Published in final edited form as:

Trends Biochem Sci. 2017 June ; 42(6): 470–482. doi:10.1016/j.tibs.2017.03.007.

Structural insights into the mechanism of group II intron splicing

Chen Zhao¹ and Anna Marie Pyle^{2,3,4,*}

¹Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520, USA

²Department of Molecular, Cellular, and Developmental Biology, Yale University, New Haven, CT 06520, USA

³Department of Chemistry, Yale University, New Haven, CT 06520, USA

⁴Howard Hughes Medical Institute, Chevy Chase, MD 20815, USA

Abstract

While the major architectural features and active-site components of group II introns have been known for almost a decade, information on individual stages of splicing has been lacking. Recent advances in crystallography and cryo-electron microscopy have provided major new insights into the structure of intact lariat introns. Conformational changes that mediate the steps of splicing and retrotransposition are being elucidated, revealing the dynamic, highly coordinated motions that are required for group II intron activity. Finally, these ribozymes can now be viewed in their larger, more natural context as components of holoenzymes that include encoded maturase proteins. These studies expand our understanding of group II intron structural diversity and evolution, while setting the stage for rigorous mechanistic analysis of RNA splicing machines.

Keywords

Group II intron; RNA structure; ribozyme; maturase; spliceosome; retrotransposition

Overview of group II introns and their protein partners

Group II introns are an ancient class of self-splicing ribozymes that are found in eubacteria, archaeobacteria and the organelles of plants, fungi and various lower eukaryotes [1–4]. These large RNA molecules were originally characterized by a distinctive arrangement of six secondary structural domains (domain 1–6, or D1–6) (Figure 1A and Figure 2A) that contain segments of highly conserved sequence [1–4]. Recent advances in structural biology have revealed that all group II introns also share an architecturally conserved tertiary structure, particularly within the catalytic core [4–8]. Based on features of their secondary structural elements, sequences of intron-encoded proteins (vide infra), and mechanisms of exon

*Correspondence: anna.pyle@yale.edu (A.M. Pyle).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

recognition, group II introns have been categorized into several families, of which IIA, IIB and IIC are the most well-studied (Figure 1A and Figure 2A) [3, 4, 9, 10]. Group II introns catalyze self-splicing via two consecutive transesterification reactions [4, 11], initiated either by an adenosine residue in D6 (branching) [12] or by a water molecule (hydrolysis) (Figure 1B) [13], resulting in liberation of lariat or linear intron molecules, respectively. Both of these pathways can be utilized *in vivo*, depending on conditions and intron subtype [12, 13]. Although splicing catalysis is mediated by the group II intron RNA, these introns encode an unusual family of protein partners called maturases, which facilitate splicing [2, 14, 15]. These proteins are indispensable during intron retrotransposition [14, 15], which involves reverse splicing of the intron into DNA targets, utilizing the same active site as that employed during forward splicing. Because of their ability to behave as mobile genetic elements, group II introns have played a critical role in genome evolution, and they are likely ancestors of eukaryotic spliceosomes, spliceosomal introns, non-LTR retrotransposons and telomerases [1–4, 16].

Despite the central importance of group II introns in evolution and modern biology, the structural features of this ribozyme class remained obscure until 2008, when the first crystal structure was obtained using a highly stable group IIC intron from the eubacterium *Oceanobacillus iheyensis* (*O.i.*) [5]. Studies on this system revealed the major architectural features of group II introns [17–19], providing new insights into their folding mechanisms [20] and a structural framework for identifying critical RNA elements within the eukaryotic spliceosome [16, 21, 22]. However, additional structures would be required to answer key questions about mechanisms of splicing and retrotransposition. For example, how does the intron adjust its conformation between the two steps of splicing and how are metal ions utilized during catalysis? What is the conformation of D6 and what does the branch-site look like? Can we visualize the structure of a maturase protein and understand how it docks within the intron scaffold? Are there additional links between group II introns and the eukaryotic spliceosome? Efforts in the past five years have clarified many of these questions and set the stage for additional mechanistic investigations.

Lariat structures of IIA and IIB introns reveal new architectural themes

Recent structures of group IIA and IIB introns have revealed major new classes of architectural motifs and additional structural states that contribute to the pathway of splicing. For example, a complete group II intron lariat structure was finally visualized through crystallographic studies on a group IIB intron from *Pylaiella littoralis* (*P.li.*) [7]. Determined to a resolution of 3.8 Å, this structure captures the *P.li.* intron in post-catalytic form, in which spliced exons are bound to the intron (Figure 2A), and it represents the first glimpse of intron D6, which was not visualized during earlier studies of the hydrolytic *O.i.* intron [7, 23]. In the *P.li.* structure, D6 projects away from D5, forming a network of interactions with D2, which include a pair of predicted tetraloop-receptor interactions (η - η' and π - π') (Figure 2B) [7, 24] that are critical for positioning reactants during the second transesterification reaction.

The group IIB lariat structure also revealed structural features and tertiary interactions that are present only in the larger, more highly evolved classes of introns such as IIB, and which

are lacking in the more ancient IIC introns [7]. For example, D3 in *P.li* is a complex 4-way junction, and two of its terminal loops form tetraloop-receptor interactions with D2 (ρ - ρ') and D5 (μ - μ') [7]. The latter interaction, along with the well-studied κ - κ' interaction, were shown to adopt structures almost identical to those predicted from chemogenetic and modeling studies on the ai5 γ group IIB intron [25–27]. The novel μ - μ' interaction connects D3 with nucleotides in the junction between D1 and D2 and a loop region of D3 interacts with D5 to extend the κ - κ' interaction (Figure 2B) [7]. Perhaps most significantly, this structure revealed the EBS2-IBS2 interaction, which is a secondary exon-recognition motif that is lacking in IIC introns (IIC introns have only the universal EBS1-IBS1 pairing) (Figure 2B) [7]. By revealing EBS2-IBS2, together with the critical pairing between the 3' exon and EBS3, the *P.li* group II intron structure provided the first complete framework for exon recognition within group II introns (Figure 2B) [7].

