Crystal structure of a eukaryotic group II intron lariat

Aaron R. Robart1, Russell T. Chan1, Jessica K. Peters1, Kanagalaghatta R. Rajashankar2 & Navtej Toor1

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

Splicing of nuclear introns results in the formation of circular RNAs having a branched lariat structure containing an unusual 2′–5′ phosphodiester bond2–4. This branched RNA product has also been found in group II introns5,6, which are self-splicing ribozymes. Defects in lariat formation result in aberrant splicing and human disease5. In eukaryotes, splicing of nuclear introns is catalysed by a large ribonucleoprotein complex called the spliceosome, which is thought to share a common ancestor with group II introns3,7.

Group II introns are catalytic RNAs with six structural domains (Extended Data Fig. 1) that splice via two transesterification reactions. In the first step of splicing, the 2′-OH of a bulged adenosine residue is activated for nucleophilic attack at the 5′ splice site to generate lariat RNA8,9. In the second step, the 3′-OH of the 5′ exon attacks the 3′ splice site to form ligated exons and excised intron lariat. The highly conserved domain V (DV) forms the group II intron active site by binding catalytic metal ions9, and domain VI (DVI) contains the bulged adenosine used as the nucleophile in the first step of splicing19.

Group II introns are divided into three structural classes: IIA, IIB and IIC10,11. Historically, the two model systems used to study group II intron structure and function have been ‘canonical’ eukaryotic IIB introns: P. littoralis LSUI2 from the brown algae *Pylaiella littoralis*12 and *aI5* from the yeast *Saccharomyces cerevisiae*. However, the only available crystal structure is of a IIC representative from the bacterium *Oceanobacillus iheyensis*9. This idiosyncratic IIC intron class is the most primitive13 and splices through hydrolysis to form linear intron16. In contrast, eukaryotic IIA and IIB introns form lariat, are evolutionarily later branching14, and therefore more closely related to the spliceosome. We targeted the *P. littoralis LSUI2* intron for structure determination since it contains a functional DVI that forms large amounts of lariat during splicing5.

Overall structure

Here we present the structure of the *P. littoralis LSUI2* intron in the post-catalytic lariat form with ligated exon product at 3.7 Å resolution (Extended Data Table 1a) solved using a Yb3+ derivative (Extended Data Fig. 2). This represents the first crystal structure of a 2′–5′ branched RNA molecule.

Reflecting the ability of IIB introns to form lariat, there are a multitude of unique tertiary interactions in the *P. littoralis LSUI2* intron compared to the *O. iheyensis* structure (Fig. 1). These newly visualized contacts include EBS2–IBS2 (Extended Data Fig. 3), μ–μ′, e–η′ and the canonical form of k–κ′. Unlike the *O. iheyensis* structure, domains II and III interact with multiple domains through long-range interactions to stabilize the overall fold of the *P. littoralis LSUI2* intron. We can now visualize the location of DVI within the intron structure (Fig. 1 and Extended Data Fig. 4), as well as the 2′–5′ lariat linkage between the first residue and the bulged adenosine.

Newly visualized tertiary interactions

One of the most highly conserved tertiary contacts in group II introns is the k–κ′ interaction between the base of the catalytic DV stem and domain I (DI)16. The conserved k sequence GAA, nucleotide A171 from near the k region, and residues from a GUAA pentaloop in DIII converge to form a pentadentate adenosine base stack (underlined residues) that inserts into the minor groove at the base of DV, rigidly placing the active site into the DI scaffold (Fig. 2a and Extended Data Fig. 5a).

The e–η′ interaction is critical for catalysis, with disruption through mutagenesis resulting in complete loss of splicing activity17. This interaction consists of nucleotides G106 and C107 pairing with C4 and G3 from the 5′ end of the intron (Fig. 2b and Extended Data Fig. 5b). The end result of these contacts is the formation of five conserved bases stacking in the following order (from bottom to top): A573, U2, G5, C4 and G3. This anchors the 5′ end in the core of the intron.

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

We can now correlate the function of DIII as a catalytic effector in group II introns. DIII interacts with the intron core through the aforementioned GUAA pentaloop, which docks into the base of DV (Fig. 2d). This μ–μ′ interaction18 serves to buttress the opposite side of DV from where catalysis takes place. Furthermore, a GAAA tetraloop from DIII interacts with the base of the DI stem (designated as t–t′) to provide...
additional reinforcement. Therefore, DIII functions as an external brace located on the outside surface to stabilize the entirety of the structure. This important role is consistent with deletion or mutagenesis of this domain rendering the intron unstable and not competent for efficient catalysis.

