UNIVERSITY OF CALIFORNIA, SAN DIEGO

Structural Insights into RNA Splicing in Group II Introns

A dissertation submitted in partial satisfaction of the

requirements for the degree

Doctor of Philosophy

in

Chemistry

by

Russell T. Chan

Committee in charge:

Professor Navtej Toor, Chair
Professor Michael Burkart
Professor Simpson Joseph
Professor Ulrich Muller
Professor Amy Pasquinelli

2016

The dissertation of Russell T. Chan is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2016

Dedication

To my wife Clara, who is my companion on this journey.

To my family, who has always believed in me.

Epigraph

When you listen and read one thinker, you become a clone…

two thinkers, you become confused…

ten thinkers, you'll begin developing your own voice…

two or three hundred thinkers, you become wise and develop your voice.

-Tim Keller

Table of Contents

List of Figures

List of Tables

Acknowledgements

I am, and will always be, grateful to Professor Navtej Toor for being my mentor and advisor throughout my graduate career. You have been a great source of inspiration and wisdom throughout my scientific journey. I have always appreciated the support you offered and the opportunities you have provided. Thank you for taking a chance on me as your first graduate student and helping me become the scientist that I am today.

I would also like to thank all the Toor lab members that I have worked with: Aaron Robart, Jessie Peters, Dan Haack, Tim Wiryaman, Ana Gomez, and Michaela Go. You have all been true friends and wonderful colleagues to me and I hope that I been to you as well. Thank you to the Muller lab and the Joseph lab as well for all the great conversations about science and life outside of the laboratory. I would specifically acknowledge Jessie and Tim for sharing an office with me. Not only did you put up with my sense of humor, but our conversations about science have been what I truly believe captured the essence of working together. Although our success seemed limited and our failures always present, I could always find empathy, understanding, and sometimes answers in our office.

To my undergraduate advisor, Dr. Joseph Ng, thank you for opening my eyes to the wonderful world of crystallography. You started me on this path as a young scientist and gave me the tools to follow my passion of research and discovery. Thank you that even though I have long since left your lab, you continually support me.

To my wonderful wife, Clara, thank you for being my rock throughout my graduate studies. You have given me so much support and encouragement throughout my graduate studies that I do not know if I could have done it without you. To my parents, Dennis and Priscilla Chan, thank you for always pushing me to be the best and never settling for what was easy. I look back on my life and I am amazed at how much I have been able to accomplish because of how you raised me. My siblings, Christopher and Kimberly, thank you for always

believing in me. You have seen me at my high and my low and you have stood by me through it all. And to the rest of my family members, thank you for the love, support and encouragement you have given me throughout my life until now.

Chapter 2, in full, is a reprint of the material as it appears in Nature Structural and Molecular Biology, **Chan, R.T.**; Robart, A.R.; Rajashankar, K.R.; Pyle, A.M.; Toor, N. (2012) Crystal structure of a group II intron in the pre-catalytic state. Nature Structure and Molecular Biology, 19(5):555-7. The dissertation author is the primary author on this paper.

Chapter 3, in full, is a reprint of the material as it appears in Nature, Robart, A.R.; **Chan, R.T.**; Peters, J.K.; Rajashankar, K.R.; Toor, N. (2014) Crystal structure of a eukaryotic group II intron lariat. Nature, 514(7521):193-7. The dissertation author is the secondary author on this paper.

Chapter 4, in part, is currently being prepared for submission for publication of the material. **Chan R.T.** and Robart A.R.; Peters, J.K.; Rajashankar, K.R.; Toor, N. The dissertation author is a co-primary author on this paper.

Chapter 5, in part, is currently in submission for publication of the material. **Chan R.T**.; Keating, K.S.; Go, M.C., Toor, N. The dissertation author is the primary author on this paper.

This work was supported, in part, by the National Institute of Health Cellular and Molecular Genetics Training Grant, the Inamori Foundation, and the ARCS Foundation.

Vita

2009	Bachelor of Science, University of California, Riverside
2013	Master of Science, University of California, San Diego
2016	Doctor of Philosophy, University of California, San Diego

Publications

Chan, Russell T.; Robart, Aaron R.; Rajashankar, Kanagalaghatta R.; Pyle, Anna M.; Toor, Navtej. “Crystal structure of a group II intron in the pre-catalytic state”. Nature Structure and Molecular Biology, 19(5):555-7, April 8th, 2012.

Robart, Aaron R.; **Chan, Russell T**.; Peters, Jessie K.; Rajashankar, Kanagalaghatta R.; Toor, Navtej. “Crystal structure of a eukaryotic group II intron lariat”. Nature, 514(7521):193-7, October 9th, 2014.

Chan, Russell T. and Peters, Jessie K.; Robart, Aaron R.; Rajashankar, Kanagalaghatta R., Toor, Navtej. “Dynamic catalytic triplex modulates RNA splicing”. Manuscript in preparation for Nature Structural & Molecular Biology, 2016.

Chan, Russell T.; Keating, Kevin S.; Go, Michaela C.; Toor, Navtej. “The GUAAY pentaloop: a novel motif in the group IIB1 intron structure”. Manuscript in submission to the Journal of Molecular Biology, 2016.

Abstract of the Dissertation

Structural Insights into RNA Splicing in Group II Introns

by

Russell T. Chan

Doctor of Philosophy in Chemistry

University of California, San Diego, 2016

Professor Navtej Toor, Chair

Group II introns are a class of ribozymes capable of self-excision from a nascent pre-mRNA sequence via two sequential transesterification reactions. Due to the mechanistic and structural similarities, it is believed that the group II intron and the spliceosome share a common evolutionary ancestor. Splicing is a critical post-transcriptional modification that can be both beneficial, through alternative splicing, and detrimental, in the case of errors in the splicing process, to organisms. The work entailed in this dissertation aims to elucidate the

structural elements that are required for accurate and productive RNA splicing. In the first part of this dissertation, the structure of a bacterial pre-catalytic intron was determined in order to understand the architecture necessary to catalyze the first step of splicing. The crystal structure of this intron prior to the first step revealed the intact 5′ splice sit that exhibited a sharp, kinked phosphate backbone, which positions the scissile phosphate alone into the active site and promote the fidelity of the splicing reaction. The second part of this dissertation focused on the crystal structure of a eukaryotic group II intron to determine what structural elements are required to activate the intron for lariat formation. Surprisingly, DII was shown to have an integral role in interacting with DVI, with the identification of a novel tertiary interaction, π-π' linking the helix of DVI adjacent to the bulged adenosine to a loop of DII. The third section details the crystal structure of the pre-2s group II intron. This structure reveals conserved junction nucleotide rearrangements within the catalytic triplex compared to previous group II intron structures. This suggests these junction nucleotides undergo rearrangements in order to facilitate the transition to the next step of splicing. Finally, the fourth project entails the biochemical and structural characterization of a novel GUAAY pentaloop identified in the lariat forming group II intron crystal structure. Based on the findings of this project, the pentaloop represents a new class of RNA tertiary interaction.

Chapter 1: Introduction

1.1 Discovery of Catalytic RNAs

Since the discovery of the first catalytic RNA in the early 1980s[1,2], the field of RNA biology has made significant discoveries impacting basic scientific research as well as health sciences. Prior to Tom Cech and Sidney Altman's ground breaking discovery, for which they received the Nobel Prize in Chemistry, RNA was only viewed as a transient intermediate of genetic information in the central dogma of molecular biology. While the capability of RNA to both store genetic information and catalyze chemical reactions had been proposed earlier[3–5], it was still a radical idea at the time. However, once the scientific community accepted the reality of catalytic RNAs, or ribozymes[6], it led to numerous discoveries that expanded the function of RNA. Such discoveries of RNA function include, but not limited to, sensing metabolites to modulate gene expression[7], silencing gene expression[8,9], self-cleavage[1,10,11], and synthesizing polypeptide chains[12–14].

Tom Cech discovered that an intron RNA, later named the group I intron, could simultaneously self-excise itself from the RNA strand and ligate the flanking exons together in the complete absence of protein [1]. This challenged not only the idea that RNA could function as a catalyst, but also the concept that introns were not simply just "junk" genetic material. In the following years, the phylogenetic analysis of auto-catalytic introns revealed two distinct categories, resulting in the "grouping" of introns into group I and group II introns[15]. Through electron micrographs, it was revealed that while the group I intron formed a linear RNA product following catalysis, the group II intron formed circular RNAs[16] that were colloquially referred to as lariat RNAs.

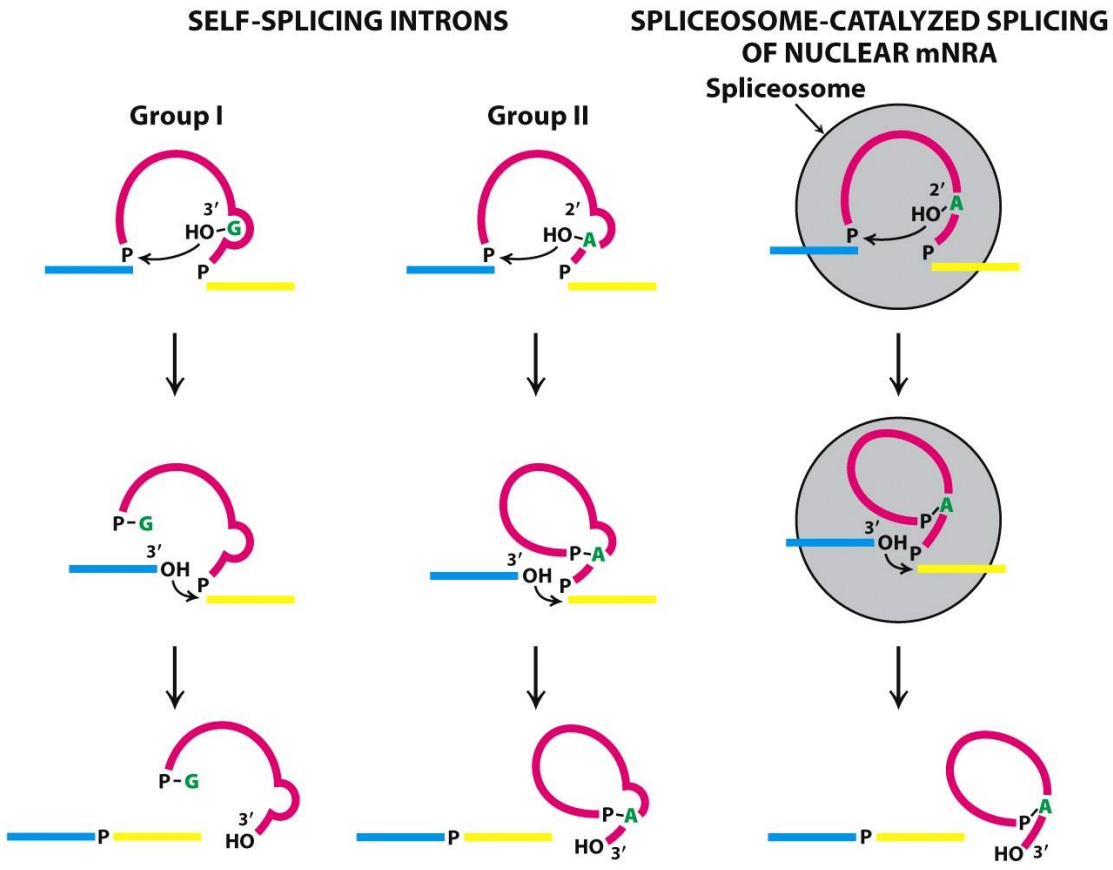

Figure 1.1: **Three systems of RNA splicing**. left) The group I intron utilizes an exogenous guanosine to facilitate the first step of splicing. The second step results in ligated exons and a free linear intron that now has an additional G at the 5′ end. middle) The group II intron does not require an exogenous co-factor, and instead uses a conserved adensonsine located at the 3′ on the intron for the first step of splicing. The second step produces ligated exons and a intron lariat. right) The spliceosome is a multicomponent-RNP that assembles on the pre-mRNA. It splices through a similar pathway to the group II intron and produces identical splicing products. Exons are denoted in blue and yellow, intron in magenta. Taken from Biochemistry, 7th Ed[17].

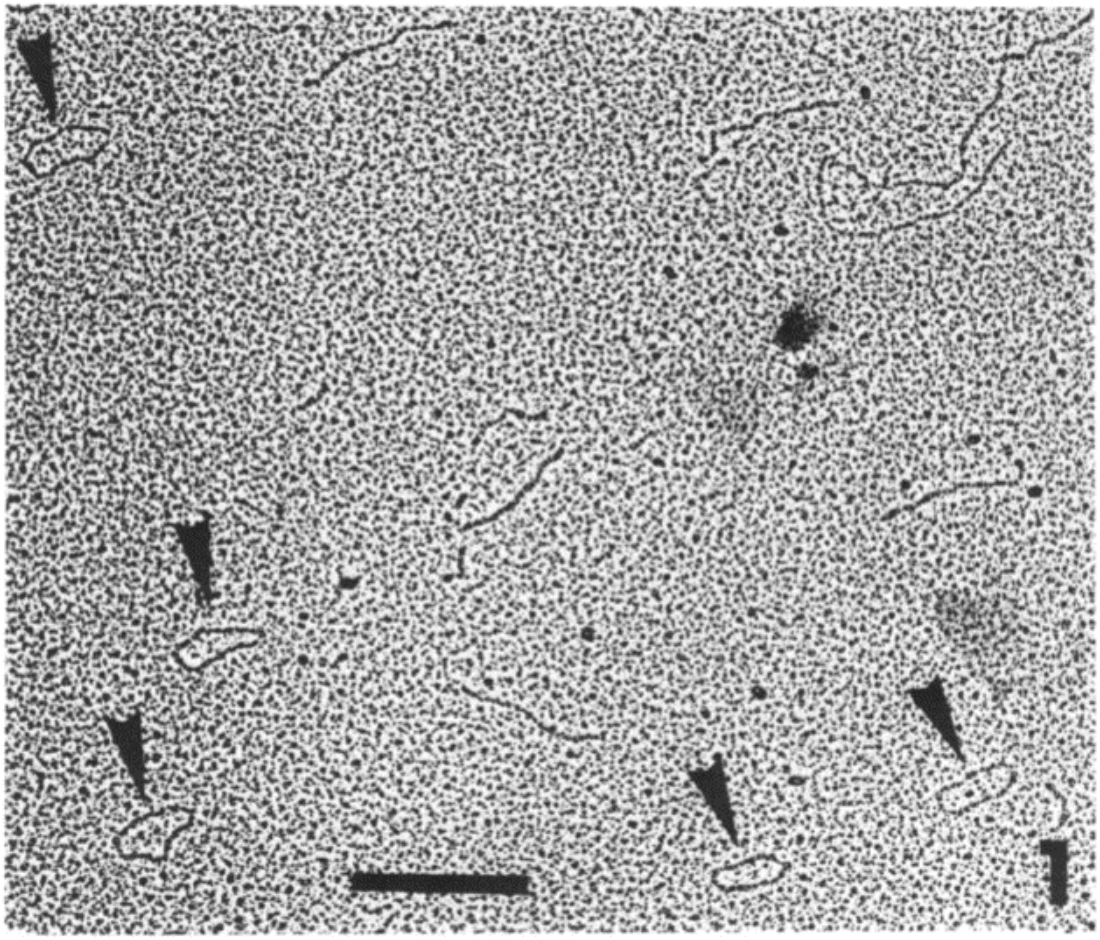

Figure 1.2: **Electron micrograph of mitochondrial RNAs show both circular and linear forms.** The authors were initially perplexed because the circular RNA withstood incubation at 80° C and 70% formamide. Arrows highlight the circular RNAs in the micrograph. Taken from Halbreich et al[16].

The group II intron catalyzes two sequential transesterification reactions in order to self-splice and ligate the flanking exon sequences together[11], and this reaction does not require any protein co-factor *in vitro* and is independent of ATP. For the first step of splicing, the 2′ hydroxyl of a conserved adenosine located near the 3′ of the intron is activated for nucleophilic attack at the 5′ splice site. This reaction cleaves the 5′ exon-intron junction and releases the 5′ exon from the phosphate backbone. It also results in an unusual 2′-5′ phosphate linkage between the adenosine and the first nucleotide, and gives rise to the

circular, lariat intron. The second step of splicing occurs once the lariat bond has left the active site, thus allowing the 3′ splice site to enter the active site. Once the 3′ splice site is correctly placed in the active site, the 3′ hydroxyl of the 5′ exon is activated for nucleophilic attack on the 3′ splice site and results in the ligation of the flanking exons and release of lariat RNA intron. Due to the reversibility of the transesterification reactions, excised group II introns are capable of acting as a genetic mobile element and can reverse splice into ligated exons[18,19].

1.2 Group II Intron Structure and Function

The group II intron is a mobile retroelement[12], with dual catalytic functionality to self-splice itself from an RNA sequence as well as reverse splice itself back into double-stranded DNA. It consists of two basic components: a ribozyme domain that adopts a specific fold capable of catalyzing the splicing reaction, and an ORF interrupting the ribozyme domain that codes for a protein known as the maturase[20]. The ribozyme component adopts a specific secondary structure that is divided into six distinct domains[21].

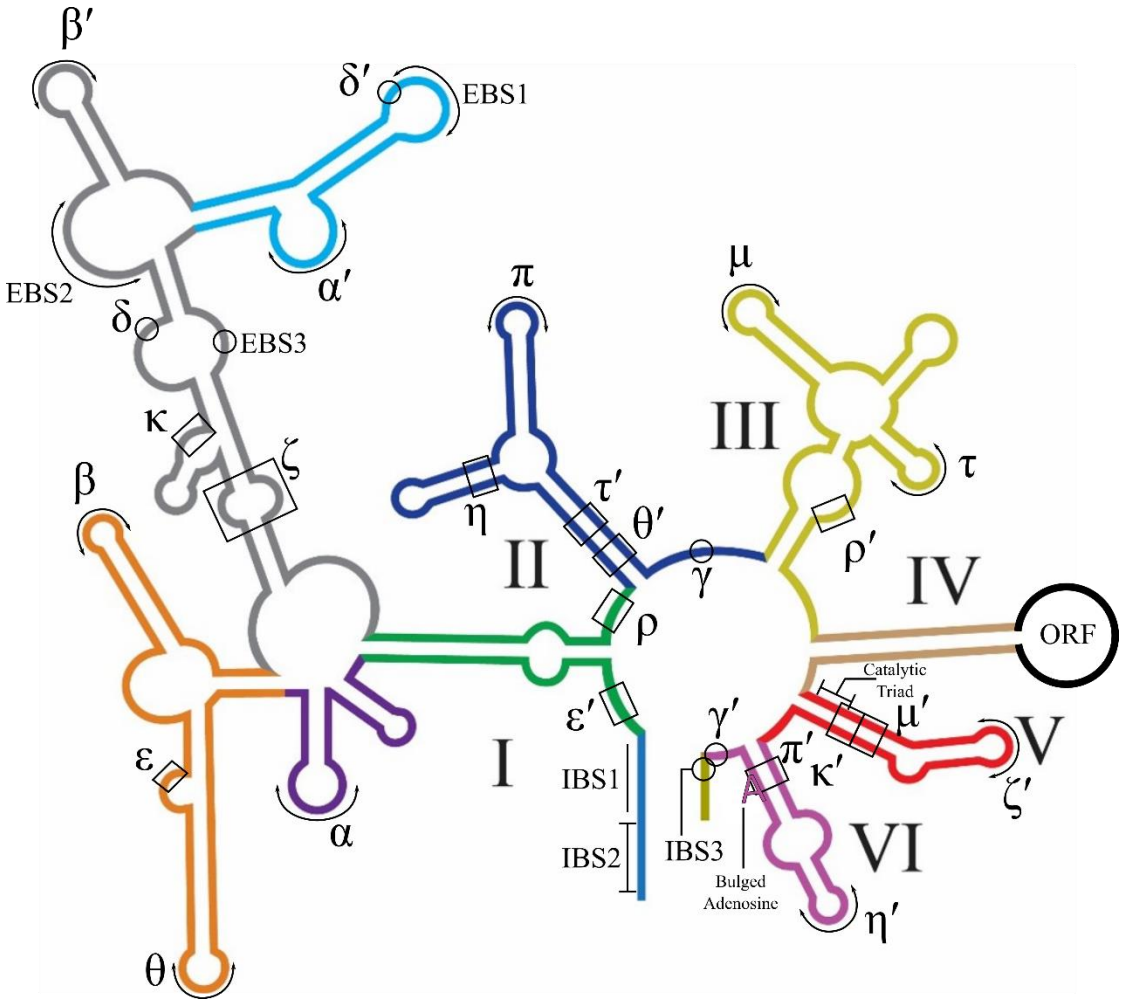

Figure 1.3**: The secondary structure of group II introns.** Colors of each domain are, in general, consistent within this body of work. DI: Green, purple, orange, grey, cyan. DI: blue. DIII: yellow. DVI: wheat. DV: red. DVI: magenta. 5′ exon: sky blue. 3′ exon: pale yellow. Courtesy of Jessica K. Peters.

The active site of the group II intron is composed of a catalytic triad as well as a two nucleotide[22] bulge that are found within domain V (DV). These residues form a metal binding pocket that is responsible for recruiting the catalytic metals, M1 and M2. The catalytic role of M1 and M2 was first proposed in the mid-1980s[23], but direct evidence did not come until the first atomic-resolution structure of the group II intron in 2008[24]. The structure revealed the

catalytic triad participating in base triples in order to recruit the catalytic metals to the active site of the group II intron.

Domain I is located on the 5′ end of the intron and makes up the majority of the ribozyme structure. Indeed, the vast majority of tertiary interactions found within the group II intron involve DI. Both α-α′[25] and β-β[26]′ are kissing loop interactions between the distal regions of DI. Interestingly, while the sequence of α-α′ is highly divergent among introns, disruption of this interaction through mutagenesis results in defective self-splicing activity[27]. It has therefore been identified to be an important structural element of the group II intron. These interactions are believed to be the first tertiary structures to be formed and therefore establish the structure of DI to set up the remaining tertiary interactions with the other domains. The sequence and length of these interactions are not strictly conserved, but their universal presence in group II introns suggest that they are critical to the overall intron architecture. ε -ε′ is a critical tertiary interaction that involves two WC base pairs between a region of DI and the third and fourth nucleotide of the intron[28]. Deletion of this interaction results in severe defects to the first step of splicing. The κ-κ′[29] and ζ-ζ′[30] interaction positions the base and tip of DV, respectively, to induce a ~70° bend in this region[24]. This bend is critical as it allows the DV 2-nt bulge to interact properly with the catalytic triad. In addition to the tertiary interactions with DV, DI contains **E**xon **B**inding **S**equences that properly position the 5′ exon (EBS1 and EBS2) and the 3′ exon (EBS3)[25]. These interactions are composed of long-range WC base pairs which allows for sequence specific pairing and contribute to the accuracy of the splicing reaction.

Domain II contains a tertiary interaction that positions the tip DVI to a helix of DII, known as η-η′[31], Domain II has been shown to not be required for the first step of splicing, with deletion of this entire domain having no effect on the formation of lariat-3′-exon intermediate, but results in a stalling of the second step and the formation of fully spliced products. This led to the hypothesis that domain II acted as an allosteric effector, which promoted DVI to be in an “up” state that corresponds to the first step and a “down” stage that corresponds to the second

step. A tertiary interaction was later proposed between a region of DI interacting with the internal loop of DVI and was named ι-ι′[32]. However, with the findings of this work, that model has come into question and a new model has been proposed (*vide infra*).

Connecting domains II and III is a highly conserved single stranded nucleotide linker, also known as J2/3. Within this linker are three critical nucleotides important for group II intron splicing. The first is the γ nucleotide[28], which is critical for the second step of splicing by recognizing the last nucleotide of the intron through a Watson-Crick base pair. This interaction with EBS3 anchors the correct 3′ splice site into the active site to promote the second step of splicing, with mutations at this position severely affecting the rate and fidelity of second step of splicing[33]. Directly upstream of the γ nucleotide are two highly conserved nucleotides that form base triples with the first two positions of the catalytic triad[34,35]. This reinforces the active site architecture, but also places the γ nucleotide adjacent to the active site. Surprisingly, deletion of these two nucleotides results in a dramatic inhibition of the second step, with only a minor effect on the first step.

Domain III is a highly conserved region within the group II intron, but its precise function was unclear. A previous study identified a single tertiary interaction, μ-μ′[36], which interacts with DV, but still failed to provide a rational for its conservation. Deletion of the entire domain resulted in an attenuation of the splicing reaction, but the addition of DIII in trans was shown to recover activity[37]. Along with other findings, this led to the hypothesis that DIII functions as a “catalytic” effector[37–39]. It was not until the findings of this present work (*vide infra*) that the precise role of DIII was elucidated.

Domain VI is critical for lariat formation[40], as it contains the conserved bulged adenosine that acts as the nucleophile for the first step of splicing. Through activation of its 2′ hydroxyl during the first step of splicing, it forms the defining lariat bond that connects the adenosine residue to the first nucleotide of the intron. The use of mutations and modified nucleotides has demonstrated a strict requirement for an adenosine at this position, with

recognition of this nucleotide driven primarily by the N6 amino group[41]. Furthermore, while other domains can be added in trans to the splicing reaction, DVI must be tethered to DV in order for the 3′ exon to be ligated properly[42].

1.3 Phylogenetic Diversity of Group II Introns

The group II intron are found in the mitochondria (IIA1 and IIA2), chloroplast (IIB1 and IIB2), and bacteria (IIA-F)[43]. Of those introns, the IIB1 intron is the best characterized class. This is because much of our understanding of group II intron splicing comes from the *αI5*γ intron from *Saccharomyces cerevisiae*[44] or the *P.li.LSUI2* from *Pylaiella littoralis*[45] (brown algae), both of which belong to the IIB1 class. *P.li.LSUI2* is a rather interesting intron due to its high in vitro splicing activity and lariat formation in rather minimal magnesium concentrations, as compared to *αI5γ*.

The bacterial class IIC introns, in contrast, represent a much different scenario. These introns were classified rather recently, and are distinct from the other intron classes due to their minimal structure[43]. They completely lack entire regions of DI and the associated tertiary interactions, such as β-β′ and IBS2-EBS2. Interestingly, this class has been shown to interact with an intrinsic transcription terminator structure, which may replace the IBS2-EBS2 interaction with the 5′ exon. In addition, the oddities of this class extend into the active site, where they are the only group II introns to contain a CGC catalytic triad sequence, as opposed to the AGC sequence that is found in other group II introns as well as the spliceosome. Furthermore, *in vitro* splicing of group IIC introns is prone to numerous errors, such as cryptic 5′ selection, and internal cleavage products. Lastly, the IIC intron is incapable of splicing through the lariat pathway and can only splice via hydrolysis. It is only with the addition of the maturase protein that the IIC intron switches from the hydrolysis pathway to the branching pathway[46].