Through advances in cryo-electron microscopy (cryo-EM), a group IIA intron lariat from *Lactococcus lactis* (*L.l.* or LI.LtrB) was solved to 3.8 Å resolution by single particle cryo-EM (Figure 2A) [8]. Despite the limited resolution, this ground-breaking structure was highly significant because it included a bound maturase protein (*vide infra*) (Figure 5B), revealing critical protein-RNA interactions and RNA tertiary interactions for the first time [8]. For example, the *L.l.* structure revealed the conserved β - β' kissing loop that helps maintain D1 architecture, but it lacks the more poorly-conserved τ - τ' interaction observed in *P.li* (Figure 2C) [8].

Despite apparent differences among structures of the IIC, IIB and IIA introns, a comparison of their structures reveals a common architecture and trends in their evolution become clear (Figure 2A and 2D). An almost identical functional core, represented by the group IIC introns, is maintained in all cases (Figure 2A and 2D). This minimal core is stabilized predominantly by interactions within D1, consistent with the view that D1 is the critical folding intermediate and the structural scaffold for all group II introns [20, 28, 29]. Specifically, the α - α' , D1_a T-loop (Figure 2D) and Z-anchor interactions crucial for the proper folding of D1 of all intron classes [5, 7, 8]. However, the ribose zipper ω - ω' is only observed in IIC introns [5], potentially because stem D1d₂ (which would contain ω') is pulled away in IIA and IIB introns through formation of the EBS2-IBS2 interaction (Figure 2D) [7, 8]. Possibly, the unique ribose zipper ω - ω' interaction in *O.i.* group IIC intron explains the low Mg²⁺ requirement for group IIC intron folding [20].

Structural insights into the mechanism of group II intron splicing

An elaborate tertiary structure enables group II introns to carry out a complex set of highly specific reactions. During self-splicing and reverse splicing, the intron catalyzes two sequential transesterification reactions [4], each of which involves passage through a set of structurally coordinated micro-states (Figure 3A). Both steps of splicing (or reverse-splicing into DNA, in the case of retrotransposition) occur within a single catalytic center that contains distinct substructures for activating the nucleophile and the phosphodiester linkage at the target splice site (Figure 3C) [4]. During the sequential stages of splicing, components pass in and out of the active-site, forming networks of transient interactions that are probably coordinated by conformational coupling between states (Figure 3A and 3C). The

architecture of the active site immediately before and after the first step of splicing (the pre- and post-first step states) and of an active-site intermediate that occurs between the steps of splicing (specifically between the post-first and pre-second steps) was revealed by a collection of *O.i.* group II intron crystal structures in which water functions as the nucleophile during the first step of splicing (Figure 3B) [4, 18, 30, 31]. Much-needed structural insights into the second step of splicing were recently revealed by the crystal structure of a group II intron chimera that was obtained by fusing D1-5 of the *O.i.* intron with D6 from an *Azotobacter vinelandii* (*A.v.*) group IIC intron, resulting in a stable intron that splices through the branching pathway (*O.i.-A.v.*, solved to 3.5Å resolution) (Figure 3B) [32, 33]. Together, these structures have provided a wealth of information on the structural states that stimulate splicing by group II introns.

A complex metal ion center within the group II intron core

In multiple states of splicing, the *O.i.* group II intron maintains a complex metal ion center at the catalytic core which includes two divalent ions (M1, M2), supported by two monovalent ions (K1, K2) (Figure 3B) [4, 6, 30, 31]. The two divalent ions (in most cases Mg^{2+} ions) are located at a sharp kink formed by the 2-nt bulge in D5 (Figure 3B) with M1 and M2 coordinating the scissile phosphate, while M2 activating the nucleophile (Figure 3B) [6, 30, 31]. These two Mg^{2+} ions are located in positions supported by functional biochemical data on group II introns [22, 34, 35], and they have similar configurations as those observed in other ribozymes and protein enzymes that utilize two-metal-ion mechanisms for phosphodiester cleavage [36]. In addition, the two site-bound monovalent ions (K1 and K2, being either K^+ or NH_4^+) help to stabilize the proper conformation of the catalytic center (Figure 3B) [6, 30, 31].

The association of these 4 metal ions (M1, M2, K1 and K2) is likely to be dynamic and cooperative. For example, in the presence of Li^+ or Na^+ , which are too small to properly occupy the K1 and K2 sites, all 4 metal ions at the *O.i.* group II intron catalytic center are released from the catalytic core and rearrangements of active-site nucleotides are observed [6, 30, 31]. It will be important to determine how these metals are arranged when the active-site is occupied by the branch-site during each step of splicing. Unfortunately, due to the low resolution and limited anomalous scattering experiments, only one divalent metal ion could be unambiguously identified in structures of the *O.i.-A.v.* core [32]. That said, basic features of the *O.i.-A.v.* active-site are quite similar to those characterized within constructs of the *O.i.* intron [5, 30, 31].

The first transesterification step of group II intron splicing

The structural state that occurs immediately before the first step of splicing (the pre-first step) has been visualized crystallographically by mutating active-site residues [18] and by trapping *O.i.* splicing precursors with Ca^{2+} ions [30, 31] (Figure 3B), which block splicing chemistry. *In vitro*, water is the nucleophile during the first step of *O.i.* splicing, and crystal structures show this water molecule aligned for attack on the phosphate at the tightly kinked 5'-splice site, which is poised to react (Figure 3B) [30, 31]. Given that Ca^{2+} is not a perfect mimic of Mg^{2+} , it will be important to obtain additional structural information on the organization of active-site components immediately prior to the first step. Unfortunately, the

pre-first step structure for the branching pathway has not yet been characterized. For the first step of branching to occur, D6 will need to insert correctly within the active-site and adopt a catalytically-active conformation that presents the 2'-OH group of the branch-site in an appropriate configuration.

