Published in final edited form as:

Curr Opin Struct Biol. 2017 December ; 47: 30–39. doi:10.1016/j.sbi.2017.05.002.

The group II intron maturase: a reverse transcriptase and splicing factor go hand in hand

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

The splicing of group II introns *in vivo* requires the assistance of a multifunctional intron encoded protein (IEP, or maturase). Each IEP is also a reverse-transcriptase enzyme that enables group II introns to behave as mobile genetic elements. During splicing or retro-transposition, each group II intron forms a tight, specific complex with its own encoded IEP, resulting in a highly reactive holoenzyme. This review focuses on the structural basis for IEP function, as revealed by recent crystal structures of an IEP reverse transcriptase domain and cryo-EM structures of an IEP-intron complex. These structures explain how the same IEP scaffold is utilized for intron recognition, splicing and reverse transcription, while providing a physical basis for understanding the evolutionary transformation of the IEP into the eukaryotic splicing factor Prp8.

Introduction

Group II introns are self-splicing ribozymes that catalyze their own excision from precursor RNA molecules and the ligation of flanking exons (Figure 1a) [1–6]. This complex set of reactions proceeds relatively slowly *in vitro*, and it has long been known that group II intron splicing is facilitated by a unique protein cofactor that is encoded within the open reading frame of group II intron Domain 4 (D4), and that this cofactor is usually required *in vivo* [7–10]. This intron-encoded protein (IEP), also known as a maturase, is a multi-domain protein that contains an N-terminal reverse transcriptase (RT) domain followed by a C-terminal X domain (Figure 1b) [8,10–12]. Some IEPs also contain a DNA-binding domain (DBD) and endonuclease (EN) domain at the C-terminus (Figure 1b) [10,12].

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

Conflict of interest

The authors declare no competing financial interests.

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.

IEPs are multi-functional proteins that participate directly in RNA recognition, splicing, reverse transcription and retro-transposition [7–10,13,14]. Each IEP recognizes its parent group II intron by forming strong and highly-specific interactions with a specific RNA segment that is located within intron D4 [15,16]. This protein-RNA interaction orients the IEP in a spatial position that facilitates splicing. After splicing, the IEP remains tightly bound to the liberated intron and the resultant IEP-intron ribonucleoprotein (RNP) complex behaves as a mobile element that can copy and paste the intron sequence into new genomic sites [7,13,17–20]. The mobility of the IEP-intron RNP depends on reverse transcriptase activity of the IEP, which is carried out by the RT and X domains [13], and it is facilitated by the DBD and EN domain in certain classes of IEPs [9,12]. Group II intron splicing and retrotransposition are likely to have played a central role in the transition from bacteria to eukaryotes because novel processes that distinguish eukaryotic organisms (such as RNA splicing and telomeric stabilization of chromosomes) involve components that derive from ancestral group II introns [3,9,10,21,22].

Despite the essential role of IEPs in group II intron biology and evolution, their reaction mechanisms have remained unclear for more than two decades due to a lack of high resolution structural information. However, two complementary studies [23,24] changed this situation dramatically by providing the first structural information on group II intron IEPs and their partners. One study provided ultra-high resolution crystal structures of RT domains from the IEPs of the eubacteria *Roseberia intestinalis* (*R.i.*, solved to 1.2 Å resolution) and *Eubacterium rectale* (*E.r.*, solved to 2.1 Å) (Figure 1b) [24]. The second study resulted in the 3.8 Å cryo-EM reconstruction of a group-II-intron-IEP RNP complex from *Lactococcus lactis* (*L.l.*) (Figure 1b) [23]. Here we will focus on the mechanistic and evolutionary implications of these IEP structures and we will relate them to the behavior of group II introns as splicing machines and retroelements.

The IEP has a polymerase scaffold that is expanded by insertion motifs

Phylogenetic analysis suggests that group II intron IEPs belong to a family of non-LTR reverse transcriptases that include telomerase, and the insect R2 and human L1 retrotransposons [10,25,26]. The presence of finger and palm subdomains (which comprise the RT domain) in the IEP has been suggested by the presence of seven motifs (RT1-7) that are conserved in all reverse transcriptases from both LTR and non-LTR families (Figure 2a) [11,13,27]. The X domain that is C-terminal to the RT domain has no sequence homology to any existing protein domains. Nevertheless, based on its location relative to the putative finger and palm domains, the X domain has been proposed to function as a polymerase thumb [8,11–13]. Indeed, the cryo-EM structure of the group II intron IEP [23] shows an overall right-hand configuration that is typical of a polymerase, and the three α -helices formed by the X domain protrude out of the palm in a manner similar to the thumb subdomain of HIV reverse transcriptase and telomerase enzymes (Figure 2b). However, the α -helices of these different thumb domains differ in lengths and sequence (Figure 2b).