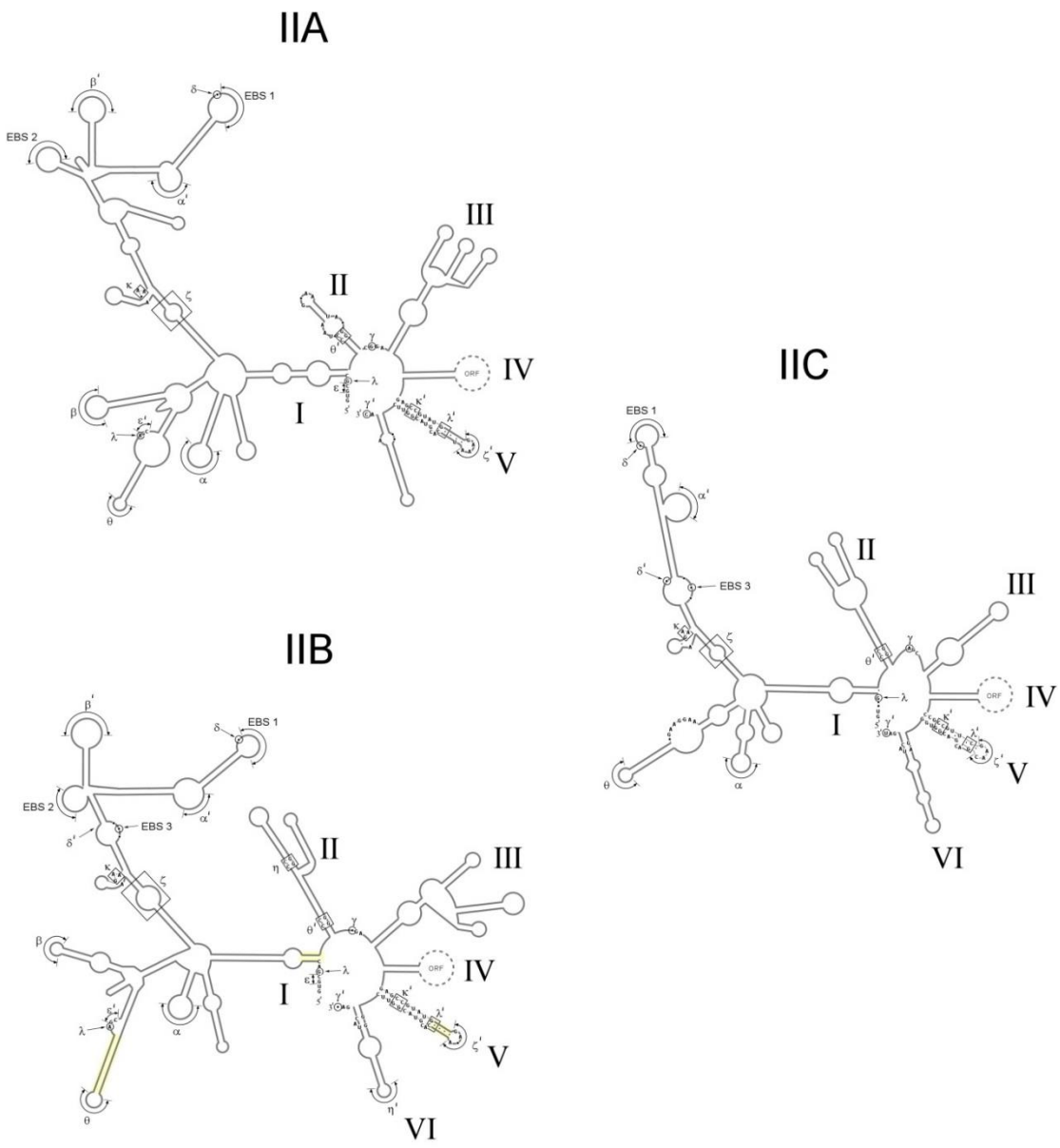

Figure 1.4: **The three group II intron subclasses.** Subclass IIA are found predominantly in mitochondrial genomes, while IIB are found in chloroplast genome. A distinction between these subclasses is how the 3′ exon is recognized, which in the former is is the δ-δ′, and the later it is EBS3-IBS3. The IIC subclass is limited to bacteria, and is differentiated by its minimal DI. Taken from Toor *et al.* 2009[47].

1.4 The Group II Intron and the Spliceosome

The spliceosome is a multimegadalton ribonucleoprotein complex that is composed of five snRNA components (U1, U2, U4, U5, and U6) and over a hundred different protein[48]. There is strong evidence that the group II intron and the spliceosome share a common ancestor. Unlike the group II intron, the spliceosome is highly dependent upon proteins to facilitate the remodeling of RNA-RNA, RNA-protein, and protein-protein interactions, even though the active site is thought to be composed of catalytic RNA. The spliceosome assembles on nascent pre-mRNA, beginning with U1 identifying the 5′-exon-intron junction[49] and U2 binding to the branch point within the intron[50]. The preassembled tri-snRNP, composed of U4/U6.U5[51–53], then joins the growing complex. The spliceosome then undergoes a major conformational rearrangement, with U1 and U4 leaving the spliceosome. At this point, the spliceosome is activated for catalysis and the splicing reaction proceeds. Several conformational changes are required to coordinate the two sequential splicing steps and are mediated by helicase-dependent rearrangements[54]. Once splicing has completed, the spliceosome leaves with the newly released intron, and the spliceosomal components are recycled and the process starts anew[55].

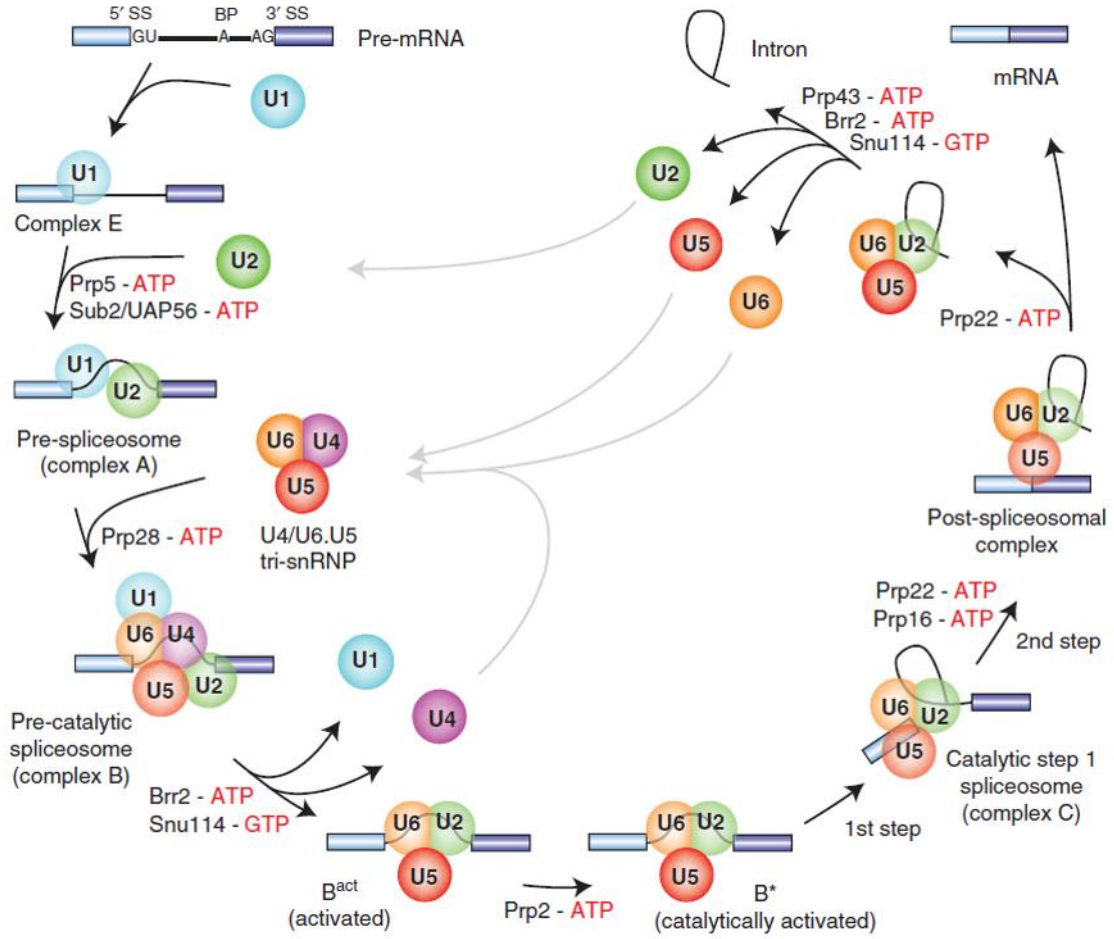

Figure 1.5: **The spliceosomal assembly pathway.** The assembly of the five snRNP components (colored circles) are shown to assemble on the pre-mRNA sequence. Conserved DExH/D-box RNA ATPases/helicase are specified at the stages where they facilitate structural rearrangements. Taken from Will and Lührmann[48].

Due to the fact that the group II intron and the spliceosome share a common catalytic mechanism[56], and exhibit sequence homology of core RNA structures, this has led to the hypothesis that they share a common evolutionary ancestor[16,21]. According to this model, various selective pressures caused the ribozyme to split into several pieces. Despite fragmentation, the ribozyme could still assemble in pieces to catalyze this trans-splicing-reaction, which is an observable phenomenon in group II introns[57]. In addition to the biochemical similarities, there exist similar structural elements between the group II intron and

the spliceosome. For example, DV and U6 of the group II intron and the spliceosome, respectively, contain almost identical secondary structural features[58]. Both contain a catalytic triad, and a two-nucleotide bulge that are separated by a conserved distance. Additionally, the RNA architecture of both systems contains analogous interactions in close proximity to the active site (Figure 1.6). Moreover, the crystal structure of Prp8 is observed to resemble the overall architecture of the maturase protein[59], providing a protein link between the two systems.

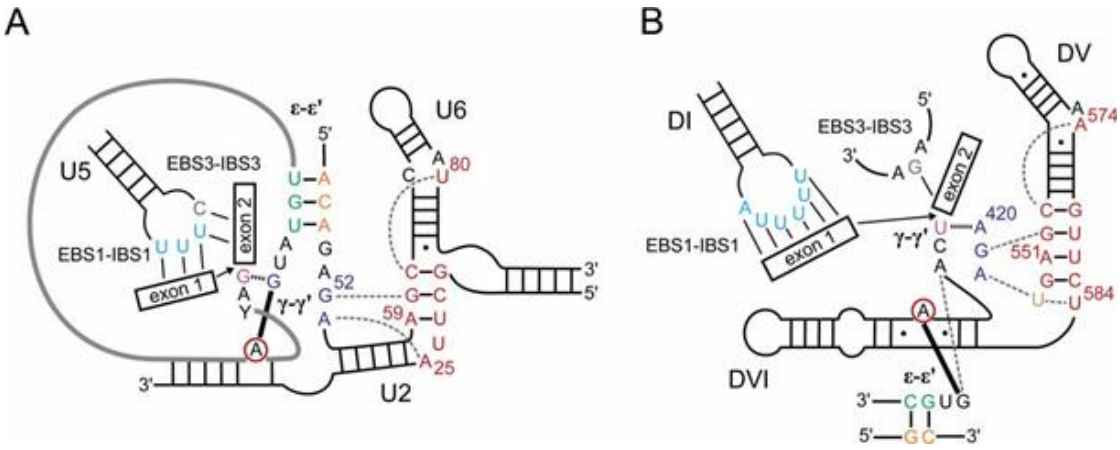

Figure 1.6: **Secondary structure comparison of the spliceosome (A) and the group II intron (B).** Both represent the RNA architecture prior to the second step of splicing. Colors used are consistent with Figure 1.3 and for the analogous regions of the snRNA components. The lariat bond is denoted by the thick black line that links the bulged adenosine (red circle) with the first nucleotide of the corresponding intron. Taken from Peters and Toor[60].

1.5 The Significance of Splicing

The process of RNA splicing must be maintained with high fidelity. While errors in splicing do occur, there are quality control mechanisms in place to minimize aberrant protein generation[61–63]. Unfortunately, these pathways can be overwhelmed or stop functioning properly, and result in disease phenotypes. It has been suggested that over a third of disease-causing mutations directly impact RNA splicing, although this number varies greatly between studies[64–68]. As a specific example, there is evidence that TDP-43 aggregates due to splicing

errors, and results in ALS-like symptoms in mouse models[69]. Furthermore, cancer is a prominent example of a splicing-related disease[70,71]. Cancer cells have been observed to have a high proportion of abnormally spliced mRNAs, and this this has been described as transcriptome instability.

Another important aspect of splicing is through alternative splice site selection. Prior to the Human Genome Project[72], it was believed that the human genome contains close to 100,000 genes[73,74]. It was, therefore, rather shocking when it was reported that the human genome had only 20,000-25,000 genes. In light of this conclusion, it became apparent that alternative splicing is a major force in creating proteome diversity by allowing one gene to code for multiple protein isoforms[75,76]. Since its discovery[77,78], it has been found to be a ubiquitous post-transcriptional event in eukaryotes. This is exemplified in the *D. melanogaster* gene Dscam, which is currently believed to have 38,016[79,80] spliced variants. This is underscored by the overall size of the genome at 15,016 genes.

1.6 Goal of this Dissertation

The first crystal structure of the group II intron provided the first direct insight into the mechanism of RNA splicing catalyzed by group II introns and the spliceosome. However, only the post-catalytic structure of the bacterial IIC intron was available at the start of this work. It was unclear based on the structural data of a single state what structural elements are necessary for specific elements of splicing. The following chapters detail key discoveries of how the group II intron is capable of promoting aspects that include, but are not limited to: the fidelity of splice site selection, lariat formation, and rearrangements to transition between steps of splicing. Due to the growing body of evidence that suggests a shared evolutionary history between the group II intron and spliceosome, the findings in this dissertation are directly applicable to eukaryotic RNA splicing.

References

1. 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, 487–496 (1981).

2. 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, 849–857 (1983).

3. Crick, F. H. C. The origin of the genetic code. *J. Mol. Biol.* 38, 367–379 (1968).

4. Jukes, T. H. *The Genetic Code. The Molecular Basis for Genetic Expression.* Carl R. Woese. *Q. Rev. Biol.* 43, 327–327 (1968).

5. Orgel, L. E. Evolution of the genetic apparatus. *J. Mol. Biol.* 38, 381–393 (1968).

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

7. Stormo, G. D. & Ji, Y. Do mRNAs act as direct sensors of small molecules to control their expression? *Proc. Natl. Acad. Sci. U. S. A.* 98, 9465–7 (2001).

8. Wightman, B., Ha, I. & Ruvkun, G. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell* 75, 855–62 (1993).

9. Lee, R. C., Feinbaum, R. L. & Ambros, V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* 75, 843–54 (1993).

10. Hutchins, C. J., Rathjen, P. D., Forster, A. C. & Symons, R. H. Self-cleavage of plus and minus RNA transcripts of avocado sunblotch viroid. *Nucleic Acids Res.* 14, 3627–40 (1986).

11. 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).

12. Nissen, P., Hansen, J., Ban, N., Moore, P. B. & Steitz, T. A. The structural basis of ribosome activity in peptide bond synthesis. *Science* 289, 920–30 (2000).

13. Muth, G. W., Ortoleva-Donnelly, L. & Strobel, S. A. A single adenosine with a neutral pKa in the ribosomal peptidyl transferase center. *Science* 289, 947–50 (2000).

14. Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. *Science* 289, 905–20 (2000).

15. Michel, F., Jacquier, A. & Dujon, B. Comparison of fungal mitochondrial introns reveals extensive homologies in RNA secondary structure. *Biochimie* 64, 867–81 (1982).

16. 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 RNA. *Cell* 19, 321–329 (1980).

17. Berg, Jeremy M.; Tymoczko, John L.; Stryer, L. in *Biochemistry* (W. H. Freeman, 2012).

18. 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).

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

20. Lambowitz, A.M., Caprara, M.G., Zimmerly S, in *The RNA World* (ed. Gesteland RF, Cech TR, A. J.) 451–485 (Cold Spring Harbor Laboratory Press, 1999).

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

22. Schmidt, U., Podar, M., Stahl, U. & Perlman, P. S. Mutations of the two-nucleotide bulge of D5 of a group II intron block splicing in vitro and in vivo: phenotypes and suppressor mutations. *RNA* 2, 1161–72 (1996).

23. 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).

24. 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).

25. 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).

26. Michel, F. & Ferat, J. L. Structure and activities of group II introns. *Annu. Rev. Biochem.* 64, 435–61 (1995).

27. 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).

28. Jacquier, A. & Michel, F. Base-pairing interactions involving the 5′ and 3′-terminal nucleotides of group II self-splicing introns. *J. Mol. Biol.* 213, 437–447 (1990).

29. 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).

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

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

32. 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).

33. 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).

34. 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).

35. 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).

36. 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).

37. 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).

38. 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).

39. 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, 3839–49 (1998).

40. 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).

41. 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).

42. Dib-Hajj, S. D., Boulanger, S. C., Hebbar, S. K., Peebles, C. L., Franzen, J. S., Perlman, P. S. Domain 5 interacts with domain 6 and influences the second transesterification reaction of group II intron self-splicing. *Nucleic Acids Res.* 21, 1797–804 (1993).

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

44. 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).

45. 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).

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

47. Toor, N., Keating, K. S. & Pyle, A. M. Structural insights into RNA splicing. *Curr. Opin. Struct. Biol.* 19, 260–6 (2009).

48. Will, C. L. & Lührmann, R. Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.* 3, (2011).

49. Yu, Y.-T., Scharl, E. C., Smith, C. M. and Steitz, J. A. in *The RNA World I1* (ed. Gesteland, R F., Atkins, J. F. and Cech, T.) 487–523 (Cold Spring Harbor Laboratory Press, 1999).

50. Parker, R., Siliciano, P. G. & Guthrie, C. Recognition of the TACTAAC box during mRNA splicing in yeast involves base pairing to the U2-like snRNA. *Cell* 49, 229–239 (1987).

51. Wan, R., Yan, C., Bai, R., Wang, L., Huang, M., Wong C. 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).

52. Nguyen, T. H. D., Galej, W. P., Bai, X., Oubridge, C., Newman, A. J., Scheres, S. H.

W., Nagai, K. Cryo-EM structure of the yeast U4/U6.U5 tri-snRNP at 3.7 Å resolution. *Nature* 530, 298–302 (2016).

53. Agafonov, D. E., Kastner, B., Dybkov, O., Hofele, R.V., Liu, W.T., Urlaub, H., Lührmann, R., Stark, H. Molecular architecture of the human U4/U6.U5 tri-snRNP. *Science* 351, 1416–20 (2016).

54. Cordin, O., Hahn, D. & Beggs, J. D. Structure, function and regulation of spliceosomal RNA helicases. *Curr. Opin. Cell Biol.* 24, 431–8 (2012).

55. Rino, J., Carvalho, T., Braga, J., Desterro, J. M. P., Lührmann, R., Carmo-Fonseca, M. A Stochastic View of Spliceosome Assembly and Recycling in the Nucleus. *PLoS Comput. Biol.* 3, e201 (2007).

56. Guthrie, C. Messenger RNA splicing in yeast: clues to why the spliceosome is a ribonucleoprotein. *Science* 253, 157–63 (1991).

57. Goldschmidt-Clermont, M., Choquet, Y., Girard-Bascou, J., Michel, F., Schirmer-Rahire, M., Rochaix J. D. A small chloroplast RNA may be required for trans-splicing in Chlamydomonas reinhardtii. *Cell* 65, 135–43 (1991).

58. 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).

59. 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).

60. Peters, J. K. & Toor, N. Group II intron lariat: Structural insights into the spliceosome. *RNA Biol.* 12, 913–7 (2015).

61. Isken, O. & Maquat, L. E. Quality control of eukaryotic mRNA: safeguarding cells from abnormal mRNA function. *Genes Dev.* 21, 1833–56 (2007).

62. Fasken, M. B. & Corbett, A. H. Process or perish: quality control in mRNA biogenesis. *Nat. Struct. Mol. Biol.* 12, 482–488 (2005).

63. Fasken, M. B. & Corbett, A. H. Process or perish: quality control in mRNA biogenesis. *Nat. Struct. Mol. Biol.* 12, 482–8 (2005).

64. Lim, K. H., Ferraris, L., Filloux, M. E., Raphael, B. J. & Fairbrother, W. G. Using positional distribution to identify splicing elements and predict pre-mRNA processing defects in human genes. *Proc. Natl. Acad. Sci. U. S. A.* 108, 11093–8 (2011).

65. Douglas, A. G. L. & Wood, M. J. A. RNA splicing: disease and therapy. *Brief. Funct. Genomics* 10, 151–64 (2011).

66. Faustino, N. A. & Cooper, T. A. Pre-mRNA splicing and human disease. *Genes Dev.* 17, 419–437 (2003).

67. Xiong, H. Y., Alipanahi, B., Lee, L. J., Bretschneider, H., Merico, D., Yuen, R. K. C., Hua, Y., Gueroussov, S., Najafabadi, H. S., Hughes, T. R., Morris, Q., Barash, Y., Krainer, A. R., Jojic, N., Scherer, S. W., Blencowe, B. J., Frey, B. J. RNA splicing. The human splicing code reveals new insights into the genetic determinants of disease. *Science* 347, 1254806 (2015).

68. Ward, A. J. & Cooper, T. A. The pathobiology of splicing. *J. Pathol.* 220, 152–63 (2010).

69. Ling, J. P., Pletnikova, O., Troncoso, J. C. & Wong, P. C. TDP-43 repression of

nonconserved cryptic exons is compromised in ALS-FTD. *Science* 349, 650–5 (2015).

70. Sveen, A., Ågesen, T. H., Nesbakken, A., Rognum, T. O., Lothe, R. A., Skotheim, R. I. Transcriptome instability in colorectal cancer identified by exon microarray analyses: Associations with splicing factor expression levels and patient survival. *Genome Med.* 3, 32 (2011).

71. Skotheim, R. I. & Nees, M. Alternative splicing in cancer: noise, functional, or systematic? *Int. J. Biochem. Cell Biol.* 39, 1432–49 (2007).

72. Lander, E. S., International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* 409, 860–921 (2001).

73. Fields, C., Adams, M. D., White, O. & Venter, J. C. How many genes in the human genome? *Nat. Genet.* 7, 345–6 (1994).

74. Antequera, F. & Bird, A. Number of CpG islands and genes in human and mouse. *Proc. Natl. Acad. Sci. U. S. A.* 90, 11995–9 (1993).

75. Pan, Q., Shai, O., Lee, L. J., Frey, B. J. & Blencowe, B. J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* 40, 1413–5 (2008).

76. Black, D. L. Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.* 72, 291–336 (2003).

77. Berget, S. M., Moore, C. & Sharp, P. A. Spliced segments at the 5’ terminus of adenovirus 2 late mRNA. *Proc. Natl. Acad. Sci. U. S. A.* 74, 3171–5 (1977).

78. Chow, L. T., Gelinas, R. E., Broker, T. R. & Roberts, R. J. An amazing sequence arrangement at the 5’ ends of adenovirus 2 messenger RNA. *Cell* 12, 1–8 (1977).

79. Zipursky, S. L., Wojtowicz, W. M. & Hattori, D. Got diversity? Wiring the fly brain with Dscam. *Trends Biochem. Sci.* 31, 581–8 (2006).

80. Neves, G., Zucker, J., Daly, M. & Chess, A. Stochastic yet biased expression of multiple Dscam splice variants by individual cells. *Nat. Genet.* 36, 240–6 (2004).

Chapter 2: Crystal structure of a group II intron in the pre-catalytic state.

2.1 Abstract

Group II introns are self-splicing catalytic RNAs that are thought to be ancestral to the spliceosome. Here we report the 3.65-Å crystal structure of the group II intron from *Oceanobacillus iheyensis* in the pre-catalytic state. The structure reveals the conformation of the 5' splice site in the catalytic core and represents the first structure of an intron prior to the first step of splicing.

2.2 Introduction

Group II introns are mobile retroelements that have a self-splicing catalytic RNA component thought to be evolutionarily related to the spliceosome.[1] Group II introns are postulated to have first originated in bacteria billions of years ago. These primitive molecules then invaded the first complex cells of eukaryotes, which subsequently evolved into plants, animals, and humans.[2] In eukaryotes, the dual functionalities of splicing and mobility became separated into the spliceosomal introns and retroelements. These genetic elements comprise at least 50% of most mammalian genomes and have contributed greatly to genomic diversity.[3] Because of this evolutionary impact, group II introns have been used as a tractable model system to gain insight into the structure and function of their eukaryotic descendants.

Group II intron RNAs catalyze splicing via two distinct transesterification events.[4] In the first step, nucleophilic attack occurs at the 5′ splice site by the 2′-OH of a bulged adenosine within the intron, or by a water molecule from solution. This results in the formation of an intron-3′-exon intermediate. This is followed by the second step in which the 3′-OH of the newly cleaved 5′ exon attacks the 3′ splice site to form ligated exons and liberated intron. During splicing, group II introns utilize exon binding sequences (EBS) called EBS1 and EBS3 to properly position the 5′ and 3′ splice sites in the catalytic core, respectively.[5,6] The 3′ splice site has the additional structural requirement that the 3′ terminal residue of the intron engages in the γ-γ′ interaction, which helps anchor the 3′-splice site in the core.[7] The γ-γ′ interaction

contains a single base pair between the terminal 3′ intron nucleotide (γ′) and a residue located between domains II and III of the intron (γ).

In this study, we set out to determine the structure of the group II intron in a pre-catalytic state directly preceding the first step of splicing. Previously, we had solved the crystal structure of a group II intron from *Oceanobacillus iheyensis* (*O. iheyensis*) in the post-catalytic state.[8,9] This structure revealed that splicing is likely to occur via a two-metal ion mechanism, catalyzed by residues within the highly conserved intron domain V. In addition, the scissile phosphate of bound ligated exon substrate was found to be positioned directly over the active site. In this post-catalytic structure, the junction between the two exons is likely to represent a relaxed state that exists after the intron completes both steps of splicing, immediately before exon release. RNA splicing is a dynamic process involving multiple conformational changes to properly position substrates in the active site. In order to understand and visualize the different stages of splicing, it is necessary to trap the intron at each stage and determine its structure using x-ray crystallography. In particular, the pre-catalytic state has never been visualized in any intron splicing system.

Here we report the 3.65 Å crystal structure of the *O. iheyensis* intron trapped in the pre-splicing state, before the onset of catalysis. This represents the first structure of an intron prior to splicing and allows us to visualize the position and conformation of the 5′ splice site within the catalytic core of the group II intron. In order to trap the intron at this stage and prevent cleavage of the 5′-splice site, a splicing precursor construct of the *O. iheyensis* intron was deactivated through mutagenesis of the central guanosine of the CGC catalytic triad. This mutation changes the G•U wobble to an A-U pair, thereby maintaining pairing within the domain V helix, and minimizing potential disruption of the RNA structure. The mutant displayed no detectable splicing activity using denaturing polyacrylamide gel electrophoresis, however primer extension analysis revealed ~15% cleavage of the 5′ exon upon extended storage of the RNA at 4˚C (see Supplementary Discussion). Intron RNA was in vitro

transcribed and purified using a native purification procedure (see Supplementary Methods). In this procedure, the intron folds during in vitro transcription and is purified in a native state using diafiltration.[8] The intron RNA crystallized in space group P212121 and the structure was refined to Rwork=20.5% and Rfree=24.9% (Supplementary Table 1, PDB accession 4DS6).