**DII positions DIV in the active site**

In the *O. iheyensis* structure, DII was truncated to a small stem loop structure, and in many previous biochemical studies of the *a15y* intron, DII was similarly shortened to study the first step of splicing. We can now visualize the intact DII substructure and find that it serves as a central hub for four different tetraloop receptor interactions (Fig. 3a). DII makes contacts with domains I, III and VI to organize a large portion of the intron structure.

DII has a ‘Y-shaped’ RNA secondary structure with two stems, D2a and D2b (Extended Data Fig. 1), coaxially stacking on top of each other such that a tetraloop receptor from D2a and a GCAA tetraloop from D2b are facing the same side (Fig. 3a). This combination provides a binding interface for DIV, which contains both a GAGA tetraloop and a tetraloop receptor. Therefore, DIV is tightly placed in the core of the intron via two tandem tetraloop receptor interactions with DII. The interaction between D2a and DIV is known as the $\kappa-\kappa'$ contact and we are designating the newly discovered interaction between D2b and DIV as $\pi-\pi'$. The $\pi-\pi'$ interaction is especially interesting due to its proximity to the bulged adenosine residue A615, which is the nucleophile for the first step of splicing. The $\pi$ tetraloop interacts with nucleotides directly adjacent to the bulged adenosine (Fig. 3a) and therefore probably has important effects upon the positioning of this nucleotide within the active site.

Mutagenesis of the GNRA tetraloops to UUCG was done to test the effects of these two interactions on splicing. Disrupting either contact significantly inhibited the second step of splicing, leading to an accumulation of lariat 3' exon and 5' exon (Fig. 3b and Extended Data Fig. 6). However, disrupting both interactions simultaneously resulted in a near complete block of the second step. This indicates that $\kappa-\kappa'$ and $\pi-\pi'$
and ligated exons. Mutagenesis of either g(WT) intron efficiently catalyses both steps of splicing and forms intron lariat

...are also found in... four large anomalous peaks in the ribozyme core (Fig. 4a). Two of these peaks (M1 and M2) are coordinated to the junction phosphate between the ligated exons, while the other two peaks (M3 and M4) are coordinated to the 5' end of the intron.

Active site metal ion configuration

Mg^{2+} ions are an absolute requirement for the catalysis of RNA splicing. To identify active site metal ions, soaks were performed using the anomalous scatterer Yb^{3+}, which exhibits the same octahedral coordination geometry as Mg^{2+} and preferentially binds to sites containing highly coordinated magnesium ions.\(^9,21\). This revealed four large anomalous peaks in the ribozyme core (Fig. 4a). Two of these peaks (M1 and M2) are also found in O. iheyensis\(^8\) and are embedded within DV to coordinate to the junction phosphate between the ligated exons, while the other two peaks (M3 and M4) are coordinated to the 5' end of the intron.

the second step, resulting in the accumulation of lariat 3' exon and 5' exon. A combination of both mutations (\(\Delta \pi \Delta \pi'\)) nearly blocks the second step of splicing with predominant intron 3' exon present.\(c\). \(F_o – F_e\) density for the 2'-5' lariat phosphodiester bond contoured at 3\(\sigma\). The nucleobase of A615 is disordered and not visualized. This map was calculated using a model deleted for A615, G1 and U2 (shown in stick format) to avoid model bias. The 5' (G1) and 3' (A620) ends form a non-canonical base pair.

Figure 4 | The core of the of P. li.LSU12 intron. a, Yb^{3+} anomalous map (orange mesh) contoured at 12\(\sigma\) reveals four large peaks, which correspond to highly coordinated magnesium ions (orange spheres). b, Comparison of the catalytic triplexes found in the O. iheyensis and P. li.LSU12 introns. O. iheyensis contains a continuous triple helix spanning the DV catalytic triad (CGC) and J2/3. In P. li.LSU12, the J2/3 residue A422 is disengaged from the catalytic triad (AGC) and forms a base triple with J4/5 and J5/6 nucleotides (U549 and U584, respectively).
M3 is located in a binding pocket formed by the highly conserved 5' end (Extended Data Fig. 8a), which has the sequence GUGCG. Nucleotides in this region exhibit a highly contorted backbone configuration that wraps around all sides of M3 due to the ε–ε' interaction involving residues G3 and C4 (Fig. 2b). Therefore, we postulate that the primary function of ε–ε' is to order a crucial metal-binding platform which structures the 5' end of the intron.