However, even in the absence of high resolution information on the pre-first step, genetic, phylogenetic and biochemical experiments have established a number of interactions that are likely to stabilize D6 in the pre-first step conformation. For example, Li *et al.* identified a receptor in D1C₁ that interacts with D6 to position it for the first step of splicing [37]. This D1–D6 interaction has been further supported by a study on group II intron mutants in which matched D6 and D1C₁ receptors lead to the highest fraction of lariat branching [33]. Interestingly, introns with unmatched D6 and D1C₁ receptors are still able to branch [33], suggesting that conformational sampling of D6 alone can be sufficient for adopting the precise conformation that is required for the first step of catalysis. Specific D1–D6 interactions may therefore shift the equilibrium and increase the intron population that is reactive for branching.

The second transesterification step of group II intron splicing

Before the second step of splicing can proceed, significant structural rearrangements occur within the core. Unfortunately, the state that exists immediately before the second transesterification reaction has not been visualized at high resolution, so it remains unclear how the reactants are aligned for reaction. However, a potential conformational intermediate that lies between the post-first step and the pre-second step has been visualized in crystal structures obtained in conditions of Li⁺ or Na⁺ ions [30, 31]. In these structures, all four metals have been released (*vide supra*) and J2/3 adopts an alternative “toggled” conformation that opens the core, providing additional space for reactants and products to exchange within the active-site [30, 31].

New crystal structures of the *O.i.-A.v.* chimera provide important insights into the configuration of the second-step active site [32]. These structures capture the lariat after it has completed the first step of reverse-splicing, which may represent the same intron structure as the pre-second step during forward splicing. These structures and supporting mutational data indicate that formation of the second step active site is coupled to a striking rearrangement in the D6 secondary structure (Figure 4A) [32]. During the first step of splicing, D6 forms a one-nucleotide bulge that only contains the nucleophile adenosine (Figure 4A), as described in previous studies [32]. However, the second step is stabilized by an isoform of D6 in which the register of base-pairings has changed, resulting in an alternative secondary structure that presents a two-nucleotide bulge at the branch-site (Figure 4C) [32]. One caveat is that the *O.i.-A.v.* chimera undergoes the second-step very slowly [32, 33], which is unusual because the second step is rapid in all other introns, and it is not observed to be rate-limiting [10, 11, 30]. In addition, the D5 and D6 rearrangements that are inferred from studies of this chimera may not be general. That said, future experiments with intact introns from diverse classes will establish whether D6 rearrangement is a general feature of all group II introns.

The mechanisms by which the core rearranges and ensures directional passage from the first to the second step of splicing remain unclear, but speculations can be drawn from existing structural information. For example, the movement that shifts the 5' exon out of its position as the leaving group and into the nucleophilic site (thereby enabling the terminal 3'-OH to attack the 3'-splice site) may be triggered by 5' splice site cleavage during the first step of splicing (Figure 3C). Since the 5' exon is likely to remain base-paired to EBS1 [20], EBS1 would need to swing inward to re-position the 5' exon. This swing motion may be induced by the abrupt relaxation in structural restraint that occurs when the 5' splice site is cleaved. This action severs the intron 5' end from the 5' exon, giving rise to an altered EBS1 conformation that is seen in the crystal structure of isolated group II intron D1 molecules (Figure 4B) [20]. A similar motion in loop 1 of U5, which pairs with the 5' exon, is proposed to mediate transitions between the two steps of splicing in spliceosomal systems [38]. Placement of the 3'-splice site phosphate into the substrate site is probably coupled to motion of the branch-site adenosine in D6, which is supported by analysis of the *O.i.-A.v.* crystal structures (Figure 4C) [32]. This conformation is stabilized by the η - η' [24, 33, 39] and π - π' interactions between D2 and D6 [7], and by the γ - γ' interaction [32] (Figure 1A). Intriguingly, although the branch site needs to move over 20 Å during this process, computational modeling suggests that this large movement can be achieved by a small rotation in D6 [25].

The first transesterification step of group II intron reverse splicing

The *O.i.-A.v.* chimera structures have revealed features that may be specifically relevant to reverse splicing, which is an obligate stage of retrotransposition by group II introns. For example, in structures with and without the 5' exon, D5 is observed to adopt different conformations (Figure 4D) [32], consistent with previous studies showing that D5 structure can be dynamic [40, 41]. The structure lacking a 5' exon is probably only relevant for retrotransposition because, during forward RNA splicing reactions, the 5' exon is always bound to EBS1. Indeed, the EBS1-exon interaction forms and persists during the earliest stages of group II intron folding [20]. In the *O.i.-A.v.* lariat that lacks the 5' exon, EBS1 becomes disordered, leading to an alternative conformation in the critical D5 receptor known as ζ' , and in surrounding regions of D1 [32] (Figure 4D). The authors propose that this conformational change may provide a checkpoint for ensuring that catalytically competent conformations of the intron can only form when the correct 5' exon is bound [32]. While this is an intriguing idea, it is worth noting that in previous crystal structures of the linear *O.i.* intron [5, 19, 30], the D5 structure remains the same, whether 5' exon is bound or not. Therefore, it is possible that this 5' exon-induced conformational change is unique to the *O.i.-A.v.* chimera, or it may only occur in lariat forms of the intron, suggesting a benefit of lariat introns over linear introns during retro-transposition [32]. However, both linear and lariat introns can readily retrotranspose [14, 42, 43], so there may be advantages to both mechanisms during different types of intron mobility.

Intron-encoded proteins tune and regulate ribozyme behavior

Despite their robust autocatalytic ribozyme activity, group II introns rarely function alone. Instead, they participate in splicing and retrotransposition as components of an RNP

holoenzyme that includes the intron RNA in complex with a family of encoded proteins, known as maturases [2, 44] (Zhao & Pyle, *Curr. Opin. Struct. Biol.*, submitted). These multifunctional proteins have been elusive targets for structural biologists, and their relative complexity has hindered biochemical investigations of their mechanism. Recent improvements in crystallographic strategy [45] and in cryo-EM [8] have finally elucidated maturase structures and provided much-needed insights into their roles as splicing and mobility cofactors.