2.3 Results

The overall fold of the pre-catalytic group II intron is remarkably similar to that of the post-catalytic state (figure 1a), indicating that no large-scale conformational changes occur in domains I to V during the transition between the two steps of splicing. In the core of the ribozyme (Figs. 1b and 1c; Supplementary Figs. 1 and 2), strong electron density is now visible for the 5′ splice site docked into the catalytic pocket of domain V. The 5′ exon is paired with EBS1 indicating that this interaction persists throughout the splicing reaction into the post-catalytic state. The junction between the 5′ exon and the intron undergoes a dramatic reversal in the direction of the RNA chain. This produces a sharp kink in the backbone that presents the scissile phosphate of the splice site to the active site in domain V. Thus, the structure reveals the requirement for a structural distortion of the 5′ splice site just prior to catalysis. This also results in an unusual arrangement of two highly distorted backbones in close proximity to each other with the two-nucleotide bulge region of domain V being very near to the 5′ splice site (Supplementary Fig. 3). The position of the 5′ splice site within this crystal structure is entirely consistent with previous crosslinking data on the unspliced (pre-catalytic) forms of the aI5γ and L.l.ltrB group II introns.[10-12] Based on the cumulative biochemical data, multiple groups had modeled the 5′ splice site in close proximity to the active site in domain V.[6,10,12-15] The kinked 5′ splice site observed in our crystal structure was first proposed and modeled by de Lencastre *et al*.[10,15] Even though the scissile phosphate is positioned in the active site, no native density is visible for the catalytic metal ions. The most plausible explanation is that the mutation of the catalytic triad abolished binding of these metals. Analysis of the backbone of domain V both in the wild-type (post-catalytic) and mutant (pre-

catalytic) introns reveals that the G359A mutation causes a slight shift in the active site that may abrogate binding of the catalytic metal ions (Supplementary Discussion and Supplementary Fig. 4). It should be emphasized that the exon-binding sequences within the intron are responsible for kinking of the 5′ splice site and these residues remain as the wild-type sequence. Therefore, binding of the catalytic metal ions is not expected to alter the conformation of the 5′ splice site.

Table 2.1: Data collection and refinement statistics

	Oi-G359A Mutant
Data collection	
Space group	$P2_12_12_1$
Cell dimensions *a*,	
b, *c* (Å)	89.88, 95.31, 226.17
α, β, γ (°)	90, 90, 90
Resolution (Å)	40.41-3.65 (3.71-3.65)
R_{sym}	5.4 (60.8)
I / *I*	23.6 (3.4)
Completeness (%)	98.5 (95.3)
Redundancy	3.9 (3.5)
Refinement	
Resolution (Å)	3.65
No. reflections	22036
R_{work} / R_{free}	20.5/24.9 (41.0/44.2)
No. atoms	
RNA	8429
Ligand/ion	16
Water	29
B-factors	
RNA	155.5
Ligand/ion	119.5
Water	132.9
R.m.s. deviations	
Bond lengths (Å)	0.008
Bond angles ()	0.983

* Values in parentheses are for highest-resolution shell.

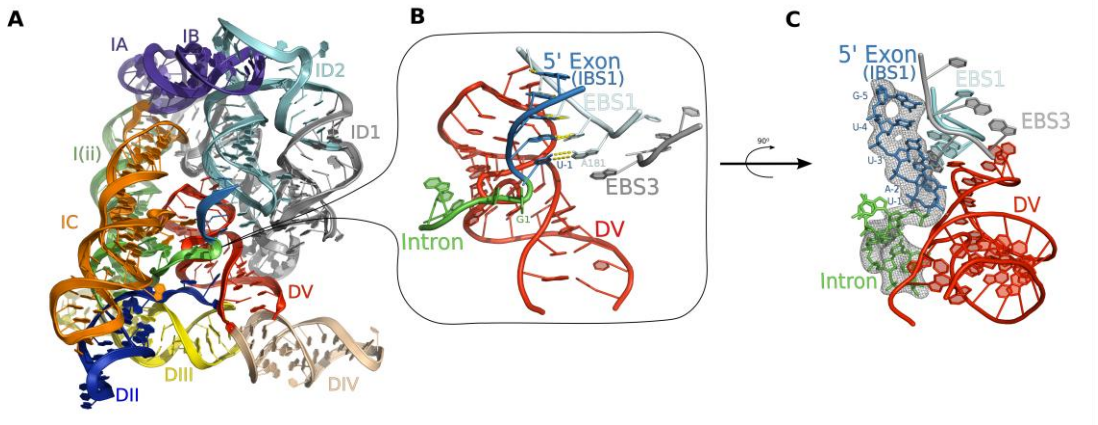

Figure 2.1: **The pre-catalytic structure of the group II intron.** a) The structure of the pre-catalytic state reveals the position of the 5′ splice site (junction between green and blue) in the center of the molecule. b) Close-up view of the 5′ splice site reveals a sharp kink in the backbone, which is positioned near the bulge and catalytic triad of domain V (red). c) F_o-F_c density for the 5′ splice site contoured at 3σ. The F_o-F_c density map was calculated using a model deleted for intron nucleotides 1-5 and the 5′ exon in order to avoid model bias.

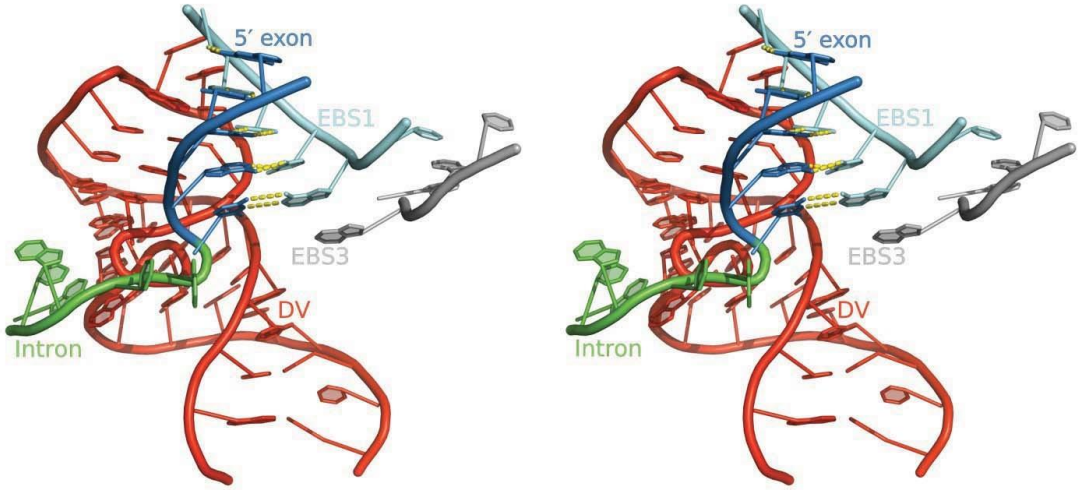

Figure 2.2: **Stereo representation of the pre-catalytic core of the group II intron.** See Figure 1b caption for details.

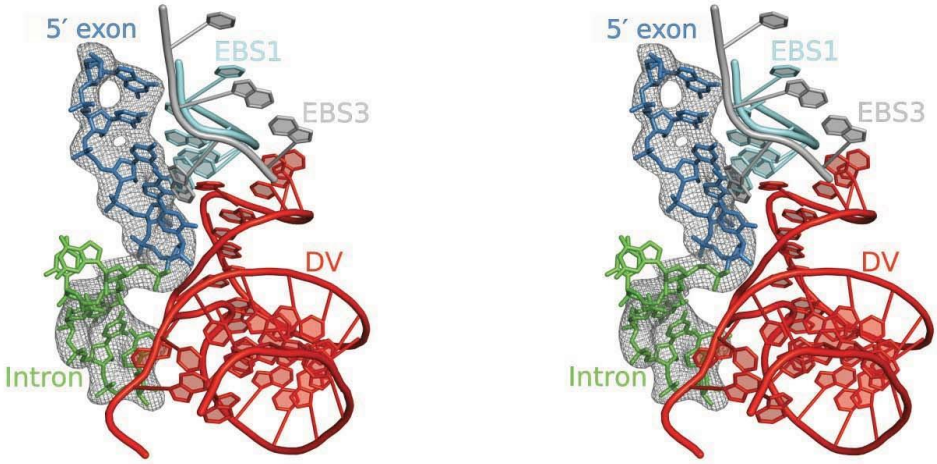

Figure 2.3: **Stereo representation of electron density for the 5 splice site.** See Figure 1c caption for details.

.

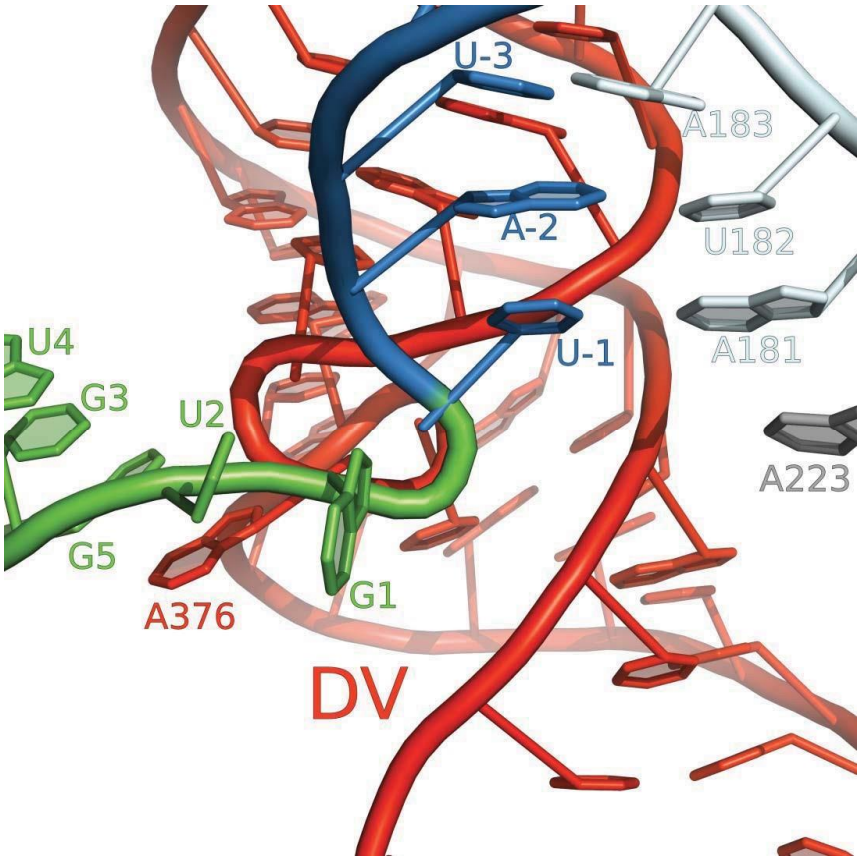

Figure 2.4: **The 5 splice site and the bulge of domain 5 are juxtaposed.** The core of the pre-catalytic intron reveals the unusual arrangement of two highly kinked backbones in close proximity to each other.

At the 5′ splice site, the first two nucleotides of the intron (G1 and U2) engage in a base stacking interaction that likely stabilizes the position of the backbone in this single-stranded region. Unlike the post-catalytic structure, there is now electron density for the nucleobase of the γ residue (A287), which forms a continuous base stack with the domain V bulge nucleotide A376 and the G5 residue from the 5′ terminus (Supplementary Fig. 5). A287 is functionally important because it pairs to the 3′ terminal nucleotide of the intron and plays a crucial role in positioning the 3′ splice site. The presence of the A287 nucleobase now enables us to model the likely position of the 3′ splice site within the catalytic core. Together with the pre- and post-catalytic structures, we have conceptualized a theoretical model of the complete group II intron splicing pathway (Fig. 2). In this model, the 5′ splice site is first drawn near the

catalytic metal ions and the kink is induced (Fig. 2a). This allows the nucleophile to attack and cleave the 5′ exon-intron junction. This is immediately followed by entry of the 3′ splice site to the active site. Nucleotides at the 3′ splice site are expected to pair simultaneously to both the γ nucleotide and EBS3 (Fig. 2b), resulting in a bifurcated set of interactions that are likely to kink the 3′ splice site. This arrangement is also consistent with previous biochemical data regarding the involvement of these tertiary interactions during 3′ splice site selection by group II introns.[6,7,16] The resulting conformation enables the 3′ splice site to approach the active site at an angle of attack of ~90 degrees from that of the 5′ splice site. As a result, the 3′-OH of the 5′ exon is located in close proximity to both the scissile phosphate of the 3′ splice site and one of the catalytic metal ions. This is consistent with the reaction mechanism whereby this hydroxyl group is activated by a metal ion and engages in nucleophilic attack at the 3′ splice site with subsequent ligation of the exons.[17] After ligation, the junction between the 5′ and 3′ exons exists in a relaxed state prior to product release (Fig. 2c).

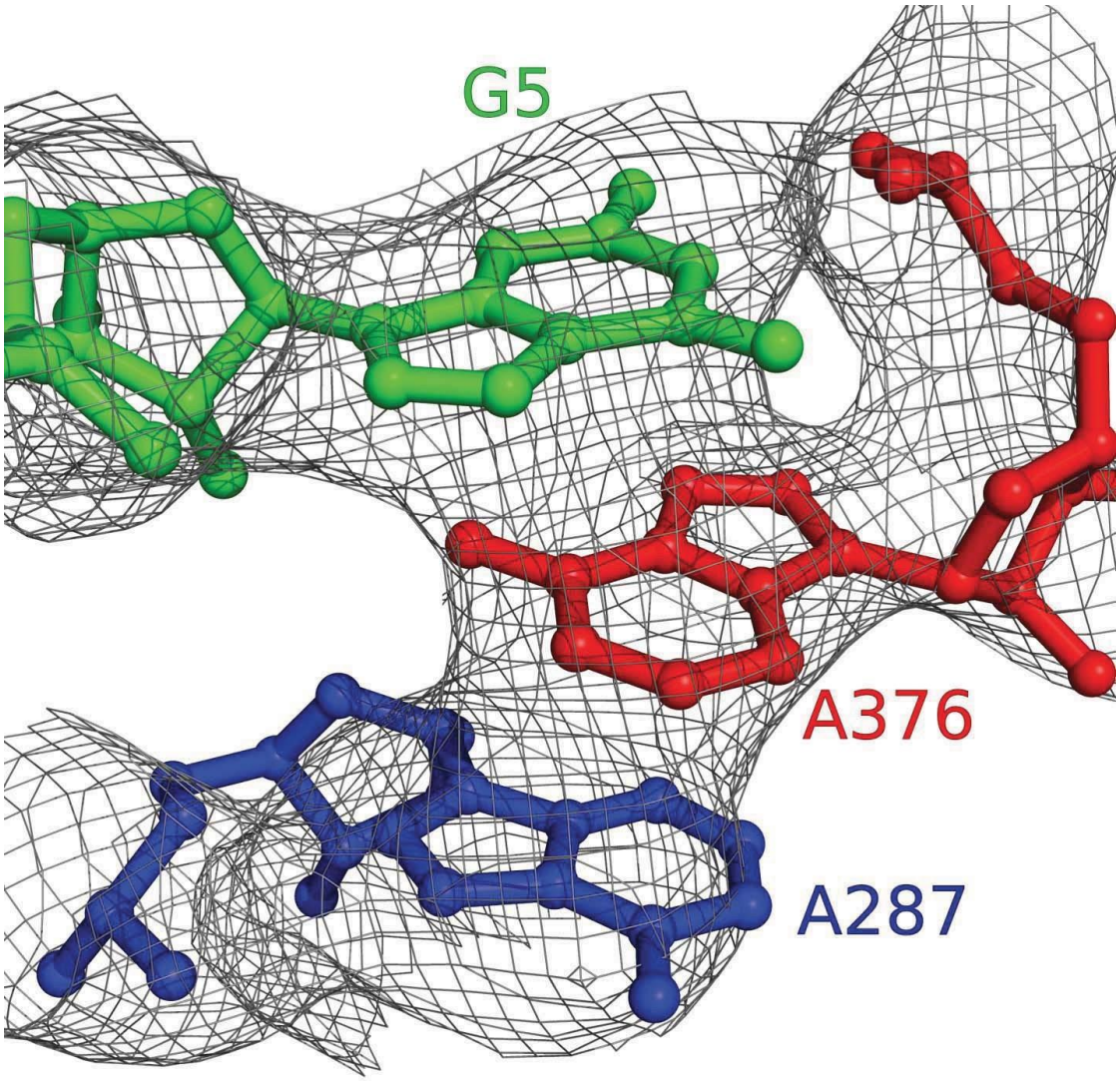

Figure 2.5: **Visualization of nucleotide A287.** Electron density contoured at 1 for the base stacking of G5, A376, and A287 (γ nucleotide).

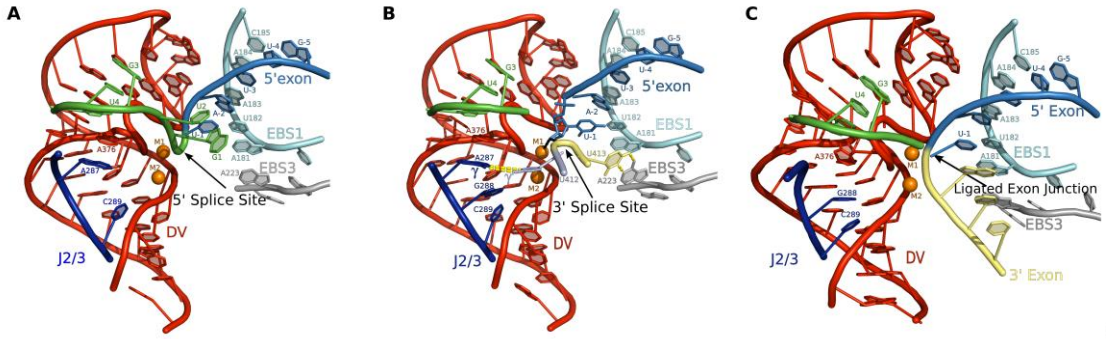

Figure 2.6**: Theoretical model for the complete group II intron splicing pathway.** a) The 5′ splice site is kinked immediately before the onset of catalysis. The kink is positioned in close proximity to the catalytic metal ions (M1 and M2). A ribose 2′-OH group or water molecule (not depicted) is activated for nucleophilic attack and cleaves the splice site. Residue 288 is not shown because there is no electron density for the base of this nucleotide in the pre-catalytic structure. b) The 3′ splice site is positioned in the active site through its interaction with EBS3 and the γ nucleotide. These tertiary interactions cause the 3′ splice site to also adopt a sharp kink which presents the scissile phosphate to the active site metal ions. The 3′-OH (shown in stick format) of the 5′ exon is in a position to coordinate to M1 as well as to the 3′ splice site (coordination indicated by black lines). G1 and U2 are not shown due to the fact that these nucleotides must depart before the 3′ splice site enters the active site. c) The 3′ splice site is cleaved, the exons ligated, and the product adopts a “relaxed” conformation prior to release by the intron. Abbreviation: J2/3 - junction sequence between domains II and III.

This theoretical model suggests that group II intron active sites are versatile and can accommodate both steps of splicing with small conformational alterations. Given their mechanistic similarities, it is likely that both group II introns and the spliceosome induce sharp bends at the splice sites for proper positioning of the scissile phosphates. The generality of this behavior is underscored by the evolutionarily unrelated group I introns, which induce a sharp kink at the 3′ splice site in constructs designed to capture the second step of splicing.[18] Therefore, the sharp structural distortion of splice sites is likely to contribute to the mechanism of all RNA splicing reactions.

References

1. Sharp, P.A. Five easy pieces. *Science* 254, 663 (1991).
2. Martin, W. & Koonin, E.V. Introns and the origin of nucleus-cytosol compartmentalization. *Nature* 440, 41-5 (2006).
3. Koonin, E.V. Intron-dominated genomes of early ancestors of eukaryotes. *J Hered* 100, 618-23 (2009).
4. Robart, A.R. & Zimmerly, S. Group II intron retroelements: function and diversity. *Cytogenet Genome Res*, 110, 589-97 (2005).
5. Jacquier, A. & Michel, F. Multiple exon-binding sites in class II self-splicing introns. *Cell* 50, 17-29 (1987).
6. Costa, M., Michel, F. & Westhof, E. A three-dimensional perspective on exon binding by a group II self-splicing intron. *EMBO J* 19, 5007-18 (2000).
7. 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).
8. 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).
9. Toor, N., Rajashankar, K., Keating, K.S. & Pyle, A.M. Structural basis for exon recognition by a group II intron. *Nat Struct Mol Biol* 15, 1221-2 (2008).
10. de Lencastre, A., Hamill, S. & Pyle, A.M. A single active-site region for a group II intron. *Nat Struct Mol Biol* 12, 626-7 (2005).
11. Noah, J.W. & Lambowitz, A.M. Effects of maturase binding and Mg^{2+} concentration on group II intron RNA folding investigated by UV cross-linking. *Biochemistry* 42, 12466-80 (2003).
12. Dai, L., Chai, D., Gu, S.Q., Gabel, J., Noskov, S.Y., Blocker, F.J., Lambowitz, A.M., Zimmerly, S. A three-dimensional model of a group II intron RNA and its interaction with the intron-encoded reverse transcriptase. *Mol Cell* 30, 472-85 (2008).
13. Boudvillain, M., de Lencastre, A. & Pyle, A.M. A tertiary interaction that links active-site domains to the 5′ splice site of a group II intron. *Nature* 406, 315-8 (2000).
14. Wang, J. Inclusion of weak high-resolution X-ray data for improvement of a group II intron structure. *Acta Crystallogr D Biol Crystallogr* 66, 988-1000 (2010).
15. 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).
16. 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).
17. Roitzsch, M., Fedorova, O. & Pyle, A.M. The 2′-OH group at the group II intron terminus acts as a proton shuttle. *Nat Chem Biol* 6, 218-224 (2010).
18. 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).

2.4 Methods

The crystallization construct consists of a 7 nt 5′ exon, the G359A mutant intron, and a 2 nt 3′ exon. Intron RNA was in vitro transcribed and purified using a minor variation of the method described in Toor et al.[1], with the only difference being that the intron RNA is incubated at 55°C after in vitro transcription instead of 60°C. The RNA was crystallized by combining one volume of intron RNA (10 mg/ml), one volume of 0.5 mM spermine, and one volume of 10 mM magnesium acetate, 50 mM MES pH 5.6, and 2.3 M ammonium sulfate (mother liquor). Crystals formed via sitting drop vapor diffusion at 30° C in three weeks. Crystals were gradually transferred to a cryoprotectant containing 10 mM magnesium acetate, 50 mM MES pH 5.6 and saturated lithium sulfate, followed by flash freezing in liquid nitrogen. Molecular replacement was done using PHENIX.[2] The post-catalytic structure (PDB accession 3IGI) was used as a search model with a deletion of intron nucleotides 1-5 and the exon substrate to avoid bias at the 5′ splice site. Minimization was performed using base pair restraints in CNS version 1.3.[3] Simulated annealing and TLS refinement was done using PHENIX. Modeling was done in Coot[4] using the RCrane[5] plugin to improve model geometry. Molprobity was used to evaluate the quality of the refinement at all stages. Figures were prepared with PYMOL.[6] Data was collected at NE-CAT beamline 24-ID-C at the Advanced Photon Source (Argonne National Laboratory, Argonne, Illinois).

2.5 Supplementary Discussion

We also observed weak density near the 5′ splice site and the J2/3 region, which corresponds to a small fraction of the intron reacting to form the post-catalytic state. Primer extension analysis of the 5′ end of the RNA at pre- and post-crystallization time points revealed that ~15% of the 5′ exon was cleaved in both samples (data not shown). The RNA is stored at 4°C in a buffer containing 10 mM Mg^{2+} prior to crystallization. Therefore, the density could represent evidence of metal ions being weakly bound to the active site and catalyzing cleavage at a very slow rate over an extended period of time. We observed no in vitro splicing

activity for the mutant intron with incubation at 55°C for 30 minutes in a splicing buffer consisting of 125 mM $MgCl_2$, 500 mM NH_4Cl, and HEPES pH 7.0. In contrast, the wild-type intron splices to completion under these conditions.[1]

We postulate that mutation of the catalytic triad leads to a lack of electron density for the catalytic metal ions in the active site. Supplementary figure 4 shows overlapping models of the catalytic domain V from both the pre- and post-catalytic structures. The post-catalytic structure represents the wild-type form of the active site which is competent for binding catalytic metal ions (metal ions not shown). The cleft formed between the backbones of the two-nucleotide bulge and catalytic triad is responsible for coordinating these ions. With the G359A mutant, one can see that there a slightly increased distance (~2 Å greater) between the backbones of the triad and the bulge. This increase in distance prevents metal ions from binding because magnesium cannot span this larger distance to simultaneously coordinate to both backbones. Therefore, the structural effects of this mutation are subtle and cause little perturbation within the active site, with the overall conformation of domain V being very similar in both pre- and post-catalytic states. It is also likely that the crystallization conditions negatively affect binding of the metal ions. The crystals only formed under very high ionic strength (2.3 M ammonium sulfate) and it is known that incubation of biological macromolecules at high salt concentrations interferes with the binding of physiologically important metal ions.

During modeling of the catalytic triad of domain V, we noticed that electron density for the mutant nucleotide A359 was consistent with that of an A•U Hoogsteen pair. We have modeled it as such; however at this level of resolution, it is not possible to conclusively discriminate between the two possibilities.

Supplementary References

1. 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).

2. 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, T.C., Zwart, P.H. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66, 213-21 (2010).

3. Brunger, A.T. Version 1.2 of the Crystallography and NMR system. *Nat Protoc* 2, 2728- 33 (2007).

4. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66, 486-501 (2010).

5. Keating, K.S. & Pyle, A.M. Semiautomated model building for RNA crystallography using a directed rotameric approach*. Proc Natl Acad Sci* U S A 107, 8177-82 (2010).

6. The PyMOL Molecular Graphics System, Version 1.3. (Schrödinger, LLC.).

Acknowledgements

We thank the staff of the NE-CAT beamline ID-24 at the Advanced Photon Source of Argonne National Laboratory. We thank Kevin Keating for valuable discussions. This work was supported by startup funds from the University of California, San Diego (N.T.) and the Howard Hughes Medical Institute (A.M.P.). R.T.C. was supported by the Cell, Molecular, and Genetics (CMG) Training Program funded by NIH predoctoral training grant T32GM007240.

Chapter 2, in full, is a reprint of the material as it appears in *Nature Structural and Molecular Biology*, **Chan, R.T.**; Robart, A.R.; Rajashankar, K.R.; Pyle, A.M.; Toor, N. (2012) Crystal structure of a group II intron in the pre-catalytic state. Nature Structure and Molecular Biology, 19(5):555-7. The dissertation author is the primary author on this paper.

Chapter 3: Crystal structure of a eukaryotic group II intron lariat.

3.1 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.

3.2 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 3.1) 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].

