Published in final edited form as:

Wiley Interdiscip Rev RNA. 2020 July ; 11(4): e1589. doi:10.1002/wrna.1589.

Retroelement Origins of Pre-mRNA Splicing

Daniel B. Haack¹, Navtej Toor^{1,*}

¹Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, CA 92093, USA

Abstract

Recent cryo-EM structures of a group II intron caught in the process of invading DNA have given new insight into the mechanisms of both splicing and retrotransposition. Conformational dynamics involving the branch-site helix domain VI are responsible for substrate exchange between the two steps of splicing. These structural rearrangements have strong parallels with the movement of the branch-site helix in the spliceosome during catalysis. This is strong evidence for the spliceosome evolving from a group II intron ancestor. We also observe other topological changes in the overall structure of the catalytic domain V that may occur in the spliceosome as well. Therefore, studying group II introns not only provides us with insight into the evolutionary origins of the spliceosome, but also may inform the design of experiments to further probe structure-function relationships in this eukaryotic splicing apparatus.

Graphical Abstract

*Correspondence should be addressed to N.T. (ntoor@ucsd.edu).

1. INTRODUCTION

The origin of introns and the eukaryotic spliceosome has been a long-standing question in the field of molecular evolution. Based on mechanistic similarities, Philip Sharp¹ and Thomas Cech² first proposed that the spliceosome evolved from a group II intron ancestor, which originated within bacteria billions of years ago. Both systems employ a lariat intermediate during RNA splicing and share similar secondary structures of the catalytic RNA components. However, group II introns have an additional functionality that allows them to act as selfish retroelements by reinserting themselves into the genome through a copy and paste mechanism known as retrotransposition^{3,4}. The implications of this dual function of group II introns are far reaching and may help explain how spliceosomal introns came to comprise 25% of the human genome, greatly increasing proteomic diversity. This process of intron expansion likely led to many of the unique characteristics seen in eukaryotes^{5,6}, including alternative splicing pathways and spurred the evolution and speciation of multicellular organisms⁶. The potential impact that RNA splicing has had on the diversification of life on Earth is striking. However, until relatively recently there was little hard biochemical evidence for this hypothesis. In this regard, structural studies of the group II intron and the spliceosome over the past decade have now converged to provide strong support for a group II intron origin of eukaryotic RNA splicing.

Group II introns are self-splicing catalytic RNAs that are 400–800 nucleotides in size⁷. They catalyze a two-step splicing mechanism in which the intervening intron excises itself followed by ligation of the neighboring exons. Group II introns have a conserved RNA secondary structure that is comprised of six domains arranged around a central pinwheel⁸ (Figure 1). Domain I provides the sequence recognition motifs that bind the flanking exons through Watson-Crick pairing^{9,10}. These are known as Exon Binding Sequences (EBS) and serve to delineate the splice sites (SS) by pairing with the Intron Binding Sequence (IBS) in the exons. Domain V forms the active site of the ribozyme through the coordination of two catalytic magnesium ions (M1 and M2)¹¹. DV contains two highly conserved motifs known as the two-nucleotide (2-nt) bulge and catalytic triad that are responsible for binding M1 and M2. Domain VI (DVI) contains the bulged adenosine that serves as the nucleophile for the first step of splicing. Domain IV (DIV) in some group II introns encodes an open reading frame (ORF) for a protein known as the maturase that promotes splicing upon binding to the intron RNA^{12–14}. The self-splicing of group II introns consists of two sequential transesterification reactions⁷. The first step of forward splicing begins after the maturase protein binds to the intron RNA. Once bound, the newly formed RNA protein complex (RNP) initiates splicing by activating the 2'-OH of the bulged adenosine from DVI for nucleophilic attack on the 5' splice site (Figure 2). This results in the formation of the hallmark 2'-5' phosphodiester bond that circularizes the intron to form the lariat RNA. In the second step, the 3'-OH of the 5' exon attacks the 3' splice site to form ligated exons and free lariat RNA.