M4 coordinates to conserved IIB intron residues A6, C7 and A341 to stabilize the 5' end further (Extended Data Fig. 8b). These residues are in close spatial proximity to the recently proposed 1 motif (Extended Data Fig. 1) that is known to promote lariat formation and is predicted to serve as a receptor for positioning DVI (ref. 22). An ε G79A point mutant was crystallized, revealing strong signals for M1 and M2; however, M3 and M4 were no longer visible (Extended Data Fig. 8c, d). Splicing assays of this mutant also show a negative effect on the first step of splicing (Extended Data Table 1b). Furthermore, the ε–ε' interaction, which forms the metal-binding platform for M3 and M4, persists throughout group II intron catalysis17. Taken together, the data suggest that M3 and M4 participate in the first step of splicing by positioning the 5' splice site in the active site to present the scissile phosphate to M1, M2 and the bulged adenosine. However, it is possible that M3 exists only in the post-catalytic state to stabilize the repositioned lariat bond. Given the rarity of highly coordinated magnesium sites in large RNAs23, the existence of four such metal ions in close proximity in the active site of P.li.LSU12 is striking. In addition, we observe a conserved monovalent ion near the M1/M2 catalytic centre (Extended Data Fig. 8c).

Catalytic triplex rearrangement

In the O. iheyensis structure, J2/3 and residues from DV form a catalytic triplex structure essential for the catalysis of splicing and that is also found in the spliceosome41. In the O. iheyensis triplex, J2/3 residues G288 and C289 (analogous to G421 and A422 in P.li.LSU12) form base triples with the first two nucleotides of the catalytic triad. Unexpectedly, we observe in the P.li.LSU12 structure that one of the J2/3 residues is completely disengaged from the catalytic triad. Specifically, A422 has moved away from the triad and stacks directly underneath the base of the DV helix to form a base triple with residues from the J4/5 and J5/6 linkers (Fig. 4b). The possibility of J2/3 participating in conformational changes has been previously postulated26. However, we can now see the specific nature of this ‘switch’ with the disengagement of J2/3 from the catalytic triad into an alternate configuration. The linkers between domains are highly conserved for each subclass of group II introns, and based on the P.li.LSU12 structure we hypothesize that they are dynamic and modulate splicing.

π–π’ is a dynamic interaction

DVI is proposed to engage in large-scale conformational changes between the two steps of splicing20,22. To test this model, a catalytic triad mutant (AGC→GAU) inactive for splicing was crystallized and solved at 7 Å (Extended Data Fig. 9). Strong electron density was observed for the η–η’ interaction between DII and DVI, indicating that this contact persists throughout both steps of splicing and that there is no large-scale change in DVI position. Therefore, a new model is required to explain how DVI mediates the transition between the two steps of splicing.

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

Evolutionary implications

This bulged adenosine of DVI is analogous to the branch site adenosine in spliceosomal introns. The branch site sequence UACUAAC (nucleophilic adenosine underlined) pairs to the U2 snRNA to extrude the adenosine from the helix26 as in DVI. Owing to the mechanistic and evolutionary similarities with group II introns, it is likely that the branch site adenosine residue and the 5' end of nuclear introns will adopt a similar spatial arrangement relative to the catalytic core of the spliceosome. We predict that the spliceosomal branch site adenosine will be sequestered after the first step of splicing with an RNA and/or protein contact analogous to the π–π’ interaction in P.li.LSU12. In fact, an interaction observed between a region just downstream of the branch site and the U5 snRNA loop may be the spliceosomal counterpart of π–π’ (ref. 27).

The P.li.LSU12 structure provides a rationale for the phylogenetic conservation of the 5' end. Group II introns have the consensus sequence GUGYG (where Y = U or C)41, which is similar to the GUAYG and GURAG (where R = G or A) sequences found at the 5' ends of yeast and mammalian introns, respectively24. In the spliceosome, the U6 snRNA pairs with the 5' end20 in an analogous manner to the ε–ε' interaction in group II introns. Therefore, these conserved sequences probably form a similar metal-binding platform in spliceosomal introns, with the 5' end forming a highly distorted backbone to coordinate magnesium ions that orient the splice site in the core.