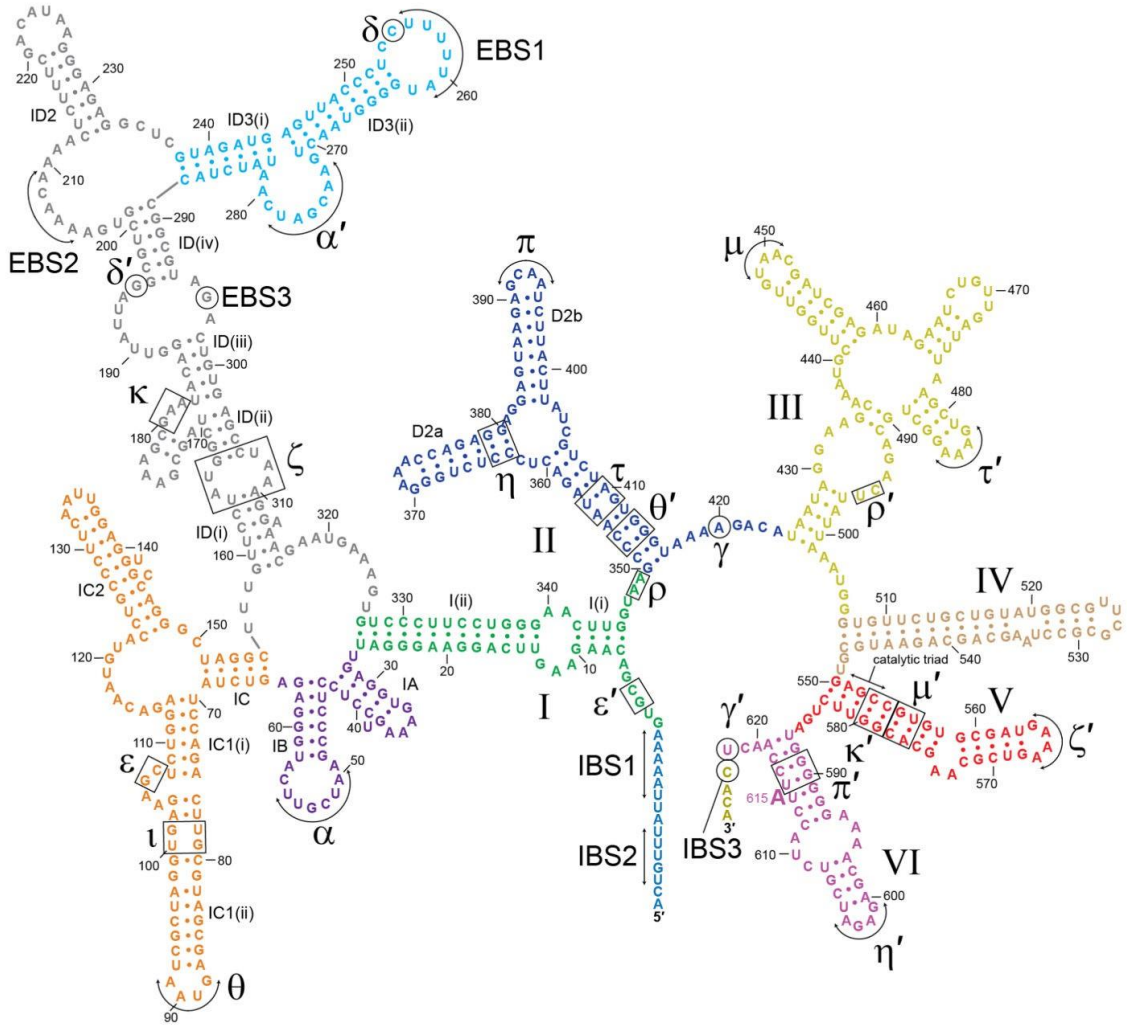

Figure 3.1: **Secondary structure of P.li.LSUI2 intron crystallization construct.** Tertiary interactions are indicated with Greek letters and domains are labeled with Roman numerals. Colouring of the individual domains is consistent with the overall view of the tertiary structure shown in Fig. 1.

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].

3.2 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 3.1) solved using a Yb^{3+} derivative (Figure 3.2). This represents the first crystal structure of a 2'-5' branched RNA molecule.

Table 3.1: **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.

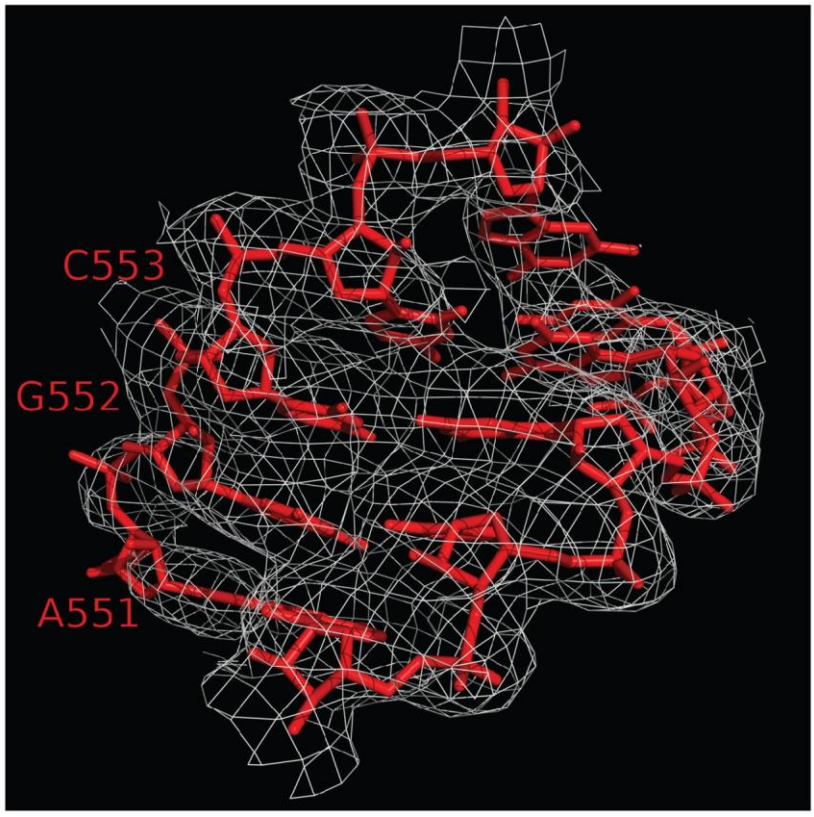

Figure 3.2: **The Yb-MAD experimental, density modified map of the portion of DV containing the catalytic triad contoured at 1.8σ**.

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 3.3). These newly visualized contacts include EBS2–IBS2 (Figure 3.4), 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 3.3 and Figure 3.5), as well as the 2'-5' lariat linkage between the first residue and the bulged adenosine.

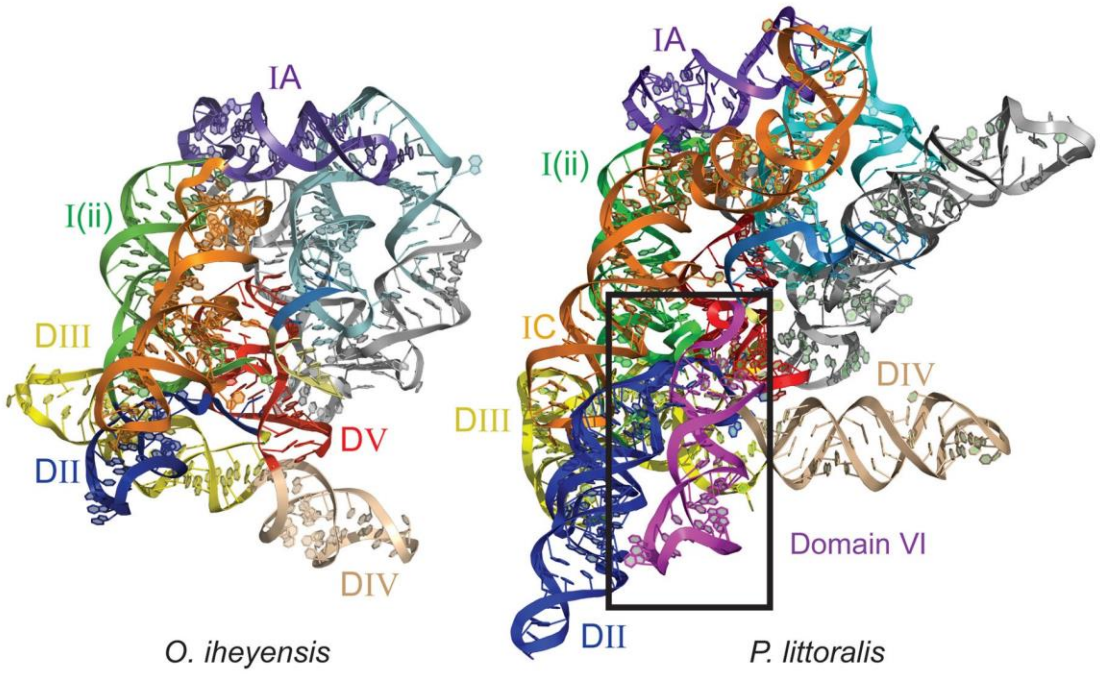

Figure 3.3: **A comparison of the tertiary structures of *O. iheyensis* and *P.li.LSUI2* group II introns.** *P.li.LSUI2* has a significantly larger and more complex structure with a correspondingly greater number of unique RNA tertiary contacts. Domain VI of *P.li.LSUI2* is highlighted within the box.

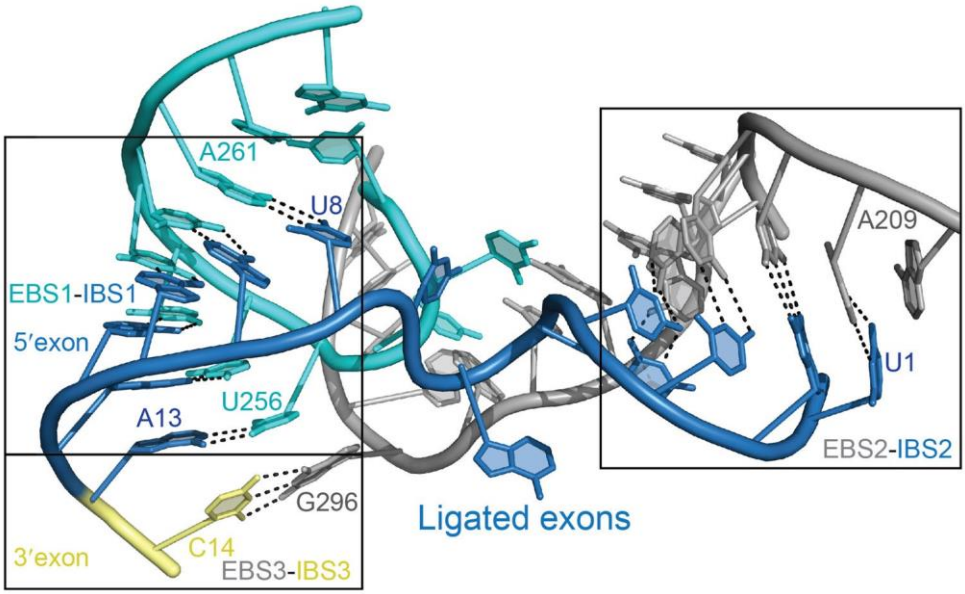

Figure 3.4: **The path of the 5′ exon through the intron structure.** The EBS1-IBS1 and EBS2-IBS2 interactions position the 5′ exon. They do not form a continuous binding interface with the presence of a highly distorted backbone at the junction between these two motifs. As a result, the helical axes 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.

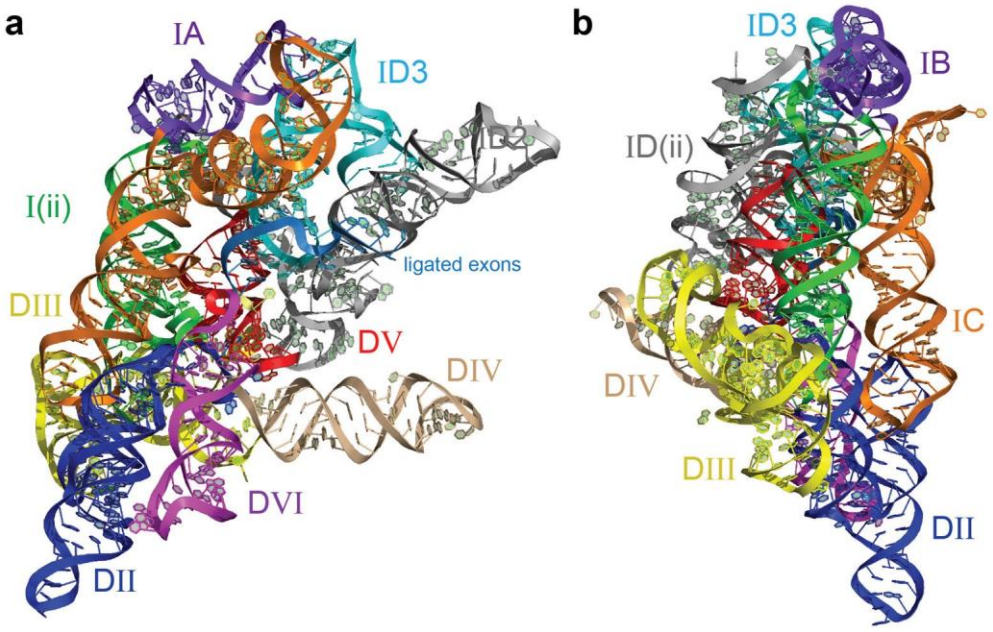

Figure 3.5: **Overall tertiary structure of the *P.li.LSUI2* intron.** Individual domains and subdomains are depicted in different colors. Domain names are labeled with Roman numerals. a and b show different rotations of the intron structure.

3.3 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 3.6a and 3.7a).

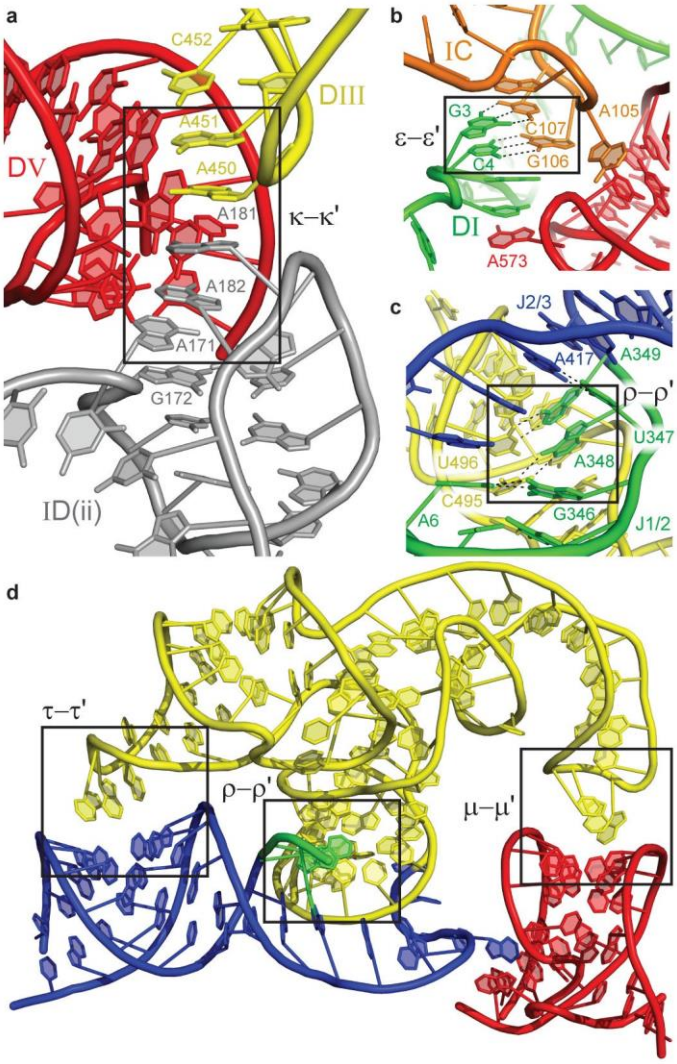

Figure 3.6: **Tertiary interactions in a IIB intron.** a, The κ loop (grey) and DIII (yellow) converge to form an extended base stack involving five adenosine residues inserting into the minor groove of DV (red). b, G106 and C107 form Watson-Crick pairs with C4 and G3 from the 5′ end to form the ε-ε′ 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).

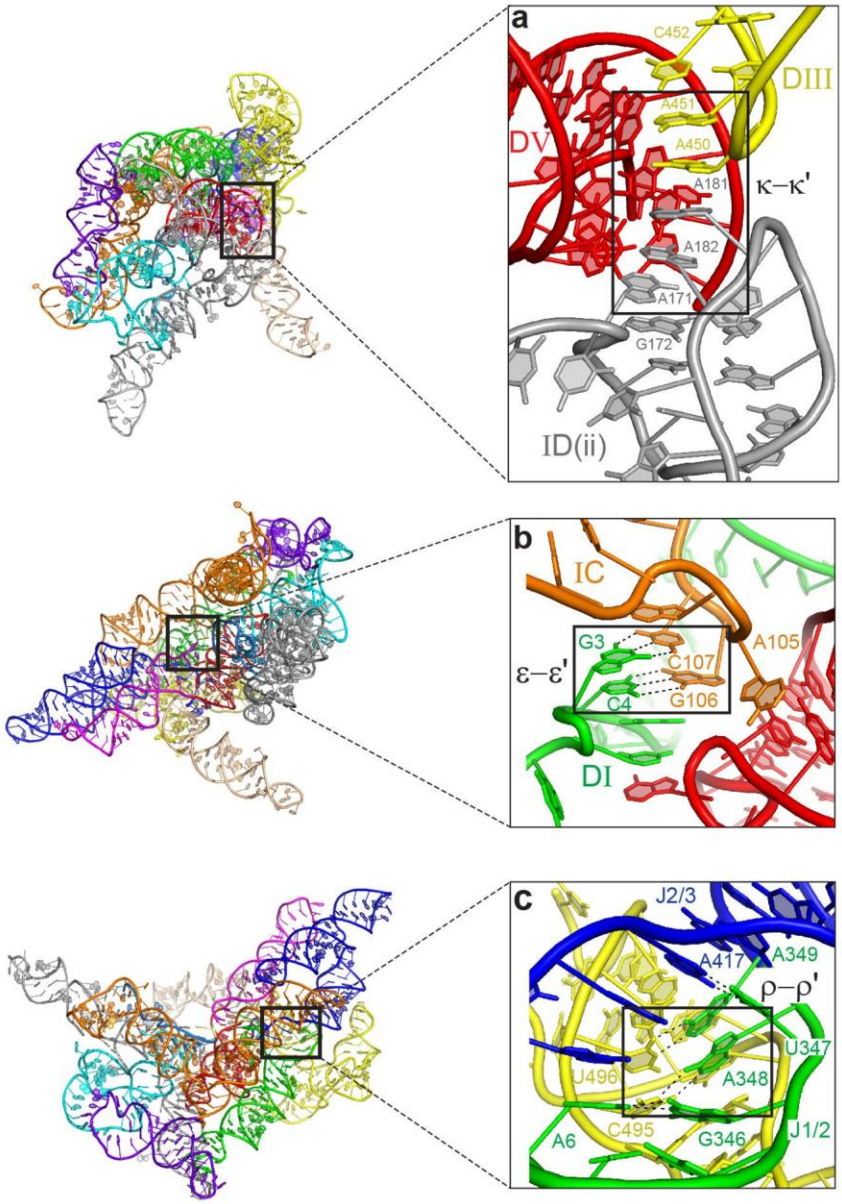

Figure 3.7: **Companion to Fig. 2 showing the location of the individual tertiary interactions relative to the overall structure.**

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 3.6b and Figure 3.7b). 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 3.6c and Figure 3.7c). 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 3.6c).

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 3.6d). This μ–μ' interaction[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].

3.4 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 3.8a). DII makes contacts with domains I, III and VI to organize a large portion of the intron structure.

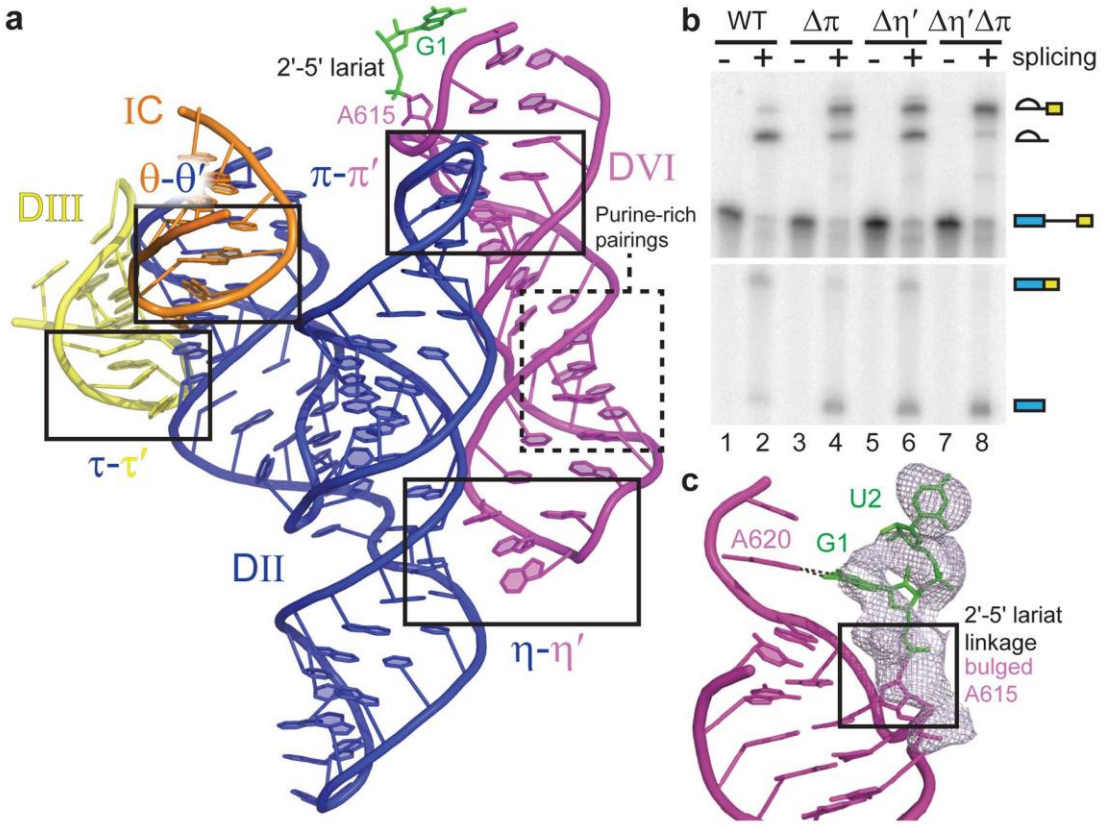

Figure 3.8: **The position of DVI within the intron structure.** a, DII forms an inverted Y-shaped structure that engages in four distinct tetraloop-receptor interactions with the IC stem (orange), DIII (yellow), and DVI (purple). DVI interacts with DII via two tandem tetraloop-receptor interactions, 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, F_o–F_c 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 3.1), 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 3.8a). 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 3.8a) 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 3.8b and Figure 3.9). 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.

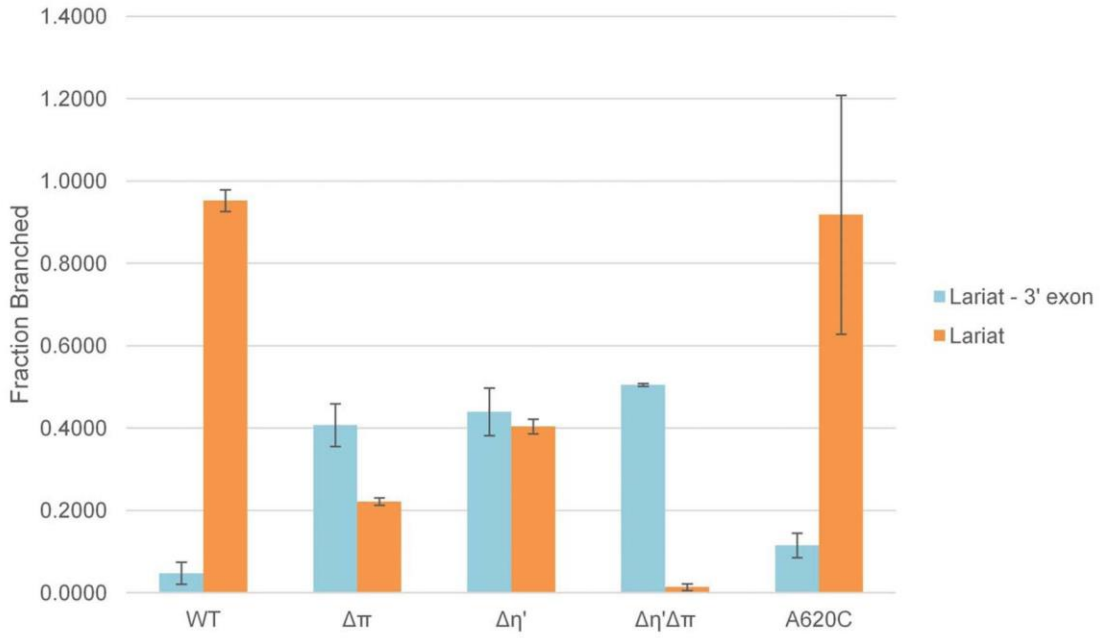

Figure 3.9: **Splicing assays for the DVI mutants showing the proportion of branched product.** Blue and orange bars correspond to lariat-3′ exon and lariat, respectively. The Δη′ and Δπ mutants accumulate large amounts of lariat-3′ exon, thus indicating a second step splicing defect. The Δη′Δπ double 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[44], indicating a certain degree of flexibility for this pairing. There is evidence for a similar interaction between the termini of nuclear introns[45] 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 3.8a). 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 3.10a). The nucleobase component of the A615 residue is disordered (Figure 3.8c and Figure 3.10b), 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 3.8c), which is important for the second step (Figure 3.9).