The splicing reaction is also completely reversible¹⁵ allowing the intron to insert into RNA and DNA targets. This type of reverse splicing is utilized by group II introns to invade dsDNA and replicate using a copy-and-paste mechanism known as retrotransposition. In order to form a group II intron complex competent for reverse splicing, it must first undergo the forward splicing reaction yielding a free intron lariat/maturase complex and ligated exons fit for translation (Figure 2). Once the intron has fully reverse spliced into the target DNA⁴, the maturase utilizes several protein domains to convert the intron RNA into an integrated DNA copy. The maturase consists of both endonuclease and reverse transcriptase (RT) domains that cut the antisense DNA strand and initiate cDNA synthesis³, respectively. Therefore, group II introns are also considered to be retroelements as they replicate via an RNA intermediate. The conservation of this mechanism as well as homology between RT domains has forged a hypothesis that group II intron are also ancestral to the LINE elements^{16,17} that comprise ~45% of mammalian genomes. If true, group II introns have had a large impact upon the evolution of eukaryotic genomes through both retrotransposition and splicing.

Recent cryo-EM structures of the spliceosome^{18,19} reinforce the hypothesis that the eukaryotic splicing machinery evolved from a group II intron ancestor. Comparison with crystal structures of the group II intron^{11,20} revealed structural homology in the active site architecture (Figure 3). DV of the group II intron forms a catalytic triplex that coordinates the catalytic metal ions in an almost identical manner as the U2/U6 snRNA pairing in the spliceosome. Combining the fact that both systems form a branched 2'-5' lariat intron has cultivated strong support for a shared ancestry. In this review, we provide further support for

this evolutionary connection with homology extending to the level of conformational dynamics during catalysis.

2. CRYO-EM STRUCTURES OF A GROUP II INTRON

Previous crystal structures of group II introns were solved in the absence of a maturase. These structures provided limited information on the conformational rearrangements during the two steps of RNA splicing. Both steps of splicing are catalyzed in a single active site; however, it was unclear how the 5' and 3' splice site substrates were exchanged during the transition between the two transesterification reactions. Given the fact that the crystal lattice is a confining environment, it is likely that the full range of group II intron dynamics were never observed.

In contrast, cryo-EM allows for the capture of different conformational states as the sample is vitrified in its native state. Therefore, this allows the RNA to sample conformational space that would normally be inhibited using a crystallographic approach. To determine the mechanism of substrate exchange during splicing, we have solved two structures of a group II intron with its cognate maturase in the process of invading a double-stranded DNA (dsDNA) substrate²¹. We were able to visualize conformational rearrangements that were previously not captured using crystallography. The two structures represent the group II intron immediately before (pre-1r) and after (pre-2r) the first step of reverse splicing into DNA.

A comparison of the group II intron structures from x-ray crystallography and cryo-EM reveals that these two techniques provide complementary insight into catalysis. For example, crystal structures of the group II intron provide more detailed visualization of metal ions and their coordination within the active site DV. Both catalytic magnesium ions and a conserved monovalent ion (either Na⁺ or K⁺) are captured within DV in multiple crystal structures^{11,20,22}. Many metal ions are also seen in the crystal structures playing a purely structural role. In both cryo-EM structures, we see only one of the catalytic metal ions (M1) within the active site DV and did not detect the monovalent ion even though Na⁺ is present in the buffer. This is likely due to the fact that it is difficult to detect metal ions using cryo-EM for unknown reasons. An equivalent crystal structure at 3 Å of a protein or RNA yields many metal ions in both the active site and surrounding regions. Therefore, crystallography is likely a more sensitive method for the detection of metal ion binding sites in macromolecules.

An additional benefit of crystallography are heavy metal soaks, which can be used in combination with anomalous dispersion to highlight metal ions within RNA structures. In crystal structures of the group IIB and IIC introns, core metal ions were highlighted using Yb³⁺ ions that exhibit a strong anomalous signal^{11,20} when using x-rays of a specific wavelength. Yb³⁺ serves as an analogue of Mg²⁺ and exhibits an identical octahedral coordination geometry. The ability to identify specific metal ions does not currently exist when using cryo-EM for structure determination, thus leaving an important place in structural biology for crystallography.