Crystal structures of maturase RT domains from the eubacteria *Roseburia intestinalis* (*R.i.*) and *Eubacterium rectale* (*E.r.*) were recently solved at 1.2 Å and 2.1 Å resolution (Figure 5A and Figure 6C), revealing a unique form of RT scaffold that is shared with non-LTR retrotransposons (like human L1), and which is distinct from the RT scaffold of retroviruses, such as HIV [45] (Zhao & Pyle, *Curr. Opin. Struct. Biol.*, submitted). Remarkably, the closest homologs to the maturase RT are the RT-like domain of spliceosomal Prp8, and the RNA-dependent RNA polymerase from flaviviruses, supporting the evolutionary relationship between group II intron and eukaryotic spliceosome, and linking their evolution with that of RNA viruses (Figure 6C) [45] (Zhao & Pyle, *Curr. Opin. Struct. Biol.*, submitted).

One face of the maturase RT domain contains a large positively-charged surface that mediates strong, highly specific interactions with intron D4 (Figure 5A) [45]. This RNA binding surface is opposite the face that encloses the reverse transcriptase active-site, suggesting an economical strategy by which these ancient proteins packed multiple functions into a tiny scaffold [45]. In all the crystal structures that have been solved thus far, the RT domains dimerize along a specific, extended interface, which is maintained in solution and is consistent with the stoichiometry reported during previous biochemical studies of maturase-assisted splicing (Figure 5D) [46–48]. Importantly, this dimerization interface does not overlap with the dimerization interface in the HIV RT p51-p66 heterodimer [49].

Of particular interest is the maturase-intron holoenzyme, which was finally visualized at 3.8 Å resolution using cryo-EM single particle reconstruction [8]. This ground-breaking structure of the full-length LtrA maturase in complex with its parent L1.LtrB intron lariat captures the post-catalytic state of the intron-maturase RNP particle, along with a fragment of spliced mRNA that remains bound to the intron EBS1 and EBS2 (Figure 5B) [8]. LtrA interacts with intron D4 via the outer surface of the RT domain (Figure 5B) [8], using the positively charged surface identified in the crystallographic investigations (Figure 5A) [45]. This same surface also interacts with peripheral loop (Id(iii)a) within D1 (Figure 5C) [8], thereby stabilizing maturase binding within the RNA scaffold. The maturase “X domain” mediates a network of major interactions with the RNA. For example, a protein motif known as the “ti-loop” inserts deeply into D1, anchoring the maturase X domain along the intron surface (Figure 5B) [8]. Additionally, an α -helical segment within the X domain (Ta3) runs parallel with the associated mRNA and buttresses the exon-EBS interactions, potentially enhancing the ability of the intron to recognize the 5'-splice site (Figure 5C) [8]. Loop regions within the DNA binding domain (DBD) also contribute to stabilization of exon-EBS interactions (Figure 5C) [8]. An intriguing feature of the available holoenzyme structure is

that the intron RNA is associated with only one molecule of maturase [8, 44], which contrasts with structural and functional evidence for a dimeric maturase form [46–48]. However, it is possible that the monomeric form is mechanistically important and that changes in maturase stoichiometry provide a switch for mediating important conformational changes within the holoenzyme [50] (Zhao & Pyle, *Curr. Opin. Struct. Biol.*, submitted).

Despite these advances in understanding group II holoenzyme structure, the precise role of the maturase in promoting group II intron reactivity remains unclear. Maturase components do not interact directly with reactants in the catalytic site and available data suggest an allosteric role for the protein [8, 51, 52] (Zhao & Pyle, *Curr. Opin. Struct. Biol.*, submitted). While stabilizing interactions between the X-domain and EBS1 may increase splice-site recognition fidelity and improve ground-state stability (Figure 5C), they do not explain the dramatic rate enhancements that are observed in the presence of protein [46, 53]. That said, stabilization of the EBS1-IBS1 pairing by the maturase X domain (Figure 5D) is likely to be critical during retrotransposition, since the DNA target strand will bind EBS1 with lower affinity than an RNA exon, which is further stabilized by tertiary contacts involving 2'OH groups along the backbone of the EBS1-IBS1 duplex [32].

Previous chemical probing [51] and protein-RNA cross linking [53] studies suggest that the maturase facilitates splicing by forming weak interactions with intron D6. However, in the current cryo-EM model with monomeric maturase, the maturase X domain is $> 6 \text{ \AA}$ away from D6 (Figure 5B) [8]. If one models a second maturase protein by aligning the dimer interfaces revealed by the maturase crystal structures [45], the second maturase monomer is positioned to interact directly with D6, although direct experimental evidence is required to support this hypothesis. Interestingly, X domain sequences are poorly conserved [54, 55], which may indicate that maturase X domains adapt to divergent specifications (eg. the length and sequence) of intron D6. However, since X domains in all maturases are highly positively charged, their low sequence conservation could suggest a promiscuous mechanism in which the maturase creates a positively-charged fence that restricts the conformational space for D6, thereby enhancing reactivity.

Parallels between group II introns and the spliceosome

The hypothesis that group II introns and the eukaryotic spliceosome share an evolutionary history was proposed almost 30 years ago [56, 57], and evidence supporting this hypothesis has accumulated steadily during the intervening years [21, 22, 38, 58–61]. Their common ancestry has almost certainly been confirmed by recent structural studies on group II intron maturases [8, 44, 45, 62] (Zhao & Pyle, *Curr. Opin. Struct. Biol.*, submitted) and the spliceosome itself [63–71].

Active-site RNA structures in group II introns and the spliceosome, particularly D5 and U6, are almost identical and they play similar roles in the catalysis of splicing reactions, as described in recent reviews (Figure 6A and 6B) [4, 67]. Striking structural similarities have also been found for critical protein components in these two systems. Indeed, recent maturase structures demonstrate that this class of proteins is closely homologous to the spliceosomal core protein Prp8 (Figure 6C), demonstrating that both protein and RNA

components of group II and spliceosomal systems share common ancestry [8, 44, 45] (Zhao & Pyle, *Curr. Opin. Struct. Biol.*, submitted).

Adaptations in Prp8 are particularly interesting, as they shed light on the evolution of a more complex splicing machine. For example, the outer surface of the RT-like domain, which is a positively-charged RNA-binding motif in maturases, has become a protein-protein interaction motif in Prp8 (Zhao & Pyle, *Curr. Opin. Struct. Biol.*, submitted). Indeed, RNA binding in Prp8 has been relegated to an auxiliary N-terminal domain, which interacts with U5 snRNA [63–66, 68, 69, 71] (Zhao & Pyle, *Curr. Opin. Struct. Biol.*, submitted). Despite this adaptation, the X/thumb domain in Prp8 inserts within the catalytic center in a manner that is almost identical to that of the maturase X/thumb domain in group II intron holoenzymes (Figure 6D) [8, 63, 65]. Considering the central role of Prp8 in pre-mRNA splicing [72, 73], this observation suggests that the contribution of Prp8 and maturases to the mechanism of splicing is the same.

The evolution of spliceosomes, which function as multiple-turnover enzymes that catalyze splicing in-trans, is one of the most important events in the genetic diversification of eukaryotes. During this transition, the peripheral structures and structural scaffolding of group II introns were replaced by spliceosomal proteins, and yet RNA components within the catalytic center remain well-preserved, perhaps because ribozymes are particularly proficient at catalyzing specific phosphodiester cleavage and ligation reactions.

Concluding remarks

Recent advances in group II intron structural biology have provided many unprecedented insights into the structures and functions of these elaborate ribozymes. Specifically, intron D6 and additional structural motifs have now been visualized in class IIA and IIB introns, and the structure of the branch-site is increasingly well-defined. Crystal structures of linear and lariat introns at different stages of the splicing cycle have provided major insights into reaction mechanism. It is now possible to visualize group II introns as components of a holoenzyme that includes the fascinating maturase proteins that contribute to splicing and retrotransposition. Finally, advances in cryo-EM single particle reconstruction have led to high resolution spliceosome structures, enabling detailed comparisons with group II introns. And yet these advances mark only the beginning of a long journey toward understanding the physical and chemical mechanism of RNA splicing and retrotransposition (see Outstanding Questions).

Acknowledgments

We thank Dr. Srinivas Somarowthu and all Pyle Lab members for discussion. Chen Zhao is supported by Gruber Science Fellowship and a Yale University Fellowship. This work is supported by the National Institute of Health (R01GM50313). Prof. Anna Marie Pyle is a Howard Hughes Medical Institute Investigator.

References

1. Pyle AM. The tertiary structure of group II introns: implications for biological function and evolution. *Crit Rev Biochem Mol Biol.* 2010; 45:215–232. [PubMed: 20446804]

2. Lambowitz AM, Belfort M. Mobile Bacterial Group II Introns at the Crux of Eukaryotic Evolution. *Microbiol Spectr*. 2015; 3
3. Zimmerly S, Semper C. Evolution of group II introns. *Mob DNA*. 2015; 6:7. [PubMed: 25960782]
4. Pyle AM. Group II Intron Self-Splicing. *Annu Rev Biophys*. 2016; 45:183–205. [PubMed: 27391926]
5. Toor N, et al. Crystal structure of a self-spliced group II intron. *Science*. 2008; 320:77–82. [PubMed: 18388288]
6. Marcia M, Pyle AM. Principles of ion recognition in RNA: insights from the group II intron structures. *RNA*. 2014; 20:516–527. [PubMed: 24570483]
7. Robart AR, et al. Crystal structure of a eukaryotic group II intron lariat. *Nature*. 2014; 514:193–197. [PubMed: 25252982]
8. Qu G, et al. Structure of a group II intron in complex with its reverse transcriptase. *Nat Struct Mol Biol*. 2016; 23:549–557. [PubMed: 27136327]
9. Michel F, et al. Comparative and functional anatomy of group II catalytic introns—a review. *Gene*. 1989; 82:5–30. [PubMed: 2684776]
10. Nagy V, et al. Predicted group II intron lineages E and F comprise catalytically active ribozymes. *RNA*. 2013; 19:1266–1278. [PubMed: 23882113]
11. Daniels DL, et al. Two competing pathways for self-splicing by group II introns: a quantitative analysis of in vitro reaction rates and products. *J Mol Biol*. 1996; 256:31–49. [PubMed: 8609612]
12. Peebles CL, et al. A self-splicing RNA excises an intron lariat. *Cell*. 1986; 44:213–223. [PubMed: 3510741]
13. Podar M, et al. Group II intron splicing in vivo by first-step hydrolysis. *Nature*. 1998; 391:915–918. [PubMed: 9495347]
14. Lambowitz AM, Zimmerly S. Group II introns: mobile ribozymes that invade DNA. *Cold Spring Harb Perspect Biol*. 2011; 3:a003616. [PubMed: 20463000]
15. Pyle, AM., Lambowitz, AM. Group II Introns: Ribozymes That Splice RNA and Invade DNA. In: Gesteland, RF., editor. *The RNA World*. 3rd. Cold Spring Harbor Laboratory Press; 2006. p. 469–505.
16. Keating KS, et al. A structural analysis of the group II intron active site and implications for the spliceosome. *RNA*. 2010; 16:1–9. [PubMed: 19948765]
17. Toor N, et al. Tertiary architecture of the *Oceanobacillus iheyensis* group II intron. *RNA*. 2010; 16:57–69. [PubMed: 19952115]
18. Chan RT, et al. Crystal structure of a group II intron in the pre-catalytic state. *Nat Struct Mol Biol*. 2012; 19:555–557. [PubMed: 22484319]
19. Toor N, et al. Structural basis for exon recognition by a group II intron. *Nat Struct Mol Biol*. 2008; 15:1221–1222. [PubMed: 18953333]
20. Zhao C, et al. Crystal structure of group II intron domain 1 reveals a template for RNA assembly. *Nat Chem Biol*. 2015; 11:967–972. [PubMed: 26502156]
21. Fica SM, et al. Evidence for a group II intron-like catalytic triplex in the spliceosome. *Nat Struct Mol Biol*. 2014; 21:464–471. [PubMed: 24747940]
22. Fica SM, et al. RNA catalyses nuclear pre-mRNA splicing. *Nature*. 2013; 503:229–234. [PubMed: 24196718]
23. Peters JK, Toor N. Group II intron lariat: Structural insights into the spliceosome. *RNA Biol*. 2015; 12:913–917. [PubMed: 26121424]
24. Costa M, et al. Multiple tertiary interactions involving domain II of group II self-splicing introns. *J Mol Biol*. 1997; 267:520–536. [PubMed: 9126835]
25. Somarowthu S, et al. Visualizing the ai5gamma group IIB intron. *Nucleic Acids Res*. 2014; 42:1947–1958. [PubMed: 24203709]
26. Somarowthu S. Progress and Current Challenges in Modeling Large RNAs. *J Mol Biol*. 2016; 428:736–747. [PubMed: 26585404]
27. Boudvillain M, Pyle AM. 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. [PubMed: 9843513]