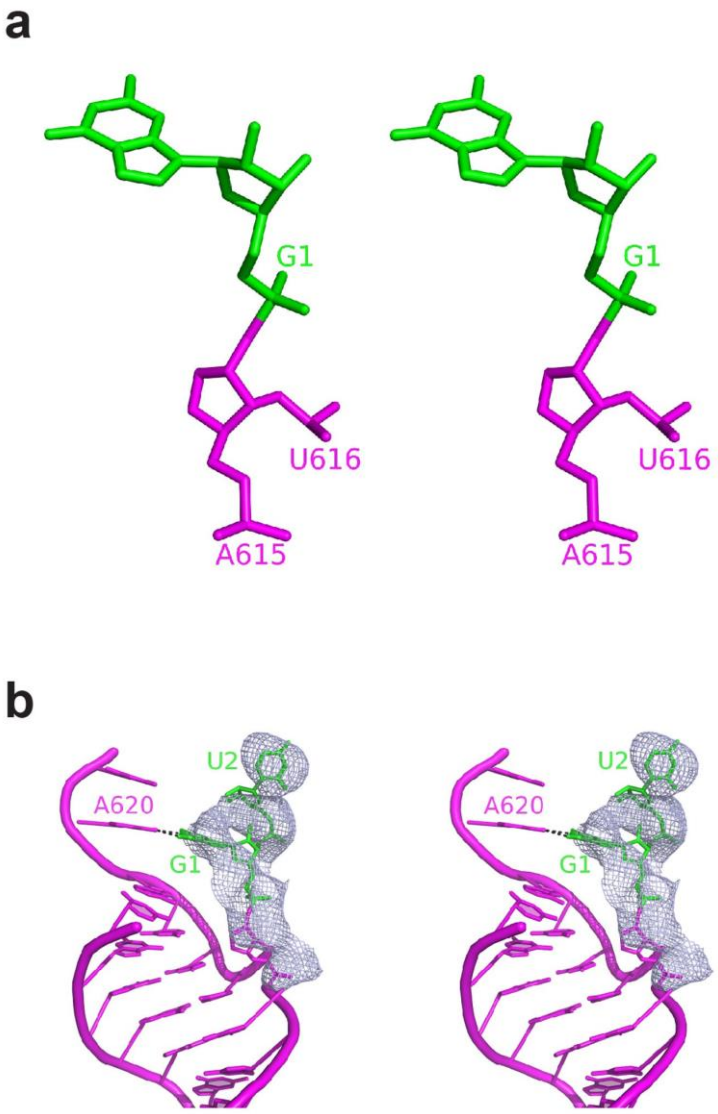

Figure 3.10: **Stereo depiction of the lariat bond.** a, The lariat 2′-5′ phosphodiester bond in wall-eyed stereo format. b, Stereo version of Fig. 3c. See Fig. 3c legend for details. F_o-F_c density for the lariat bond contoured at 3σ.

3.5 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[9,21]. This revealed four large anomalous peaks in the ribozyme core (Figure 3.11a). 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.

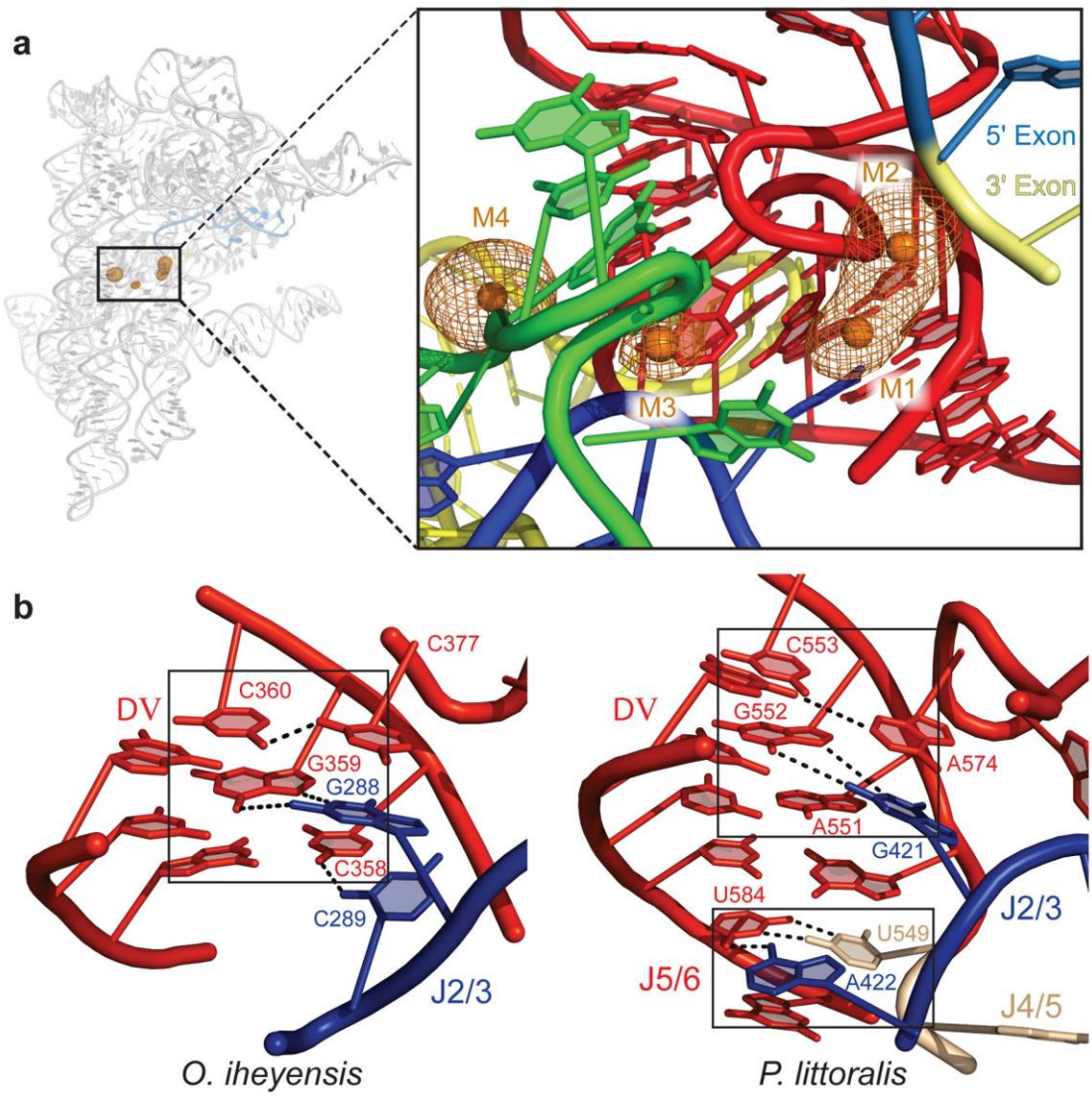

Figure 3.11: **The core of the of *P.li.LSUI2 intron*.** a, Yb^{3+} anomalous map (orange mesh) contoured at 12σ reveals four large peaks, which correspond to highly-coordinated magnesium ions (orange spheres). b, Comparison of the catalytic triplexes found in the *O. iheyensis* and *P.li.LSUI2* introns. *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 3.12a), 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 3.6). 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 3.12b). These residues are in close spatial proximity to the recently proposed ι motif (Figure 3.1) that is known to promote lariat formation and is predicted to serve as a receptor for positioning DVI (ref. 22). An ι G79A point mutant was crystallized, revealing strong signals for M1 and M2; however, M3 and M4 were no longer visible (Figure 3.12c, d). Splicing assays of this mutant also show a negative effect on the first step of splicing (Table 3.2). 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[23], 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 3.12e).

Table 3.2: **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

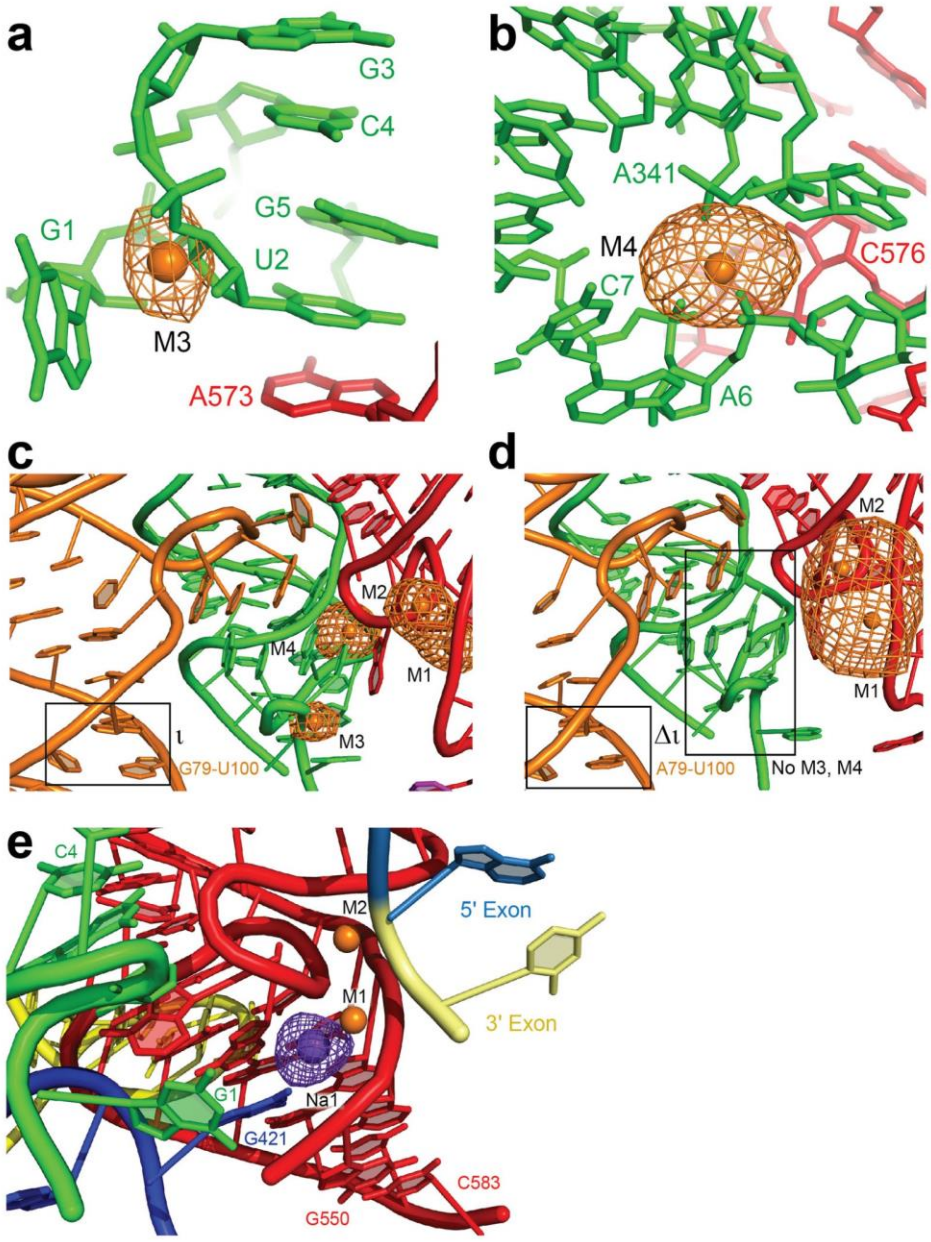

Figure 3.12: **Anomalous peaks in the core of the intron.** a and b depict the RNA ligands surrounding metals M3 and M4, respectively. Yb^{3+} anomalous map contoured at 9σ. c, Yb^{3+} anomalous map for WT contoured at 9σ. d, Compared with the WT intron, the Yb^{3+} anomalous map for the ι G79A mutant (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[46]. 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.

3.6 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[24]. 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 3.11b). The possibility of J2/3 participating in conformational changes has been previously postulated[25]. 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.

3.7 π–π' is a dynamic interaction

DVI is proposed to engage in large-scale conformational changes between the two steps of splicing[20,22]. To test this model, a catalytic triad mutant (AGC→GAU) inactive for splicing was crystallized and solved at 7 Å (Figure 3.13). 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.

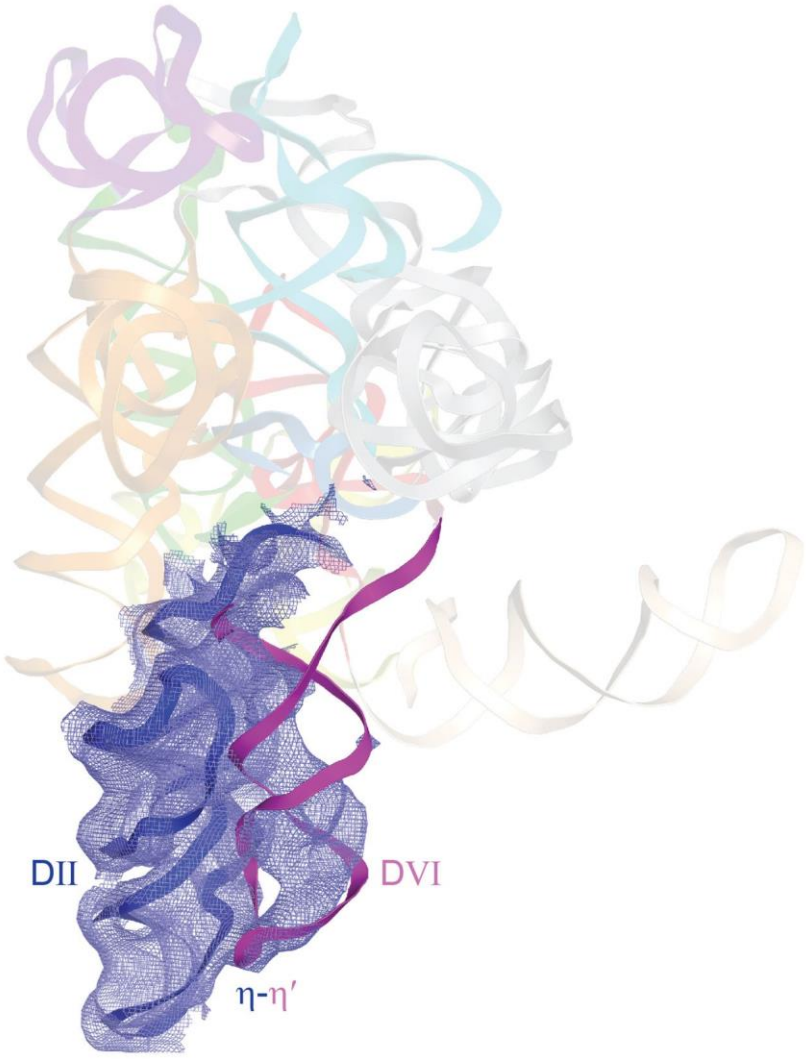

Figure 3.13: **$2F_o$-F_c density for DVI in the pre-catalytic structure contoured at 1σ**. The η-η' interaction persists throughout the splicing reaction and is visible in the pre-catalytic state. The weaker density for the central region of DVI suggests a partially disordered, dynamic region with possible 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 3.8b), 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 3.14). 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.

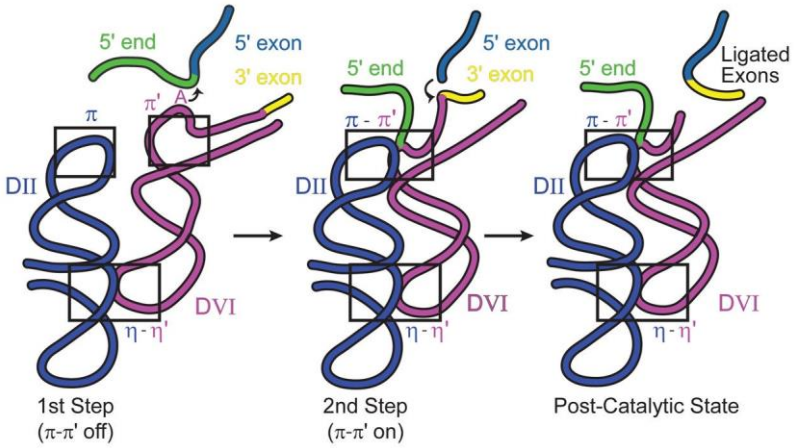

Figure 3.14: **Model for DVI as the conformational switch for splicing.** π-π' mediates the transition between the two steps of catalysis. See text for details.

3.8 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[26] 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'[27].

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[28]. In the spliceosome, the U6 snRNA pairs with the 5' end[29] 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[30]. These reverse splicing reactions are the mechanism through which group II introns are able to invade DNA[31] 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[32], 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[33]. 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.

3.9 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 vitro 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[34], heavy metal sites were identified with SHELXD[35] and phasing done using Phenix[36] and SHELXE[37]. RNA nucleotides were modeled using COOT[38] and the RCrane plugin[39]. The phylogenetically predicted secondary structure (Extended Data Fig. 1) guided modeling into the electron density. Structure refinement was done using Buster[40], Phenix[36], DEN[41], and Phenix.Erasser[42]. The 2′-5′ phosphodiester bond was restrained using a cif restraint file in phenix.refine. All software was compiled by SBGrid[43].

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.

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*. 1984; 37:415–427.

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*. 1984; 225:898–903.

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*. 1985; 313:552–557.

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*. 1986; 44:213–223.

5. 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*. 1986; 44:225–234.

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*. 2004; 24:440.

7. Galej, W.P., Oubridge, C., Newman, A.J., Nagai, K. Crystal structure of Prp8 reveals active site cavity of the spliceosome. *Nature*. 2013; 493:638–643.

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*. 2013; 503:229–234.

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

10. Lambowitz, A.M., Zimmerly, S. Group II introns: mobile ribozymes that invade DNA. *Cold Spring Harb Perspect* Biol. 2011; 3

11. Toor. N., Hausner. G, Zimmerly, S. Coevolution of group II intron RNA structures with their intronencoded reverse transcriptases. *RNA*. 2001; 7:1142–1152.

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

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*. 1997; 274:353–364.

14. Rest, J.S., Mindell, D.P. Retroids in archaea: phylogeny and lateral origins. *Mol Biol Evo*l. 2003; 20:1134–1142.

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*. 2006; 34:6461–6471.

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*. 1998; 17:7091–7104.

17. Jacquier, A., Michel, F. Base-pairing interactions involving the 5′ and 3′-terminal nucleotides of group II self-splicing introns. *J Mol Biol*. 1990; 213:437–447.

18. Fedorova, O., Pyle, A.M. A conserved element that stabilizes the group II intron active site. *RNA*. 2008; 14:1048–1056.

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*. 2003; 330:197–209.

20. Chanfreau, G., Jacquier, A. An RNA conformational change between the two chemical steps of group II self-splicing. *EMBO J*. 1996; 15:3466–3476.

21. 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*. 2004; 430:45–50.

22. 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*. 2011; 30:3040–3051.

23. Klein, D.J., Moore, P.B., Steitz, T.A. The contribution of metal ions to the structural stability of the large ribosomal subunit. *RNA*. 2004; 10:1366–1379

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*. 2014; 21:464–471.

25. Marcia, M., Pyle, A.M. Visualizing group II intron catalysis through the stages of splicing. *Cell*. 2012; 151:497–507.

26. Query, C.C., Moore, M.J., Sharp, P.A. Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model. *Genes Dev*. 1994; 8:587–597.

27. 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*. 2013; 32:2804–2818.

28. Parker, R., Guthrie, C. A point mutation in the conserved hexanucleotide at a yeast 5′ splice junction uncouples recognition, cleavage, and ligation. *Cell*. 1985; 41:107–118.

29. Lesser, C.F., Guthrie, C. Mutations in U6 snRNA that alter splice site specificity: implications for the active site. *Science*. 1993; 262:1982–1988.

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

31. Yang, J., Zimmerly, S., Perlman, P.S., Lambowitz, A.M. Efficient integration of an intron RNA into double-stranded DNA by reverse splicing. *Nature*. 1996; 381:332–335.

32. Lynch, M., Richardson, A.O. The evolution of spliceosomal introns. *Curr Opin Genet Dev*. 2002; 12:701–710.

33. Tseng, C.K., Cheng, S.C. Both catalytic steps of nuclear pre-mRNA splicing are reversible. *Science*. 2008; 320:1782–1784.

34. Otwinowski, Z., Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods in Enzymology*. 1997; 276:307–326.

35. Schneider, T.R., Sheldrick, G.M. Substructure solution with SHELXD. *Acta Crystallogr D Biol Crystallogr*. 2002; 58:1772–1779.

36. 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, T.C., Zwart, P.H. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr*. 2010; 66:213–221

37. Sheldrick, G.M. A short history of SHELX. *Acta Crystallogr A*. 2008; 64:112–122.

38. Emsley, P., Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr*. 2004; 60:2126–2132.

39. Keating, K.S., Pyle, A.M. RCrane: semi-automated RNA model building. *Acta Crystallogr D Biol Crystallogr*. 2012; 68:985–995.

40. 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*. 2004; 60:2210–2221.

41. Schroder, G.F., Levitt, M., Brunger, A.T. Super-resolution biomolecular crystallography with low resolution data. *Nature*. 2010; 464:1218–1222.

42. Chou, F.C., Sripakdeevong, P., Dibrov, S.M., Hermann, T., Das, R. Correcting pervasive errors in RNA crystallography through enumerative structure prediction. *Nat Methods*. 2013; 10:74–76.

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

44. Chanfreau, G., Jacquier, A. Interaction of intronic boundaries is required for the second splicing step efficiency of a group II intron. *EMBO J.* 1993; 12:5173–5180.

45. 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*. 1993; 361:660–662.

46. 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*. 1998; 5:986–992.

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 3, in full, is a reprint of the material as it appears in Nature, Robart, A.R.; **Chan, R.T.**; Peters, J.K.; Rajashankar, K.R.; Toor, N. (2014) Crystal structure of a eukaryotic group II intron lariat. Nature, 514(7521):193-7. The dissertation author is the secondary author on this paper.

Chapter 4: Dynamic catalytic triplex that modulates RNA splicing

4.1 Abstract

Group II introns are self-splicing catalytic RNAs that are thought to be ancestral to the spliceosome. Both group II introns and the spliceosome share a common active site architecture containing multiple base triples that are essential for catalysis. However, the precise nature of the role and configuration of this catalytic triplex during the progression of splicing was unknown. Here we show that structural rearrangements within the catalytic triplex directly mediate the transition between the pre-second step and post-catalytic stages of RNA splicing. We report the 3.6 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 substrate is aligned in the active site adjacent to the catalytic metal ions and the 3′-OH nucleophile of the 5′ exon. 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 alternative 5′ and 3′ splice site recognition 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. Based on this data, we propose a detailed mechanism for the second step of RNA splicing encompassing a dynamic catalytic triplex that has implications for the evolutionarily related spliceosome.

4.2 Introduction

Group II introns are self-splicing ribozymes that catalyze two transesterification reactions to excise themselves from a pre-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. Introns comprise ~25% of the genomes of mammals[1] and contribute greatly to organismal complexity through the process of alternative splicing which allows increased proteomic diversity[2,3]. There is strong phylogenetic, biochemical, and structural evidence supporting the hypothesis that group II introns are ancestral to the eukaryotic spliceosomal machinery[4]. Recently, a cryo-EM structure of the spliceosome was determined at 3.6 Å 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[5]. In addition, the RT/maturase protein component of group II introns has extensive structural and sequence homology with the core Prp8 protein from the spliceosome[6]. Therefore, the cumulative evidence strongly suggests an evolutionary link between group II introns and the spliceosome.

Group II introns have an RNA secondary structure consisting of six domains[7] with domain V coordinating the two catalytic metal ions, M1 and M2[8], and domain VI containing the bulged adenosine used as the nucleophile in the first step. The active site of group II introns resides within DV that has a secondary structure consisting of a stem loop containing a two-nucleotide bulge and a conserved catalytic triad separated by five base pairs. The crystal structure of a primitive group IIC intron from the bacterium *Oceanobacillus iheyensis* (*O. iheyensis*)[9] first revealed that the CGC catalytic triad (AGC sequence in eukaryotes) forms three consecutive base triples with the two-nucleotide bulge of domain V (DV) and the junction of domains II and III (J2/3). Yeast genetic experiments also confirmed analogous base triple

interactions within the U2/U6 snRNA, which forms the active site of the spliceosome[10]. This 'catalytic triplex' is also present in the cryo-EM structure of the spliceosome. Given the high degree of conservation, it is likely that probing the functional role of the catalytic triplex in group II introns will give insight into the analogous spliceosomal core. Interestingly, the structure of the eukaryotic group IIB intron from *Pylaiella littoralis* (*P.li.LSUI2*)[11] revealed an additional configuration of the catalytic triplex different from that observed in the bacterial group IIC intron. This new configuration has one base triple disengaged and shifted downward within the DV helix. This was unexpected given the high degree of sequence conservation seen in this core region of group II introns from bacteria to higher eukaryotes[12]. Given the fact that there are at least two different configurations of base triples within the catalytic domain V, we hypothesized that these have a functional linkage with distinct stages of splicing. Consistent with this observation, Marcia *et al*.[13] had observed that J2/3 disengaged from the core after the first step in the bacterial IIC intron. However, this construct did not allow the full transition to the second step to be visualized as domain VI and the 3' splice site are missing from this bacterial structure.

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 (Table 4.1). This represents the first structure of a group II intron at an intermediate stage directly preceding the second step of splicing. The intact 3′ splice site is aligned in the active site adjacent to the catalytic metal ions and the 3′-OH nucleophile of the 5′ exon (Figure 4.1). Comparisons with the post-catalytic structure reveal extensive rearrangements in the catalytic triplex. As splicing progresses, the conserved junction sequences between DII and DIII (J2/3), DIV and DV (J4/5), and DV and DVI (J5/6) engage in a "shuffling" action within the catalytic triad of domain V to facilitate the transition towards the subsequent step of splicing. Mutagenesis of these conserved linker sequences results in aberrant splicing, consistent with a role in positioning exon substrates in the core. Furthermore, we propose each stage of splicing is associated with a specific, unique arrangement of the catalytic triplex.

Table 4.1: 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 dimensions		
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)
$I / \sigma I$	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 (Å)	0.002	0.009
Bond angles (°)	0.516	1.080

*Values in parentheses are for highest-resolution shell.

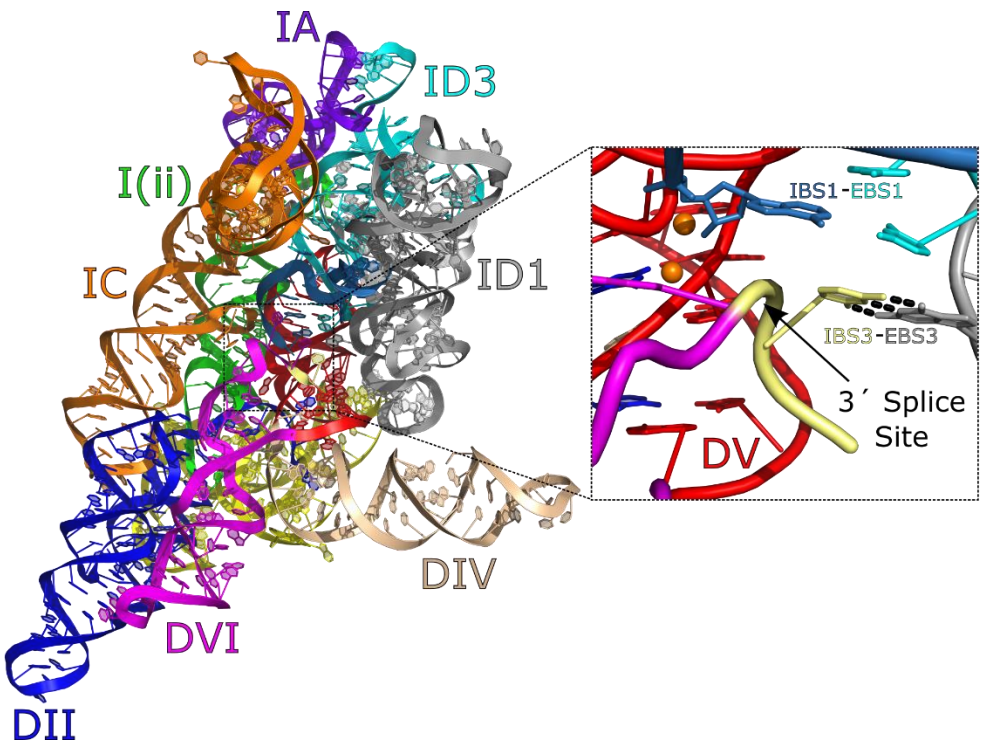

Figure 4.1: **The overall structure of the pre-2s group II intron**. The overall structure is similar to the post-catalytic structure, with the large exception of the intact 3′ splice site at the core of the intron (inset). The 3′ exon is help in place by EBS3.