3. DYNAMICS WITHIN THE ACTIVE SITE OF THE GROUP II INTRON

Previous crystal structures of the group II intron have provided evidence that the 2-nt bulge and catalytic triad of DV form a dynamic active site that engages in local nucleotide dynamics that effect the coordination of the catalytic metals M1 and M2. Marcia and Pyle (2012) found that one of the bulge nucleotides of DV undergoes a base flipping event that affects the coordination of the catalytic metal ions and nucleophile placement in the first step of splicing²². Chan et al. (2018) found additional dynamics in the catalytic triplex within DV. In this work, the base triples in DV were found to exist in different configurations before and after the second step of splicing²³. These base triples were hypothesized to assist in properly positioning the of 3' splice site during the second transesterification reaction. The insights gleaned from crystallography are important parts of the dynamic puzzle of group II introns. However, the constraints imposed on dynamics by the process of crystallization meant there was a strong possibility that allosteric events that modulated catalysis remained unseen.

In the cryo-EM structures, we see larger-scale dynamics involving the overall structure of DV. In particular, there is a lengthwise expansion of DV in the transition from the pre-1r to the pre-2r states that is equivalent to a stretching of more than one base-pair register²¹. This expansion of DV is also associated with an 8 Å movement of metal M1. These conformational changes were completely unexpected as DV was not seen to engage in such large-scale dynamics in all previous crystal structures of group II introns from multiple species. We hypothesize that these conformational changes originate from cleavage of the DNA substrate in the active site. We observe an expansion of the metal-binding pocket for M1 in the pre-2r state. This expansion is likely spurred after the cleavage of the scissile phosphate results in the loss of a coordinating oxygen ligand for M1 and stereochemical inversion of another. Concurrent with DV expansion, the newly characterized ψ - ψ' and ϕ - ϕ' tertiary interactions disengage (Figure 4). In particular, the ψ - ψ' interaction is in close proximity to the two-nt bulge of DV and M1; therefore, this contact is more greatly affected by the dynamics of the metal-binding pocket of DV. We hypothesize that ψ - ψ' is the first stage of an allosteric network that relays the status of the active site to the rest of the RNA structure during splicing. In other words, ψ - ψ' serves as a sensor to monitor the progress of catalysis to initiate further downstream conformational rearrangements.

U6 snRNA also contains a catalytic triplex²⁴ that comprises the active site of the spliceosome that exhibits a high degree of homology to the DV catalytic triplex within the group II intron. Genetic evidence also supports the existence of conformational dynamics within the catalytic triplex²⁴ and analogous U2/U6 snRNA pairing²⁵. Eysmont et al. (2019) found that destabilization of the lower region of U6 ISL is required for proper splicing activity. This is consistent with our observation of rearrangement of the catalytic triplex during group II intron splicing²³. It is also possible that the expansion of DV seen in the cryo-EM structure of the group II intron may represent the stable intermediates of this dynamic domain during splicing. The U6 snRNA has so far not been seen to engage in similar conformational dynamics in all currently available structures of the spliceosome. Given the similarity in mechanism and branch-site dynamics (see below), we strongly suspect that these conformational rearrangements, analogous to those seen for DV, also

occur during catalysis in the spliceosome. However, visualizing these changes with cryo-EM may require mutagenesis of spliceosomal protein co-factors and/or snRNAs. Numerous mutations are already known that can trap the spliceosome at different stages of assembly/catalysis. This may require the screening of additional constructs that would be appropriate for cryo-EM structure determination to detect these dynamics in the U2/U6 core.

4. CONSERVATION OF BRANCH-SITE HELIX DYNAMICS IN THE GROUP II INTRON AND SPLICEOSOME

Domain VI has long been hypothesized to engage in conformational dynamics in the transition between the first and second steps of splicing²⁶. However, the precise nature of molecular dynamics involving DVI had evaded detection through x-ray crystallography. In the cryo-EM structures of the *Tet* group II intron, we found that DVI engaged in a large-scale 90° swinging movement in the process of reverse splicing into DNA (Figure 5). Specifically, the maturase forms an RNA-protein contact between the RT thumb domain and DVI in the pre-2r state that brings the lariat bond into the active site in preparation for the second step of reverse splicing. This contact, termed matX-DVI, is also essential for forward splicing as mutagenesis of the protein component of this interaction results in a loss of splicing activity. Therefore, this suggests that the matX-DVI interaction is essential for both the forward and reverse reactions and the associated movement of DVI effectively exchanges splice site substrates within the active site.