IEPs are members of a specialized polymerase family that includes non-LTR retrotransposons, which are an important class of mobile elements found in humans and other animals [28]. Members of this family contain an N-terminal extension that is generally

referred to as RT0 (Figure 1b, Figure 2a and Figure 3a) [12,25]. The RT0 region ranges from ~60–200 residues in length, depending on the family of IEP. In the high-resolution crystal structures of the IEP RT domains [24], the last 56 amino acids in RT0 form two sets of anti-parallel α -helices that dock onto the outer surface of the finger subdomain (Figure 1a and 2a) [24]. Based on genetic analysis of *L.l.* IEP [29] and mutational studies of *B.m.* R2 [30], the RT0 region is responsible for specific RNA recognition. In fact, the surface formed by the last ~56 amino acids in the RT0 region is highly positively charged (Figure 4b) [24], and it forms extensive interactions with a region in intron D4 (D4A) in the cryo-EM structure of the IEP-intron complex (Figure 3a) [23]. The amino acids at the very N-terminus of RT0 are disordered in both the crystal structures [24] and the cryo-EM structure [23]. Because these residues are enriched in positive charges, the disordered region is likely to interact with RNA, thereby improving the overall stability of the IEP-intron complex.

Another distinguishing feature of group II intron IEPs are the specialized motifs that are inserted between canonical regions of the RT domain (Figure 2a) [9–12]. For example, there are long linkers between RT2 and RT3 (linker 2a), between RT3 and RT4 (linker 3a), and sometimes between regions RT4 and RT5 (linker 4a, Figure 2a) [11,12]. Crystal structures of the IEP RT domains [24] show that insertion 2a forms an α -helix that buttresses the active-site β -hairpin containing the catalytic YADD motif (Figure 2c). Although direct hydrogen bonding is not involved, insertion 2a appears to stabilize the active-site β -hairpin by hydrophobic and van der Waals interactions (Figure 2c). While a similar configuration is also observed in telomerase [31,32], the HIV reverse transcriptase deviates [33,34], as the same region is occupied by an extended loop that extends away from the β -hairpin (Figure 2c). Although the precise role of the α -helix in motif 2a has not been tested experimentally, it may influence catalysis by stabilizing the active site.

Insertion 3a has two structural roles. On the one hand, it contributes an α -helix to the finger domain (IFD) (Figure 2d) that is also observed in telomerase [31,32] and HCV RNA polymerase [24,35]. In addition, insertion 3a contains a loop structure (the α -loop) that encloses the reverse transcriptase active site, as observed in crystal structures of the *E.r.* and *R.i.* IEP RT domains (Figure 2d) [24]. This α -loop is unique to group II intron IEPs, and preliminary data suggests it is involved in reverse transcriptase processivity (Zhao & Pyle, unpublished data). However, the conformational state of insertion 3a differs within the cryo-EM structure of *L.l.* IEP, in which the 3a region forms a short β -hairpin that swings away from the active-site by ~120° (Figure 2d) [23]. The two states revealed by available structures suggest that insertion 3a is flexible and capable of opening and closing in response to the presence of RNA templates during reverse transcription.

Insertion 4a is an auxiliary RNA-binding element that is not present in all IEPs. Based on sequence alignment, insertion 4a is tiny in *E.r.* and *R.i.* IEPs, but it is about 60 residues long in *L.l.*, although only the first 14 residues are visible in the cryo-EM structure (Figure 2a and 2e) [23]. These 14 residues extend an α -helix within the palm subdomain, enabling the IEP to contact RNA stem d(iii)a in intron domain 1 (D1) (Figure 2e) [23], and suggesting its function as an auxiliary RNA-binding element. Intriguingly, the *E.r.* and *R.i.* group II introns lack stem d(iii)a within intron D1, and correspondingly, their IEPs lack its interaction

partner insertion 4a. This observation provides a powerful example of the co-evolution between the group II introns and their encoded IEPs [36].

High-resolution crystal structures of the IEP RT domain [24] provide a starting point for understanding the structure and function of RT domains in non-LTR retro-transposons such as the L1 element [37], which are a major cause of genomic instability and sporadic cancer in humans [38,39]. Based on secondary structure prediction of L1 RT and sequence alignment with the IEP, some of the α -helices in RT0 of the L1 RT are interrupted by proline residues, although both insertions 2a and 3a could form structures similar to their counterparts in the IEP (Figure 2f). Importantly, the similarity of insertion 3a suggests that the reverse transcriptase processivity of L1 RT may also be augmented by an α -loop as in the group II intron IEP, further suggesting that L1 contains a reactive, high-processivity RT. Additionally, as in the case of *R.i.* and *E.r.* IEP, insertion 4a in L1 RT is so short that it is insignificant, consistent with its proposed function as an auxiliary RNA interaction motif (Figure 2f). In fact, a homology model of the L1 RT domain, predicted by I-TASSER [40] with restraints from the 1.2 Å structure of IEP RT domain [24], suggests that the L1 RT domain adopts a structure that is very similar to that of the IEP (Figure 2f).

The IEP is a splicing factor

In addition to the reverse transcriptase activity that is provided by its polymerase domains, the IEP is also a splicing factor that can substantially stimulate the splicing efficiency of host group II introns [18]. Interestingly, its role in intron splicing requires the complete polymerase scaffold, including both the RT domain and the X domain [14]. To identify the residues responsible for IEP-assisted splicing, it is important to distinguish residues that affect IEP recruitment and positioning from residues that directly modulate group II intron catalysis.

The recruitment and positioning of the IEP within its host intron requires both the RT domain and the X domain. As mentioned earlier, the RT domain associates strongly and specifically with group II intron D4A through the RT0 region, and this RT0-D4A interaction is the primary interface for stabilizing the RNA-protein complex (Figure 3a) [15,16,23,24,29]. For a subgroup of holoenzymes such as *L.l.*, the 4a insertion motif extends to D1 helix d(iii)a, further stabilizing the RT-intron interaction (Figure 3a) [23]. However, because there is a flexible linker connecting the RT domain and X domain, fixing the RT domain alone would not be sufficient for holding the entire IEP in the correct orientation for catalysis. Therefore, an insertion loop in the X domain, which is denoted as the “ti-loop” [27], penetrates deeply inside intron D1 to anchor the X domain along the 5'-exon-EBS duplex (Figure 3b) [23]. The sequence of this ti-loop is poorly conserved among IEP species, but it is generally 5–15 residues long and is enriched in positively charged residues (Figure 3b).

Direct modulation of group II intron catalysis is probably mediated by the X domain, however the precise mechanism of this modulation is still unclear. RNA probing [41] and IEP-intron cross-linking studies [27] suggest that the IEP modulates splicing by weakly interacting with the intron active site and with D6, which contains the branch-site adenosine.

This model is reasonable because the D6 conformational changes that are required for correct orientation of the branch-site are likely to be rate-limiting, and weak interactions with the IEP may shift the conformational equilibrium of D6, thereby speeding up the reaction rate. However, in the cryo-EM structure of *L.I.* IEP-intron complex, the IEP does not form any direct interactions with intron D6, and the closest IEP residues are ~6–10 Å away (Figure 3c) [23]. However, it remains possible that the modulation of RNA conformational equilibrium is not achieved by specific protein-RNA interactions, but rather by overall structural restraints imposed by protein binding and the strong positive charge on the X domain. This idea is consistent with previous observations that the association of IEP reduces the Mg^{2+} requirement for group II intron catalysis [41,42], but it does not explain why the IEP protects nucleotide bases in D6 from chemical probes and footprinting agents [27]. Finally, it remains possible that the IEP modulates chemical catalysis by the intron active-site. For example, the essential EYSC motif [14] is located within an α -helix that is proximal to the 5'-exon-intron duplex (Figure 3c) [23]. However, the EYSC motif may simply position the IEP in a manner that reduces dynamic behavior of the 5'-exon-intron duplex.

Is the IEP a dimer?

Biochemical studies on the stoichiometry of IEP binding and its effect on catalysis indicate that the IEP forms a stable dimer with group II intron RNA [18,43,44]. Consistent with this, the IEP RT domain forms a dimer in solution and in crystals of the protein (Figure 3d) [24]. Paradoxically, the cryo-EM structure revealed only a single IEP in complex with intron RNA (Figure 3d) [23,45]. This inconsistency in IEP stoichiometry may be an experimental artifact, or it may indicate that the IEP-intron complex adopts different stoichiometries at different stages of splicing, and that this may have functional ramifications.

A model for dimer formation would reconcile some of the apparent conflicts between the cryo-EM structure and biochemical data on the same *L.I.* system. For example, dimerization of the IEP that is mediated by the interface observed in the crystal structures [24] would bring a second IEP monomer into direct contact with intron D6, which would explain the IEP-induced protection sites that are observed on D6 [27]. Additionally, in the *L.I.* IEP-intron complex, the active sites for endonuclease (EN) and reverse transcriptase components are 45 Å apart [23], which raises the question of how the free 3'OH group generated by the EN domain is fed into the active site of the reverse transcriptase during retrotransposition. Dimerization of the IEP might solve this problem by placing the EN domain from one monomer next to the reverse transcriptase active site of the second monomer [46].

However, it is not entirely clear how the dimer would form in the context of the IEP, or how it would function. For example, when a second IEP monomer is modeled into the cryo-EM structure [23] using the same dimerization interface observed in the crystal structures [24], steric clashes are observed in the absence of mitigating conformational changes [45]. These would be resolved, however, with simple hinge motions in linker domains of the IEP (Figure 3d). A dimeric form of the IEP also raises questions about regulation of the second active-site. For example, retrotransposition presumably only requires a single active site for reverse

transcriptase and endonuclease function. If the IEP forms a dimer, the second active site might lead to undesired side-reactions or a loss of fidelity.

Evolutionary impact: The transformation of an IEP into a general splicing factor

Group II introns have diversified the genomes in all domains of life, and they are almost certainly progenitors of the eukaryotic spliceosome. Evidence for their common ancestry is apparent from similarities between recent structures of the eukaryotic spliceosome [47–51] and structures of group II intron IEPs [23,24] and isolated group II introns [2,4,5,23]. Shared features in RNA sequence, metal ion binding sites, RNA secondary structural elements [52] and parallels in splicing mechanism [53,54] have been discussed extensively in previous reviews [3,9,10,45,55]. Two major events would have facilitated the transformation of group II introns (or a common ancestor) into spliceosomes. The first would be fragmentation of a group II intron RNA into pieces (such as the bipartite and tripartite group II introns observed in plants) [10,56,57], ultimately resulting in functional RNAs that would operate on splice-sites in-trans. The second event, which is mechanistically dependent on the first, would be the ability of group II intron components to function as multiple-turnover enzymes that catalyze splicing at numerous sites [10,57].

This process is likely to have occurred in stages, and vestiges of this transition are evident in splicing events that contribute to plant respiration. For example, there are two group II intron IEPs in plants (matR in mitochondria [57] and matK in chloroplasts [58]) that act not only on their parent introns, but also on a large set of additional group II introns within plant organelles, acting as multiple-turnover splicing factors. In addition, there is an IEP-like protein encoded within plant nuclei (nMat) that is translated and targeted to the mitochondria [57], where it facilitates the splicing of another set of group II introns.

The availability of atomic IEP structures now makes it possible to envision the physical mechanism for evolutionary transformation of an IEP into a general splicing factor in the spliceosome and in plant organelles. For example, the transition between the IEP and Prp8 appears to have been facilitated by mutation of the basic RT0 residues into acidic amino acids that facilitate regulation and protein binding (Figure 4b). Perhaps to compensate for the resultant loss of RNA binding affinity, Prp8 has acquired an N-terminal auxiliary domain (NTD) that interacts with U5 snRNA, which positions it within the catalytic center (Figure 4c). In plant systems, the IEPs matK (plant chloroplasts) and matR (plant mitochondria) contain a truncation of the N-terminal region that is normally involved in sequence-specific RNA recognition (Figure 4a) [10,57,58], consistent with their ability to splice numerous different group II introns. Similarly, the RT0 region of nMAT contains amino acids expected to abrogate the formation of specific RNA interactions (Figure 4a). Intriguingly, the IEP-related proteins encoded in plant nuclei have the same domain organization as a bacterial IEP, whereas IEP-related proteins encoded in plant organelles have a degenerated RT domain, suggesting divergent evolution (Figure 4a) [59,60]. While it is likely that eukaryotic complexity has been facilitated by the emergence of multiple-turnover splicing machines

such as the spliceosome, other RNA processing machines have clearly evolved from IEPs, and their hallmarks remain evident in modern organisms.

Implications for reverse transcriptase enzymes

Recent advances in IEP structural biology have important ramifications for the evolution and function of reverse-transcriptase enzymes. Crystal structures of the IEP RT domain revealed not only striking homology with Prp8, but a particularly close homology with viral RNA-dependent RNA polymerases (RdRPs) such as NS5B from Hepatitis C Virus [24]. This close similarity in sequence and three-dimensional structure, which is consistent with predictions from previous phylogenetic analyses [25], suggests a shared common ancestor for IEPs and viral RdRPs. Importantly, this relationship contrasts sharply with the lack of structural homology between IEPs and retroviral RT enzymes [24]. Indeed, the structural differences between retroviral reverse transcriptases (such as HIV RT) and IEPs/RdRPs suggests two separate origins for RT enzymes. Given that human L1 elements and IEPs share a common ancestor [25], these findings underscore the need for caution in using HIV RT as a model for understanding behavior of the L1 element.

The IEP RT structures have also confirmed a shared evolutionary heritage with telomerase [26,45] and with bacterial group II intron-like RTs [61]. The latter RTs are associated with CRISPR elements and the acquisition of bacterial immunity from RNA [62]. Interestingly, the CRISPR-associated group II intron-like RT contains only the polymerase finger and palm but not the thumb (RT domain). Similar deletion constructs of group II intron IEPs retain catalytic activity but lose processivity [24], suggesting that the CRISPR-associated RTs perform short primer extension reactions.

Conclusions and perspectives

Through advances in crystallography and single particle cryo-EM, it is now possible to visualize a unique class of proteins represented by IEPs and non-LTR retrotransposons. These first IEP structures provide much-needed insights into the mechanistic role of IEPs in RNA recognition, splicing, reverse transcription and retrotransposition. The IEP structures have cemented the evolutionary relationship between group II introns and the spliceosome, and they have profound implications for the evolution of reverse transcriptases in general. But it is important to emphasize that these initial structures provide only the first step in mechanistic understanding of IEP function. For example, the physical mechanism by which the IEP stimulates RNA splicing and reverse-splicing into DNA remains to be elucidated, and the stoichiometry of the IEP-intron states must be clarified. With the help of single particle cryo-EM, structures of intron-IEP complexes at different stages will ultimately reveal a clear picture of IEP-assisted splicing and retrotransposition, thereby elucidating processes that are essential for metabolism in plants, fungi and yeast, and providing new insights into the role of Prp8 during spliceosomal processing.

Acknowledgments

We thank 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 and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Peebles CL, Perlman PS, Mecklenburg KL, Petrillo ML, Tabor JH, Jarrell KA, Cheng HL. A self-splicing RNA excises an intron lariat. *Cell*. 1986; 44:213–223. [PubMed: 3510741]
2. Toor N, Keating KS, Taylor SD, Pyle AM. Crystal structure of a self-spliced group II intron. *Science*. 2008; 320:77–82. [PubMed: 18388288]
3. 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]
4. Marcia M, Pyle AM. Visualizing group II intron catalysis through the stages of splicing. *Cell*. 2012; 151:497–507. [PubMed: 23101623]
5. Robart AR, Chan RT, Peters JK, Rajashankar KR, Toor N. Crystal structure of a eukaryotic group II intron lariat. *Nature*. 2014; 514:193–197. [PubMed: 25252982]
6. Pyle AM. Group II Intron Self-Splicing. *Annu Rev Biophys*. 2016; 45:183–205. [PubMed: 27391926]
7. Zimmerly S, Guo H, Eskes R, Yang J, Perlman PS, Lambowitz AM. A group II intron RNA is a catalytic component of a DNA endonuclease involved in intron mobility. *Cell*. 1995; 83:529–538. [PubMed: 7585955]
8. Matsuura M, Saldanha R, Ma H, Wank H, Yang J, Mohr G, Cavanagh S, Dunny GM, Belfort M, Lambowitz AM. A bacterial group II intron encoding reverse transcriptase, maturase, and DNA endonuclease activities: biochemical demonstration of maturase activity and insertion of new genetic information within the intron. *Genes Dev*. 1997; 11:2910–2924. [PubMed: 9353259]
9. Lambowitz AM, Belfort M. Mobile Bacterial Group II Introns at the Crux of Eukaryotic Evolution. *Microbiol Spectr*. 2015:3.
- 10•. Zimmerly S, Semper C. Evolution of group II introns. *Mob DNA*. 2015; 6:7. A comprehensive summary of group II intron IEP classes and a discussion of the roles of group II introns in eukaryotic evolution. [PubMed: 25960782]
11. Zimmerly S, Hausner G, Wu X. Phylogenetic relationships among group II intron ORFs. *Nucleic Acids Res*. 2001; 29:1238–1250. [PubMed: 11222775]
12. Blocker FJ, Mohr G, Conlan LH, Qi L, Belfort M, Lambowitz AM. Domain structure and three-dimensional model of a group II intron-encoded reverse transcriptase. *RNA*. 2005; 11:14–28. [PubMed: 15574519]
13. Zimmerly S, Guo H, Perlman PS, Lambowitz AM. Group II intron mobility occurs by target DNA-primed reverse transcription. *Cell*. 1995; 82:545–554. [PubMed: 7664334]
14. Cui X, Matsuura M, Wang Q, Ma H, Lambowitz AM. A group II intron-encoded maturase functions preferentially in cis and requires both the reverse transcriptase and X domains to promote RNA splicing. *J Mol Biol*. 2004; 340:211–231. [PubMed: 15201048]
15. Wank H, SanFilippo J, Singh RN, Matsuura M, Lambowitz AM. A reverse transcriptase/maturase promotes splicing by binding at its own coding segment in a group II intron RNA. *Mol Cell*. 1999; 4:239–250. [PubMed: 10488339]
16. 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*. 2004; 10:1433–1443. [PubMed: 15273321]

17. Cousineau B, Smith D, Lawrence-Cavanagh S, Mueller JE, Yang J, Mills D, Manias D, Dunny G, Lambowitz AM, Belfort M. Retrohoming of a bacterial group II intron: mobility via complete reverse splicing, independent of homologous DNA recombination. *Cell*. 1998; 94:451–462. [PubMed: 9727488]
18. Saldanha R, Chen B, Wank H, Matsuura M, Edwards J, Lambowitz AM. RNA and protein catalysis in group II intron splicing and mobility reactions using purified components. *Biochemistry*. 1999; 38:9069–9083. [PubMed: 10413481]
19. Cousineau B, Lawrence S, Smith D, Belfort M. Retrotransposition of a bacterial group II intron. *Nature*. 2000; 404:1018–1021. [PubMed: 10801134]
20. Pyle, AM., Lambowitz, AM. *The RNA World*. 3. Gesteland RF: Cold Spring Harbor Laboratory Press; 2006. Group II Introns: Ribozymes That Splice RNA and Invade DNA; p. 469-505.
21. Sharp PA. On the origin of RNA splicing and introns. *Cell*. 1985; 42:397–400. [PubMed: 2411416]
22. Cech TR. The generality of self-splicing RNA: relationship to nuclear mRNA splicing. *Cell*. 1986; 44:207–210. [PubMed: 2417724]
- 23••. Qu G, Kaushal PS, Wang J, Shigematsu H, Piazza CL, Agrawal RK, Belfort M, Wang HW. Structure of a group II intron in complex with its reverse transcriptase. *Nat Struct Mol Biol*. 2016; 23:549–557. The first structure of a group-II-intron-IEP complex, which was solved by single-particle cryo-EM reconstruction, to a resolution of 3.8 Å. It reveals unprecedented details of intron-IEP architecture, and provides insights into the mechanism of IEP-assisted splicing and retro-transposition. [PubMed: 27136327]
- 24••. 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. The first crystal structure of a group II intron IEP RT domain, solved to 1.2 Å resolution. This structure clearly demonstrates the structural homology of the group II intron IEP with spliceosomal Prp8. Additionally, it reveals a potential dimerization interface for group II intron IEPs and a positively-charged RNA binding surface. [PubMed: 27136328]
25. Xiong Y, Eickbush TH. Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J*. 1990; 9:3353–3362. [PubMed: 1698615]
26. Dlakic M, Mushegian A. Prp8, the pivotal protein of the spliceosomal catalytic center, evolved from a retroelement-encoded reverse transcriptase. *RNA*. 2011; 17:799–808. [PubMed: 21441348]
27. Dai L, Chai D, Gu SQ, Gabel J, Noskov SY, Blocker FJ, Lambowitz AM, Zimmerly S. 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]
28. Han JS. Non-long terminal repeat (non-LTR) retrotransposons: mechanisms, recent developments, and unanswered questions. *Mob DNA*. 2010; 1:15. [PubMed: 20462415]
29. Gu SQ, Cui X, Mou S, Mohr S, Yao J, Lambowitz AM. Genetic identification of potential RNA-binding regions in a group II intron-encoded reverse transcriptase. *RNA*. 2010; 16:732–747. [PubMed: 20179150]
30. Jamburuthugoda VK, Eickbush TH. Identification of RNA binding motifs in the R2 retrotransposon-encoded reverse transcriptase. *Nucleic Acids Res*. 2014; 42:8405–8415. [PubMed: 24957604]
31. Gillis AJ, Schuller AP, Skordalakes E. Structure of the *Tribolium castaneum* telomerase catalytic subunit TERT. *Nature*. 2008; 455:633–637. [PubMed: 18758444]
32. Mitchell M, Gillis A, Futahashi M, Fujiwara H, Skordalakes E. Structural basis for telomerase catalytic subunit TERT binding to RNA template and telomeric DNA. *Nat Struct Mol Biol*. 2010; 17:513–518. [PubMed: 20357774]
33. Ding J, Das K, Hsiou Y, Sarafianos SG, Clark AD Jr, Jacobo-Molina A, Tantillo C, Hughes SH, Arnold E. 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]
34. Huang H, Chopra R, Verdine GL, Harrison SC. Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science*. 1998; 282:1669–1675. [PubMed: 9831551]

35. Lesburg CA, Cable MB, Ferrari E, Hong Z, Mannarino AF, Weber PC. Crystal structure of the RNA-dependent RNA polymerase from hepatitis C virus reveals a fully encircled active site. *Nat Struct Biol.* 1999; 6:937–943. [PubMed: 10504728]
36. Toor N, Hausner G, Zimmerly S. Coevolution of group II intron RNA structures with their intron-encoded reverse transcriptases. *RNA.* 2001; 7:1142–1152. [PubMed: 11497432]
37. Malik HS, Burke WD, Eickbush TH. The age and evolution of non-LTR retrotransposable elements. *Mol Biol Evol.* 1999; 16:793–805. [PubMed: 10368957]
38. Beck CR, Garcia-Perez JL, Badge RM, Moran JV. LINE-1 elements in structural variation and disease. *Annu Rev Genomics Hum Genet.* 2011; 12:187–215. [PubMed: 21801021]
39. Rodic N, Steranka JP, Makohon-Moore A, Moyer A, Shen P, Sharma R, Kohutek ZA, Huang CR, Ahn D, Mita P, et al. Retrotransposon insertions in the clonal evolution of pancreatic ductal adenocarcinoma. *Nat Med.* 2015; 21:1060–1064. [PubMed: 26259033]
40. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: protein structure and function prediction. *Nat Methods.* 2015; 12:7–8. [PubMed: 25549265]
41. Matsuura M, Noah JW, Lambowitz AM. Mechanism of maturase-promoted group II intron splicing. *EMBO J.* 2001; 20:7259–7270. [PubMed: 11743002]
42. 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]
43. Rambo RP, Doudna JA. Assembly of an active group II intron-maturase complex by protein dimerization. *Biochemistry.* 2004; 43:6486–6497. [PubMed: 15157082]
44. Molina-Sanchez MD, Garcia-Rodriguez FM, Toro N. Functionality of In vitro Reconstituted Group II Intron RmInt1-Derived Ribonucleoprotein Particles. *Front Mol Biosci.* 2016; 3:58. [PubMed: 27730127]
- 45•. Agrawal RK, Wang HW, Belfort M. Forks in the tracks: Group II introns, spliceosomes, telomeres and beyond. *RNA Biol.* 2016; This review summarizes the recent group II intron IEP structures and discusses the role of the IEP in evolution. doi: 10.1080/15476286.2016.1244595:1-5
46. Piccirilli JA, Staley JP. Reverse transcriptases lend a hand in splicing catalysis. *Nat Struct Mol Biol.* 2016; 23:507–509. [PubMed: 27273636]
47. Yan C, Hang J, Wan R, Huang M, Wong CC, Shi Y. Structure of a yeast spliceosome at 3.6-angstrom resolution. *Science.* 2015; 349:1182–1191. [PubMed: 26292707]
48. Galej WP, Wilkinson ME, Fica SM, Oubridge C, Newman AJ, Nagai K. Cryo-EM structure of the spliceosome immediately after branching. *Nature.* 2016; doi: 10.1038/nature19316
49. Rauhut R, Fabrizio P, Dybkov O, Hartmuth K, Pena V, Chari A, Kumar V, Lee CT, Urlaub H, Kastner B, et al. Molecular architecture of the *Saccharomyces cerevisiae* activated spliceosome. *Science.* 2016; 353:1399–1405. [PubMed: 27562955]
50. Wan R, Yan C, Bai R, Huang G, Shi Y. Structure of a yeast catalytic step I spliceosome at 3.4 Å resolution. *Science.* 2016; doi: 10.1126/science.aag2235
51. Yan C, Wan R, Bai R, Huang G, Shi Y. Structure of a yeast catalytically activated spliceosome at 3.5 Å resolution. *Science.* 2016; doi: 10.1126/science.aag0291
52. 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]
- 53••. Fica SM, Tuttle N, Novak T, Li NS, Lu J, Koodathingal P, Dai Q, Staley JP, Piccirilli JA. RNA catalyzes nuclear pre-mRNA splicing. *Nature.* 2013; 503:229–234. Using metal ion specificity swap experiments and mutational analysis, this work shows that the active-sites for group II intron and the spliceosome catalysis are almost identical. This is considered to be the first concrete evidence that the spliceosome is a ribozyme. [PubMed: 24196718]
54. Fica SM, Mefford MA, Piccirilli JA, Staley JP. Evidence for a group II intron-like catalytic triplex in the spliceosome. *Nat Struct Mol Biol.* 2014; 21:464–471. [PubMed: 24747940]
- 55•. Nguyen TH, Galej WP, Fica SM, Lin PC, Newman AJ, Nagai K. CryoEM structures of two spliceosomal complexes: starter and dessert at the spliceosome feast. *Curr Opin Struct Biol.* 2016; 36:48–57. This review compares the catalytic center of group II intron and spliceosome,

which strongly supports the evolutionary relationship between these two systems. [PubMed: 26803803]

56. Ritlop C, Monat C, Cousineau B. Isolation and characterization of functional tripartite group II introns using a Tn5-based genetic screen. *PLoS One*. 2012; 7:e41589. [PubMed: 22876289]
57. Brown GG, Colas des Francs-Small C, Ostersetzer-Biran O. Group II intron splicing factors in plant mitochondria. *Front Plant Sci*. 2014; 5:35. [PubMed: 24600456]
58. Zoschke R, Nakamura M, Liere K, Sugiura M, Borner T, Schmitz-Linneweber C. An organellar maturase associates with multiple group II introns. *Proc Natl Acad Sci U S A*. 2010; 107:3245–3250. [PubMed: 20133623]
59. Mohr G, Lambowitz AM. Putative proteins related to group II intron reverse transcriptase/maturases are encoded by nuclear genes in higher plants. *Nucleic Acids Res*. 2003; 31:647–652. [PubMed: 12527773]
60. Sultan LD, Mileshina D, Grewe F, Rolle K, Abudraham S, Glodowicz P, Khan Niazi A, Keren I, Shevtsov S, Klipcan L, et al. The reverse-transcriptase/RNA-maturase protein MatR is required for the splicing of various group II introns in Brassicaceae mitochondria. *Plant Cell*. 2016; This work presents the first characterization of a mitochondria IEP-related protein and proves that it is a trans-splicing factor. doi: 10.1105/tpc.16.00398
61. Simon DM, Zimmerly S. A diversity of uncharacterized reverse transcriptases in bacteria. *Nucleic Acids Res*. 2008; 36:7219–7229. [PubMed: 19004871]
62. Silas S, Mohr G, Sidote DJ, Markham LM, Sanchez-Amat A, Bhaya D, Lambowitz AM, Fire AZ. Direct CRISPR spacer acquisition from RNA by a natural reverse transcriptase-Cas1 fusion protein. *Science*. 2016; 351:aad4234. This work demonstrates that bacterial immunity can be acquired by information from RNA molecules using a reverse-transcriptase that is related to the group II intron IEP. [PubMed: 26917774]

Highlights

- The intron encoded protein (IEP) is a reverse transcriptase that functions in splicing.
- The IEP has a canonical polymerase scaffold that is elaborated by insertions.
- As a splicing factor, the finger and thumb domains recruit the IEP to its intron target.
- As a splicing factor, the thumb domain stabilizes exon-intron interaction.
- The intron-specific IEP transformed into a general splicing factor during evolution.

Figure 1.

Structure of group II intron and its IEP. **(a)** Secondary and tertiary structure of group II introns. The secondary structural domains (D1–6, left) are color coded to match the tertiary structure of a group II intron lariat RNA (PDB: 4R0D, right), and the open reading frame (ORF) in D4 is indicated. **(b)** Domain organization of group II intron IEPs showing that IEPs from *Eubacterium retale* (*E.r.*) and *Roseburia intestinalis* (*R.i.*) contain only the RT domain and X domain. On the crystal structure of the *R.i.* RT domain (bottom left, PDB: 5HHJ), the dotted oval indicates the expected position of the X domain that is not included in the crystal structure. The IEP from *Lactococcus lactics* (*L.l.*) contains DNA binding domain (DBD) and endonuclease (EN) domains following the RT and X domains (bottom right, PDB: 5G2X). A large portion of its EN domain is disordered is not visualized in the structure.

Figure 2.

Group II intron IEPs are constructed on a polymerase scaffold. **(a)** Domain organization of HIV RT and group II intron IEPs. The seven conserved regions in all RTs are boxed and labeled, and insertions in group II intron IEPs are indicated as 2a, 3a, 4a and 7a. Insertions 2a, 3a and 4a are mapped onto the tertiary structures of *R.i.* (below left, PDB: 5HHJ) and *L.l.* RT domains (below right, PDB: 5G2X) using the same color code. **(b)** Comparison of the IEP X domain and polymerase thumb domains. The X (thumb) domain of *L.l.* IEP (PDB: 5G2X) is compared with the thumb domains of HIV polymerase (PDB: 2HMI).

telomerase (TERT, PDB: 3KYL). In all structures, the analogous regions are colored in red, and the rest of thumb domain, if present, is colored in pink. **(c)** Comparison of insertion 2a in IEPs (PDB: 5HHJ) and the analogous location in HIV reverse transcriptase (PDB: 2HMI). **(d)** Comparison of insertion 3a in *E.r/R.i.* (PDB: 5HHJ) with *L.l.* (PDB: 5G2X) IEPs. The insertion in finger domain (IFD) motif contains two α -helices that are indicated by a black box. **(e)** Location of insertion 4a. In *L.l.* IEP (PDB: 5G2X), insertion 4a contacts helix d(iii)a in intron D1. *E.r/R.i.* IEPs lack insertion 4a. **(f)** Homology model for the RT domain of human retro-transposon L1. The domain organization of L1 ORF2 and partial sequence alignment between RT domains from *R.i.* IEP and L1 ORF2 are shown at left. The predicted structure of L1 ORF RT domain is shown at right. It was produced by I-TASSER [40], using the secondary structure topology of the target sequence and restraints from the crystal structure of *R.i.* RT domain. The model shown here has an expected TM-score of 0.67 ± 0.13 (>0.5 indicates correct topology) and an expected RMSD of 7.3 ± 4.2 Å.

Figure 3.

Interaction of an IEP with its cognate group II intron RNA. **(a)** Overall scaffold of the *L.I.* IEP-intron complex (PDB: 5G2X). The IEP contacts the intron RNA via 3 anchor points (black boxes). Anchor points 1 and 2 are mediated by the RT domain, whereas anchor point 3 is mediated by the X domain. **(b)** Structure and sequence conservation of anchor point 3, whereby the "ti-loop" in the X-domain penetrates deeply into intron D1 within the *L.I.* IEP-intron complex structure (PDB: 5G2X). The consensus sequence for ti-loop is shown below. **(c)** Location of the EYSC motif. **(d)** Potential dimerization of the group II intron IEP. Observed dimerization of the IEP RT domain (right, PDB: 5HHJ). A hypothetical IEP dimerization site within the *L.I.* IEP-intron complex structure (orange oval, PDB: 5G2X) suggests a location for the additional IEP monomer if the IEP were to dimerize within the holoenzyme complex.

Figure 4.

Loss of RNA-IEP binding specificity through the course of evolution. **(a)** Domain organization of group II intron (GII) IEP, spliceosomal Prp8 and group II intron trans-splicing factors in mitochondria (MatR), chloroplasts (MatK) and nuclear (nMat). The shading on RT0 indicates its loss of amino acids associated with RNA binding affinity and/or specificity. MLS represents a mitochondrial localization signal. **(b)** Electrostatic surface electrostatic potential of the RT0 regions (dotted orange ovals) in group II intron IEP (PDB: 5HHJ) and spliceosomal Prp8 (PDB: 5LJ3). **(c)** Spliceosomal Prp8 (PDB: 5LJ3), which uses an auxiliary NTD domain (rather than RT0) to interact with snRNA U5.