4.3 Iridium (III) hexamine promotes 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 hexamine prior to data collection[11]. Inspection of the major groove of DVI revealed two iridium hexamine molecules and a magnesium hexahydrate (Figure 4.2) in the vicinity of 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. Therefore, this interaction 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. We initiated the current study based on the hypothesis that the presence of these three outer-shell coordinated metal ions could “lock down” this dynamic π-π′ region into the observed conformation with

biochemical consequences for the splicing reaction. This is supported by splicing reactions of *P.li.LSUI2* in the presence or absence of iridium hexamine in a limiting splicing conditions of $MgCl_2$ and NH_4Cl (Figure 4.3). Conditions lacking iridium hexamine are unable to support splicing (Figure 4.3a). In contrast, supplementation with iridium hexamine at concentrations as low as 50 μM (~80-fold lower than $[Mg^{2+}]$) is able to promote splicing (Figure 4.3b). The low concentration of iridium hexamine required 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 hexamine. 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[14]. This is in contrast to the splicing reaction at 5.4 mM Mg in the absence of iridium hexamine which results in the significant accumulation of lariat-3' exon intermediate. The fact that this intermediate is not visible in the presence of iridium hexamine suggests that this metal complex has a stimulatory effect on the second step to favor the post-catalytic state. Group II intron splicing is known to be reversible[15–17], and it is possible that iridium hexamine locks DVI into a conformation that shifts the reaction equilibrium towards exon ligation. This would prevent the reversal of the splicing reaction and lead to an accumulation of fully spliced lariat RNA in the presence of iridium hexamine and no visible lariat 3'-exon intermediate. Further evidence that iridium hexamine favors the post-catalytic state is provided by an analysis of the 3' end of the RNA present in native crystals and those soaked in iridium hexamine. End mapping was done to determine cleavage at the 3' splice site (Figure 4.4). This confirms the presence of an intact 3′ splice site in the native crystals and spliced, post-catalytic intron in the iridium hexamine derivative.

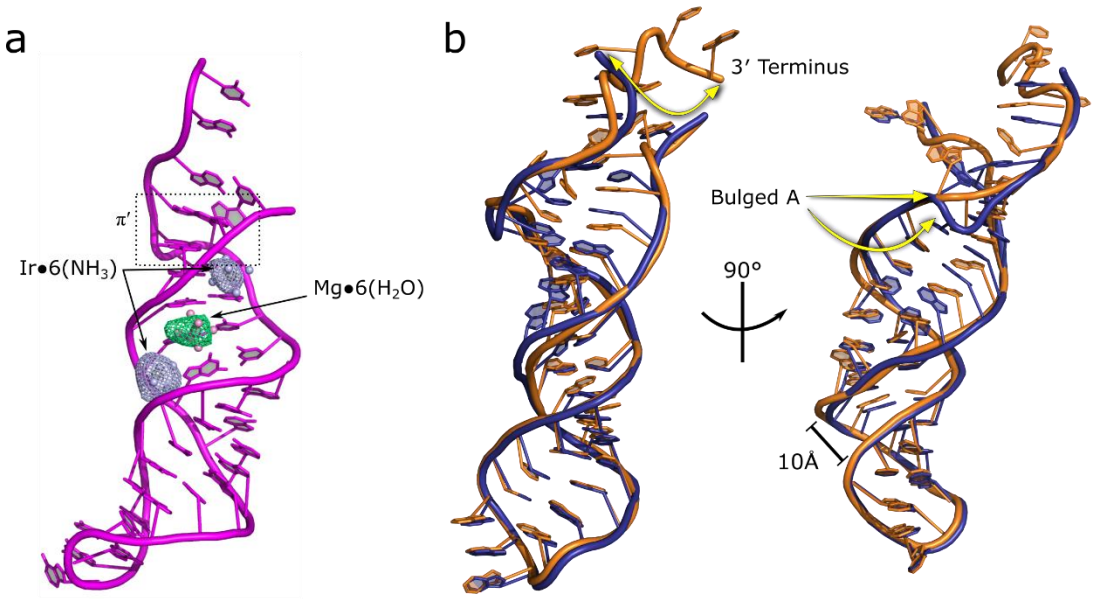

Figure 4.2**: Iridium hexamine bound to the major groove of DVI.** a) Two iridium hexamines and a magnesium hexahydrate are bound to major groove of DVI in the postcatalytic structure. Ir^{+} anomolous maps (light blue) controured to 2.0σ, $Mg{\cdot}6H_2O$ F_o-F_c map (green) contoured to 3.6σ. b) The post-catalytic (blue) and pre-2s (orange) conformation of DVI overlayed. The 3′ terminus of the intron undergoes a rearrengeme between this step. The backbone of the bulged adenosine is highly distroted in the presense of $Ir{\cdot}6NH_4$.

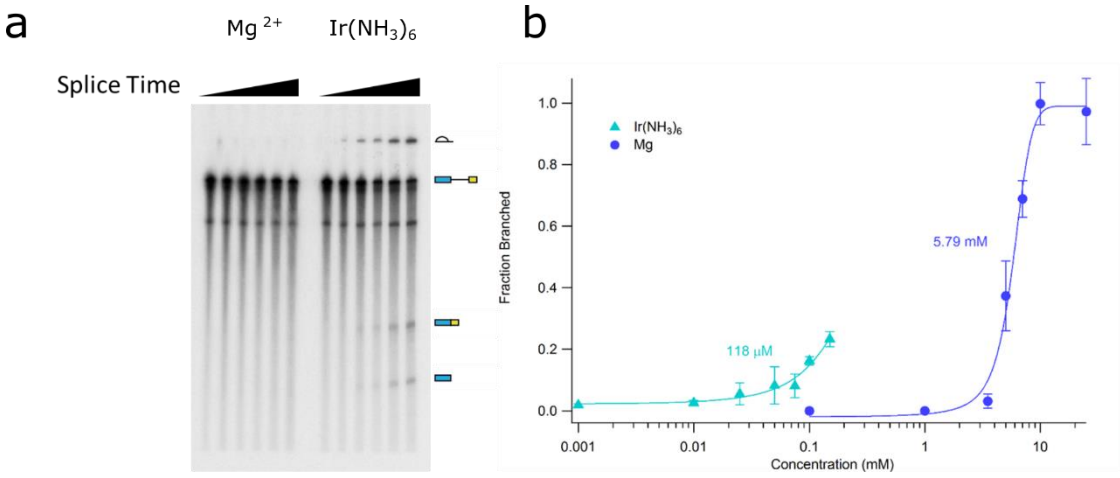

Figure 4.3: **Iridium hexamine shifts equilibrium towards the post-catalytic state.** a) *P.li.LSUI2* spliced in limiting conditions with Mg^{2+} (4 mM $Mgcl_2$, 50 mM NH_4Cl, left) or Ir^{3+} (50 µM $Ir{\cdot}6NH_4$). Kinetic curves describing the effects of Mg^{2+} and Ir^{3+} on splicing.

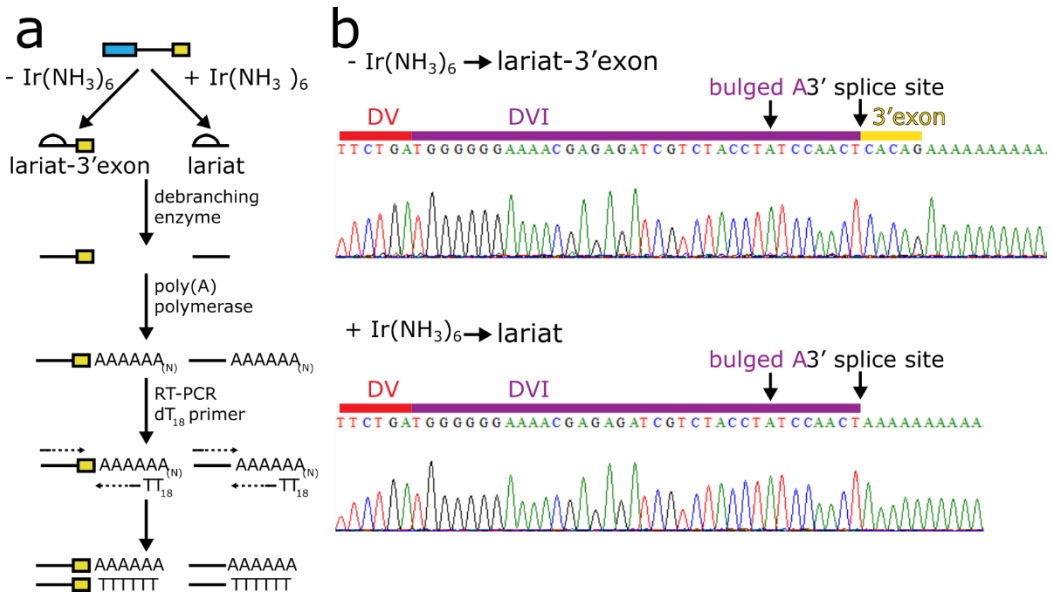

Figure 4.4: **3′ end mapping of *P.li.LSUI2* crystals.** a) Diagram describing the 3′ end mapping reaction. Crystals were harvested and pooled into two fractions for analysis of the effect of Iridium hexamine. b) RT-PCR products were cloned and sequenced. 90% of sequences without Iridium hexamine incubation contained an intact 3′ splice site. 100% of sequences soaked with iridium hexamine were fully spliced.

4.4 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=24.5%) using diffraction data collected from native crystals in the absence of iridium hexamine (Table 4.1). In order to avoid model bias, F_o-F_c omit maps were calculated by deleting nucleotides spanning the 3′ splice site (Figure 4.5) for both the pre-2s and post-catalytic states. This reveals density for an intact 3′ splice site in the native crystals and confirms that this structure represents the pre-2s state. A comparison of both structures reveals significant differences within DV, 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 4.1, Figure 4.5a). The scissile phosphate junction at the 3′ splice site is highly distorted and kinked as predicted by Chan *et al*. (2012)[18]. 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. The 3′ splice site is positioned in the core via domain 6 (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 in this structure is located ~10 Å away from the catalytic metals M1 and M2; therefore, it must undergo a shift towards the active site in order to be cleaved in the second step. This suboptimal positioning of the scissile phosphate accounts for the intact state of the 3′ splice site in the structure.

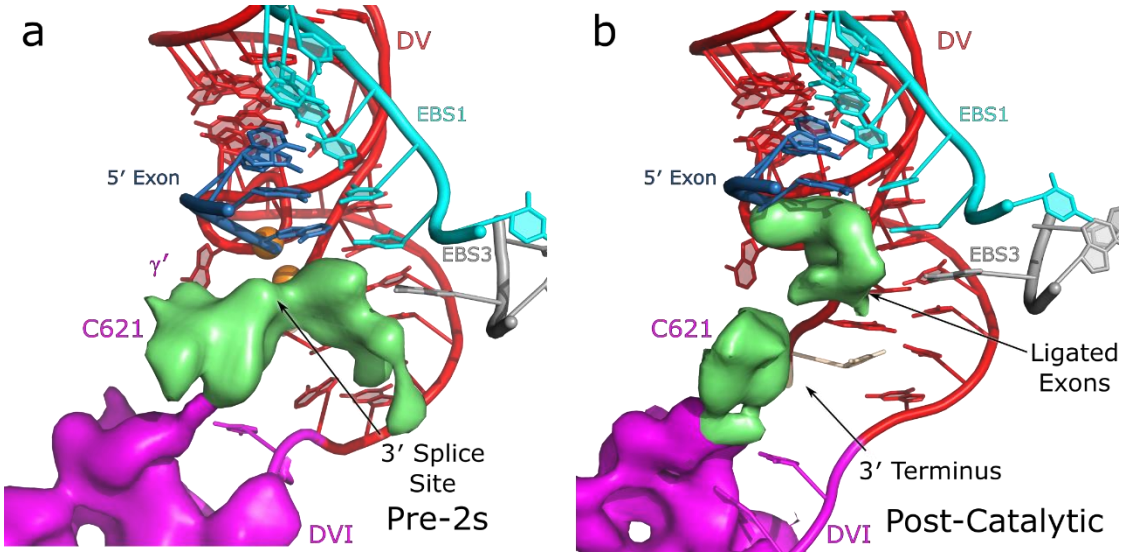

Figure 4.5: **Omit maps of the 3′ terminus of *P.li.LSUI2*.** a) The pre-2s map shows density corresponding to a sharp bend at the 3′ splice site. b) The post-catalytic map shows two discrete regions of density that are non-continuous between the 3′ end of the intron and the ligated exons (the last two nucleotides of the exons were deleted to generate the maps). The F_o-F_c omit map (green) were contoured to 2.7σ. The 2F_o-F_c map (magenta) was contoured to 0.8σ.

This structure also 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. Evidence for this hypothesis comes from the fact that the γ-γ′ interaction is not engaged in the observed state. This interaction consists of a single Watson-Crick pair between the 3′ terminal nucleotide of the intron and a residue from J2/3. The γ-γ′ interaction is critical for 3’ splice site capture for the second step[19–21]. In the lariat 3′-exon structure, 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 (Fig 4.6). There are only two instances of the γ nucleotide engaged in an active conformation, with the J2/3 nucleotides having been altered in each case. In the bacterial pre-catalytic structure[18], a mutation was made to the central catalytic triad position that destabilized the J2/3 catalytic triplexes. While being captured in a pre-catalytic state with the 5′ exon still covalently attached to the intron, this triplex destabilization allowed the γ nucleotide to toggle to in an active conformation. This was further corroborated with various metal soaks perturbing the structure of the bacterial intron[13]. The lack of this interaction accounts for the fact that the 3′ splice site is not optimally positioned for cleavage in this structure and indicates that γ-γ′ is a transient and dynamic contact. We propose that γ is coupled to the J2/3 nucleotide disengaging from the triplex, and acts as a quality control mechanism to ensure that only the correct 3′ splice site can enter into the active site between the first and second steps of splicing.

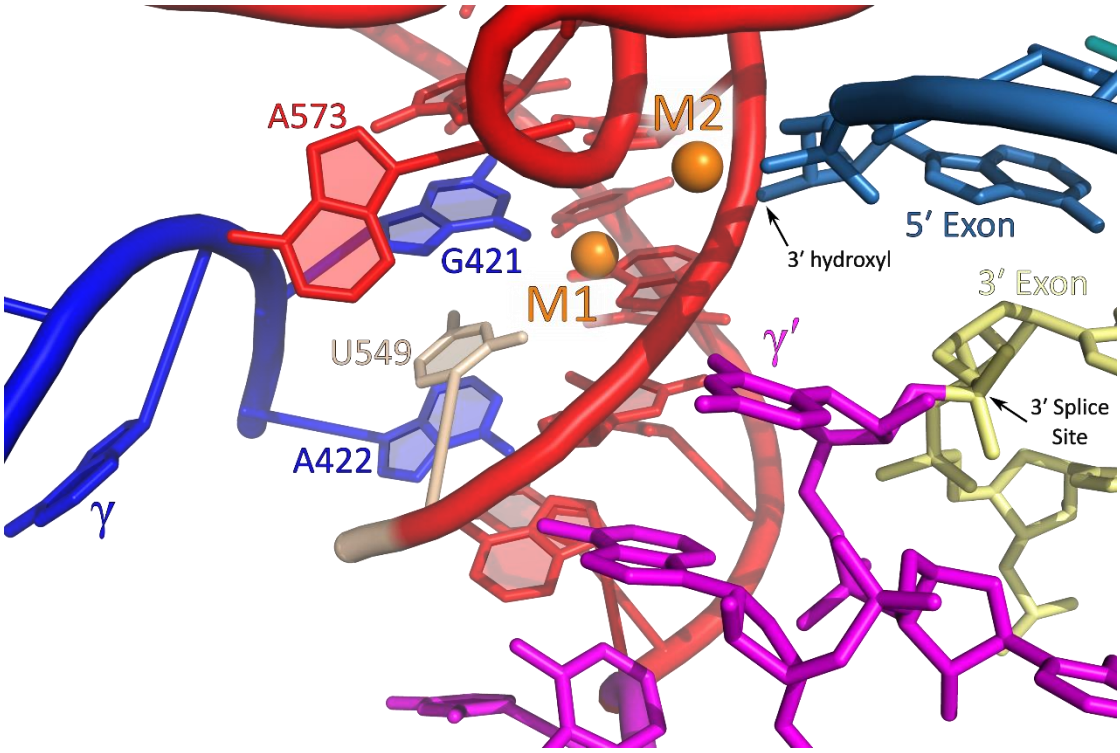

Figure 4.6: **The γ nucleotide in an inactive conformation.** While the γ′ nucleotide is poised to pair, the γ nucleotide is rotated away from the active site and not base-stacking with A573.

4.5 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 J2/3 (Figure 4.7). A comparison of the structures of the post-catalytic IIB with the pre- and post-catalytic IIC intron revealed significant differences in the patterns of base pairing within the catalytic triplex (Figure 4.7b, Figure 4.7c). In the pre- and post-catalytic structures of the bacterial *O. iheyensis* intron (Figure 4.6a), three base triples are observed involving the catalytic triad, the DV two-nucleotide bulge, and J2/3.

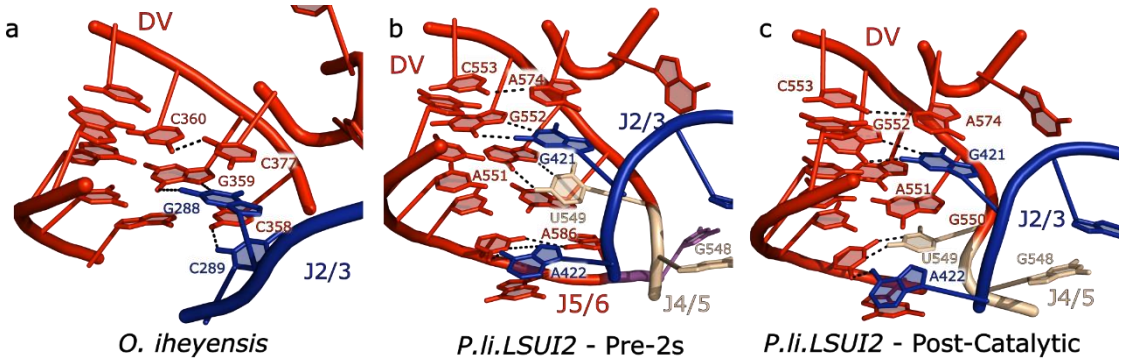

Figure 4.7: **The three distinct catalytic triplexes of group II introns.** a) The catalytic triplex arrentment observed in the bacterial intron *O.i.* contains two J2/3 nucleotides participating in base tripples. It's inability to properly catalyze the second step of splicing indicates that this arrangement corresponds to the active site directly prior to the first step of splicing. The base triples highlighted are between C358- G386-C289, analogous to A551-U582-G421, respectively. b) The pre-2S catalytic triplex of *Pli.LSUI2* displaces a J2/3 residue with a J4/5 residue in one of the catalytic triplex position. This is initiated by the formation of π-π', which moves the basal stem of DVI to promote the rearrentments. c) The post-catalytic triplex conformation of *Pli.LSUI2* has only catalytic triplexs formed, reflecting an inactive state once splicing has completed.

However, in the post-catalytic eukaryotic IIB intron, we saw the first evidence for additional junction nucleotides (other than J2/3) participating in core base triples. Specifically, we observed 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. We hypothesize that this triplex configuration represents the 'off state' following the completion of splicing to generate ligated exons. Surprisingly, the pre-2s structure reveals yet another configuration of the catalytic triplex. 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 of the triad. 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. This inference is made based upon the fact that the bacterial intron is not observed to engage in these alternative configurations since it is a truncated, linear form lacking domain VI and the 3′ splice site that is the second step substrate[13,18]. In addition, it is also missing the J5/6 linker, which participates in the catalytic triplex in the eukaryotic group IIB intron[12]. As a result, the

linkers in the bacterial structure are likely not able to reorient themselves into catalytically relevant configurations as observed in the group IIB intron. Therefore, it is not possible to capture the true nature of the transition to the second step in this bacterial model system. However, there is structural evidence for J2/3 disengaging from the catalytic triad after the first step of hydrolytic splicing, which is consistent with the remodeling of the catalytic triplex observed in this study.

The cumulative data suggests that there are at least three distinct, specific configurations of the catalytic triplex associated with different stages of splicing (Figure 4.6). We propose that these different triplex conformations allow the intron active site to choreograph the correct placement of splice site substrates within the catalytic core to ensure high fidelity of splice site selection.

4.6 Catalytic triplex nucleotides modulate RNA splicing

Previous work by Boulanger *et al*.[22] tested the effects of mutations and changes in length of the J5/6 linker in the related aI5γ group IIB intron. Both *P.li.LSUI2* and aI5γ share a highly conserved secondary structure that is typical of the chloroplast-like class 1 phylogenetic subgrouping. Therefore, this allows biochemical data from aI5γ to be directly correlated to the *P.li.LSUI2* intron. Boulanger *et al*. found that different alterations in the length and sequence of the J5/6 linker resulted in significant effects upon lariat formation, exon ligation, and 3' splice site selection. Shortening of the linker resulted in significant defects in lariat formation.

To test the function of the analogous junction nucleotides in the *P.li.LSUI2* intron, we altered the sequence and length of the J4/5 and J5/6 linkers. All of the mutants resulted in aberrant splicing with inhibition of either the first or second step of splicing (dependent upon the specific mutation) (Figure 4.8a, 4.8b). Shortening of the J5/6 linker (J5/6(1)) results in complete inhibition of lariat formation. Instead, this mutant engages in hydrolytic cleavage at a cryptic site within the 5′ exon. Therefore, this has a large negative effect upon the first step of splicing with the intron attempting to use an alternative 5′ splice site. Lengthening of J5/6 to

four or five residues results in a significant accumulation of lariat-3′ exon, which indicates a second step splicing defect. These mutants also engage in alternative 3′ splice site selection up to 16 nucleotides downstream of the wild-type 3′ splice site. Single-nucleotide substitutions within J5/6 also result in a second step splicing defect. The cumulative data are consistent with a model in which conformational changes in the catalytic triplex are involved in a complex interplay of substrate positioning events in the active site that are essential for proper splice site selection. It is also possible that there exist additional configurations of the catalytic triplex that may be more transient in nature and not yet visualized.

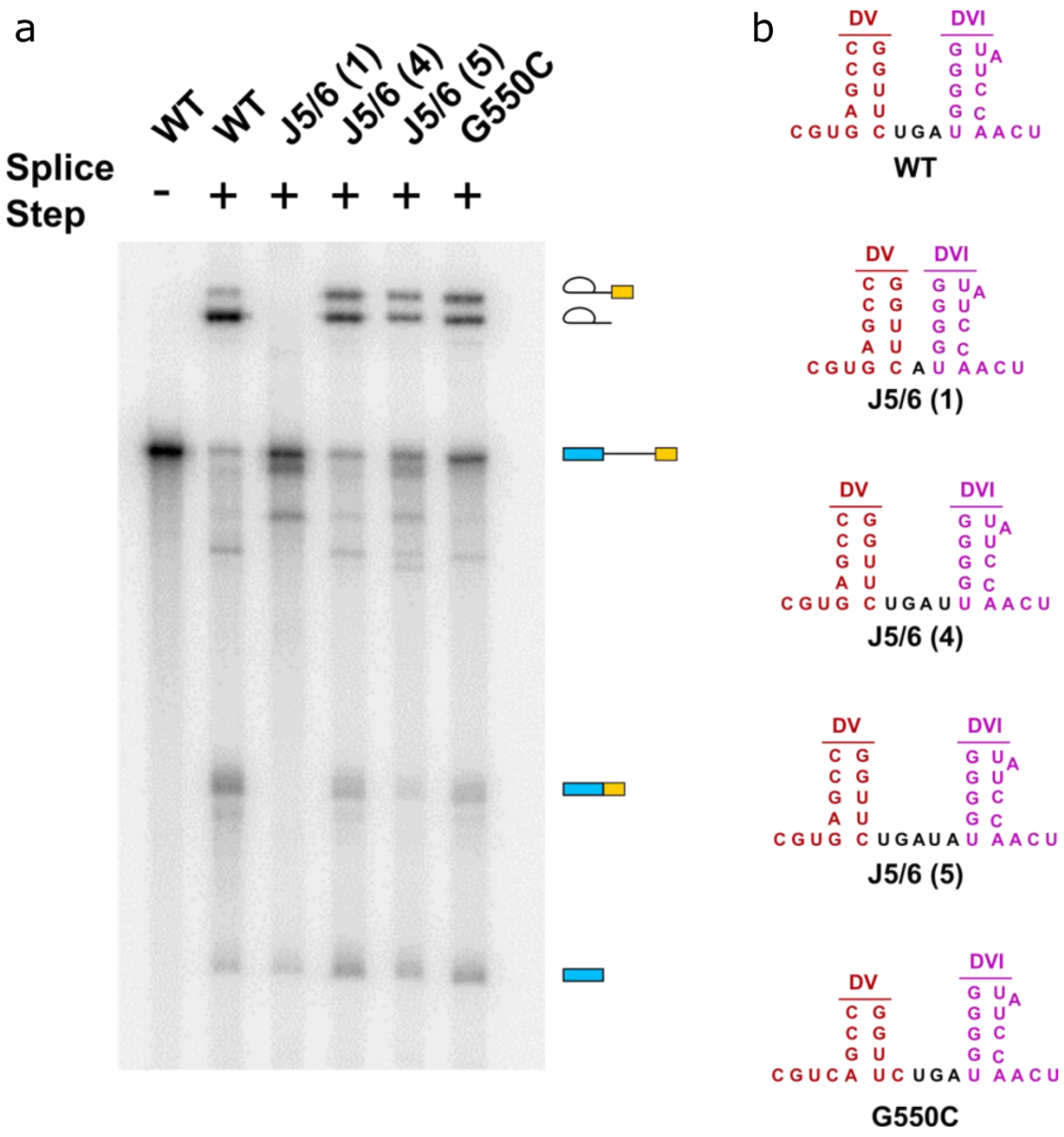

Figure 4.8**: Junction mutations affect splicing equilibrium.** a) In vitro splicing assays of *P.li.LSUI2* after 30 minutes. WT has reached equilibrium, with the final lariat product as the predominant band. Mutating the length of the J5/6 linker to four, or five nucleotides results in an accumulation of pre-2s intron. b) Secondary structures depicting the mutations done. G550C changes the length of J5/6 to four nucleotides by unpairing the first base pair of DV.