This swinging of DVI is likely to be directly facilitated by the disengagement of several tertiary interaction (π - π' , η - η' , and ϕ - ϕ') which helps to anchor DVI prior to this motion. As a result, we hypothesize that ψ - ψ' first disengages that leads to a chain reaction that results in π - π' , η - η' , and ϕ - ϕ' turning off allowing DVI to swing into the pre-2r state (Figure 4). This model directly relates the status of the active site to the position of DVI with ψ - ψ' being poised to sense the expansion of the metal-binding pocket upon catalysis to trigger this cascade of allosteric events emanating from the catalytic core.

A comparison with the branch-site movement in the spliceosome reveals striking parallels between the two systems. The branch-site helix also undergoes a ~90° swinging in the transition between the C and P states of the spliceosome^{18,27,28} (Figure 5), which are the analogous stages of catalysis to the conformations observed in the cryo-EM structures of the group II intron²¹. Therefore, the group II intron and the spliceosome use an identical mechanism to shuffle between the 5' and 3' splice site substrates. This further cements the hypothesis that the dynamics associated with RNA splicing first evolved in a group II intron ancestor.

5. OUTSTANDING QUESTIONS

These cryo-EM structures only represent the events surrounding the first step of reverse splicing and the mechanism of substrate exchange. Further work remains to gain insight into the second step of reverse splicing. It is also still unknown how the bottom strand of the DNA is cleaved by the endonuclease domain of the maturase. Density for the endonuclease domain was absent in our cryo-EM structures. Lastly, we have not captured the mechanism

of RT engagement on the bottom primer strand for cDNA synthesis. We hypothesize that additional conformational rearrangements are required to engage the endonuclease and RT domains to begin the cDNA integration process. It is likely this will require the collection of a much larger dataset to gain more particles in lowly populated 3D classes.

6. SPLICEOSOME EVOLVED FROM A RETROELEMENT

The cumulative evidence strongly suggests that the spliceosome evolved from a group II intron retroelement. Group II introns likely first evolved in bacteria billions of years ago. In prokaryotes, ~99% of group II introns are located in intergenic regions between genes and function more as retroelements rather than introns²⁹. During the endosymbiont event, bacteria were engulfed by a primitive archaeal host and evolved into the mitochondria and chloroplasts (Figure 6). In lower eukaryotes such as yeast and fungi, a large proportion of group II introns are located within conserved housekeeping genes such as cytochrome oxidase²⁹. Therefore, in lower eukaryotes, group II introns have begun the process of colonizing genes and are starting to function more as introns and less like retroelements³⁰. In plants and other higher eukaryotes, group II introns are almost entirely found within genes³¹. In addition, some group II introns in these organisms are found to be fragmented into multiple RNAs³² as seen for the snRNAs of the spliceosome.

This provides strong support for the “introns late” hypothesis as opposed to “introns early”, which has been a longstanding question in evolutionary biology. The introns late hypothesis postulates that spliceosomal introns first evolved in eukaryotes and continued to insert into new locations in the genome throughout evolutionary history³³. The fact that there is now extensive evidence for the evolutionary connection between group II introns and the spliceosome strongly supports the idea that eukaryotic intronogenesis began with the endosymbiont event that resulted in the formation of mitochondria and chloroplasts^{5,6}. The group II introns found in the engulfed bacteria formed part of the organellar genomes. These group II introns then migrated to the nucleus through retrotransposition to eventually form spliceosomal introns. The invasion of eukaryotic genomes by group II/spliceosomal introns likely had dramatic effects of the genomic stability of early eukaryotes and is even thought to have led to the formation of the nuclear membrane as a defense mechanism to retain mRNA integrity⁵. The nuclear membrane likely evolved to spatially separate RNA splicing in the nucleus from translation by the ribosome in the cytosol. This would prevent the translation of intron-containing pre-mRNAs, which would have deleterious effects upon the cell. As a result, pre-mRNAs are first spliced in the nucleus followed by export to the cytoplasm for translation into proteins. Additional quality control mechanism also likely evolved out of intron invasion such as nonsense mediated decay and protein ubiquitination⁶. The implications of these evolutionary scenarios involving group II introns are immense. It is quite likely that group II introns are responsible for the genesis of many of the unique characteristics seen in modern eukaryote. Alternative splicing in particular is one of the defining features of eukaryotes and allows for the production of multiple protein isoforms from a single gene. Therefore, group II introns not only led to the formation of eukaryotes, but also additional speciation and organismal complexity.