28. Pyle AM, et al. Folding of group II introns: a model system for large, multidomain RNAs? *Trends Biochem Sci.* 2007; 32:138–145. [PubMed: 17289393]
29. Donghi D, et al. The structural stabilization of the kappa three-way junction by Mg(II) represents the first step in the folding of a group II intron. *Nucleic Acids Res.* 2013; 41:2489–2504. [PubMed: 23275550]
30. Marcia M, Pyle AM. Visualizing group II intron catalysis through the stages of splicing. *Cell.* 2012; 151:497–507. [PubMed: 23101623]
31. Marcia M, et al. Now on display: a gallery of group II intron structures at different stages of catalysis. *Mob DNA.* 2013; 4:14. [PubMed: 23634971]
32. Costa M, et al. Crystal structures of a group II intron lariat primed for reverse splicing. *Science.* 2016; 354
33. Monachello D, et al. Activating the branch-forming splicing pathway by reengineering the ribozyme component of a natural group II intron. *RNA.* 2016; 22:443–455. [PubMed: 26769855]
34. Sigel RK, et al. Metal ion binding sites in a group II intron core. *Nat Struct Biol.* 2000; 7:1111–1116. [PubMed: 11101891]
35. Gordon PM, Piccirilli JA. Metal ion coordination by the AGC triad in domain 5 contributes to group II intron catalysis. *Nat Struct Biol.* 2001; 8:893–898. [PubMed: 11573097]
36. Yang W, et al. Making and breaking nucleic acids: two-Mg²⁺-ion catalysis and substrate specificity. *Mol Cell.* 2006; 22:5–13. [PubMed: 16600865]
37. Li CF, et al. Linking the branchpoint helix to a newly found receptor allows lariat formation by a group II intron. *EMBO J.* 2011; 30:3040–3051. [PubMed: 21712813]
38. Hetzer M, et al. Trans-activation of group II intron splicing by nuclear U5 snRNA. *Nature.* 1997; 386:417–420. [PubMed: 9121561]
39. Chanfreau G, Jacquier A. An RNA conformational change between the two chemical steps of group II self-splicing. *EMBO J.* 1996; 15:3466–3476. [PubMed: 8670849]
40. Pechlaner M, et al. Structure and conformational dynamics of the domain 5 RNA hairpin of a bacterial group II intron revealed by solution nuclear magnetic resonance and molecular dynamics simulations. *Biochemistry.* 2013; 52:7099–7113. [PubMed: 24001362]
41. Pechlaner M, et al. Protonation-Dependent Base Flipping at Neutral pH in the Catalytic Triad of a Self-Splicing Bacterial Group II Intron. *Angew Chem Int Ed Engl.* 2015; 54:9687–9690. [PubMed: 26119804]
42. Roitzsch M, Pyle AM. The linear form of a group II intron catalyzes efficient autocatalytic reverse splicing, establishing a potential for mobility. *RNA.* 2009; 15:473–482. [PubMed: 19168748]
43. Zhuang F, et al. Linear group II intron RNAs can retrohome in eukaryotes and may use nonhomologous end-joining for cDNA ligation. *Proc Natl Acad Sci U S A.* 2009; 106:18189–18194. [PubMed: 19833873]
44. Agrawal RK, et al. Forks in the tracks: Group II introns, spliceosomes, telomeres and beyond. *RNA Biol.* 2016:1–5.
45. Zhao C, Pyle AM. Crystal structures of a group II intron maturase reveal a missing link in spliceosome evolution. *Nat Struct Mol Biol.* 2016; 23:558–565. [PubMed: 27136328]
46. Saldanha R, et al. RNA and protein catalysis in group II intron splicing and mobility reactions using purified components. *Biochemistry.* 1999; 38:9069–9083. [PubMed: 10413481]
47. Rambo RP, Doudna JA. Assembly of an active group II intron-maturase complex by protein dimerization. *Biochemistry.* 2004; 43:6486–6497. [PubMed: 15157082]
48. Molina-Sanchez MD, et al. Functionality of In vitro Reconstituted Group II Intron RmInt1-Derived Ribonucleoprotein Particles. *Front Mol Biosci.* 2016; 3:58. [PubMed: 27730127]
49. Ding J, et al. Structure and functional implications of the polymerase active site region in a complex of HIV-1 RT with a double-stranded DNA template-primer and an antibody Fab fragment at 2.8 Å resolution. *J Mol Biol.* 1998; 284:1095–1111. [PubMed: 9837729]
50. Piccirilli JA, Staley JP. Reverse transcriptases lend a hand in splicing catalysis. *Nat Struct Mol Biol.* 2016; 23:507–509. [PubMed: 27273636]
51. Matsuura M, et al. Mechanism of maturase-promoted group II intron splicing. *EMBO J.* 2001; 20:7259–7270. [PubMed: 11743002]

52. Noah JW, Lambowitz AM. Effects of maturase binding and Mg²⁺ concentration on group II intron RNA folding investigated by UV cross-linking. *Biochemistry*. 2003; 42:12466–12480. [PubMed: 14580192]
53. Dai L, et al. A three-dimensional model of a group II intron RNA and its interaction with the intron-encoded reverse transcriptase. *Mol Cell*. 2008; 30:472–485. [PubMed: 18424209]
54. Blocker FJ, et al. Domain structure and three-dimensional model of a group II intron-encoded reverse transcriptase. *RNA*. 2005; 11:14–28. [PubMed: 15574519]
55. Zimmerly S, et al. Phylogenetic relationships among group II intron ORFs. *Nucleic Acids Res*. 2001; 29:1238–1250. [PubMed: 11222775]
56. Sharp PA. On the origin of RNA splicing and introns. *Cell*. 1985; 42:397–400. [PubMed: 2411416]
57. Cech TR. The generality of self-splicing RNA: relationship to nuclear mRNA splicing. *Cell*. 1986; 44:207–210. [PubMed: 2417724]
58. Madhani HD, Guthrie C. A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. *Cell*. 1992; 71:803–817. [PubMed: 1423631]
59. Butcher, SE. Structural and Catalytic Roles of Metal Ions in RNA. The Royal Society of Chemistry; 2011. 8 The Spliceosome and Its Metal Ions; p. 235-251.
60. Sashital DG, et al. U2-U6 RNA folding reveals a group II intron-like domain and a four-helix junction. *Nat Struct Mol Biol*. 2004; 11:1237–1242. [PubMed: 15543154]
61. Sontheimer EJ, et al. Metal ion catalysis during group II intron self-splicing: parallels with the spliceosome. *Genes Dev*. 1999; 13:1729–1741. [PubMed: 10398685]
62. Gu SQ, et al. Genetic identification of potential RNA-binding regions in a group II intron-encoded reverse transcriptase. *RNA*. 2010; 16:732–747. [PubMed: 20179150]
63. Yan C, et al. Structure of a yeast catalytically activated spliceosome at 3.5 Å resolution. *Science*. 2016
64. Wan R, et al. Structure of a yeast catalytic step I spliceosome at 3.4 Å resolution. *Science*. 2016
65. Galej WP, et al. Cryo-EM structure of the spliceosome immediately after branching. *Nature*. 2016
66. Yan C, et al. Structure of a yeast spliceosome at 3.6-angstrom resolution. *Science*. 2015; 349:1182–1191. [PubMed: 26292707]
67. Nguyen TH, et al. CryoEM structures of two spliceosomal complexes: starter and dessert at the spliceosome feast. *Curr Opin Struct Biol*. 2016; 36:48–57. [PubMed: 26803803]
68. Yan C, et al. Structure of a yeast step II catalytically activated spliceosome. *Science*. 2016
69. Nguyen TH, et al. The architecture of the spliceosomal U4/U6.U5 tri-snRNP. *Nature*. 2015; 523:47–52. [PubMed: 26106855]
70. Rauhut R, et al. Molecular architecture of the *Saccharomyces cerevisiae* activated spliceosome. *Science*. 2016; 353:1399–1405. [PubMed: 27562955]
71. Wan R, et al. The 3.8 Å structure of the U4/U6.U5 tri-snRNP: Insights into spliceosome assembly and catalysis. *Science*. 2016; 351:466–475. [PubMed: 26743623]
72. Galej WP, et al. Crystal structure of Prp8 reveals active site cavity of the spliceosome. *Nature*. 2013; 493:638–643. [PubMed: 23354046]
73. Turner IA, et al. Dissection of Prp8 protein defines multiple interactions with crucial RNA sequences in the catalytic core of the spliceosome. *RNA*. 2006; 12:375–386. [PubMed: 16431982]

Outstanding Questions

What is the structure of the intron and how are the reactants oriented before each step of splicing? Can we visualize the pre-first step and pre-second step of splicing?

What is the sequence of conformational changes that occur between the steps of splicing?

How does the branch-site adenosine, along with the 5' exon and the 3' exon, move in and out of the active site?

How are metal ions organized within the active site during branching and during the second step of splicing?

Does the secondary structure of D6 rearrange in all introns, providing a general mechanism for switching between first- and second-step active sites?

Given that the individual steps of splicing are reversible, what is the driving force for pushing splicing to completion? Are individual steps of splicing coupled?

How does the maturase facilitate splicing by group II introns? Does it interact directly with intron D6 and active-site components or is its role allosteric?

Are both the monomeric and dimeric forms of the maturase functionally relevant to specific stages of splicing and retrotransposition?

Trends

Group II introns from different classes share a common core but have diverse peripheral structures.

In the post-catalytic state, group II intron domain 6, which contains the branch point, is stabilized by two tetraloop-receptor interactions that facilitate the transition between steps of splicing.

Group II intron domain 6 has two secondary structures in equilibrium, each of which may be involved in a specific step of splicing.

Structural studies on the maturase protein show that it is an unusual class of reverse-transcriptase that is related to spliceosomal Prp8 protein and viral RNA-dependent RNA polymerases.

The maturase is recruited to its host group II intron by specific RNA-protein interactions, where it inserts a specialized domain (the X/thumb domain) into the active-site.

Group II introns and spliceosomes share close homology in both RNA and protein components.

Figure 1.

Secondary structure and splicing pathways of group II introns. (A) Secondary structure diagrams of three group II introns from class IIC (left), IIB (middle) and IIA (right). The secondary structures were extracted from available tertiary structure models determined by crystallography or cryo-electron microscopy (PDB IDs: 3IGI for IIC, 4R0D for IIB and 5G2X for IIA). The open reading frame (ORF) is shown in D4. Long range interactions are denoted with Greek letters and are color coded. Exon binding sites (EBS) and intron binding sites (IBS) are shown in their respective locations. (B) Primary pathways for group II intron self-splicing. The branching pathway (top) uses the 2'OH group in an adenosine (branch site) in D6 as the first step nucleophile, and the hydrolysis pathway (bottom) uses an external water molecule as the first step nucleophile. Reversibility of each step is indicated by the double arrows.

Figure 2.

Tertiary structures of group II introns. (A) Tertiary structures of group II introns from IIC (left), IIB (middle) and IIA (right) (PDB IDs: 4E8K for IIC, 4R0D for IIB and 5G2X for IIA). (B) Novel interactions visualized in the crystal structure of a group IIB intron lariat (PDB ID: 4R0D). Top left: D2 (blue) and D6 (purple) interacts with two “anti-parallel” tetraloop-receptor interactions (η - η' and π - π'). Top right: D3 (orange) interacts with the base of D5 (red) through the μ - μ' interaction, which extends the canonical κ - κ' interaction between D5 (red) and D1 (grey). Bottom: exon recognition in the group IIB intron. The

5' exon recognition pairs with IBS1 (intron binding site 1, in cyan) via the EBS1 (exon binding site 1, in grey) and with IBS2 (purple) via EBS2 (grey). The 3' exon recognition involves only a single base-pair between IBS3 (blue) and EBS3 (black). (C) The β - β' interaction visualized in the cryo-EM reconstruction of a group IIA intron lariat (PDB ID: 5G2X). Only the D1 portion of the intron lariat is shown. The bottom right corner provides a close-up of the β - β' kissing loop interaction. (D) Comparison of the core structure of group II introns from different classes.

Figure 3.

The stages of group II intron splicing. (A) Individual states along the splicing pathway. During reverse splicing, which occurs during retrotransposition by the intron, DNA is targeted in-trans. Therefore, the double arrows connecting forward and reverse splicing indicate that intron RNA has the same conformation during both types of transformations, despite the fact that the chemical composition of the nucleic acid target is different. Notably, while reverse-splicing can be initiated by linear intron, it cannot complete the second step of reverse-splicing [43]. (B) Crystal structures of the group II intron active site at specific stages of splicing. The left panel shows the active site of the pre-catalytic state of the first step of splicing (PDB ID: 4FAQ), the middle panel shows the state immediately after the first step of splicing (PDB ID: 4FAR), and the right panel shows the state of a lariat intron before the first step of reverse splicing (coordinates are in the supplementary PDB file in the associated publication [32]). All phosphorus atoms are colored in orange, and the nucleophilic oxygen is colored in green. The scissile phosphate is shown in a “ball and

stick” representation. (C) Occupancy of the group II intron active site at different stages of splicing.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 4.

Conformational dynamics during group II intron splicing. (A) Isoforms of the D6 secondary structure. The regions colored in red have alternative conformations during the two steps, and arrows indicate the direction for the base pairs to slide in order to form alternative conformations. (B) Comparison of 5' exon-EBS1 conformation in isolated D1 molecules (PDB ID: 4Y1O) and in the full-length intron (PDB ID: 4FAQ). The left panel shows the full-length intron structure (grey) superimposed with the 5' exon-EBS1 from isolated D1 (green). In the swing state (green), the 3' end of the 5' exon swings 9.2 Å away, coordinated by the movement of EBS1. (C) Interactions that stabilize D6 in the lariat intron chimera (coordinates were obtained from supplementary PDB file in the associated publication [32]). D6 is docked by interactions with D1 and J2/3, and this configuration leads to the formation of second step active site. In this way, the 2'-5' branch formation is coupled to the second step of splicing. (D) 5' exon binding modulates conformation of intron active site. The left panel is the active site of *O.i.* group II intron (PDB ID: 4E8M) in the absence of 5' exon, the middle panel is the active site of *O.i.-A.v.* chimera group II intron (PDB ID: 5J02) in the presence of 5' exon, while the right panel is the active site of *O.i.-A.v.* chimera group II intron (PDB ID: 5J01) in the absence of 5' exon. In the absence of 5' exon, EBS1 is disordered, leading to alternative conformations of ζ' in D5, A72 in D1C₁ and A106 in λ that interacts with the 2'OH group of residue -2 in the 5' exon.

Figure 5.

Structures of group II intron maturases. (A) Structures and electrostatic surfaces of group II intron maturases. In the top row, left panel shows the crystal structure of the RT domain of a group IIC intron maturase (PDB ID: 5HHJ), and the right panel shows the cryo-EM structure of the RT domain of a group IIA intron maturase (LtrA, PDBID: 5G2Y). Their corresponding electrostatic surfaces are shown in the bottom row. (B) The overall architecture of a group II intron RNP solved by cryo-EM (PDB ID: 5G2Y), showing insertion of the Ti-loop into D1. (C) The protein-RNA interactions that may influence catalysis, with RNA and protein domains annotated. (D) The maturase RT domain forms a dimer in crystals and solution. Crystal structure of a maturase RT domain dimer (PDB ID: 5HHJ) is shown (left panel). This RT dimer is stabilized by an extended interface (cyan, right).

Figure 6.

Comparison of group II introns and the yeast spliceosome structures (A) Comparison of the catalytic RNA secondary structures in the active sites of the group II intron (left) and spliceosome (right). Dots indicate base-pairs while arrows indicate triple helix interactions. (B) Comparison of the RNA tertiary structure within the group II intron (PDB ID: 4FAR) and spliceosome (PDB ID: 5LJ3, C-complex) active-sites. Both structures represent the state immediately after the first step of splicing. (C) Comparison of the structure of protein components in group II introns and the spliceosome. Left panel is the crystal structure of RT domain (finger and palm) from a group IIC intron (PDB ID: 5HHJ), middle panel is the cryo-EM structure of the maturase from a group IIA intron (PDB ID: 5G2Y), and the right panel is the crystal structure of the large domain of spliceosomal protein Prp8 (PDB ID: 4I43). Domain organization of group II intron maturase and spliceosomal Prp8 are shown in the bottom left corner. RT: reverse transcriptase. IFD: insertion in finger domain. DBD: DNA binding domain. EN: endonuclease. (D) Comparing position of X/thumb domain in the active sites of the group II intron and the spliceosome. The EN domain is not shown.