These biochemical results also may explain the aberrant splicing observed in vitro for group IIC introns, which have been shown to undergo alternative splicing and/or promiscuous intramolecular hydrolytic cleavage(s)[23]. This may be correlated with the fact that these introns undergo splicing through hydrolysis, which implies that DVI and J5/6 are not properly positioned. As a result, both splice site fidelity and lariat formation is compromised in a manner similar to that seen in the J5/6 mutants of the eukaryotic group IIB intron. However, it is known that addition of the corresponding maturase rescues lariat formation, likely through stabilization of DVI and the junction linkers[24].

4.7 Model for the second step of RNA splicing

Given the pre-2s and post-catalytic structures of the eukaryotic group IIB intron, we are now able to construct a detailed model for the second step of splicing. Following 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 A422 disengaging from A551 (Figure 4.9). Displacement of the lariat bond from the core allows the 3′ terminus to enter the active site and remodels J5/6. 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.

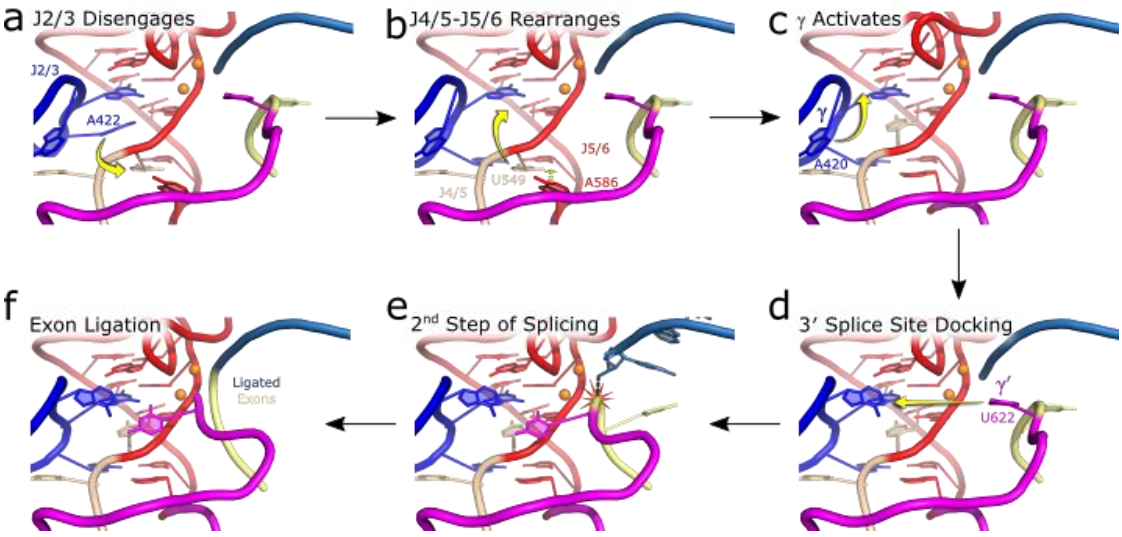

Figure 4.9**: Model of the complete pathway for the second step of splicing.** a) A422 disengages from the catalytic triad, following the first step of splicing. b) The formation of π-π′ causes the basal stem of DVI to shift, prompting A586 of J5/6 to remodel and activate U549 of J4/5 to engage in the catalytic triplex. c) A420 (γ) rotates ~180° to an active conformation to pair with U622 (γ′). d) U622 (γ′) pairs with A420, which moves the 3′ splice site into the active site. e) The second step of splicing takes place, with the 3′ OH of the 5′ exon activated for nucleophilic attack on the 3′ splice site. f) The exons are ligated together, and the active site moves to an inactive state, with only 2 of the three catalytic triplexes engaged.

4.8 Discussion

There is now extensive evidence that the spliceosome is also a ribozyme and has an active site configuration similar to that of the group II intron[5,10]. Based on the conservation of junction nucleotides between both systems, this line of thought can be extended there as well (Figure 4.10). Here, we describe core rearrangements of the catalytic triplex in a eukaryotic group II intron that are associated with the progression of the different stages of splicing. We further show that this triplex is essential for the efficiency and fidelity of splice site selection with mutagenesis of the junction nucleotides having profound effects on catalysis. Taken together, the data suggests a model in which dynamic rearrangements of the catalytic triplex are responsible for the positioning and removal of exon substrates from the active site for the two steps of splicing. This model is also consistent with biochemical effects observed for these linker sequences in the literature.

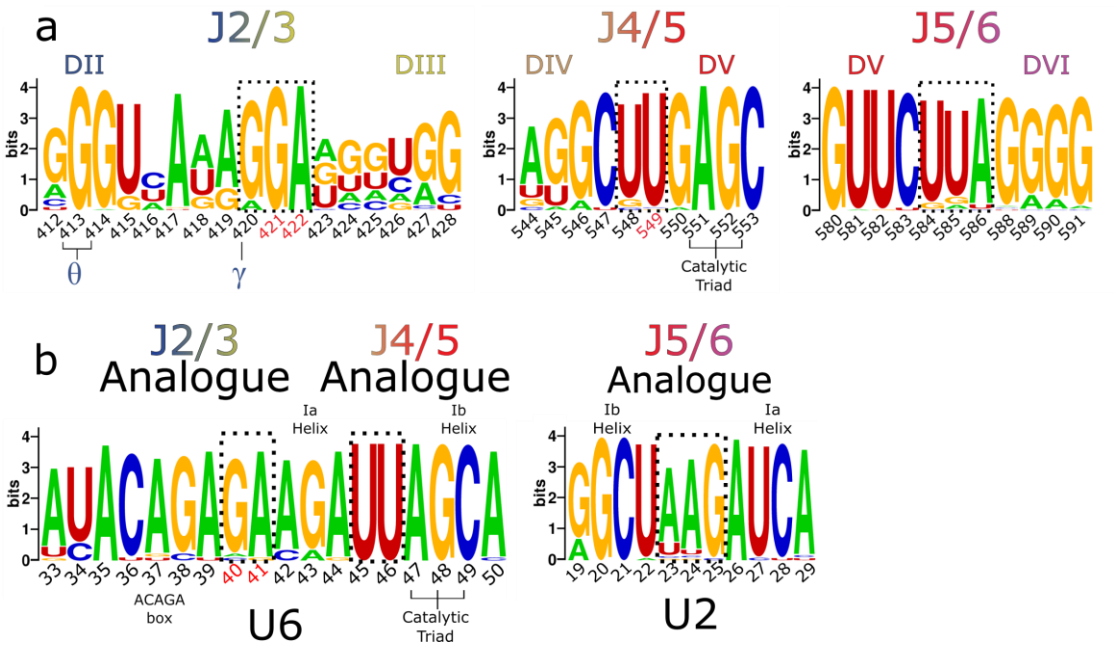

Figure 4.10: **Sequence conservation of junction nucleotides.** a) The junction nucleotides of the group II intron are highly conserved, and correlate to their observed function in the pre-2s structure. b) The Spliceosome junction nucleotides identified by either their proximity to the catalytic triad (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*.

Given that there is evidence for a group II intron-like catalytic triplex in the spliceosome, we propose that it also utilizes similar core RNA rearrangements to properly select the correct splice sites to facilitate catalysis. This is supported by an analysis of the cryo-EM structure of the spliceosome, which reveals the existence of sequences in the U2/U6 snRNA catalytic core that are analogous to the J2/3, J4/5, and J5/6 linkers from the eukaryotic group IIB intron. A prominent feature of this structure is the location of the spliceosome protein prp45, which has an extended conformation and encircles a significant portion of the spliceosome (Figure 4.11). Interestingly, prp45 interacts with the phosphate backbone of the U6 snRNA and is in close proximity to the catalytic triplex. We hypothesize that this region of prp45 functions as the 'triplex sensor' domain that can sense the conformation of the triplex configuration in the active site and relay this information to other regions of the spliceosome to

regulate splicing. This would allow conformational rearrangements in the spliceosome to be synchronized with either the first or second step of splicing.

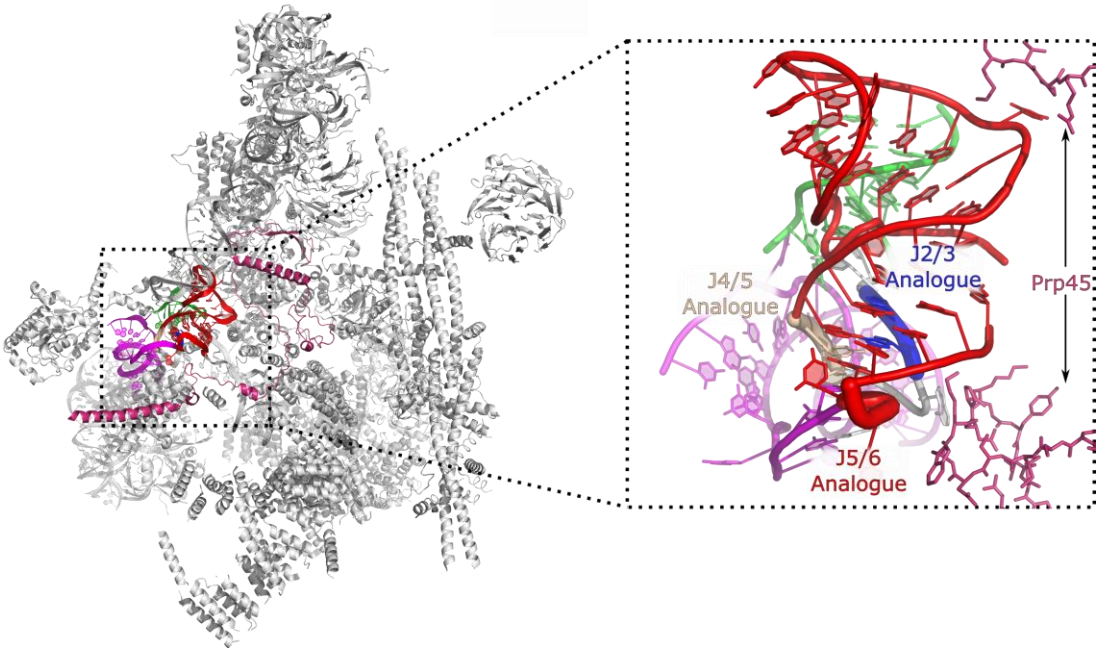

Figure 4.11: **The core architecture of the spliceosome.** The spliceosome structure contains analogous junction nucleotides to the group II intron (highlighted with thicker backbone representations). Prp45 (magenta) spans a large distance into the overall structure and makes contacts with the analogous J5/6 region. PDB ID 3JB9 used in this figure. Chains A,B,K,h (spp42, cwf10, prp5, cwf15, respectively) not shown in the overall figure so RNA core can be visualized.

4.9 Methods

Crystallization of *Pylaiella littoralis* LSU I2 intron RNA

The crystallization construct of *P.li.LSUI2* RNA contained the identical construct prepared previously[11]. Crystallization utilized to same conditions to grow the rod-like crystals that grew within two days 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), 0.5 mM spermine, and 100 mM NaCl; followed by flash freezing in liquid nitrogen. It is important to note that no iridium hexamine was present at any step of crystallization.

Structure determination

The crystal structure of the pre-2s *P.li.LSUI2* was solved at 3.7Å. 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[25]. The post-catalytic structure of *P.li.LSUI2* (PDBID 4R0D) was used as a search model for molecular replacement using PHENIX[26]. 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[27] and the RCrane plugin[28]. Structure refinement was done using Buster[29], Phenix[26], DEN[30], and Phenix.Erasser[31]. SBGrid compiled all the software used[32]. Figures were prepared using Pymol[33], and sequence logos using Weblogos[34].

In vitro self-splicing assays

All *in vitro* splicing assays were as described in Robart *et al* (2014). Kinetic curves were analyzed and drawn using Igor Pro[35].

References

1. Sakharkar, M. K., Chow, V. T. K. & Kangueane, P. Distributions of exons and introns in the human genome. *In Silico Biol.* 4, 387–93 (2004).
2. Pan, Q., Shai, O., Lee, L. J., Frey, B. J. & Blencowe, B. J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* 40, 1413–5 (2008).
3. Black, D. L. Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.* 72, 291–336 (2003).
4. Sharp, P. A. Five easy pieces. *Science* 254, 663 (1991).
5. Hang, Jing; Wan, Ruixue; Yan, Chuangye; Shi, Y. Structural basis of pre-mRNA splicing. *Science (80-.).* (2015).
6. 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).
7. Michel, F., Umesono, K. & Ozeki, H. Comparative and functional anatomy of group II catalytic introns--a review. *Gene* 82, 5–30 (1989).
8. 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).

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

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

14. 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).

15. 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, 473–82 (2009).

16. Mörl, M., Niemer, I. & Schmelzer, C. New reactions catalyzed by a group II intron ribozyme with RNA and DNA substrates. *Cell* 70, 803–10 (1992).

17. Mörl, M. & Schmelzer, C. Integration of group II intron bI1 into a foreign RNA by reversal of the self-splicing reaction in vitro. *Cell* 60, 629–36 (1990).

18. 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, 555–557 (2012).

19. Costa, M., Michel, F. & Westhof, E. A three-dimensional perspective on exon binding by a group II self-splicing intron. *EMBO J.* 19, 5007–18 (2000).

20. 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).

21. Jacquier, A. & Michel, F. Base-pairing interactions involving the 5′ and 3′-terminal nucleotides of group II self-splicing introns. *J. Mol. Biol.* 213, 437–447 (1990).

22. Boulanger, S. C., Faix, P.H., Yang, H., Zhuo, J., Franzen, J.S., Peebles, C.L., Perlman, P.S. Length changes in the joining segment between domains 5 and 6 of a group II intron inhibit self-splicing and alter 3’ splice site selection. *Mol. Cell. Biol.* 16, 5896–904 (1996).

23. 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).

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

25. Otwinowski, Z., Minor, W. & Jr, C. C. W. Processing of X-ray Diffraction Data Collected in Oscillation Mode, Methods in Enzymology Vol. 276 Part A. 276, 307–326 (1997).

26. 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, T.C., Zwart, P.H. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr.* 66, 213–21 (2010).

27. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D. Biol. Crystallogr.* 60, 2126–32 (2004).

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

29. 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).

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

31. 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).

32. 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).

33. The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.

34. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: A Sequence Logo Generator.

35. Gomez, J. F., Brioso, M. A., Machado, J. D., Sanchez, J. L. & Borges, R. New approaches for analysis of amperometrical recordings. *Ann. N. Y. Acad. Sci.* 971, 647–54 (2002).

Acknowledgements

Chapter 4, in part, is currently being prepared for submission for publication of the material. **Chan R.T.** and Robart A.R.; Peters, J.K.; Rajashankar, K.R.; Toor, N. The dissertation author is a co-primary author on this paper.

Chapter 5: The GUAAY pentaloop: a novel motif in the group IIB1 intron structure

5.1 Abstract

Large RNAs often utilize tetraloops as structural elements to stabilize the overall tertiary fold. In contrast, only one pentaloop has been implicated to participate in an RNA-RNA interaction as revealed through nucleotide analogue interference mapping (NAIM). Here we show that the conserved GUAAY pentaloop found in domain II of group IIB1 introns participates in a novel class of RNA tertiary interaction. This pentaloop is highly conserved within the IIB1 class and interacts with the minor groove of the catalytic DV. The loop architecture and insertion orientation is distinctive from the typical GNRA tetraloop-receptor interaction, with in vitro splicing data indicating that a tetraloop is incompatible at this position. We therefore propose, based on phylogenetic, structural, and biochemical data, that the GUAAY pentaloop represents a novel RNA structural motif.

5.2 Introduction

One of the most ubiquitous RNA structures is the tetraloop [1,2], a four nucleotide sequence that caps an A-form double helix. While numerous tetraloop motifs have been identified, three are the most prevalent: CUUG[3], UNCG[4], and GNRA[5], with the latter most demonstrating a robust ability to form RNA tertiary structures. GNRA tetraloops comprise a large majority of tetraloop-tetraloop-receptor interactions, where the loop nucleotides interact with the minor groove of an RNA helix. Additionally, different GNRA-loop sequences have varying affinities for receptor sequences, allowing this tertiary structure motif to be sequence specific[6–8].

While pentaloops have been demonstrated to have a wide range of functions[9–11], only one has been postulated to function as an RNA tertiary element[12]. The GUAAY pentaloop is found within group IIB1 introns and participates in the μ-μ′ interaction. Nucleotide Analogue

Interference Mapping (NAIM), it was identified that the μ-μ′ interaction involves the pentaloop of the DIIIa helix interacting directly with the catalytic DV (Figure 1a). This μ-μ′ interaction occurs directly adjacent to the κ-κ′ interaction in the group IIB1 intron crystal structure[13], and results in a quintuple adenosine base stack inserting into the minor groove (Figure 5.1b. and Figure 5.1c). Additionally, the ζ-ζ′ is a canonical GNRA tetraloop-receptor interaction that induces a 70° bend to establish the active site. Here we present phylogenetic, structural, and biochemical evidence that distinguishes the GUAAY pentaloop from the ubiquitous GNRA tetraloop and reveals a new class of RNA tertiary interactions.

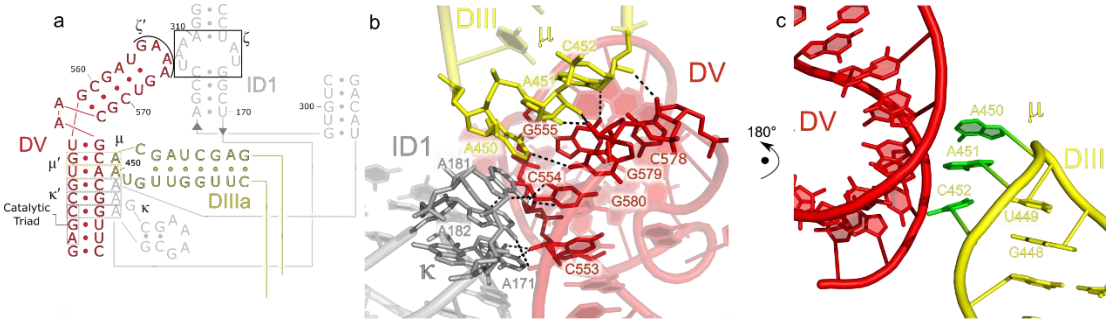

Figure 5.1: **DV acts as a receptor to two tertiary interactions.** . a) Secondary structure show the three tertiary interactions involving DV. μ and κ insert five adenosines into the minor groove or binding face of DV. b) The tertiary structure of the minor groove of DV. The five adenosines base stack with one another, perpendicular to the direction of the minor groove. c) The μ-μ′ interaction positions three nucleotides into the minor groove of DV. The first two nucleotides initiate the loop, while the final three nucleotides (green) contact the receptor.

5.3 Sequence Conservation

Alignment of group II introns within the same phylogenetic class[14] as *P.li.LSUI2* reveal that the first, second, third, and fourth nucleotide of the pentaloop sequence is highly conserved as GUAA (Table 1). At the fifth position, a pyrimidine residue is always observed, with a U and C representing 71% and 19% of the sequences, respectively, giving rise to the GUAAY consensus sequence. Additionally, the loop is closed off by a conserved U-G wobble pair (85%), but there are no observed occurrences of the inverse G-U wobble pair. This differs from GNRA tetraloops, which tend to have a preference for C-G base pairs at this position[15,16]. In addition to the sequence conservation, the loop is attached to the DIIIa helix which has a highly conserved length of seven base-pairs (Figure 5.1a).

Table 5.1**. Sequence Alignment of the GUAAY pentaloop in group IIB1 introns.** *P.li.I2* (*P.li.LSUI2*) sequence is bolded. Sequence positions that are 100% conserved are denoted with a black background, while positions that are >75% conserved are denoted with a light grey background. Positions that were >95% pyrimidine and purine are denoted in light blue and light green, respectively.

Intron name	N	Y	U	G	U	A	A	Y	G	R	N
P.li.I2	**G**	**U**	**U**	**G**	**U**	**A**	**A**	**C**	**G**	**A**	**U**
Sh.fr.I1	A	U	U	G	U	A	A	C	G	A	U
Vi.an.I1	A	U	U	G	U	A	A	C	G	A	U
Vi.ha.I1	A	U	U	G	U	A	A	C	G	A	U
Wo.sp.I1	U	U	U	G	U	A	A	C	G	A	A
Wo.sp.I3	U	U	U	G	U	A	A	C	G	A	A
A.v.I3	C	C	U	G	U	A	A	C	G	G	G
A.v.I1	C	C	U	G	U	A	A	C	G	G	G
Th.e.I1	U	C	U	G	U	A	A	C	G	G	A
Al.sh.I1	C	C	U	G	U	A	A	C	G	G	G
Cu.me.I1	G	C	U	G	U	A	A	U	G	G	U
Cu.me.I2	G	C	U	G	U	A	A	U	G	G	U
Ag.r.I1	G	C	U	G	U	A	A	U	G	G	U
Al.bo.I1	C	C	U	G	U	A	A	U	G	G	G
Al.sh.I2	C	C	U	G	U	A	A	U	G	G	G
B.t.I1	U	C	U	G	U	A	A	U	G	G	G
B.vu.I1	C	C	U	G	U	A	A	U	G	G	G
B.vu.I2	U	C	U	G	U	A	A	U	G	G	G
Ci.ro.I1	G	C	U	G	U	A	A	U	G	G	U
E.c.I5	U	C	U	G	U	A	A	U	G	G	A
P.st.I1	C	C	U	G	U	A	A	U	G	G	G
H.s.I1	A	C	U	G	U	A	A	U	G	G	U
Hm.ar.I1	G	C	U	G	U	A	A	U	G	G	U
Kl.pn.I2	U	C	U	G	U	A	A	U	G	G	A
Le.pn.I1	U	C	U	G	U	A	A	U	G	G	A
Le.pn.I2	G	C	U	G	U	A	A	U	G	G	C
P.st.I2	C	C	U	G	U	A	A	U	G	G	G
P.p.I2	C	C	U	G	U	A	A	U	G	G	G
P.p.I4	G	C	U	G	U	A	A	U	G	G	C
Po.sp.I3	U	C	U	G	U	A	A	U	G	G	A
Po.sp.I4	U	C	U	G	U	A	A	U	G	G	A
Po.sp.I1	U	C	U	G	U	A	A	U	G	G	A
Po.sp.I2	U	C	U	G	U	A	A	U	G	G	A
Re.sp.I1	G	C	U	G	U	A	A	U	G	G	C
Sh.sp.I2	U	C	U	G	U	A	A	U	G	G	A

Table 5.1**. Sequence Alignment of the GUAAY pentaloop in group IIB1 introns, cont'd.**

Intron name	N	Y	U	G	U	A	A	Y	G	R	N
Th.e.I7	U	C	U	G	U	A	A	U	G	G	A
X.f.I1	C	C	U	G	U	A	A	U	G	G	G
Th.e.I3	U	C	U	G	U	A	A	U	G	G	A
Eu.re.I1	C	U	U	G	U	A	A	U	G	A	G
An.pr.I1	C	U	U	G	U	A	A	U	G	A	G
Eu.si.I2	C	U	U	G	U	A	A	U	G	A	G
Fa.pr.I1	C	U	U	G	U	A	A	U	G	A	G
S.ce.I5	A	U	U	G	U	A	A	U	G	A	U
Sh.ba.I1	A	U	U	G	U	A	A	U	G	A	U
Sh.pi.I1	G	U	U	G	U	A	A	U	G	A	U
Sb.mo.I1	U	U	U	G	U	A	A	U	G	A	A
Sh.sp.I1	A	U	U	G	U	A	A	U	G	A	U
Vi.ha.I2	A	U	U	G	U	A	A	U	G	A	U
Sh.se.I1	A	U	U	G	U	A	A	U	G	A	U
E.c.I8	G	C	U	G	U	A	A	U	A	A	C
T.e.I8	U	G	U	G	U	A	A	U	A	C	A
R.pi.I1	G	C	U	G	U	A	A	U	C	G	U
P.ae.I2	G	C	U	G	U	A	A	U	C	G	U
Ch.lu.I1	C	C	U	G	U	A	A	U	C	G	G
Ch.ph.I1	C	C	U	G	U	A	A	U	C	G	G
Pr.ae.I4	C	C	G	G	U	A	A	U	G	G	G
S.eq.I1	G	C	A	G	U	A	A	U	G	G	C
Hp.au.I1	U	C	C	G	U	A	A	U	C	G	A

5.3 Pentaloop Structure

The first nucleotide of the pentaloop (G448) stacks directly on top of the capping U-G wobble pair (Figure 5.2a). Interestingly, instead of stacking on U447, which would be expected in an A-form helix, it stacks almost entirely on G553, which draws the backbone of the loop in the direction of the minor groove of DV. This may explain the preference for a U-G over G-U wobble pair, as the later would orient the loop further from its receptor interaction. The second pentaloop nucleotide (U449) continues the base stack above the first nucleotide. The Watson-Crick face of these two nucleotides are oriented in a similar direction, which keeps the phosphate backbone in an A-form helix-like conformation. In addition, the Watson-Crick face

of U449 is positioned towards the backbone of C452. In this orientation, N3 of U449 forms a hydrogen bond with a non-bridging oxygen. A cytosine substitution would abolish this hydrogen bond, while a purine substitution would result in either a steric clash or move the backbone into a sub-optimal loop orientation. Altogether, the configuration of the second nucleotide in context within the overall loop structure provides a rationale for its conservation. The third loop nucleotide (A450) initiates the uridine-like turn[17] and causes its Watson-Crick face to deviate from an A-form helix rotation. Instead, the orientation of the glycosidic bond is rotated ~180° which allows the insertion of its nucleobase into the minor groove of DV. This contrasts with the GNRA tetraloop, where a uridine-like turn conformation is initiated between nucleotides one and two (**G|N**RA)[18]. The fourth pentaloop nucleotide (A451) stacks beneath the previous nucleotide and also inserts its nucleobase into the minor groove of DV. Finally, the last nucleotide of the pentaloop (C542) finishes the base stack below A451. It makes a single hydrogen bond between O2 and N2 of G555 from the receptor. Because O2 is present in either uracil or cytosine, this provides a rationale for the conservation of a pyrimidine at this position.

A structural motif search was done to determine if similar loop structures existed in the PDB, but were unannotated. Using the AMIGOS II program[19], we initiated an (η,θ) search, with the pentaloop values corresponding to (182.1, 218.3), (175.2, 256.4), (18.7, 231.2), (162.9, 223.6), and (167.8, 273.0). Based on this search, which included the recent spliceosome and U4/U6.U5 tri-snRP structures, no match was identified, indicating that this loop is a novel RNA structure.

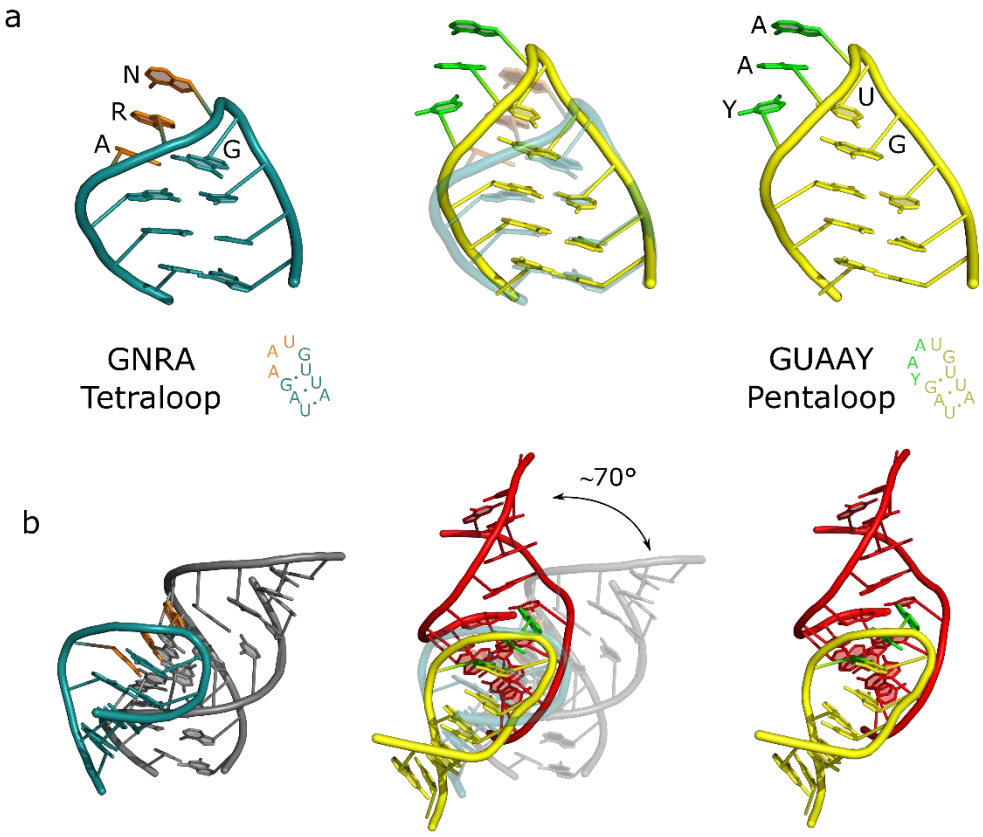

Figure 5.2: **Contrasting strucure of the GNRA tetraloop vs GUAAY Pentaloop.** The GNRA tetraloop (left) and the GUAAY pentaloop (right) are connected to a three base-pair helix and aligned to the bottom of the stem. This results in the GUAAY pentaloop extending approximately one base pair higher and at a different orientation. Nucleotides of the G**NRA** tetraloop and GU**AAY** pentaloop that interact with their cognate receptor are colored in orange and green, respectively b) The GNRA tetraloop (left) nucelotides insert parallel to the base planes of the receptor, while the GUAAY pentaloop nucleotides are nearly perpendicular to the receptor base planes. This would foreseeably allow more nucleotides to insert into the minor groove.

There is some commonality between the GNRA tetraloop and the GUAAY pentaloop. Both loop structures allow the last three nucleobases to splay out in a manner that allows them to interact with a receptor (Figure 5.2b). However, the base planes of the loops represent a significant difference between the two. In the **G**NR**A** tetraloop, the first and last nucleotide are on the same plane and form a non-canonical base pair between the guanosine sugar edge and the Hoogsteen face of the adenosine. In contrast, the guanosine of the pentaloop is the only base on its base plane. Rather, the second nucleotide (U449) is located

on the same base plane as the final nucleotide (G**U**AA**Y**). Furthermore, the second and last nucleobase are not interacting with each other. Instead, the second nucleobase interacts with the phosphate backbone and Watson-Crick face of the last nucleotide is oriented in the complete opposite direction. Because the first and second nucleotide stack with one another, this allows the pentaloop to extend one base plane higher, with respect to the GNRA-tetraloop (Figure 5.2a).

The last three nucleotides of the G**NRA** and GU**AAY** loop are positioned so that their nucleobases are ordered and oriented outward from the helix. This allows the nucleobases to insert into their respective receptors, but the orientation of insertion differs greatly. The inserting nucleotides of the GNRA tetraloop are approximately planar with respect to the receptor base planes (Figure 5.2b). In contrast, the GUAAY nucleotides insert into the minor groove of DV ~70° relative to the receptor nucleotides. While this orientation precludes the formation of canonical A-minor motifs, it allows the possibility of more nucleotides to stack above and below without clashing with the phosphate backbone.

5.4 *In vitro* splicing assays

The initial discovery of μ-μ′ indicated a strict requirement for the loop to be a pentaloop[12]. Due to the lack of structural information at the time, only two tetraloop substitutions were conducted at the time: mutating the entire loop to UUCG (Δμ), and deletion of the last loop nucleotide to a GUAA sequence. With the information obtained from the crystal structure of the group IIB1 intron[13], the later loop mutation actually results in a double mutation in regards to the sequence register of the loop structure. Altering the loop from five to four nucleotides results in a substitution of a uracil in a position that has an adenosine inserting into the receptor. In the same study, a point mutation at that position from an adenosine to uracil resulted in defects in splicing activity. Utilizing the structural data as a guide, we set out to fully characterize the sequence and spatial requirements that would promote the formation of μ-μ′ (Figure 5.3). While the *in vitro* splicing assay is an indirect measure for the formation of

μ-μ′, the only known function of this tertiary interaction is to stabilize the active conformation of the active site. We are therefore using splicing as an indication of whether the μ-μ′ interaction has formed or not.

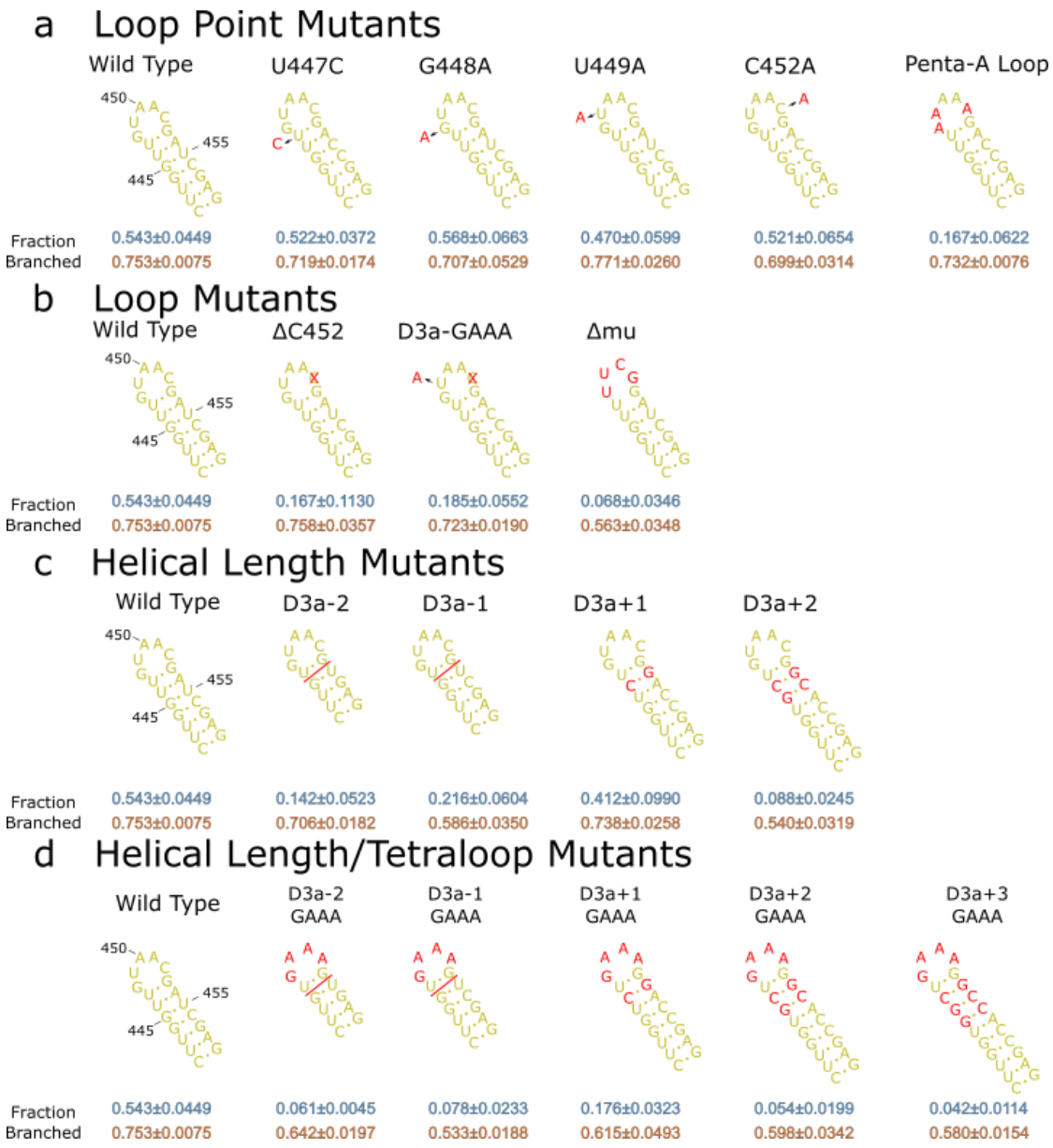

Figure 5.3: **Secondary structure of *P.li.LSUI2* D3a mutants.** Mutants were grouped into four categories: a) Loop point mutants, b) Loop mutants, c) helical length mutants, and d) helical length/tetraloop mutants. Red letters indicate changes to the wild type sequence. Red lines indicate where helical deletions took place. Values below the mutants correspond to Figure 5.4 for in vitro splicing assays at 2.5mM (blue) and 10mM (orange) $MgCl_2$.

Magnesium is an essential cation for ribozymes, both as a structural element[20–22] to stabilize the negative phosphate backbone as well as for catalytic function[23,24]. In the case of the group II intron, two catalytic metals are bound in the active site. One serves to activate the nucleophile for attack at the scissile phosphate, while the other stabilizes the transition state of the leaving group. In standard splicing conditions for *P.li.LSUI2*[25], 10 mM $MgCl_2$ is used to stimulate splicing *in vitro*. However, at this concentration we observed no significant difference in splicing activity for any of the D3a mutants (Figure 5.4). This led us to believe that this magnesium concentration is capable of overcoming structural defects caused by mutagenesis, as seen in Figure 5.4 (orange bars). This is especially apparent in the Δμ mutant, which should be incapable of forming the loop-receptor interaction (Figure 4b). To address this, $MgCl_2$ concentrations were lowered to 2.5mM. However, it is still useful to use the 10 mM $MgCl_2$ conditions in order to validate the retention of activity in the mutants assayed in this study.

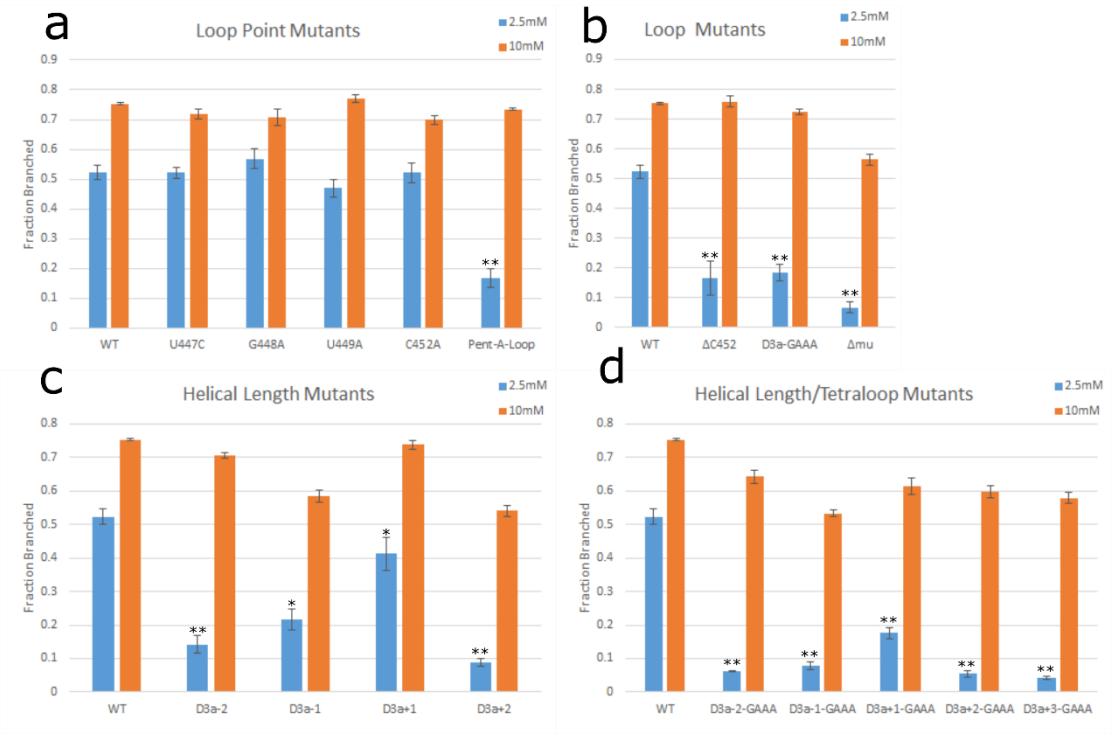

Figure 5.4: **In vitro splicing assay of *P.li.LSUI2* D3a mutants.** a) Point mutants near the pentaloop to assess their individual contribution. The single point mutants retained activity comprable to WT. The exception is mutating the loop entirely to adenosines, resulting a drastic reduction in activity. b) Mutating the loop to any of the tetraloop sequences results in severe defects in splicing activity. c) Changing the helical length from the native seven base-pairs perturbs activity. d) The defects caused by the GAAA-tetraloop cannot be rescued with different helical length mutations. 30 minute time points taken for all splice experiments. * <0.05 by two-tailed Student's t-test. ** < 0.001 by two-tailed Student's t-test.

The high level of conservation of the GUAAY pentaloop is remarkable for group IIB1 introns, with the first four positions being completely conserved across all the sequences. Despite this, mutants with single base substitutions seemed to retain splicing activity comparable to wild type (Figure 5.4a). By comparison, mutating the loop entirely to adenosines leads to ~3-fold difference. This would suggest that there is some level of cooperativity between the loop nucleotides that contributes to the overall structure.

While the previous study deleted the terminal pyrimidine to test the effects of a GUAA-tetraloop sequence, a GAAA sequence was also tested. The latter allows for a canonical

GNRA-tetraloop fold to still maintain an adenosine at the very tip of the loop to potentially interact with the DV minor groove. Both of these mutants result in ~3-fold reduction in splicing activity (Figure 5.4b), consistent with the findings of Fedorova *et al*.[12] This still does not rule out the possibility that a tetraloop is incompatible at this position, because a tetraloop substation would effectively lower the loop by one base plane (*vide supra*). Different helical lengths were tested (-2 to +2) to favor docking of this tetraloop and affect splicing activity. Increasing the helix by 1 base pair had a modest reduction on the effect of splicing, however the other helical changes (-2, -1, +2) resulted in significant defects in splicing (Figure 5.4c). The addition of a GAAA-loop with varying helical lengths (-2 to +3) does not rescue splicing activity, rather it exacerbated the defect further to levels comparable to Δμ (Figure 5.4d).

5.5 Discussion

The GUAAY-pentaloop sequence found in the μ-μ′ of group IIB1 introns represents a new, distinct RNA tertiary motif. Consistent with previous studies, this loop cannot be substituted for a tetraloop. In addition, the defect caused by a tetraloop substitution cannot be rescued by changing the length of the attached helix. The helix attached to the μ-loop emanates from a highly conserved four-way junction, with two other tertiary interactions in close proximity to this junction. It could be possible to adapt a GNRA-tetraloop at this position, but would likely require the remodeling of the aforementioned junction. It is probable, however, that a GNRA tetraloop is completely incompatible in this position. Due to the close proximity of κ-κ′, the receptor must be able to accommodate five adenosines in the minor groove. Rather than inserting planar with respect to the receptor nucleotides, the loop nucleotides insert at an unusual angle that allows the base stacking plane to be parallel to the direction of the minor groove. This orientation would allow five, and potentially more, nucleotides to base stack effectively into the minor groove without clashing with the phosphate backbone. Additionally, there may be other stable variants of the pentaloop, similar to the GUAAY, but unable to accommodate splicing activity in the specific context of the μ-μ′ interaction.

5.6 Methods

IIB1 intron sequences were obtained from the Intron Database[26]. Alignment was done using Bioedit[27]. Sequence logos were generated using the WebLogo[28]. AMIGOS II[19] motif search was done on the following data set: PDBs accessed on September 2015, with a minimum of 30 RNA nucleotides, and a resolution greater then 3.5Å.

The constructs 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. Overlapping PCR was used to generate the mutant constructs, then cloned back into pUC57 with EcoRV and T4 DNA ligase. Plasmid was linearized using HindIII and used for in vitro transcription with T7 RNA polymerase. Radiolabeled transcripts were prepared using 10 µCi [α-32P] UTP (3000 Ci/mmole), 0.5 mM UTP, 1 mM other NTPs, and 2.5mM $MgCl_2$. 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, 1M NH_4Cl, 40 mM Tris-HCl (pH 7.5), and 0.02% SDS, and either 2.5 mM or 10 mM $MgCl_2$. Reactions were stopped by addition of 2.5V 95% EtOH. 0.5V 3M NaOAc pH 5.2 was added to precipitate the RNA at -80° C for 30 mins. Precipitations were spun at max speed at RT for 15 mins. Pellets were resuspended in 5ul of 0.5x TE, 50% formamide. Splicing products were resolved using a denaturing 4% polyacrylamide (19:1)/8 M urea gels. Gels were frozen at -20° C, and phosphor screens exposed for 16hrs. The Bio-Rad PMI was used to image and quantitate gels. All splicing assays were done in triplicate.

References

1. Cate, J. H., Gooding, A.R., Podell, E., Zhou, K., Golden, B.L., Kundrot, C.E., Cech, T.R., Doudna, J.A. Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science* 273, 1678–85 (1996).

2. 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).

3. Jucker, F. M. & Pardi, A. Solution structure of the CUUG hairpin loop: a novel RNA tetraloop motif. *Biochemistry* 34, 14416–27 (1995).

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. Jaeger, L., Michel, F. & Westhof, E. Involvement of a GNRA tetraloop in long-range RNA tertiary interactions. *J. Mol. Biol.* 236, 1271–6 (1994).

6. Fiore, J. L. & Nesbitt, D. J. An RNA folding motif: GNRA tetraloop-receptor interactions. *Q. Rev. Biophys.* 46, 223–64 (2013).

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. Seetharaman, M., Eldho, N. V, Padgett, R. A. & Dayie, K. T. Structure of a self-splicing group II intron catalytic effector domain 5: parallels with spliceosomal U6 RNA. *RNA* 12, 235–47 (2006).

9. Stefl, R. & Allain, F. H.-T. A novel RNA pentaloop fold involved in targeting ADAR2. *RNA* 11, 592–7 (2005).

10. Lapouge, K. Perozzo, R., Iwaszkiewicz, J., Bertelli, C., Zoete, V., Michielin, O., Scapozza, L., Haas, D. RNA pentaloop structures as effective targets of regulators belonging to the RsmA/CsrA protein family. *RNA Biol.* 10, 1031–41 (2013).

11. Ilgu, M., Fulton, D.B., Yennamalli, R.M., Lamm, M.H., Sen, T.Z., Nilsen-Hamilton, M.. An adaptable pentaloop defines a robust neomycin-B RNA aptamer with conditional ligand-bound structures. *RNA* 20, 815–24 (2014).

12. 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).

13. 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).

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

15. Moody, E. M., Feerrar, J. C. & Bevilacqua, P. C. Evidence that folding of an RNA tetraloop hairpin is less cooperative than its DNA counterpart. *Biochemistry* 43, 7992–8 (2004).

16. Antao, V. P., Lai, S. Y. & Tinoco, I. A thermodynamic study of unusually stable RNA and DNA hairpins. *Nucleic Acids Res.* 19, 5901–5 (1991).

17. Quigley, G. J. & Rich, A. Structural domains of transfer RNA molecules. *Science* 194, 796–806 (1976).

18. Jucker, F. M. & Pardi, A. GNRA tetraloops make a U-turn. *RNA* 1, 219–22 (1995).

19. Wadley, L. M., Keating, K. S., Duarte, C. M. & Pyle, A. M. Evaluating and learning from RNA pseudotorsional space: quantitative validation of a reduced representation for RNA structure. *J. Mol. Biol.* 372, 942–57 (2007).

20. Cate, J. H., Hanna, R. L. & Doudna, J. A. A magnesium ion core at the heart of a ribozyme domain. *Nat. Struct. Biol.* 4, 553–558 (1997).

21. DRAPER, D. E. A guide to ions and RNA structure. *RNA* 10, 335–343 (2004).

22. Misra, V. K. & Draper, D. E. On the role of magnesium ions in RNA stability. *Biopolymers* 48, 113–35 (1998).

23. 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).

24. Bowman, J. C., Lenz, T. K., Hud, N. V & Williams, L. D. Cations in charge: magnesium ions in RNA folding and catalysis. *Curr. Opin. Struct. Biol.* 22, 262–272 (2012).

25. 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).

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

27. Tippmann, H.-F. Analysis for free: comparing programs for sequence analysis. *Brief. Bioinform.* 5, 82–7 (2004).

28. Crooks, G. E., Hon, G., Chandonia, J.M. & Brenner, S. E. WebLogo: A Sequence Logo Generator. *Genome Res* (2004).

Acknowledgements

We thank Timothy Thomas Wiryaman for comments on the manuscript and for the numerous books that he read. R.T.C. was supported by the Cell, Molecular, and Genetics (CMG) Training Program funded by NIH predoctoral training grant 5T32GM007240. This work was supported by NIH grant 5R01GM102216 awarded to N.T.

Chapter 5, in part, is currently being prepared for submission for publication of the material. **Chan R.T**.; Keating, K.S.; Toor, N. The dissertation author is the primary author on this paper.

Chapter 6: Conclusion

6.1 Summary

The work contained in this dissertation has uncovered significant structural elements that pertain to RNA splicing. The first study utilized a point mutant to inactivate a bacterial, linear-forming group II intron to determine the structural elements required for the first step of splicing. The crystal structure of this pre-catalytic group II intron found that the 5′ splice site adopts a sharp kinked phosphate backbone conformation. This presents a single phosphate into the active site and would thereby ensure the fidelity of the splicing reaction. The same phosphate backbone conformation at the 3′ is observed in the pre-2s structure of the phylogenetically unrelated group I intron and suggests that this is a universal conformation used for RNA splicing. In addition, this work supports the idea that only one active site is present in the group II intron and is readily capable of catalyzing both steps of splicing.

The second project focused on elucidating the structural requirements necessary to promote lariat formation. Therefore, the eukaryotic intron from *Pylaiella littoralis* was targeted due to its high *in vitro* activity and propensity to splice through the lariat forming pathway. The crystal structure of this intron revealed an extensive network of RNA tertiary structures that stimulated the lariat forming pathway and also explained its high level of splicing activity. Satisfyingly, this body of work was able to reconcile decades of biochemical data of group II intron splicing. In addition, novel tertiary structures were identified, which include ρ-ρ′, τ-τ′, and π-π′. Of particular interest is the π-π′ interaction, which links a loop of DII to the helix of DVI that is directly adjacent to the bulged adenosine. Due to the proximity of this interaction to the bulged adenosine, and therefore the lariat bond, we suggest that this interaction is dynamic and is critical for the removal of the lariat following the first step and bringing the 3′ splice site to the active site prior to the second step.

The third project centered on obtaining a splicing intermediate of the group II intron to determine what, if any, conformational changes are required to transition from one step of

splicing to another. To this end, we determined that a crystal solution additive, iridium hexamine, was shifting equilibrium towards the post-catalytic, fully splice state. Crystals grown in the absence of iridium hexamine were determined to have an intact 3′ splice site, indicative of the pre-2s state. Comparison of the pre-2s structure with the post-catalytic structure revealed nucleotide rearrangements within the catalytic triplex between the highly conserved junction nucleotides. Mutagenesis of these linker nucleotides resulted in severe defects in either the first or second step of splicing. Base on the structural and biochemical data, we propose that each catalytic triplex conformation corresponds to a distant stage of splicing.

The goal of the fourth project was to characterize a novel pentaloop identified in the lariat forming group IIB1 intron. The GUAAY pentaloop is involved in the μ-μ′ interaction that links the loop to the minor groove of DV. This loop is highly conserved in its phylogenetic class and is distinctive of the more common GNRA tetraloop. Various mutations on the loop and proximal regions to determine if it was possible to substitute the GUAAY pentaloop for a GNRA tetraloop. Based on the in vitro splicing assays, a GNRA tetraloop is incompatible in this position and no additional mutations were able to rescue this defect. We therefore conclude that the GUAAY pentaloop is a distinct RNA structural element with respect to the GNRA tetraloop and represents a new class of RNA structural motif.

The recent cryo-EM structure of the spliceosome revealed the heart of the multimegadalton complex to be made up entirely of RNA. Furthermore, the authors concluded that the active site architecture is identical to that of the group II intron, further supporting the evolutionary link between the two systems (Figure 6.1). Additionally, several of the predictions pertaining to eukaryotic splicing based on the lariat structure have been corroborated by the aforementioned cryo-EM structure. This further supports the evolutionary link between the group II intron and the spliceosome, with the former serving as a proto-spliceosome that gave rise to the later. Therefore, the conclusions made in this dissertation pertaining to RNA splicing in the group II intron are readily applicable to the spliceosome.

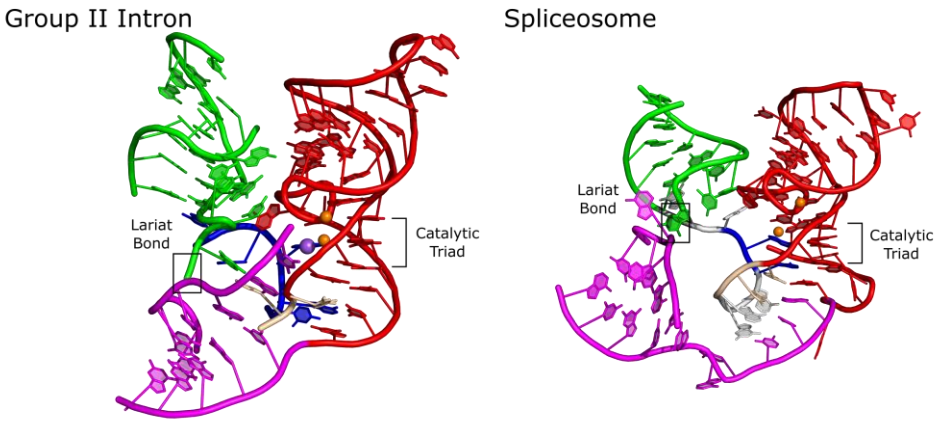

Figure 6.1: **Comparison of the active site architecture of the group II intron (left) and the spliceosome (right).** Despite many differences, the RNA components that make up the core of these enzymes remain markedly similar. A significant difference between the two is the position of the catalytic metals. In the group II intron, the distance between M1 and M2 is ~4.5Å, whereas the spliceosome it is ~7Å. This may be due to ambiguous nature of assigning metals in electron microscopy and the exact identify of M1 is not experimentally possible. Colors are consistent with the colors used for the group II intron domains. Orange spheres correspond to magnesium ions, the purple sphere is a monovalent ion determined through anomalous metal soaks.