The architecture of the spliceosome echoes this evolutionary relationship. The core spliceosomal protein prp8 has structural homology to the group II intron maturase with RT and endonuclease domains³⁴. In addition, the U6 snRNA forms a catalytic triplex as seen in the group II introns^{18,19} with similar requirements for conformational flexibility to promote splicing. Therefore, both the ribozyme and protein components of the spliceosome are derived from group II introns. With this evolutionary pedigree, the spliceosome may have retained some type of retroelement activity. It is also possible that the spliceosome may interact with LINE retroelements for intron dispersal. In this hypothetical model, the spliceosome could catalyze reverse splicing reactions of an intron into a non-cognate mRNA. A LINE element RT could then convert this into a cDNA that would undergo recombination into the host genome. This would result in the insertion of a new copy of an intron. Group II introns containing endonuclease-deficient maturase proteins have also been shown to efficiently promote retrotransposition through DNA replication forks³⁵. This may be a possible mechanism for the spliceosome to reverse splice directly into DNA. In this model, the LINE element RT would synthesize the cDNA copy of the integrated RNA. A possible third scenario could be the use of a spliceosomal helicase to unwind a double-stranded DNA substrate to prepare for reverse splicing of the intron. It is likely that one of these mechanisms is responsible for intron proliferation in eukaryotes to such an extent that human genes each have an average of ~8 introns.

7. CONCLUSION

Pre-mRNA splicing in eukaryotes likely originated from a group II intron ancestor during eukaryogenesis^{5,6,36}. This also provides an evolutionary link with the process of retrotransposition. Group II introns are postulated to be ancestral to the non-LTR retroelements, such as the mammalian LINE elements. LINE elements and their derived sequences are thought to comprise ~45% of the human genome with introns contributing another 25%. Therefore, ~70% of the human genome is comprised of genetic elements that have evolved from group II introns. It is likely that group II introns led to the diversification of eukaryotic life into the many forms and species that we see today. In addition, the core of the spliceosome consists of the catalytic U6 snRNA and the RT homolog prp8 protein, which is analogous to the group II intron RNA in complex with its maturase. Therefore, the spliceosome can be thought of a group II intron retroelement core that picked up additional protein co-factors during the course of evolution to form the modern-day eukaryotic splicing apparatus.

Funding Information

This work was supported by NIH grant 1R01GM123275 awarded to N.T.

References

1. Sharp PA "Five easy pieces". *Science* 254, 663 (1991). [PubMed: 1948046]
2. Cech TR The generality of self-splicing RNA: relationship to nuclear mRNA splicing. *Cell* 44, 207–10 (1986). [PubMed: 2417724]
3. Zimmerly S, Guo H, Perlman PS & Lambowitz AM Group II intron mobility occurs by target DNA-primed reverse transcription. *Cell* 82, 545–54 (1995). [PubMed: 7664334]