In regards to the evolutionary rationale for the 2'-5' phosphodiester bond, it is known that the lariat is required for full reversibility of group II intron splicing20. These reverse splicing reactions are the mechanism through which group II introns are able to invade DNA33 and disperse throughout genomes. We hypothesize that the lariat bond pre-organizes the core of the intron structure to facilitate this reversibility. Reverse splicing has also been proposed as a vehicle for the prolific expansion of nuclear introns34, which comprise ~25% of mammalian genomes. There is biochemical support for this hypothesis in that the spliceosome has recently
been shown to catalyse reverse splicing reactions\textsuperscript{31}. The lariat probably has a similar role in splicingosmal introns by allowing reverse splicing to occur, thus accounting for its phylogenetic conservation across the kingdoms by enabling ‘selfish introns’ to replicate. Therefore, the P.\textit{li.LSU12} crystal structure has provided the first glimpse of the branched lariat linkage that was probably crucial for intron proliferation in eukaryotes.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 11 April; accepted 22 August 2014.**

**Published online 24 September 2014.**

1. Grabowski, P. J., Padgett, R. A., & Sharp, P. A. Messenger RNA splicing in vitro: an excised intervening sequence and a potential intermediate. Cell 37, 415–427 (1984).
2. Padgett, R. A., Konarska, M. M., Grabowski, P. J., Hardy, S. F. & Sharp, P. A. Lariat RNA's as intermediates and products in the splicing of messenger RNA precursors. Science 225, 898–903 (1984).
3. Konarska, M. M., Grabowski, P. J., Padgett, R. A., & Sharp, P. A. Characterization of the branch site in lariat RNAs produced by splicing of mRNA precursors. Nature 313, 552–557 (1984).
4. Peebles, C. L. et al. A self-splicing RNA excises an intron lariat. Cell 44, 213–223 (1986).
5. van der Veen, R. et al. Excised group II introns in yeast mitochondria are lariats and can be formed by self-splicing in vitro. Cell 44, 225–234 (1986).
6. Di Leo, E. et al. A point mutation in the lariat branch point of intron 6 of NPC1 as the cause of abnormal pre-mRNA splicing in Niemann–PICK type C disease. Hum. Mutat. 24, 440 (2004).
7. Galej, W. P. Oubridge, C., Newman, A. J. & Nagai, K. Crystal structure of Prp8 reveals active site cavity of the spliceosome. Nature 493, 638–643 (2013).
8. Fica, S. M. et al. RNA catalyses nuclear pre-mRNA splicing. Nature 503, 229–234 (2013).
9. Toor, N., Keating, K. S., Taylor, S. D. & Pyle, A. M. Crystal structure of a self-spliced group II intron. Science 320, 77–82 (2008).
10. Lambowitz, A. M. & Zimmerly, S. Group II introns: mobile ribozymes that invade DNA. Cold Spring Harb. Perspect. Biol. 3, a003616 (2011).
11. Toor, N., Haasner, G. & Zimmerly, S. Coevolution of group II intron RNA structures with their intron-encoded reverse transcriptions. RNA 7, 1142–1152 (2001).
12. Michel, F., Umesono, K. & Ozeki, H. Comparative and functional anatomy of group II catalytic introns—a review. Gene 82, 5–30 (1989).
13. Costa, M., Fontaine, J. M., Loiseaux-de Goër, S. & Michel, F. A group II self-splicing intron from the brown alga Fylaella littoralis is active at unusually low magnesium concentrations and forms populations of molecules with a uniform conformation. J. Mol. Biol. 274, 353–364 (1997).
14. Rest, J. S. & Mindell, D. P. RetroDs in archaea: phylogeny and lateral origins. Mol. Biol. Evol. 20, 1134–1142 (2003).
15. Toor, N., Robart, A. R., Christianson, J. & Zimmerly, S. Self-splicing of a group II intron: 5’ exon recognition and alternative 5’ splicing events implicate the stem-loop motif of a transcriptional terminator. Nucleic Acids Res. 34, 6461–6471 (2006).
16. Boudvillain, M. & Pyle, A. M. Defining functional groups, core structural features and inter-domain tertiary contacts essential for group II intron self-splicing: a NAIM analysis. EMBO J. 17, 7091–7104 (1998).
17. Jacquier, A. & Michel, F. Base-pairing interactions involving the 5’ and 3’-terminal nucleotides of group II self-splicing introns. J. Mol. Biol. 213, 437–447 (1990).
18. Fedorova, O. & Pyle, A. M. A conserved element that stabilizes the group II intron active site. RNA 14, 1048–1056 (2008).
19. Fedorova, O., Mitros, T. & Pyle, A. M. Domains 2 and 3 interact to form critical elements of the group II intron active site. J. Mol. Biol. 330, 197–209 (2003).
20. Chamfrey, G. & Jacquier, A. An RNA conformational change between the two chemical steps of group II self-splicing. EMBO J. 15, 3466–3476 (1996).
21. Adams, P. L., Stahlley, M. R., Kosiek, A. B., Wang, J. & Strobel, S. A. Crystal structure of a self-splicing group I intron with both exons. Nature 430, 45–50 (2004).
22. Li, C. F., Costa, M. & Michel, F. Linking the branchpoint helix to a newly found receptor allows lariat formation by a group II intron. EMBO J. 30, 3040–3051 (2011).
23. Klein, D. J., Moore, P. B. & Steitz, T. A. The contribution of metal ions to the structural stability of the large ribosomal subunit. RNA 10, 1366–1379 (2004).
24. Fica, S. M., Melford, M. A., Picciulli, J. A. & Staley, J. P. Evidence for a group II intron-like catalytic triplex in the spliceosome. Nature Struct. Mol. Biol. 21, 464–471 (2014).
25. Marcia, M. & Pyle, A. M. Visualizing group II intron catalysis through the stages of splicing. Cell 151, 497–507 (2012).
26. Query, C. C., Moore, M. J. & Sharp, P. A. Branch nucleophile selection in pre-mRNA splicing: evidence for the bi-catalytic duplex model. Genes Dev. 8, 587–597 (1994).
27. Anokhina, M. et al. RNA structure analysis of human spliceosomes reveals a compact 3D arrangement of snRNAs at the catalytic core. EMBO J. 32, 2804–2818 (2013).
28. Parker, R. & Guthrie, C. A point mutation in the conserved hexanucleotide at a yeast 5’ splice junction uncouples recognition, cleavage, and ligation. Cell 41, 107–118 (1985).
29. Lesser, C. F. & Guthrie, C. Mutations in U6 snRNA that alter splice site specificity: implications for the active site. Science 262, 1982–1988 (1993).
30. Roitzsch, M. & Pyle, A. M. The linear form of a group II intron catalyzes efficient autocatalytic reverse splicing, establishing a potential for mobility. RNA 15, 473–482 (2009).
31. Yang, J., Zimmerly, S., Perlman, P. S. & Lambowitz, A. M. Efficient integration of an intron RNA into double-stranded DNA by reverse splicing. Nature 381, 322–335 (1996).
32. Lynch, M. & Richardson, A. O. The evolution of splicingosmal introns. Curr. Opin. Genet. Dev. 12, 701–710 (2002).
33. Tseng, C. K. & Cheng, S. C. Both catalytic steps of nuclear pre-mRNA splicing are reversible. Science 320, 1782–1784 (2008).

**Acknowledgements** We thank S. Banerjee and the staff of the NE-CAT beamlines at the Advanced Photon Source (APS) of Argonne National Laboratory. We thank P. Ghosh, S. Joseph, G. Ghosh, R. Doolittle, Y. Tor, D. Donoghue and T. Wiryaman for comments on the manuscript. We thank R. Das and F.-C. Chou for assistance with phenix.erasser for structure refinement and G. Briegle for advice on Buster refinement. We also thank N. T. Schirle for preliminary biochemical characterization of the P.\textit{li.LSU12} intron, R.T.C. was supported by the Cell, Molecular, and Genetics Training Program funded by NIH predoctoral training grant 5T32GM008326. NE-CAT is supported by NIH grant 8P41GM103403-10 and APS is supported by the US DOE under contract number DE-AC02-06CH11357. This work was supported by a Hellman Foundation Fellowship and NIH grant 5R01GM102216 awarded to N.T.

**Author Contributions** A.R.R. and J.K.P. performed the experiments. A.R.R., R.T.C., J.K.P. and N.T. designed the experiments. A.R.R., R.T.C., J.K.P., K.R.R. and N.T. analysed the data. A.R.R., R.T.C. and N.T. wrote the manuscript with input from all authors.

**Author Information** Coordinates and structure factors have been deposited in the Protein Data Bank under accession code 4R0D. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.T. (ntor@ucsd.edu).
METHODS

Cloning and preparation of Pyliella littoralis P.li.LSU2 intron RNA. DNA was synthesized (Genscript) corresponding to the second intron interrupting the large ribosomal subunit (LSU) in the mitochondria of the brown alga Pyliella littoralis (P.li.LSU2). The crystallization construct contains a 15-nucleotide 5’ exon and a 5-nucleotide 3’ exon followed by a HindIII restriction site. The DIV open reading frame (ORF) was removed from the P.li.LSU2 intron and replaced with a UUCG tetraloop. This was cloned into the EcoRV site of pUC57. Plasmid was linearized by HindIII digestion before in vitro transcription. Non-conserved loops in domains 1, 2 and 4 were changed from the wild-type sequence. Most significantly, changing the sequence UAUUUUAU to UCGACAUAGG in the ID2 stem loop improved both crystallization rate and diffraction. The final construct retained wild-type splicing activity. Transcription was performed overnight at 37°C using T7 RNA polymerase in 25 mM MgCl₂, 2 mM spermidine, 5 mM DTT, 40 mM Tris-HCl pH 7.5, 0.05% Triton X-100, 2.5 mM of each NTP, and thermostable inorganic pyrophosphatase (New England Biolabs). CaCl₂ was added to a final concentration of 1.2 mM, treated with DNase I for 45 min, followed by proteinase K digestion for 1 h. The intron reacted to completion during in vitro transcription and was subjected to a native purification procedure previously used in the structure determination of the O. iheyensis intron⁹. Spliced intron RNA was repeatedly washed with 10 mM MgCl₂ and 5 mM sodium cacodylate pH 6.5 and concentrated to 10 mg ml⁻¹ using a 100kDa molecular mass cutoff Amicon Ultra-15 column.

Crystallization. The native crystals were grown in sitting drops by vapour diffusion at 30°C. Equal volumes of RNA (10 mg ml⁻¹) were mixed with 0.4 mM spermine, 21% 2-methyl-2,4-pentanediol (MPD), 175 mM magnesium acetate tetrahydrate and 90 mM MES monohydrate (pH 5.6). Rod-like crystals appeared within 2 days, and grew to a maximum size of 50 × 50 × 900 μm. Crystals were gradually exchanged into 21% MPD, 100 mM magnesium acetate tetrahydrate, 50 mM MES monohydrate (pH 5.6), 3 mM iodide hexammine, 0.5 mM spermine and 100 mM NaCl, followed by flash freezing in liquid nitrogen. Iodine hexammine was used as an additive for the native crystals as it reduced mosaicity. Pre-catalytic intron RNA was obtained by mutating the AGC catalytic triad of DV to GAU. The mutation maintained stem structure determinant. The crystal structure of the P.li.LSU2 intron was solved in vitro self-splicing assays. The construct used for the in vitro self-splicing assays contained wild-type P.li.LSU2 sequence with DIV ORF removed and a 250-nucleotide 5’ exon and 75-nucleotide 3’ exon. This was cloned into the pUC57 plasmid. Plasmid was linearized using HindIII and used for in vitro transcription with T7 RNA polymerase. Radiolabelled transcripts were prepared as above using 10 μCi [³²P]UTP (3,000 Ci mmol⁻¹), 0.5 mM UTP, 1 mM other NTPs, and 10 mM MgCl₂. Transcripts were gel purified on a 4% polyacrylamide (19:1)/8 M urea gel, RNA was recovered by diffusion into 300 mM NaCl, 0.01% SDS, 1 mM EDTA. Self-splicing experiments were performed for 30 min at 45°C in a splicing buffer containing 10 mM MgCl₂, 1 M NaCl, 40 mM Tris-HCl (pH 7.5) and 0.02% SDS. Reactions were stopped by addition of EDTA to a final concentration of 20 mM. Splicing products were resolved using a denaturing 4% polyacrylamide (19:1)/8 M urea gels. Rate constants for the wild-type and G79A mutant were derived from curves fit to a biphasic exponential equation. All splicing assays were done in triplicate.

34. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. 276, 307–326 (1997).
35. Schneider, T. R. & Sheldrick, G. M. Substructure solution with SHELXD. Acta Crystallogr. D 58, 1772–1779 (2002).
36. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D 66, 213–221 (2010).
37. Sheldrick, G. M. A short history of SHELX. Acta Crystallogr. A 64, 112–122 (2008).
38. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. Acta Crystallogr. D 60, 2126–2132 (2004).
39. Keating, K. S. & Pyle, A. M. Rcrane: semi-automated RNA model building. Acta Crystallogr. D 68, 985–995 (2012).
40. Blanc, E. et al. Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. Acta Crystallogr. D 60, 2210–2221 (2004).
41. Schroder, G. F., Levitt, M. & Brunger, A. T. Super-resolution biomolecular crystallography with low-resolution data. Nature 464, 1218–1222 (2010).
42. Chou, F. C., Sripakdeevong, P., Dibrov, S. M., Hermann, T. & Das, R. Correcting pervasive errors in RNA crystallography through enumerative structure prediction. Nature Methods 10, 74–76 (2013).
43. Morin, A. et al. Collaboration gets the most out of software. eLife 2, e01456 (2013).
44. Chanfreau, G. & Jacquier, A. Interaction of intronic boundaries is required for the second splicing step efficiency of a group II intron. EMBO J. 12, 5173–5180 (1993).
45. Parker, R. & Siliciano, P. G. Evidence for an essential non-Watson-Crick interaction between the first and last nucleotides of a nuclear pre-mRNA intron. Nature 361, 660–662 (1993).
46. Basu, S. et al. A specific monovalent metal ion integral to the ΑΑ platform of the RNA tetraloop receptor. Nature Struct. Biol. 5, 986–992 (1998).
Extended Data Figure 1 | Secondary structure of \( P.\text{li.}\text{LSU2} \) intron crystallization construct. Tertiary interactions are indicated with Greek letters and domains are labelled with Roman numerals. Colouring of the individual domains is consistent with the overall view of the tertiary structure shown in Fig. 1.
Extended Data Figure 2 | The Yb-MAD experimental, density-modified map of the portion of DV containing the catalytic triad contoured at 1.8σ.
Extended Data Figure 3 | The path of the 5′ exon through the intron structure. The EBS1–IBS1 and EBS2–IBS2 interactions position the 5′ exon. They do not form a continuous binding interface with the presence of a highly distorted backbone at the junction between these two motifs. As a result, the helical axes of the EBS1–IBS1 and EBS2–IBS2 pairings are positioned ~90° relative to each other. The EBS3–IBS3 interaction places the 3′ exon in the active site.
Extended Data Figure 4 | Overall tertiary structure of the *P.hLSU12* intron.
Individual domains and subdomains are depicted in different colours. Domain names are labelled with Roman numerals. a and b show different rotations of the intron structure.
Extended Data Figure 5 | Companion to Fig. 2 showing the location of the individual tertiary interactions relative to the overall structure. a, κ–κ′. b, ε–ε′. c, ρ–ρ′.
Extended Data Figure 6 | Splicing assays for the DVI mutants showing the proportion of branched product. Blue and orange bars correspond to lariat 3’ exon and lariat, respectively. The Δη’ and Δπ mutants accumulate large amounts of lariat 3’ exon, thus indicating a second-step splicing defect. The Δη’Δπ double mutant is almost completely blocked before the second step. The A620C mutant shows 2.4-fold greater accumulation of lariat 3’ exon compared to the wild-type intron, indicating that the interaction between G1 and A620 is important for the second step. In the yeast al5Δ intron, G1 instead interacts with the penultimate residue44, indicating a certain degree of flexibility for this pairing. There is evidence for a similar interaction between the termini of nuclear introns45 involving nearby (but not exactly equivalent) residues, which also has a significant effect upon the second step of splicing. Therefore, the 5’ and 3’ ends of nuclear introns may have a similar arrangement within the spliceosome.
Extended Data Figure 7 | Detailed view of the lariat bond.  

a. The lariat 2’-5’ phosphodiester bond in wall-eyed stereo format.

b. Stereo version of Fig. 3c. See Fig. 3c legend for details. $F_o - F_c$ density for the lariat bond contoured at 3$\sigma$. 
Extended Data Figure 8 | Anomalous maps identifying core metal ions.
a, b, Depiction of the RNA ligands surrounding metals M3 and M4, respectively. Yb$^{3+}$ anomalous map contoured at 9σ. c, Yb$^{3+}$ anomalous map for wild type contoured at 9σ. d, Compared with the wild-type intron, the Yb$^{3+}$ anomalous map for the G79A mutant (contoured at 4σ) is lacking the peaks corresponding to M3 and M4, even at a lower contour level. e, Tl$^{+}$ was used as a probe for monovalent ions in the RNA structure. The Tl$^{+}$ anomalous map (purple mesh contoured at 5.5σ) revealed a strong peak located 3.8 Å from M1 that coordinates to the nucleobase of J2/3 residue G421 and the backbone of DV nucleotide G550. This sodium ion Na1 (purple sphere) is significantly closer to M1 than the equivalent K$^{+}$ ion found in O. iheyensis. Otherwise, this monovalent ion binding site is relatively conserved between these two introns.
Extended Data Figure 9 | $2F_o - F_c$ density for DVI in the pre-catalytic structure contoured at 1σ. The $\eta-\eta'$ interaction persists throughout the splicing reaction and is visible in the pre-catalytic state. The weaker density for the central region of DVI suggests a partially disordered, dynamic region with possible helical remodelling in the conserved internal loop during splicing. The general pattern of side-by-side packing of domains II and VI persists between the two steps. Catalytic triad mutation consisted of an AGC→GAU substitution.
Extended Data Table 1 | X-ray and kinetic data for P.ii.LSUI2

**a** Data collection and refinement statistics.

|                          | P.ii.LSUI2 native | DV Triad Mutant | G79A Mutant | Yb⁺⁺ | Ti⁺⁺ |
|--------------------------|-------------------|----------------|-------------|------|------|
| **Data collection**      |                   |                |             |      |      |
| Space group              | C222₁             | C222₁          | C222        | C222₁| C222₁|
| Cell                     |                   |                |             |      |      |
| a, b, c (Å)              | 163.7,            | 161.9,         | 211.7,      | 164.6,| 161.2,|
|                          | 255.4,            | 264.5,         | 457.2,      | 257.7,| 257.7,|
|                          | 136.8             | 137.5          | 179.5       | 138.0 | 136.3 |
| α, β, γ (°)              | 90, 90, 90        | 90, 90, 90     | 90, 90, 90  | 90, 90, 90 | 90, 90, 90 |
| Wavelength (eV)          | 11218.5           | 12662          | 8949.5      | 8949.5| 8946.5| 12657.5 |
| Resolution (Å)           | 150.0-3.68        | 50.0-7.25      | 50.0-9.75   | 50.0-4.65| 50.0-4.49| 50.0-4.48 |
|                          | (3.74-3.68)       | (7.37-7.25)    | (9.92-9.75) | (4.73-4.65)| (4.57-4.49)| (4.56-4.48) |
| Rₘᵦₜ                   | 14.9 (>100)       | 9.5 (77.9)     | 8.6 (97.0)  | 8.8 (73.9) | 9.2 (72.8) | 8.8 (74.9) |
| I/σI                    | 6.4 (0.6)         | 13.2 (1.7)     | 14.3 (1.8)  | 16.1 (2.1) | 9.2 (1.2)  | 21.8 (2.2) |
| Completeness            | 99.9 (99.9)       | 94.6 (82.5)    | 96.3 (96.9) | 99.9 (99.6) | 97.9 (88.7) | 99.3 (99.6) |
| Redundancy              | 6.8 (3.8)         | 5.7 (4.7)      | 6.5 (6.0)   | 4.4 (3.7)  | 3.6 (2.8)  | 7.8 (6.4)  |
| CC*                     | (0.743)           |                |             |      |      |      |
| **Refinement**           |                   |                |             |      |      |      |
| Resolution (Å)           | 81.8-3.68         |                |             |      |      |      |
| No. reflections          | 31107             |                |             |      |      |      |
| Rₘᵦₜ, Rₚᵦₜ             | 23.9/27.4         |                |             |      |      |      |
| No. atoms                | 13979             |                |             |      |      |      |
| RNA                      | 13471             |                |             |      |      |      |
| Ligand/ion               | 393               |                |             |      |      |      |
| Water                    | 115               |                |             |      |      |      |
| B-factors                |                    |                |             |      |      |      |
| RNA                      | 201.6             |                |             |      |      |      |
| Ligand/ion               | 181.5             |                |             |      |      |      |
| Water                    | 177.0             |                |             |      |      |      |
| R.m.s.                   |                    |                |             |      |      |      |
| Bond lengths             | 0.017             |                |             |      |      |      |
| Bond angles              | 1.483             |                |             |      |      |      |

*Highest resolution shell is shown in parenthesis.

**b** Comparison of wild-type (WT) and G79A splicing rates.

|        | k₉₅₈ (min⁻¹) | k₉₅₁ (min⁻¹) |
|--------|--------------|--------------|
| WT     | 9.51 ± 1.09  | 0.136 ± 0.0246 |
| G79A   | 2.97 ± 1.57  | 0.0767 ± 0.00942 |