine must be in close proximity to M1, M2 and the 5′ splice site to engage in nucleophilic attack. However, π–π′ places the bulged adenosine ~20 Å from the active site in the post-catalytic structure. Mutagenesis of π–π′ also has no significant effect on the first step of splicing (Figure 5.8b), and is unlikely to be engaged for lariat formation. Furthermore, DII is likely to remain largely stationary during catalysis due to the strong anchoring effect of multiple tetraloop receptor interactions with this domain. On the basis of these observations, we propose that π–π′ is a dynamic interaction that toggles DVI between two different states to mediate the transition between the first and second steps of splicing (Figure 5.14). In the first step, the bulged adenosine is engaged in the active site for nucleophilic attack at the 5′ splice site. At this stage, π–π′ exists in the 'off' state where DII is disengaged from the base of DVI. The DVI helix would also presumably exist in a relaxed conformation due to the lack of constraint provided by π–π′. Following lariat formation, DVI probably engages in remodeling of its central internal loop adjacent to A615, causing helical compression to turn 'on' the π–π′ interaction, thus sequestering the bulged adenosine away from the active site. A second possible model is that the base pairs between the G6 sequence (residues 588 to 593) and a pyrimidine-rich tract (612 to 614; 616 to 618) at the proximal side of the DVI stem rearrange to reposition the lariat phosphate and engage π–π′. Both models serve to empty the active site of the 5′ end and allow entry of the 3′ splice site, which is directly attached to the end of DVI. In fact, the primary function of the lariat may be to covalently attach to the 5′ end to provide an attachment point for this pulling action.

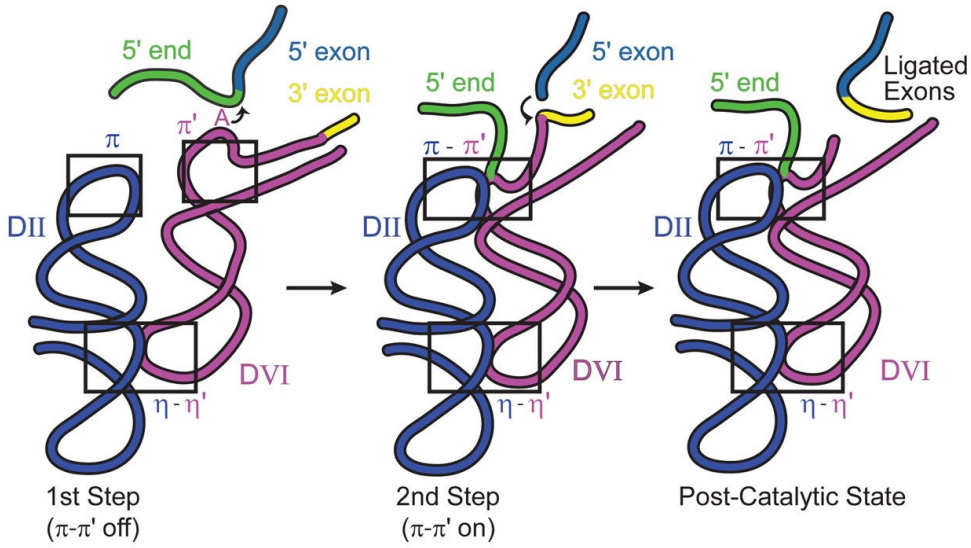

Figure 5.14: **Model for DVI as the conformational switch for splicing.** π-π' mediates the transition between the two steps of catalysis. See text for details.

5.9 Evolutionary implications

This bulged adenosine of DVI is analogous to the branch site adenosine in spliceosomal introns. The branch site sequence UACUAAC (nucleophilic adenosine underlined) pairs to the U2 snRNA to extrude the adenosine from the helix[29] as in DVI. Owing to the mechanistic and evolutionary similarities with group II introns, it is likely that the branch site adenosine residue and the 5′ end of nuclear introns will adopt a similar spatial arrangement relative to the catalytic core of the spliceosome. We predict that the spliceosomal branch site adenosine will be sequestered after the first step of splicing with an RNA and/or protein contact analogous to the π–π′ interaction in *P.li.LSUI2*. In fact, an interaction observed between a region just downstream of the branch site and the U5 snRNA loop may be the spliceosomal counterpart of π–π′[30].

The *P.li.LSUI2* structure provides a rationale for the phylogenetic conservation of the 5′ end. Group II introns have the consensus sequence GUGYG (where Y = U or C)[11], which is similar to the GUAYG and GURAG (where R = G or A) sequences found at the 5′ ends of yeast and mammalian introns, respectively[31]. In the spliceosome, the U6 snRNA pairs with the 5′ end[32] in an analogous manner to the ε–ε′ interaction in group II introns. Therefore, these conserved sequences probably form a similar metal-binding platform in spliceosomal introns, with the 5′ end forming a highly distorted backbone to coordinate magnesium ions that orient the splice site in the core.

In regards to the evolutionary rationale for the 2′-5′ phosphodiester bond, it is known that the lariat is required for full reversibility of group II intron splicing[33]. These reverse splicing reactions are the mechanism through which group II introns are able to invade DNA[34] and disperse throughout genomes. It is likely that the lariat bond pre-organizes the core of the intron structure to facilitate this reversibility. Reverse splicing has also been proposed as a vehicle for the prolific expansion of nuclear introns[35], which comprise ~25% of mammalian genomes. There is biochemical support for this hypothesis in that the spliceosome has recently been shown to catalyze reverse splicing reactions[36]. The lariat probably has a similar role in spliceosomal introns by allowing reverse splicing to occur, thus accounting for its phylogenetic conservation across the kingdoms by enabling ‘selfish introns’ to replicate. Therefore, the *P.li.LSUI2* crystal structure has provided the first glimpse of the branched lariat linkage that was probably crucial for intron proliferation in eukaryotes.

5.10 Methods

Cloning and preparation of *Pylaiella littoralis* LSU I2 intron RNA

DNA was synthesized (Genscript) corresponding to the second intron interrupting the large ribosomal subunit (LSU) in the mitochondria of the brown alga *Pylaiella littoralis* (*P.li.LSUI2*). The crystallization construct contains a 15-nt 5′ exon and a 5-nt 3′ exon followed by a Hind III restriction site. The DIV open reading frame (ORF) was removed from the *P.li.LSU2* intron and replaced with a UUCG stem tetraloop. This was cloned into the EcoRV site of pUC57. Plasmid was linearized by HindIII digestion prior to *in vitr*o transcription. Non-conserved loops in domain 1, 2 and 4 were changed from the wild type sequence. Most significantly, changing the sequence UAUUUAUA to UCGACAUAAGG in the ID2 stem loop improved both crystallization rate and diffraction. The final construct retained wild-type splicing activity. Transcription was performed overnight at 37°C using T7 RNA polymerase in 25 mM $MgCl_2$, 2 mM spermidine, 5 mM DTT, 40 mM Tris-HCl pH 7.5, 0.05% Triton X-100, 2.5 mM of each NTP, and thermostable inorganic pyrophosphatase (New England Biolabs). $CaCl_2$ was added to a final concentration of 1.2 mM, treated with DNase I for 45 minutes, followed by proteinase K digestion for 1 hour. The intron reacted to completion during *in vitro* transcription and was subjected to a native purification procedure previously used in the structure determination of the *O. iheyensis* intron[9]. Spliced intron RNA was repeatedly washed with 10 mM MgCl2 and 5 mM sodium cacodylate pH 6.5 and concentrated to 10 mg/ml using a 100 kDa molecular weight cut-off Amicon Ultra-15 column.

Crystallization

The native crystals were grown in sitting drops by vapor diffusion at 30°C. Equal volumes of RNA (10 mg/ml) were mixed with 0.4 mM spermine, 21% 2-Methyl-2,4-pentanediol (MPD), 175 mM magnesium acetate tetrahydrate, and 90 mM MES monohydrate (pH 5.6). Rod-like crystals appeared within two days, and grew to a maximum size of 50×50×900 μm. Crystals were gradually exchanged into 21% MPD, 100 mM magnesium acetate tetrahydrate, 50 mM MES monohydrate (pH 5.6), 3 mM iridium hexammine, 0.5 mM spermine, and 100 mM NaCl; followed by flash freezing in liquid nitrogen. Iridium hexammine was used as an additive for the native crystals as it reduced mosaicity. Pre-catalytic intron RNA was obtained by mutating the AGC catalytic triad of DV to GAU. The mutation maintained stem pairing but completely inhibited splicing activity. Crystallization of this mutant was done by microseeding under conditions described above. The ι G79A mutant crystallized in 0.2 M ammonium acetate, 10 mM calcium chloride, 50 mM sodium cacodylate pH 6.5, and 10% w/v polyethylene glycol 4000. Resulting crystals were cryoprotected through gradual transfer to the crystallization solution supplemented with 30% ethylene glycol.

Structure determination

The crystal structure of the *P.li.LSUI2* intron was solved at 3.7 Å resolution using multi-wavelength anomalous dispersion (MAD) with crystals soaked in 0.5 mM ytterbium (III) chloride. Yb^{3+} soaks were performed in 21% MPD, 100 mM magnesium acetate tetrahydrate, 50 mM MES monohydrate (pH 5.6), 0.5 mM spermine, 100 mM NaCl, and 0.5 mM $YbCl_3$ for 3 hours at room temperature. Tl+ soaks were performed

in the same manner with 10 mM thallium acetate. X-ray data sets were collected at NE-CAT's 24-ID-C beamline at the Advanced Photon Source (Argonne National Laboratory, Argonne, Illinois). Data was processed using HKL-2000[37], heavy metal sites were identified with SHELXD[38] and phasing done using Phenix[39] and SHELXE[40]. RNA nucleotides were modeled using COOT[41] and the RCrane plugin[42]. The phylogenetically predicted secondary structure (Figure 5.1) guided modeling into the electron density. Structure refinement was done using Buster[43], Phenix[39], DEN[44], and Phenix.Erasser[45]. The 2′-5′ phosphodiester bond was restrained using a cif restraint file in phenix.refine. All software was compiled by SBGrid[46].

***In vitro* self-splicing assays**

The construct used for the *in vitro* self-splicing assays contained wild-type *P.li.LSUI2* sequence with DIV ORF removed and a 250-nt 5′ exon and 75-nt 3′ exon. This was cloned into the pUC57 plasmid. Plasmid was linearized using HindIII and used for *in vitro* transcription with T7 RNA polymerase. Radiolabeled transcripts were prepared as above using 10 μCi [α-32P] UTP (3000 Ci/mmole), 0.5 mM UTP, 1 mM other NTPs, and 10mM $MgCl_2$. Transcripts were gel purified on a 4% polyacrylamide (19:1)/8 M urea gel, RNA was recovered by diffusion into 300 mM NaCl, 0.01% SDS, 1 mM EDTA. Self-splicing experiments were performed for 30 min at 45°C in a splicing buffer containing 10 mM $MgCl_2$, 1M NH_4Cl, 40 mM Tris-HCl (pH 7.5), and 0.02% SDS. Reactions were stopped by addition of EDTA to a final concentration of 20 mM. Splicing products were resolved using a denaturing 4% polyacrylamide (19:1)/8 M urea

gels. Rate constants for the WT and G79A mutant were derived from curves fit to a biphasic exponential equation. All splicing assays were done in triplicate.

5.11 Acknowledgements

We thank Surajit Banerjee and the staff of the NE-CAT beamlines at the Advanced Photon Source (APS) of Argonne National Laboratory. We thank Partho Ghosh, Simpson Joseph, Gourisankar Ghosh, Russell Doolittle, Yitzhak Tor, Dan Donoghue, and Timothy Wiryaman for comments on the manuscript. We thank Rhiju Das and Fang-Chieh Chou for assistance with phenix.erasser for structure refinement and Gerard Bricogne for advice on Buster refinement. We also thank Nicole T. Schirle for preliminary biochemical characterization of the *P.li.LSUI2* intron. R.T.C. was supported by the Cell, Molecular, and Genetics Training Program funded by NIH predoctoral training grant 5T32GM007240. J.K.P. was supported by the UCSD Molecular Biophysics Training Program funded by NIH predoctoral training grant 5T32GM008326. NE-CAT is supported by NIH grant 8P41GM103403-10 and APS is supported by the U.S. DOE under Contract No. DE-AC02-06CH11357. This work was supported by a Hellman Foundation Fellowship and NIH grant 5R01GM102216 awarded to N.T.

Chapter 5, in full, is a reprint with permission of the material as it appears in *Nature*: Robart, A.R., Chan, R.T., Peters, J.K., Rajashankar, K.R. & Toor, N. Crystal structure of a eukaryotic group II intron lariat. *Nature* **514**, 193-197 (2014). Copyright 2014, Nature Publishing Group. The dissertation author is the third author of this paper.

5.12 References

1. Grabowski, P. J., Padgett, R. A. & Sharp, P. A. Messenger RNA splicing in vitro: an excised intervening sequence and a potential intermediate. *Cell* **37,** 415–27 (1984).

2. Padgett, R. A., Konarska, M. M., Grabowski, P. J., Hardy, S. F. & Sharp, P. A. Lariat RNA's as intermediates and products in the splicing of messenger RNA precursors. *Science* **225,** 898–903 (1984).

3. Konarska, M. M., Grabowski, P. J., Padgett, R. A. & Sharp, P. A. Characterization of the branch site in lariat RNAs produced by splicing of mRNA precursors. *Nature* **313,** 552–7

4. Peebles, C. L., Perlman, P. S., Mecklenburg, K. L., Petrillo, M. L., Tabor, J. H., Jarrell, K. A., Cheng, H. L. A self-splicing RNA excises an intron lariat. *Cell* **44,** 213–23 (1986).

5. van der Veen, R., Arnberg, A. C., can der Horst, G., Bonen, L., Tabak, H. F., Grivell, L. A. Excised group II introns in yeast mitochondria are lariats and can be formed by self-splicing in vitro. *Cell* **44,** 225–34 (1986).

6. Di Leo, E., Panico, F., Tarugi, P., Battisti, C., Federico, A., Calandra, S. A point mutation in the lariat branch point of intron 6 of NPC1 as the cause of abnormal pre-mRNA splicing in Niemann-Pick type C disease. *Hum. Mutat.* **24,** 440 (2004).

7. Galej, W. P., Oubridge, C., Newman, A. J. & Nagai, K. Crystal structure of Prp8 reveals active site cavity of the spliceosome. *Nature* **493,** 638–43 (2013).

8. Fica, S. M., Tuttle, N., Novak, T., Li, N.-S., Lu, J., Koodathingal, P., Dai, Q., Staley, J. P, Piccirilli, J. A. RNA catalyses nuclear pre-mRNA splicing. *Nature* **503,** 229–34 (2013).

9. Toor, N., Keating, K. S., Taylor, S. D. & Pyle, A. M. Crystal structure of a self-spliced group II intron. *Science* **320,** 77–82 (2008).

10. Lambowitz, A. M. & Zimmerly, S. Group II Introns: Mobile Ribozymes that Invade DNA. *Cold Spring Harb. Perspect. Biol.* **3,** a003616 (2011).

11. Toor, N., Hausner, G. & Zimmerly, S. Coevolution of group II intron RNA structures with their intron-encoded reverse transcriptases. *RNA* **7,** 1142–52 (2001).

12. Michel, F., Umesono, K. & Ozeki, H. Comparative and functional anatomy of group II catalytic introns--a review. *Gene* **82,** 5–30 (1989).

13. Costa, M., Fontaine, J. M., Loiseaux-de Goër, S. & Michel, F. A group II self-splicing intron from the brown alga Pylaiella littoralis is active at unusually low magnesium concentrations and forms populations of molecules with a uniform conformation. *J. Mol. Biol.* **274,** 353–64 (1997).

14. Rest, J. S. & Mindell, D. P. Retroids in Archaea: Phylogeny and Lateral Origins. *Mol. Biol. Evol.* **20,** 1134–1142 (2003).

15. Toor, N., Robart, A. R., Christianson, J. & Zimmerly, S. Self-splicing of a group IIC intron: 5' exon recognition and alternative 5' splicing events implicate the stem-loop motif of a transcriptional terminator. *Nucleic Acids Res.* **34,** 6461–71 (2006).

16. Boudvillain, M. & Pyle, A. M. Defining functional groups, core structural features and inter-domain tertiary contacts essential for group II intron self-splicing: a NAIM analysis. *EMBO J.* **17,** 7091–104 (1998).

17. Jacquier, A. & Michel, F. Base-pairing interactions involving the 5' and 3'-terminal nucleotides of group II self-splicing introns. *J. Mol. Biol.* **213,** 437–47 (1990).

18. Fedorova, O. & Pyle, A. M. A conserved element that stabilizes the group II intron active site. *RNA* **14,** 1048–56 (2008).

19. Fedorova, O., Mitros, T. & Pyle, A. M. Domains 2 and 3 interact to form critical elements of the group II intron active site. *J. Mol. Biol.* **330,** 197–209 (2003).

20. Chanfreau, G. & Jacquier, A. An RNA conformational change between the two chemical steps of group II self-splicing. *EMBO J.* **15,** 3466–76 (1996).

21. Chanfreau, G. & Jacquier, A. Interaction of intronic boundaries is required for the second splicing step efficiency of a group II intron. *EMBO J.* **12,** 5173–80 (1993).

22. Parker, R. & Siliciano, P. G. Evidence for an essential non-Watson-Crick interaction between the first and last nucleotides of a nuclear pre-mRNA intron. *Nature* **361,** 660–2 (1993).

23. Adams, P. L., Stahley, M. R., Kosek, A. B., Wang, J. & Strobel, S. A. Crystal structure of a self-splicing group I intron with both exons. *Nature* **430,** 45–50 (2004).

24. Li, C.-F., Costa, M. & Michel, F. Linking the branchpoint helix to a newly found receptor allows lariat formation by a group II intron. *EMBO J.* **30,** 3040–51 (2011).

25. Klein, D. J., Moore, P. B. & Steitz, T. A. The contribution of metal ions to the structural stability of the large ribosomal subunit. *RNA* **10,** 1366–79 (2004).

26. Basu, S., Rambo, R. P., Strauss-Soukup, J,., Cate, J. H., Ferré-D'Amaré, A. R., Strobel, S. A., Doudna, J. A. A specific monovalent metal ion integral to the AA platform of the RNA tetraloop receptor. *Nat. Struct. Biol.* **5,** 986–92 (1998).

27. Fica, S. M., Mefford, M. A., Piccirilli, J. A. & Staley, J. P. Evidence for a group II intron-like catalytic triplex in the spliceosome. *Nat. Struct. Mol. Biol.* **21,** 464–71 (2014).

28. Marcia, M. & Pyle, A. M. Visualizing group II intron catalysis through the stages of splicing. *Cell* **151,** 497–507 (2012).

29. Query, C. C., Moore, M. J. & Sharp, P. A. Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model. *Genes Dev.* **8,** 587–97 (1994).

30. Anokhina, M., Bessonov, S., Miao, Z., Westhof, E. Hartmuth, K., Lührmann, R. RNA structure analysis of human spliceosomes reveals a compact 3D arrangement of snRNAs at the catalytic core. *EMBO J.* **32,** 2804–18 (2013).

31. Parker, R. & Guthrie, C. A point mutation in the conserved hexanucleotide at a yeast 5' splice junction uncouples recognition, cleavage, and ligation. *Cell* **41,** 107–18 (1985).

32. Lesser, C. F. & Guthrie, C. Mutations in U6 snRNA that alter splice site specificity: implications for the active site. *Science* **262,** 1982–8 (1993).

33. Roitzsch, M. & Pyle, A. M. The linear form of a group II intron catalyzes efficient autocatalytic reverse splicing, establishing a potential for mobility. *RNA* **15,** 473–482 (2009).

34. Yang, J., Zimmerly, S., Perlman, P. S. & Lambowitz, A. M. Efficient integration of an intron RNA into double-stranded DNA by reverse splicing. *Nature* **381,** 332–5 (1996).

35. Lynch, M. & Richardson, A. O. The evolution of spliceosomal introns. *Curr. Opin. Genet. Dev.* **12,** 701–10 (2002).

36. Tseng, C.-K. & Cheng, S.-C. Both Catalytic Steps of Nuclear Pre-mRNA Splicing Are Reversible. *Science* **320,** 1782–1784 (2008).

37. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–26 (1997).

38. Schneider, T. R. & Sheldrick, G. M. Substructure solution with SHELXD. *Acta Crystallogr. D. Biol. Crystallogr.* **58,** 1772–9 (2002).

39. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, H. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, R. C., Zwart, P. H. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr.* **66,** 213–21 (2010).

40. Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr. A.* **64,** 112–22 (2008).

41. Emsley, P. & Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60,** 2126–2132 (2004).

42. Keating, K. S. & Pyle, A. M. RCrane: semi-automated RNA model building. *Acta Crystallogr. D. Biol. Crystallogr.* **68,** 985–95 (2012).

43. Blanc, E., Roversi, P., Vonrhein, C., Flensburg, C., Lean, S. M., Bricogne, G. Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr. D. Biol. Crystallogr.* **60,** 2210–21 (2004).

44. Schröder, G. F., Levitt, M. & Brunger, A. T. Super-resolution biomolecular crystallography with low-resolution data. *Nature* **464,** 1218–22 (2010).

45. Chou, F.-C., Sripakdeevong, P., Dibrov, S. M., Hermann, T. & Das, R. Correcting pervasive errors in RNA crystallography through enumerative structure prediction. *Nat. Methods* **10,** 74–6 (2013).

46. Morin, A., Eisenbraun, B., Key, J., Sanschagrin, P. C., Timony, M. A., Ottaviano, M., Sliz, P. Collaboration gets the most out of software. *Elife* **2,** e01456 (2013).

Chapter 6: Conclusion

6.1 Introduction

Group II introns are self-splicing ribozymes found in all three domains of life. They are predominantly found in bacteria as well as in the mitochondria and chloroplasts of some eukaryotes, and rarely in archaea[1]. The secondary structure of group II introns can be subdivided into six domains that fold into an ordered three-dimensional structure that catalyzes splicing via two transesterification reactions. In the first step, the 2′-hydroxyl of a bulged adenosine in domain VI attacks the 5′ splice site resulting in the formation of a branched RNA containing a 2′-5′ phosphodiester bond (known as the 'lariat') and a free 3′-hydroxyl of the newly cleaved 5′ exon[2,3]. In the second step, this 3′-hydroxyl attacks the 3′ splice site resulting in ligated exons and the release of intron lariat. The stereochemistry of this reaction is identical to that seen in the spliceosome, which removes nuclear introns from pre-mRNAs. The group II intron and spliceosome are postulated to share a common ancestor based on similarities in conserved core RNA structures and mechanism. Recent biochemical evidence has also shown that the spliceosome is a ribozyme that uses a two-metal-ion mechanism of catalysis[4]. Two spliceosomal RNAs, U2 and U6, comprise the active site and are capable of performing a splicing-related reaction in the absence of spliceosomal proteins[5,6].

Of the three structural classes of group II introns (IIA, IIB, and IIC), IIC introns are thought to be the most primitive[7]. IIA and IIB introns typically form large amounts of lariat RNA during *in vitro* self-splicing reactions, whereas IIC introns primarily splice through hydrolysis *in vitro* to form linear intron[8]. Previously, a crystal structure of a

hydrolytic IIC intron from the bacterium *Oceanobacillus iheyensis* (*O. iheyensis*) revealed that the highly conserved DV forms an active site containing two catalytic magnesium ions[9]. However, this structure lacked domain VI (DVI), which contains the bulged adenosine nucleophile required for lariat formation in the first step. The recent crystal structure of a post-catalytic, eukaryotic group IIB intron from the brown algae *Pylaiella littoralis* (*P.li.LSUI2*) contains a functional DVI[10]. Given that this intron catalyzes lariat formation[11], it has contributed to a new understanding of the structural requirements for splicing in higher eukaryotes.

6.2 Newly visualized DVI

DVI contains the bulged adenosine residue analogous to the branch site adenosine found in spliceosomal introns. Given its importance in lariat formation, the location and orientation of DVI within the group II intron has been a long-standing question in the field of RNA splicing. There have been several proposed models for DVI, with some postulating a side-by-side packing with the catalytic DV, and others placing DVI at ~90˚ from DV in the first step[12,13]. The structure of the intron lariat finally allowed visualization of DVI and the tertiary interactions that position it in the catalytic core. In this structure, the two distal stems of domain II (DII), D2a and D2b, coaxially stack and form two tetraloop-receptor interactions with DVI (Figure 6.1a). The GAGA tetraloop in DVI docks into the receptor at the base of D2a (η-η′), and the GCAA tetraloop in D2b docks into a receptor directly adjacent to the bulged adenosine in DVI (π-π′). These two tetraloop-receptor interactions position DVI in the core of the

intron at a ~180˚ angle from DV in a conformation in which these two helices are not coaxially stacked on each other (Figure 6.1b). Mutagenesis of either of these GNRA tetraloops to a non-interacting UUCG results in an accumulation of intermediate splicing products consisting of lariat-3′ exon and 5′ exon. Furthermore, a construct containing both mutated tetraloops results in an almost complete block of the second step[10]. The discovery of the π-π′ interaction is interesting due to its proximity to the bulged adenosine residue A615. This interaction provides the means for removal of the lariat bond from the active site after the first step to allow entry of the 3′ splice site for the second step.

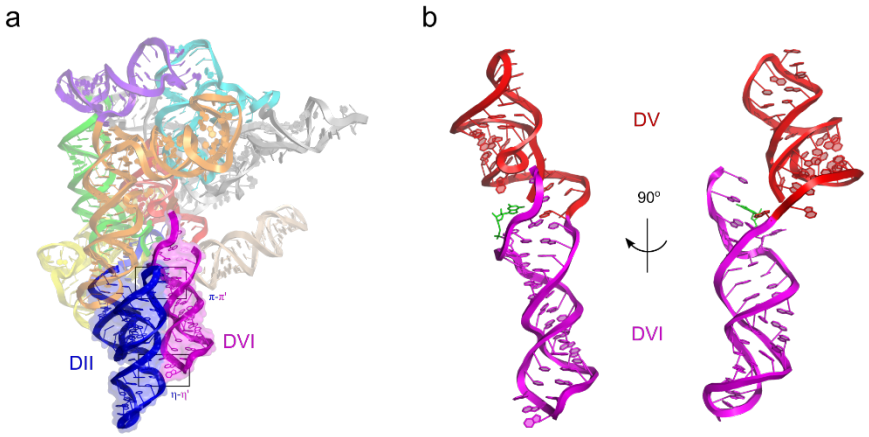

Figure 6.1: **Overall tertiary structure of *P.li.LSUI2*.** (a) Interactions between DII and DVI. The two distal stems of DII (blue) coaxialy stack and align parallel to DVI (purple). The two tetraloop-receptor interactions between domains II and VI, η-η' and π-π', are indicated with black boxes. Individual domains and subdomains are shown in different colors. (b) Arrangement of DV and DVI. DV (red) and DVI (purple) are oriented 180˚ relative to each other. The image on the left shows the two domains from the same angle as (a). The image on the right is rotated 90˚ about the y-axis to show that the two domains are not coaxially stacked. G1 is shown in green and interacts with A620.

Previous models proposed a large-scale "swinging-arm" movement of the distal end of DVI (containing the tetraloop) with η-η′ being disengaged for the first step of splicing[12,13]. However, the pre-catalytic state of *P.li.LSUI2* at 7 Å revealed strong electron density for the η-η′ interaction[10], demonstrating that this contact persists through both steps of splicing and suggesting that large scale movement of DVI is unlikely. Therefore, we propose a more subtle conformational change in which DVI remains 180° from DV with π-π′ being disengaged for the first step of splicing and engaged for the second step. This allows the bulged adenosine to reach the active site in the first step. Upon lariat formation, π-π′ engages, removing the bulged adenosine from the active site while simultaneously pulling the attached 3′ splice site into the core.

An analogous interaction sequestering the adenosine nucleophile after the first step likely exists in the spliceosome. A possible spliceosomal counterpart could be an RNA-RNA interaction between the U5 snRNA and a region just downstream of the branch site[14]. However, this may also be accomplished via an RNA-protein interaction. Another possibility for a π-π′ analogue is the spliceosomal p14 protein, which crosslinks directly to the branch site adenosine[15]. It has been proposed that this RNA-protein interaction must be disrupted to allow for the first step of splicing[16,17], in a manner similar to π-π′. The conserved protein Prp8 could be yet another candidate for engaging in an equivalent π-π′ interaction since it has been also shown to crosslink to the branch site as well as other key regions in the spliceosomal core, such as the 5′ and 3′ exons[18–20]. Mutations in RNA binding regions of Prp8 result in both first and second step defects,

indicating an essential role in positioning substrates in the active site[21]. There is also structural evidence suggesting that Prp8 is derived from a group II intron encoded protein (IEP), also known as a maturase[22]. Maturases assist in the proper folding of group II introns *in vivo*[1]. In a similar manner, Prp8 likely functions as an assembly platform in the spliceosome to mediate intron removal by snRNAs.

6.3 Active site dynamics

In previous IIC intron structures, the single-stranded linker between domains II and III (J2/3) forms base triple interactions with catalytic residues from DV[9,23]. There is also evidence for similar base triples in the active site of the spliceosome[24]. Interestingly, in the *P.li.LSUI2* structure, one of the J2/3 residues (A422) is not engaged in a base triple with the catalytic triad, as was previously observed in the IIC structures. Instead, A422 is forming a base triple with residues from J4/5 (U549) and J5/6 (U584) (Figure 6.2). This newly observed interaction of A422 introduces the possibility that the triple helix architecture in the catalytic core may change configuration between the two steps of splicing. Additionally, 7-deaza, 2,6-diaminopurine, and N6-methyl NAIM interferences were observed for the yeast *aI5γ* intron residue equivalent to A422[25]. However, only the N6 amino group of A422 interacts with nearby bases in the post-catalytic *P.li.LSUI2* structure, suggesting that this nucleotide may engage in additional transient interactions during the splicing reaction. Therefore, the observation that additional functional group identities of residue A422 are important supports the hypothesis of a dynamic role for the J2/3 linker between splicing steps likely modulating

catalysis. This hypothesis is also supported by crystal structures of the IIC intron showing that J2/3 disengages from the catalytic triad in the transition from the first to second step[23]. J4/5 and J5/6 are also likely intimately involved in conformational rearrangements within the catalytic triad in supporting roles for J2/3. A similar interaction(s) may also be occurring within the spliceosome. Mutagenesis of the U584 analogue in yeast (A25 in U2 snRNA) resulted in a 50% decrease in efficiency of the second step[26]. Therefore, A25 plays a similar role in the second step as U584 in the *P.li.LSUI2* intron. The lack of a first step splicing defect upon mutagenesis of A25 suggests that rearrangement of the base triples occurs in the transition between the two steps. This also alludes to the possibility that every stage of catalysis may have a unique arrangement of base triples in the catalytic core.

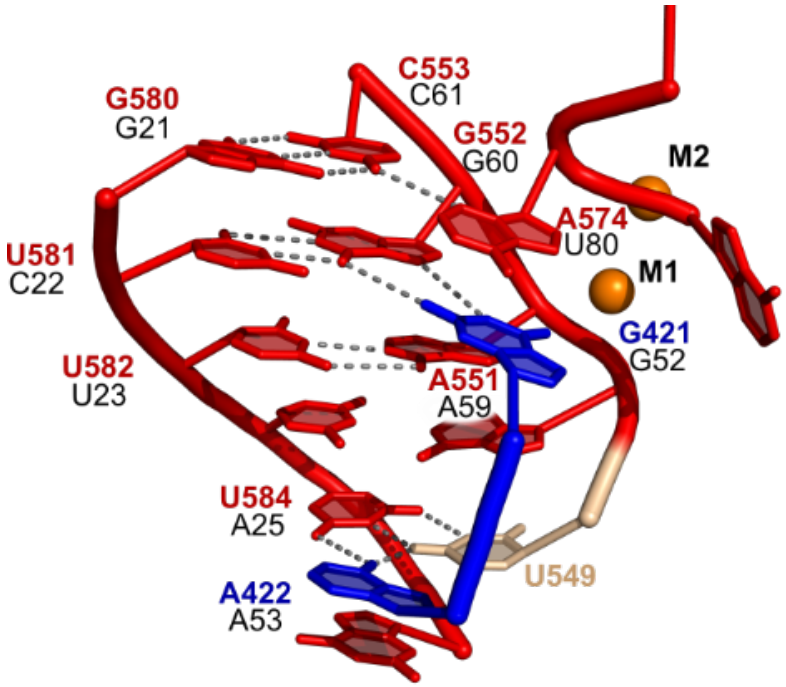

Figure 6.2: ***P.li.LSUI2* core interactions and the proposed analogous spliceosomal residues.** Coloring is consistent with the domains in Figure 6.1. *P.li.LSUI2* residue numbers are shown in red, blue, and tan with proposed analogous U2 and U6 numbering in black. Spliceosomal numbering is from *S. cerevisiae*. In *P.li.LSUI2*, J2/3 residue A422 is disengaged from the catalytic triad (AGC) and instead forms a base triple with residues U549 and U584. Catalytic metals are shown as orange spheres and labeled M1 and M2.

6.4 3′ splice site selection in the second step

The EBS1-IBS1 and EBS2-IBS2 interactions in IIB introns function to correctly place the 5′ splice site in the active site for the first step. As is seen in the *P.li.LSUI2* post-catalytic structure, these interactions remain engaged through the second step to position the 3′-OH nucleophile of the 5′ exon in the active site for the second step. It has been previously shown that deleting EBS1 from a group IIB intron and loop 1 from the U5 snRNA in yeast spliceosomes both result in an accurate and efficient first step, but

stalling of the second step[27,28]. This stalling can be relieved in group II introns upon the addition in trans of EBS1 or snRNA U5 loop 127. This suggests that both elements have a similar function in the group II intron and the spliceosome to tether the free 5′ exon to the active site between the two steps of splicing. The U5 snRNA has also been shown to base pair with the first two nucleotides of the 3′ exon (Figure 6.3a)[29,30]. This interaction is equivalent to the EBS3-IBS3 interaction in group II introns (Figure 6.3b) that forms prior to the second step to aid in the placement of the 3′ splice site, thus further extending the parallels between the two splicing systems.

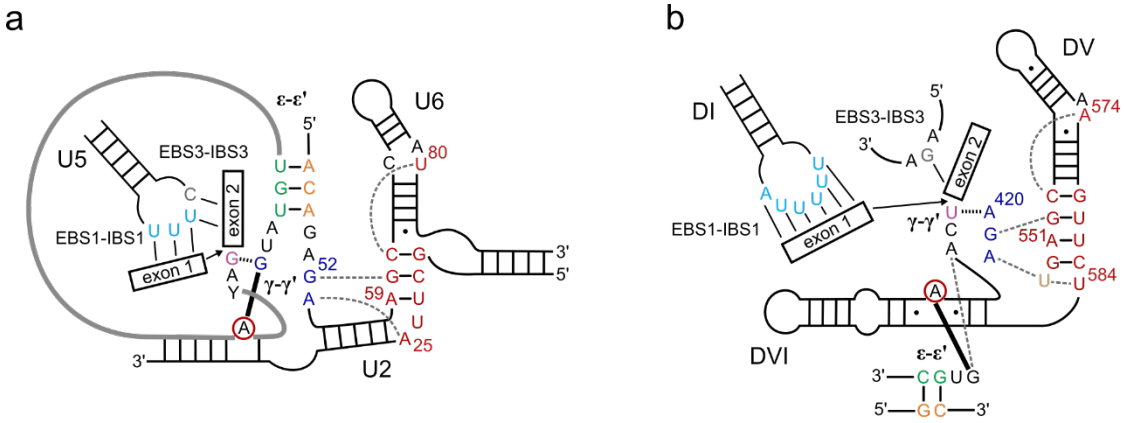

Figure 6.3: **Comparison of spliceosomal and group II intron core structures prior to the second step.** Analogous residues are shown in the same color and the color scheme is from ref. 10. DV residues are shown in red, J2/3 in blue, J4/5 in tan, EBS1 in light blue, EBS3 in gray, ε in orange, ε' in green, and γ' in purple. The conserved branch point residue is circled in red. Dotted gray lines denote base triples, solid black lines designate base pairing, the dashed black line indicates the γ- γ' base pair, and the thick black line represents the 2'-5' bond between the branch point residue and the first residue of the intron. Numbering is as in Figure 6.2. (a) Predicted spliceosomal core elements. Numbering is according to the *S. cerevisiae* U2 and U6 sequences. The intron is shown as a thick gray line. (b) Group II intron core arrangement. Numbering is as in Figure 6.2.

Prior to the second step, the γ-γ′ interaction must also form in group II introns to aid in positioning of the 3′ splice site. This highly conserved contact forms between the last nucleotide of the intron and the residue immediately 5′ to the J2/3 residues participating in base triples with the catalytic triad. This interaction brings the 3′ splice site in close spatial proximity to the catalytic metals in DV. In the post-catalytic *P.li.LSUI2* structure, γ-γ′ is found to be disengaged indicating that it is a transient interaction that only exists during the presentation of the scissile phosphate of the 3′ splice site to the catalytic metals. Furthermore, the structure reveals a non-canonical pairing between the highly conserved G1 residue from the 5′ end and A620, which is the third nucleotide from the 3′ end of the intron. Therefore, the combination of both γ-γ′ and G1-A620 cooperate to precisely position the 3′ splice site within the core. In this case, the terminal nucleotides are not in direct contact with each other. However, in the spliceosome, it has been observed that the last nucleotide of the intron forms a non-canonical G-G base pair with the first nucleotide[31,32]. As a result, the functions of γ-γ′ and G1-A620 are combined into a single entity involving direct interaction between the 5′ and 3′ ends.

Atomic resolution spliceosome structures immediately after branching[33,34] (C complex) and directly preceding exon ligation[35,36] (C* complex) reveal an additional conserved architectural feature with group II introns. Because the same active site is utilized for both steps of splicing, first-step products must be removed from the active site after branch formation before exon ligation can take place. In group II introns, this rearrangement is likely promoted by dynamic contacts within DVI[10]. As the

spliceosome does not contain an analogous DVI, and therefore π-π′, proteins have evolved to take over the role of displacing the bulged adenosine from the active site. Despite the different means of transitioning from branching to exon ligation, the distance the bulged adenosine is removed from the active site remains consistent (Figure 6.4). This conservation of bulged adenosine displacement distance is additional evidence of an evolutionary link between group II introns and the spliceosome.

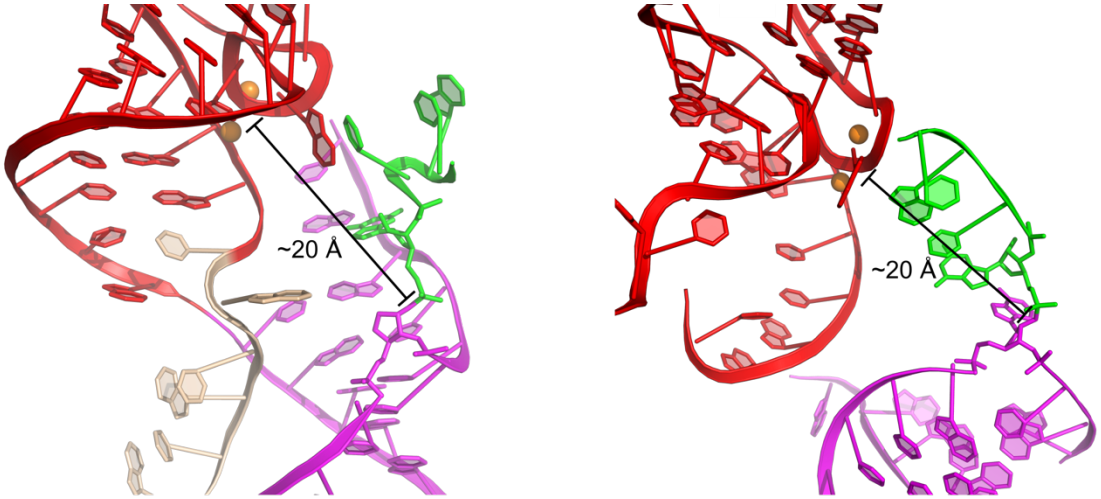

Figure 6.4: **Comparison of group II intron and spliceosome active sites immediately preceding exon ligation**. In the *P.li.LSUI2* lariat-3′ exon crystal structure, the conserved adenosine is located ~20 Å away from the catalytic metal ions in the active site. The same distance is observed in the spliceosome C* complex at the same stage of splicing (PDB: 5MPS).

6.5 Future Directions

The crystal structure of the *P.li.LSUI2* intron in the post-catalytic lariat and intermediate lariat-3′ exon forms have advanced our understanding of the structural requirements for group II intron splicing. However, this only represents the end product state and does not capture the entire spectrum of intermediates along the splicing pathway. For example, the pre-catalytic state would reveal the rationale for the conservation of the bulged adenosine nucleophile between group II introns and the spliceosome. At this stage of splicing, there must exist a defined binding pocket for this nucleobase that precisely positions the 2′-OH of the ribose sugar directly above the 5′ splice site. This structure would also reveal the specific nature of the conformational rearrangements involving DVI. Given that the distal tetraloop of DVI forms the static η-η′ interaction, the proximal region containing the bulged adenosine must either unwind or shift in order to engage π-π′ for the second step. Another intermediate of interest is the state prior to the second step of splicing, in which the 5′ exon is poised to attack the 3′ splice site. These intermediates may yet reveal additional unique catalytic triplex configurations for each stage of splicing. These base triple rearrangements would not be captured in the truncated IIC intron structure given that it was missing J5/6 and DVI. The triplex may have a similar dynamic nature in the spliceosome and assist with the forward progression of splicing.

The crystal structure of Prp8 has unexpectedly revealed an RT-like fold similar to that expected for group II intron maturases, with Prp8 forming a cavity for the binding of the catalytic U2/U6 snRNA[22]. The next major goal is to determine the structure of a

group II intron ribonucleoprotein (RNP) complex containing the cognate maturase. This protein facilitates splicing and allows retrotransposition reactions to occur in which the group II intron is able to insert into double-stranded DNA targets[1]. The structure of a group II intron RNP would also allow a direct comparison with Prp8 to definitively establish structural homology. There is now increasing evidence that group II introns and the spliceosome share a common ancestor. These advancements in the study of group II intron structure and function can now be applied to further our understanding of the nuclear splicing machinery through investigation of the analogous interactions in the spliceosome.

6.6 Acknowledgements

Chapter 6 is adapted with permission from Peters, J.K. & Toor, N. Group II intron lariat: structural insights into the spliceosome. *RNA Biology* **95**, 9244-9249 (2015). Copyright 2015, Taylor & Francis, LLC (http://www.tandfonline.com). The dissertation author is the first author of this paper.

6.7 References

1. Lambowitz, A. M. & Zimmerly, S. Group II Introns: Mobile Ribozymes that Invade DNA. *Cold Spring Harb. Perspect. Biol.* **3,** a003616 (2011).

2. Peebles, C. L., Perlman, P. S., Mecklenburg, K. L., Petrillo, M. L., Tabor, J. H., Jarrell, K. H. A self-splicing RNA excises an intron lariat. *Cell* **44,** 213–23 (1986).

3. van der Veen, R., Arnberg, A. C., van der Horst, G., Bonen, L., Tabak, H. F., Grivell, L. A. Excised group II introns in yeast mitochondria are lariats and can be formed by self-splicing in vitro. *Cell* **44,** 225–34 (1986).

4. Fica, S. M., Tuttle, N., Noval, T., Li, N.-S., Lu, J., Koodathingal, P., Dai, Q., Staley, J. P., Piccirilli, J. A. RNA catalyses nuclear pre-mRNA splicing. *Nature* **503,** 229–34 (2013).

5. Valadkhan, S., Mohammadi, A., Wachtel, C. & Manley, J. L. Protein-free spliceosomal snRNAs catalyze a reaction that resembles the first step of splicing. *RNA* **13,** 2300–11 (2007).

6. Valadkhan, S., Mohammadi, A., Jaladat, Y. & Geisler, S. Protein-free small nuclear RNAs catalyze a two-step splicing reaction. *Proc. Natl. Acad. Sci.* **106,** 11901–6 (2009).

7. Rest, J. S. & Mindell, D. P. Retroids in Archaea: Phylogeny and Lateral Origins. *Mol. Biol. Evol.* **20,** 1134–1142 (2003).

8. Toor, N., Robart, A. R., Christianson, J. & Zimmerly, S. Self-splicing of a group IIC intron: 5' exon recognition and alternative 5' splicing events implicate the stem-loop motif of a transcriptional terminator. *Nucleic Acids Res.* **34,** 6461–71 (2006).

9. Toor, N., Keating, K. S., Taylor, S. D. & Pyle, A. M. Crystal structure of a self-spliced group II intron. *Science* **320,** 77–82 (2008).

10. Robart, A. R., Chan, R. T., Peters, J. K., Rajashankar, K. R. & Toor, N. Crystal structure of a eukaryotic group II intron lariat. *Nature* **514,** 193–197 (2014).

11. Costa, M., Fontaine, J. M., Loiseaux-de Goër, S. & Michel, F. A group II self-splicing intron from the brown alga Pylaiella littoralis is active at unusually low magnesium concentrations and forms populations of molecules with a uniform conformation. *J. Mol. Biol.* **274,** 353–64 (1997).

12. Chanfreau, G. & Jacquier, A. An RNA conformational change between the two chemical steps of group II self-splicing. *EMBO J.* **15,** 3466–76 (1996).

13. Li, C.-F., Costa, M. & Michel, F. Linking the branchpoint helix to a newly found receptor allows lariat formation by a group II intron. *EMBO J.* **30,** 3040–51 (2011).

14. Anokhina, M., Bessonov, S., Miao, Z., Westhof, E. Hartmuth, K., Lührmann, R. RNA structure analysis of human spliceosomes reveals a compact 3D arrangement of snRNAs at the catalytic core. *EMBO J.* **32,** 2804–18 (2013).

15. Query, C. C., Strobel, S. A. & Sharp, P. A. Three recognition events at the branch-site adenine. *EMBO J.* **15,** 1392–402 (1996).

16. Bessonov, S., Anokhina, M., Will, C. L., Urlaub, H. & Lührmann, R. Isolation of an active step I spliceosome and composition of its RNP core. *Nature* **452,** 846–50 (2008).

17. Schellenberg, M. J., Dul, E. L. & MacMillan, A. M. Structural model of the p14/SF3b155 · branch duplex complex. *RNA* **17,** 155–65 (2011).

18. Turner, I. A., Norman, C. M., Churcher, M. J. & Newman, A. J. Dissection of Prp8 protein defines multiple interactions with crucial RNA sequences in the catalytic core of the spliceosome. *RNA* **12,** 375–86 (2006).

19. Teigelkamp, S., Newman, A. J. & Beggs, J. D. Extensive interactions of PRP8 protein with the 5' and 3' splice sites during splicing suggest a role in stabilization of exon alignment by U5 snRNA. *EMBO J.* **14,** 2602–12 (1995).

20. Ismaïli, N., Sha, M., Gustafson, E. H. & Konarska, M. M. The 100-kda U5 snRNP protein (hPrp28p) contacts the 5' splice site through its ATPase site. *RNA* **7,** 182–93 (2001).

21. Query, C. C. & Konarska, M. M. Suppression of multiple substrate mutations by spliceosomal prp8 alleles suggests functional correlations with ribosomal ambiguity mutants. *Mol. Cell* **14,** 343–54 (2004).

22. Galej, W. P., Oubridge, C., Newman, A. J. & Nagai, K. Crystal structure of Prp8 reveals active site cavity of the spliceosome. *Nature* **493,** 638–43 (2013).

23. Marcia, M. & Pyle, A. M. Visualizing group II intron catalysis through the stages of splicing. *Cell* **151,** 497–507 (2012).

24. Fica, S. M., Mefford, M. A., Piccirilli, J. A. & Staley, J. P. Evidence for a group II intron-like catalytic triplex in the spliceosome. *Nat. Struct. Mol. Biol*. **21,** 464–71 (2014).

25. Fedorova, O. & Pyle, A. M. Linking the group II intron catalytic domains: tertiary contacts and structural features of domain 3. *EMBO J.* **24,** 3906–16 (2005).

26. Madhani, H. D. & Guthrie, C. Randomization-selection analysis of snRNAs in vivo: evidence for a tertiary interaction in the spliceosome. *Genes Dev.* **8,** 1071–86 (1994).

27. Hetzer, M., Wurzer, G., Schweyen, R. J. & Mueller, M. W. Trans-activation of group II intron splicing by nuclear U5 snRNA. *Nature* **386,** 417–20 (1997).

28. O'Keefe, R. T., Norman, C. & Newman, A. J. The invariant U5 snRNA loop 1 sequence is dispensable for the first catalytic step of pre-mRNA splicing in yeast. *Cell* **86,** 679–89 (1996).

29. Newman, A. J. & Norman, C. U5 snRNA interacts with exon sequences at 5' and 3' splice sites. *Cell* **68,** 743–54 (1992).

30. O'Keefe, R. T. & Newman, A. J. Functional analysis of the U5 snRNA loop 1in the second catalytic step of yeast pre-mRNA splicing. *EMBO J.* **17,** 565–574 (1998).

31. Parker, R. & Siliciano, P. G. Evidence for an essential non-Watson-Crick interaction between the first and last nucleotides of a nuclear pre-mRNA intron. *Nature* **361,** 660–2 (1993).

32. Deirdre, A., Scadden, J. & Smith, C. W. Interactions between the terminal bases of mammalian introns are retained in inosine-containing pre-mRNAs. *EMBO J.* **14,** 3236–46 (1995).

33. Galej, W. P., Wilkinson, M. E., Fica, S. M, Oubridge, C., Newman, A. J., Nagai, K. Cryo-EM structure of the spliceosome immediately after branching. *Nature* **537,** 197–201 (2016).

34. Wan, R., Yan, C., Bai, R., Huang, G. & Shi, Y. Structure of a yeast catalytic step I spliceosome at 3.4 Å resolution. *Science* **353,** 895–904 (2016).

35. Fica, S. M., Oubridge, C., Galej, W. P., Wilkinson, M. E., Bai, X.-C., Newman, A. J., Nagai, K. Structure of a spliceosome remodelled for exon ligation. *Nature* **542,** 377–380 (2017).

36. Yan, C., Wan, R., Bai, R., Huang, G. & Shi, Y. Structure of a yeast step II catalytically activated spliceosome. *Science* **355,** 149–155 (2017).