4. Zimmerly S et al. A group II intron RNA is a catalytic component of a DNA endonuclease involved in intron mobility. *Cell* 83, 529–38 (1995). [PubMed: 7585955]
5. Martin W & Koonin EV Introns and the origin of nucleus-cytosol compartmentalization. *Nature* 440, 41–5 (2006). [PubMed: 16511485]
6. Koonin EV The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate? *Biol Direct* 1, 22 (2006). [PubMed: 16907971]
7. Lambowitz AM & Zimmerly S Group II introns: mobile ribozymes that invade DNA. *Cold Spring Harb Perspect Biol* 3, a003616 (2011). [PubMed: 20463000]
8. Michel F, Umesono K & Ozeki H Comparative and functional anatomy of group II catalytic introns--a review. *Gene* 82, 5–30 (1989). [PubMed: 2684776]
9. Jacquier A & Michel F Multiple exon-binding sites in class II self-splicing introns. *Cell* 50, 17–29 (1987). [PubMed: 3297351]
10. 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). [PubMed: 10990464]
11. Toor N, Keating KS, Taylor SD & Pyle AM Crystal structure of a self-spliced group II intron. *Science* 320, 77–82 (2008). [PubMed: 18388288]
12. Watanabe K & Lambowitz AM High-affinity binding site for a group II intron-encoded reverse transcriptase/maturase within a stem-loop structure in the intron RNA. *RNA* 10, 1433–43 (2004). [PubMed: 15273321]
13. Matsuura M, Noah JW & Lambowitz AM Mechanism of maturase-promoted group II intron splicing. *EMBO J* 20, 7259–70 (2001). [PubMed: 11743002]
14. Singh RN, Saldanha RJ, D'Souza LM & Lambowitz AM Binding of a group II intron-encoded reverse transcriptase/maturase to its high affinity intron RNA binding site involves sequence-specific recognition and autoregulates translation. *J Mol Biol* 318, 287–303 (2002). [PubMed: 12051838]
15. Roitzsch M, Fedorova O & Pyle AM The 2'-OH group at the group II intron terminus acts as a proton shuttle. *Nat Chem Biol* 6, 218–224 (2010). [PubMed: 20118939]
16. Xiong Y & Eickbush TH Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J* 9, 3353–62 (1990). [PubMed: 1698615]
17. Malik HS, Burke WD & Eickbush TH The age and evolution of non-LTR retrotransposable elements. *Mol Biol Evol* 16, 793–805 (1999). [PubMed: 10368957]
18. Galej WP et al. Cryo-EM structure of the spliceosome immediately after branching. *Nature* 537, 197–201 (2016). [PubMed: 27459055]
19. Hang J, Wan R, Yan C & Shi Y Structural basis of pre-mRNA splicing. *Science* 349, 1191–8 (2015). [PubMed: 26292705]
20. Robart AR, Chan RT, Peters JK, Rajashankar KR & Toor N Crystal structure of a eukaryotic group II intron lariat. *Nature* 514, 193–7 (2014). [PubMed: 25252982]
21. Haack DB et al. Cryo-EM Structures of a Group II Intron Reverse Splicing into DNA. *Cell* 178, 612–623 (2019). [PubMed: 31348888]
22. Marcia M & Pyle AM Visualizing group II intron catalysis through the stages of splicing. *Cell* 151, 497–507 (2012). [PubMed: 23101623]
23. Chan RT et al. Structural basis for the second step of group II intron splicing. *Nat Commun* 9, 4676 (2018). [PubMed: 30410046]
24. Fica SM, Mefford MA, Piccirilli JA & Staley JP Evidence for a group II intron-like catalytic triplex in the spliceosome. *Nat Struct Mol Biol* 21, 464–71 (2014). [PubMed: 24747940]
25. Eysmont K, Matylla-Kuliska K, Jaskulska A, Magnus M & Konarska MM Rearrangements within the U6 snRNA Core during the Transition between the Two Catalytic Steps of Splicing. *Mol Cell* 75, 538–548.e3 (2019). [PubMed: 31229405]
26. Chanfreau G & Jacquier A An RNA conformational change between the two chemical steps of group II self-splicing. *EMBO J* 15, 3466–76 (1996). [PubMed: 8670849]
27. Fica SM, Oubridge C, Wilkinson ME, Newman AJ & Nagai K A human postcatalytic spliceosome structure reveals essential roles of metazoan factors for exon ligation. *Science* 363, 710–714 (2019). [PubMed: 30705154]

28. Fica SM et al. Structure of a spliceosome remodelled for exon ligation. *Nature* 542, 377–380 (2017). [PubMed: 28076345]
29. Dai L & Zimmerly S Compilation and analysis of group II intron insertions in bacterial genomes: evidence for retroelement behavior. *Nucleic Acids Res* 30, 1091–102 (2002). [PubMed: 11861899]
30. Toor N, Hausner G & Zimmerly S Coevolution of group II intron RNA structures with their intron-encoded reverse transcriptases. *RNA* 7, 1142–52 (2001). [PubMed: 11497432]
31. Hausner G et al. Origin and evolution of the chloroplast trnK (matK) intron: a model for evolution of group II intron RNA structures. *Mol Biol Evol* 23, 380–91 (2006). [PubMed: 16267141]
32. Knoop V, Altwasser M & Brennicke A A tripartite group II intron in mitochondria of an angiosperm plant. *Mol Gen Genet* 255, 269–76 (1997). [PubMed: 9268017]
33. Cavalier-Smith T Intron phylogeny: a new hypothesis. *Trends Genet* 7, 145–8 (1991). [PubMed: 2068786]
34. Galej WP, Oubridge C, Newman AJ & Nagai K Crystal structure of Prp8 reveals active site cavity of the spliceosome. *Nature* 493, 638–43 (2013). [PubMed: 23354046]
35. Martínez-Abarca F, Barrientos-Durán A, Fernández-López M & Toro N The RmInt1 group II intron has two different retrohoming pathways for mobility using predominantly the nascent lagging strand at DNA replication forks for priming. *Nucleic Acids Res* 32, 2880–8 (2004). [PubMed: 15155857]
36. Madhani HD The frustrated gene: origins of eukaryotic gene expression. *Cell* 155, 744–9 (2013). [PubMed: 24209615]

Figure 1.

Representative secondary structure of a group II intron. Group II introns have a conserved secondary structure with six domains labeled with Roman numerals I–VI. The exon binding sequences (EBS 1–3) are used to recognize the 5' and 3' splice sites (SS) by Watson-Crick pairing to the intron binding sequences (IBS 1–3). Relevant tertiary interactions are labeled with Greek symbols.

Figure 2.

Replication cycle of a group II intron retroelement. Group II introns are first transcribed from a DNA sequence, after which the open reading frame (ORF) for the maturase protein is translated. The resulting maturase protein then binds to the intron RNA and promotes forward splicing. The products of this process are an excised lariat/maturase complex as well as ligated exons fit for translation. The intron lariat complex is then free to diffuse and bind to dsDNA that contains a sequence that is partially complementary to the EBS found within the RNA. Once bound, the RNA undergoes reverse splicing, which covalently attaches the RNA on both ends to the DNA. The maturase protein then uses its endonuclease (En) domain to cleave the bottom strand of DNA and its reverse transcriptase (RT) domain to synthesize a cDNA copy of the intron RNA. Host repair and recombination pathways complete the insertion of the group II intron copy.

Figure 3. Conservation of sequence and structure between group II introns and the spliceosome. A) Left. A secondary structure of the core components is shown for the *Te4h* group II intron from the cyanobacterium *Thermosynechococcus elongatus*. The main features that make up the catalytic triplex are shown. Right. The catalytic triplex from PDB 6ME0 is shown in detail. Hydrogen bonds are represented with yellow dashes and the catalytic metal ions (M1 and M2) are shown as orange spheres. B) The core features of the human spliceosome are conserved when compared to the group II intron. U2 and U6 pair to form the catalytic triplex (PDB 6QDV) with an almost identical tertiary structure.

Figure 4.

ψ and ϕ interactions facilitate DVI dynamics. In the pre-1r state, four tertiary interactions are engaged (ψ - ψ' , ϕ - ϕ' , π - π' , and η - η'). To facilitate the transition to the pre-2r state, all four of these interactions must disengage. This process likely initiates with the disengagement of ψ - ψ' , which causes structural perturbations that propagate down DV and results in the disruption of ϕ - ϕ' , π - π' , and η - η' . Once free, DVI can sample conformational space and is eventually captured by the matX-DVI and ι - ι' interactions.

Figure 5.

The dynamics of RNA splicing are conserved. In both the group II intron (PDB 6ME0 and 6MEC) and the spliceosome (PDB 6QDV and 5LJ5), the mechanism of substrate exchange is conserved. The helix containing the bulged adenosine (DVI in group II introns and the branch helix in the spliceosome) undergo a large 90° swinging motion. This motion helps to remove the lariat bond from the active site during forward splicing, allowing the complex to proceed with the second transesterification reaction.

Figure 6.

The retroelement origin of the spliceosome. During the endosymbiont event, a eubacterium is hypothesized to have been phagocytosed by an archaeobacterium. Once engulfed, the archaeobacterium began to digest the eubacteria, leading to a partial degradation of the eubacterial membrane. This allowed group II introns to escape from the eubacterium and invade the genome of the archaeobacterium. The sudden addition of group II introns within coding regions of the archaeobacterial genome led to a selective pressure to decouple transcription from translation. The evolutionary solution to this problem was the formation of the nuclear membrane. Meanwhile, group II introns began to degenerate becoming less

active as retroelements. This degeneration continued as the group II intron became fragmented, eventually forming the modern-day spliceosome.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript