



HAL
open science

Speciation of a group I intron into a lariat capping ribozyme

Mélanie Meyer, Henrik Nielsen, Vincent Olieric, Pierre Roblin, Steinar D. Johansen, Eric Westhof, Benoît Masquida

► **To cite this version:**

Mélanie Meyer, Henrik Nielsen, Vincent Olieric, Pierre Roblin, Steinar D. Johansen, et al.. Speciation of a group I intron into a lariat capping ribozyme. *Proceedings of the National Academy of Sciences of the United States of America*, 2014, 111 (21), pp.7659-7664. 10.1073/pnas.1322248111 . hal-02633667

HAL Id: hal-02633667

<https://hal.inrae.fr/hal-02633667>

Submitted on 27 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Speciation of a group I intron into a lariat capping ribozyme

Mélanie Meyer^{a,1}, Henrik Nielsen^b, Vincent Oliéric^c, Pierre Roblin^{d,e}, Steinar D. Johansen^f, Eric Westhof^a, and Benoît Masquida^{a,2,3}

^aArchitecture et Réactivité de l'ARN, Unité Propre de Recherche 9002, Institut de Biologie Moléculaire et Cellulaire, Université de Strasbourg, Centre National de la Recherche Scientifique, 67084 Strasbourg, France; ^bDepartment of Cellular and Molecular Medicine, The Panum Institute, University of Copenhagen, DK-2200 Copenhagen, Denmark; ^cSwiss Light Source at Paul Scherrer Institute, 5232 Villigen, Switzerland; ^dSynchrotron Soleil, 91190 Saint-Aubin, France; ^eInstitut National de la Recherche Agronomique, Unité Biopolymères, Interactions, Assemblages, 44316 Nantes, France; and ^fDepartment of Medical Biology, University of Tromsø, N-9019 Tromsø, Norway

Edited* by Thomas R. Cech, University of Colorado, Boulder, CO, and approved April 14, 2014 (received for review December 3, 2013)

The lariat-capping (LC) ribozyme is a natural ribozyme isolated from eukaryotic microorganisms. Despite apparent structural similarity to group I introns, the LC ribozyme catalyzes cleavage by a 2',5' branching reaction, leaving the 3' product with a 3-nt lariat cap that functionally substitutes for a conventional mRNA cap in the downstream pre-mRNA encoding a homing endonuclease. We describe the crystal structures of the precleavage and postcleavage LC ribozymes, which suggest that structural features inherited from group I ribozymes have undergone speciation due to profound changes in molecular selection pressure, ultimately giving rise to an original branching ribozyme family. The structures elucidate the role of key elements that regulate the activity of the LC ribozyme by conformational switching and suggest a mechanism by which the signal for branching is transmitted to the catalytic core. The structures also show how conserved interactions twist residues, forming the lariat to join chemical groups involved in branching.

RNA structure | RNA catalysis | GIR1 | crystallography | SAXS

Various mechanisms critically regulate splicing, thus controlling the fate of host gene products. Most splicing that relies solely on autocatalytic RNA introns appears to be unregulated, but control can be achieved at the RNA level because the structural versatility of RNA allows for the fusing of functional modules that can work in concert (1). Examples of riboswitches adjoined to spliceosomal (2) or group I (3) introns illustrate this concept and suggest that uncharacterized splicing regulation mechanisms may exist.

An elaborate example of RNA-regulated splicing is the group I twin-ribozyme introns found in the small subunit (SSU) ribosomal precursor in several protists (Fig. 1A). These twin-ribozyme introns are composed of a conventional group I splicing ribozyme (GIR2), into which is inserted a cassette composed of a branching ribozyme upstream from a homing endonuclease (HE) gene (4). The branching activity results in cleavage and concomitant formation of a 3-nt lariat capping the 5' end of the HE pre-mRNA (Fig. 1B) (5), hence the name “lariat capping” (LC) ribozyme. The lariat cap appears to act as a substitute for a conventional mRNA m⁷G cap in a situation in which an mRNA is expressed from within a RNA polymerase I (polI) gene (6). The cross-talk between GIR2 and the LC ribozyme is supported by the existence of three distinct processing pathways of the rRNA precursor, depending on environmental conditions. Under favorable conditions, splicing by GIR2 takes place first, followed by branching by the LC ribozyme and further processing, and eventually translation of the HE mRNA (7, 8). The transcriptional order of the ribozymes implies that the activity of the branching ribozyme is specifically repressed until splicing has taken place. Cellular stress conditions induce formation of full-length intron circles by a circularization pathway (9) that leaves the ribosomal exons unligated. Finally, starvation conditions induce branching by the LC ribozyme without splicing taking place (8). Because splicing only enables ligation of ribosomal

exons, the interplay between GIR2 and LC ribozymes influences the fate of both ribosomes and HE mRNA production.

Here, we report the crystal structures of two different forms of the LC ribozyme from *Didymium iridis* (DirLC). The structure of the wild-type unreacted form of the LC ribozyme that allowed characterization of the branching reaction (5) was solved to a resolution of 3.85 Å. A higher level of detail was achieved in the 2.5-Å crystal structure of an inactive form of the LC ribozyme based on a circular permutation (CP) over the scissile bond and closure of the natural ends using a stable 5'-UUCG-3' loop (Fig. 1C). Both structures adopted the same overall architecture (Fig. S1), as supported by the presence of the same structural features within the core, as well as the same tertiary interactions. Originally, the LC ribozyme was described as a group I-like ribozyme (GIR1) based on striking similarities at both sequence and secondary structure levels (10–12) (Fig. 1C–E). However, the present fold departs so drastically from group I intron structures that it justifies the creation of an original ribozyme family for LC ribozymes (Fig. 2). The structures reveal how the

Significance

We report the crystal structures of precleavage and postcleavage forms of the lariat-capping (LC) ribozyme. The structures show how domains from an ancestral group I ribozyme have evolved due to loss of selection pressure for self-splicing. Instead, a branching activity has been selected, resulting in capping the downstream mRNA by a 3-nt lariat stabilized by the ribozyme core. The LC ribozyme constitutes an original ribozyme family with an unexpected 3D structure that departs significantly from that of group I introns. The structures also elucidate the regulatory domain's role in transmitting a signal for cleavage to the ribozyme. The characterization of this natural evolutionary RNA speciation event is, to our knowledge, the first described at such an intricate level.

Author contributions: M.M., H.N., S.D.J., and B.M. designed research; M.M., H.N., V.O., P.R., S.D.J., and B.M. performed research; M.M., H.N., V.O., P.R., S.D.J., and B.M. analyzed data; and H.N., S.D.J., E.W., and B.M. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

Freely available online through the PNAS open access option.

Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.pdb.org (PDB ID codes 4P9R, 4P9S, and 4P8Z).

¹Present address: Institut de Génétique et de Biologie Moléculaire et Cellulaire, Centre National de la Recherche Scientifique, Université de Strasbourg, Unité Mixte de Recherche 7104, Illkirch, France.

²Present address: Génétique Moléculaire Génomique Microbiologie, Unité Mixte de Recherche 7156, Institut de Physique et de Chimie Biologiques, Université de Strasbourg, Centre National de la Recherche Scientifique, 67084 Strasbourg, France.

³To whom correspondence should be addressed. E-mail: b.masquida@unistra.fr.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1322248111/-DCSupplemental.

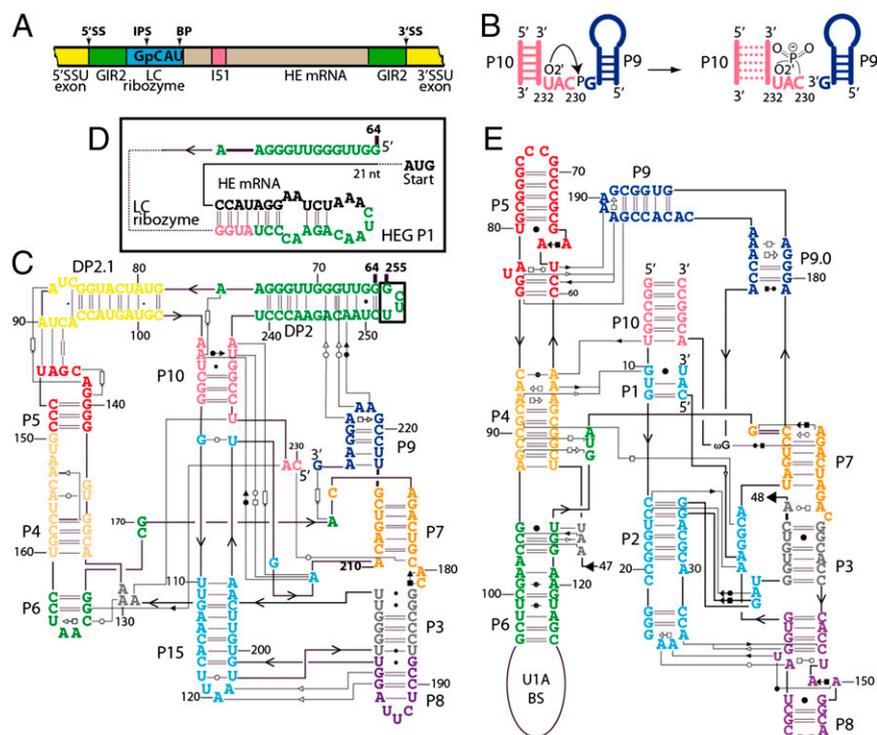


Fig. 1. Organization of the *D. iridis* SSU pre-rRNA reactivity and secondary structures of the active and inactive forms of the LC ribozyme compared with the *Azoarcus* group I intron (20). (A) The LC ribozyme is followed by a HE mRNA, which contains a 51-nt spliceosomal intron (I51). Nonobligatory introns organized identically have also been identified in *Naegleria* (44) and *Allovahlkampfia* species. This composite RNA is embedded in a canonical group I intron (GIR2) interrupting the SSU rRNA precursor. The internal processing site (IPS) of LC ribozyme is only 3 nt away from the branch point (BP). Positions of 5' and 3' splice sites (SS) are indicated. (B) Schematic representation of the branching reaction leading to the formation of the 3-nt lariat. (C–E) Secondary structure diagrams of the active form of the DirLC ribozyme (C), the alternative structure of the 3' domain of the LC ribozyme involved in cleavage inhibition (D), and group IC3 ribozyme from *Azoarcus* (E). (D) Black nucleotides correspond to the HE mRNA. A PCR artifact introduced a C residue in place of U203, which may stabilize the ribozyme as shown previously for the *Naegleria* LC ribozyme (45). Tertiary interactions are displayed using the Leontis–Westhof nomenclature (46). Stacking interactions are indicated with an arrow pointing to a rectangle.

DP2–DP2.1 domain connecting the LC ribozyme to GIR2 is organized and interacts with the catalytic core by mediating two sets of tertiary interactions. This domain has been shown to exert a regulatory role by switching on the LC ribozyme when a 3' hairpin named HEG P1 (homing endonuclease gene paired segment 1), formed during transcription to prevent branching before splicing by GIR2 (Fig. 1D and Fig. S2) (13), undergoes a conformational change leading to the formation of DP2. The structures also highlight how a residue from the core docks within an RNA pocket that was folded after the formation of DP2, to transmit a signal for branching to the catalytic center. Moreover, we uncover the way in which residues forming the lariat are twisted by means of conserved interactions, to join chemical groups involved in branching. Finally, we demonstrate that most of the LC ribozyme structural elements have diverged from group I introns, and, although they have apparently equivalent secondary structures, they have acquired alternate functions and structures in the course of evolution.

Results

The Wild-Type and CP Forms of the LC Ribozyme Adopt Identical Folds.

The CP and wild-type forms of the DirLC crystallize in space group $P2_12_12_1$ and present one molecule per asymmetric unit (Table S1). Because the wild-type ribozyme crystals only diffracted at low resolution, a CP construct with optimal conformational homogeneity was used. To achieve this goal, residues on either side of the cleavage site were defined as the 5' and 3' ends. To circumvent the poor transcription initiation capacity of the new 5' end sequence (5'-CAU-3') by T7 RNA polymerase, flanking ribozymes were added to the construct (14), and segments connecting DirLC to the splicing ribozyme were joined by a UUCG loop (Fig. S2). The 5' hammerhead and the 3' hepatitis delta virus ribozymes accurately generate 5'-hydroxyl and 2',3' cyclic phosphodiester ends (15). The structure of the CP DirLC ribozyme was solved by multiwavelength anomalous diffraction (MAD) experiments and was further used to solve the structure of the wild-type DirLC by the molecular replacement method. The observed pairing of DP2 strands in both the wild-type and

CP forms of DirLC crystal structures demonstrated that the UUCG loop tethers the ends of DP2 without forcing its formation. The presence of the UUCG loop does not perturb the structure because another construct permuted on the loop of P8 retains catalytic activity (Fig. S3). The construct is thus locked into the active fold (13), although the opening of the scissile bond prevents the branching reaction from occurring.

The LC Ribozyme Structure Presents Unique Structural Features.

Like group I introns, DirLC is organized around stacks of helices that form three distinct domains (Fig. 1C and E). DirLC harbors the group I typical P3–P7 pseudoknot that, together with an additional pseudoknot (P15–P3) and the associated three-way junction (3wj; P15–P3–P8), leads to a highly constrained, double pseudoknotted (P15–P3–P7) core (16). Distal from P15, the DP2.1 and DP2 elements, tethered to P10, reach P5 and P9, respectively (Fig. 1C). The curvature imposed by these tertiary interactions provokes stretching of the P4–P6 domain and results in an overall characteristic and striking signet ring shape in which the DP2.1 and the P4–P6 domains align perpendicular to the core (Fig. 2).

Two specific tertiary interactions are responsible for this characteristic shape. The GAAA tetraloop at the tip of P9 (L9) interacts with the shallow groove of the DP2 stem where two Watson–Crick purine–purine (R–R) pairs are observed (Figs. 2B and 3A). The tertiary interaction propagates through a ribose zipper between G221 and G68. The functional importance of this interaction is supported by an experiment in which a base pair was inserted or deleted within P9, which modifies the way the A residues from L9 interact in the shallow groove from DP2 (Fig. S4). Analysis of an analogous interaction between L9 and the regulatory domain in the *Naegleria* LC ribozyme (17) corroborates these observations. The second tertiary interaction involves the DL2.1 and the L5 loops, which form three *cis* WC base pairs between residues 5'-A90UC-3' from DL2.1 and 5'-G144AU-3' from L5 (Figs. 2D and 3B). This kissing complex is conserved at the sequence level among the known examples of LC ribozymes (Fig. S5). Moreover, both the CP and wild-type DirLC

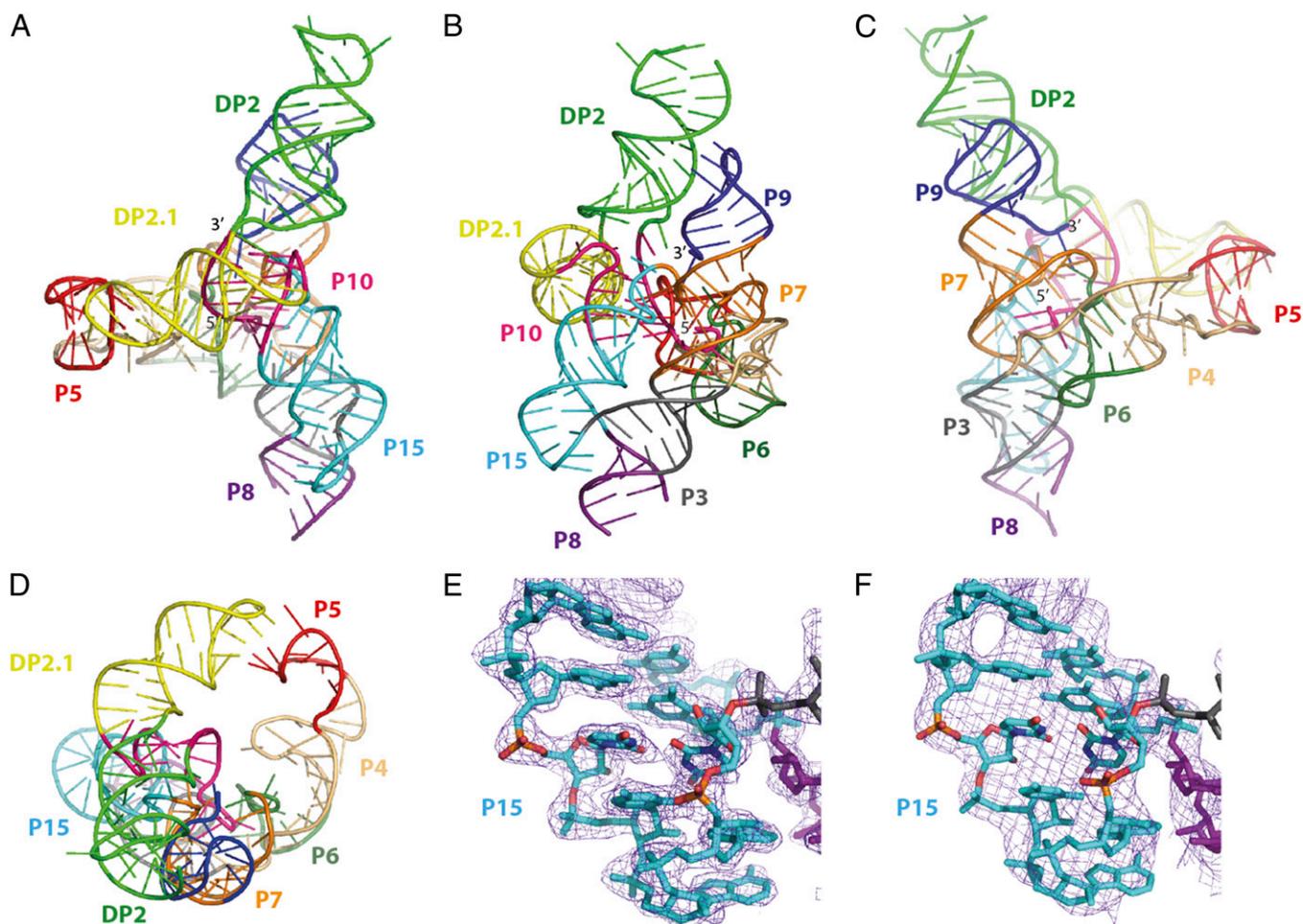


Fig. 2. Different views of the structure of the CP and wild-type forms of the DirLC ribozyme. (A–C) Three different views rotated by 90° clockwise on the vertical axis. (B) DirLC is represented according to the orientation used for group I introns with P9 in the upper right corner and P15 (equivalent to the P1 and P2 stems in group I ribozymes) in the lower left corner. (D) Rotation of DirLC from B rotated by 90° on a horizontal axis shows the signet ring shape resulting from the orientation of P4–P6 and DP2–DP2.1 domains. Discontinuities in the backbone are due to the absence of some residues in electron density maps. (E and F) Weighted electron-density maps of a representative region corresponding to the P15 loop contoured at 1.0 σ for the 2.5-Å (E) and 3.85-Å (F) resolution crystal structures. Statistics on data collection and refinement are summarized in Table S1.

ribozyme forms show very similar shapes in solution as deduced from small angle X-ray scattering experiments (SAXS) (Fig. S6).

The Fold of the Internal Loop Between P4 and P5 Allows for the L5–L2.1 Loop–Loop Interaction. The J5/4 junction between P4 and P5 adopts a zigzag motif, which is distant from the core and bends the P4–P6 domains so that L5 can form a kissing interaction with the loop of DP2.1 (Figs. 2 C and D and 3C). The J5/4 zigzag is made by two loop-like conformations that cap and unstack the flanking helices. The opposite J4/5 strand adopts a typical A-form helical conformation. U136 locks the zigzag by forming a Watson–Crick/Watson–Crick *trans* base pair with A155 that is stacked on the unpaired G137 (Fig. S7). Mutational analysis shows that A153 from the zigzag motif is critical for catalysis despite its distant location relative to the active site (12). A153 resides at the hinge of the motif where it participates in a tertiary contact through an A-minor interaction (18) to the apical G=C pair of P4, with A152 providing the second A-minor contact with the second base pair of P4 (Fig. S7C).

The DP2–DP2.1 Bridge Participates in the Folding of the Ribozyme Core. The tertiary interactions between the regulatory domain and the core lock the 3wj involving P10. P10 folds only in the active state of the LC ribozyme. Indeed, the folding of the HEG P1 hairpin characteristic of the inactive state relies on residues

from P10 as well as residues downstream from DP2 (Fig. 1 C and D and Fig. S3). The absence of the 3' residues necessary for HEG P1 folding in the crystallized constructs explains why only DP2 can be formed. At the P10 immediate 5' end, the residues from the lariat fold are stabilized by formation of the 3wj. Specific interactions (Fig. S8) position the nucleophile U232(O2') 5.1 Å away from the C230(O5') (Fig. 4A), despite the absence in the CP construct of the phosphate group involved in branching in the wild-type LC ribozyme. However, the lack of electron density for C230 and A231 in the 3.85-Å crystal structure prevents a comparison from being made with what is occurring in the wild-type LC ribozyme.

Stacking interactions between five unpaired adenosines (residues 76, 77, 103, 104, and 238) mostly stabilize the 3wj at the P10 interface (Fig. 4B). One side of A76 and A103 stacks on the last base pair of DP2 and DP2.1, respectively. A77 stacks with A104, which stabilizes the sharp turn between DP2.1 and P10, and A238 stacks upon the unpaired U239 at the inlet of the 3' strand of DP2. This architecture forms a cavity where a residue from the single strand joining P15 to P7 (J15/7), A209, fits snugly (Fig. 4C). A209 is intercalated between the G–U pair of P10 and A238. The Hoogsteen edge of A209 contacts the Watson–Crick edge of U237. Moreover, the A209 *syn* orientation permits additional H-bonds with A104(O2', N3). Thus, the 3wj acts as a recognition receptor for A209. A209 is transferred to its recognition pocket

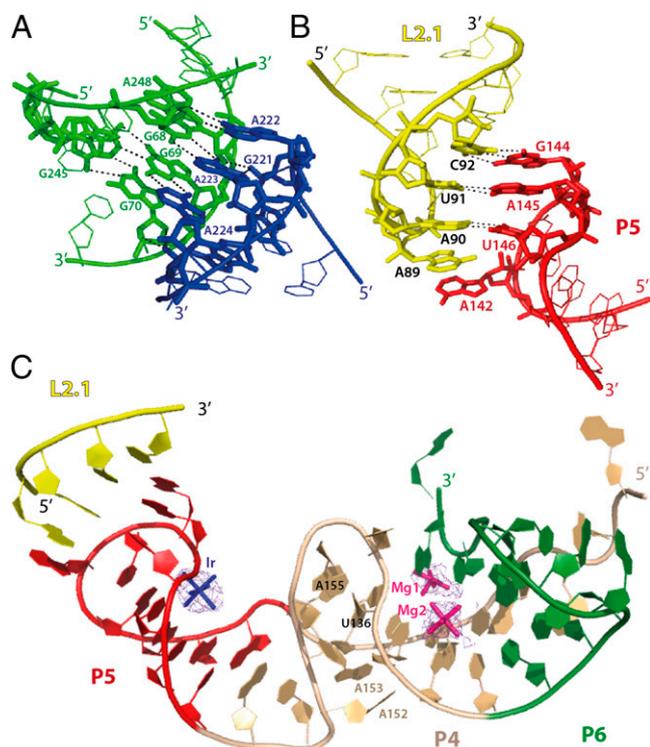


Fig. 3. Tertiary interactions promoted by a zigzag of the backbone in J5/4 stabilize the 3wj involving P10. (A) The DP2/L9 tertiary interaction involves the Watson–Crick edge of residues A222, A223, and A224 from the L9 tetraloop (deep blue), which interact in the shallow groove of A248, G68, and G69 in DP2 (green), respectively. (B) The DL2.1 loop (yellow) and the L5 loop (red) form Watson–Crick base pairs between residues 5′-A₉₀UC-3′ from DL2.1 and 5′-G₁₄₄AU-3′ from L5. (C) Overall view of the P4/P6 zigzag motif. A pentahydrated Mg²⁺ ion (Mg1) is bound to the O2P atom from G170 with the coordinated water molecules H-bonding with O6 positions of bases G134 and G135. The second Mg²⁺ ion (Mg2) is hexa-hydrated and bridges the bases of C159 and G169 from P4 and J6/7, respectively. The iridium hexammine ion in the deep groove of P5 stabilizes the loop by binding to the unpaired G141 and the preceding G=C base pair. Detailed interactions taking place in the zigzag are represented in Fig. S7. The inflection of the backbone corresponds to the A-minor interactions mediated by A152 and A153 to the G=C pairs within P4.

by means of a S-turn (19) mediated by the *trans* conformation of the wobble pair G109oU207 closing P15. This unusual geometry places the O2′ group from U207 in a position opposite to the

scissile bond (Fig. 4B). Downstream from A209, A210 loops back toward P7 and contacts the lariat fold as it passes. Moreover, an iridium hexammine ion in the vicinity of the lariat fold could indicate the presence of a metal ion site involved in catalysis (Fig. S8). Thus, the binding of A209 in its recognition pocket seems to signal to the core that the tertiary interactions mediated by the regulatory domain have taken place.

Structural Comparison of the LC and Group I Ribozymes. In group I ribozymes, the three domains from the core are invariably clamped by interactions of the loops closing P2 and P9 with stems P8 and P5, respectively (20). Strikingly, P3 and P7 have the same number of base pairs in both the LC and group I ribozymes, and the manner in which P3, P7, P8, and P9 are connected to other domains is identical. In DirLC, the P2/P8 interaction is replaced by the 3wj encompassing P15–P3–P8, resulting from the double pseudoknot P3–P7–P15. This 3wj shows how two Watson–Crick U–U pairs provide stacking continuity. In this context, P15 is equivalent to stems P2 and P1 from group I introns. Both P1 and P15 have as a first base pair a wobble G–U, albeit *cis* or *trans*, respectively (Fig. 5A). At the second step of splicing, the P1 substrate docks along the junction between P4 and P5 (J4/5) to allow the O3′ group from the catalytic U to attack the bond between the intron and the 3′ exon. Consequently, in group I ribozymes, the P4–P6 domain docks along the core and forms a tertiary interaction with P9, due to the conserved presence of a structural module that bends P9 toward P5 (Fig. 5B). In DirLC, P4, P5, and P6 present the same topology as in group I ribozymes, although the domain they form adopts a completely different structure. The reorganization of P4–P6 together with the L2.1/L5 loop–loop interaction prevent any contact between P15 and J4/5, which is averted by the absence of the characteristic triple interactions which take place in group I introns between P4 and J6/7.

In the second step of splicing, the scissile phosphate is brought in close proximity of the nucleophile by the docking of ωG to the P7 G-binding site (Fig. 5C). Mutational analysis of the LC ribozyme has shown G229, which is homologous to ωG, to be a critical nucleotide (16). G229 was expected to bind at the G-binding site before branching. However, in the ribozyme crystal structures, G229 is not bound to P7 and is instead located ~10 Å from the nucleophile in the CP crystals. G216 does not form the first base triple with either C171 or A172, a standard characteristic of P7 in group I splicing (Fig. 5C). P7 is also relaxed in group I ribozymes after cleavage (21), which could indicate that the crystal structure of the CP form reflects the postcatalytic state of the LC ribozyme.

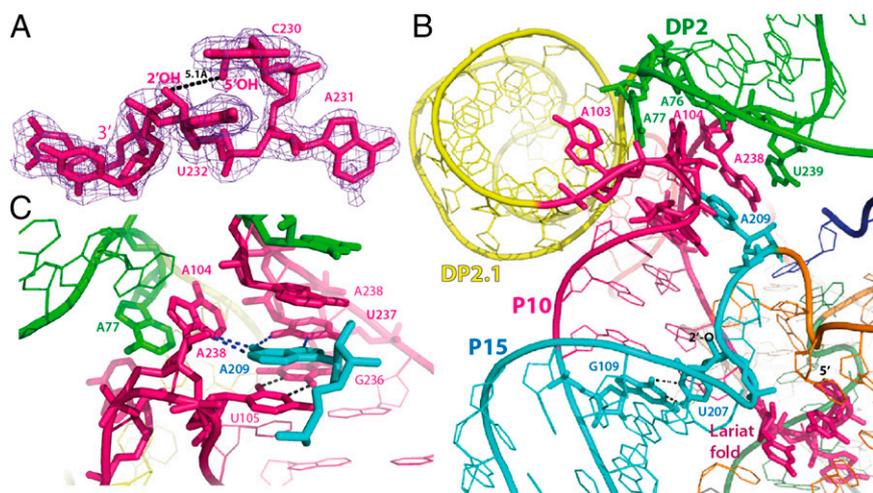


Fig. 4. Structure of the lariat fold and of the P10 3wj. (A) The lariat fold inscribed in a weighted $2F_o - F_c$ map contoured at 1.0σ shows a very constrained backbone with bases popping out to interact with specific residues (Fig. S8). The 5′-OH group of C230 is at 5.1 Å from the O2′ group of U232 despite the absence of the phosphate group missing because of the design of the CP construct. (B) Overall view of the ribozyme centered on the P10 3wj shows how five adenosines (76, 77, 103, 104, and 238) and a U residue (U239) mediate stacking continuity to stabilize the A209 binding pocket (cyan). The 2′ and 5′ hydroxyl groups from U207 and from C230 are labeled. (C) A close-up shows how A209 mediates stacking with neighboring residues (U105, A238) as well as how its Hoogsteen and Watson–Crick edges form H-bonds with residues in the same plane (A104, U237).

Experimental Procedures

Preparation of the Circularly Permuted Form of the LCz Ribozyme. RNA transcripts were obtained by in vitro transcription using T7-RNA polymerase (37). Cleavage by the 5' hammerhead ribozyme in the course of the transcription reaction was observed while further incubation at 60 °C was necessary to complete H₆V cleavage. Thus, the RNA of interest harbors a 5' hydroxyl and a 2',3' cyclic phosphodiester ends.

Crystallization, X-Ray Data Collection, Structure Determination, and Refinement.

Crystal growth was completed between a few weeks and few months at 20 °C in 400-nL droplets composed of 200 nL of a 100 μM solution of purified LC RNA transcripts mixed with the same volume of crystallization solutions from various high-throughput (HT) kits from Hampton or Jena BioSciences set up by using a mosquito robot (TPP Labtech). The CP LC ribozyme crystallized in IndexHT-D11-F11-F12-H6, NatrixHT-E11, and PEGRxHT-F1-F3, and the wild-type ribozyme crystallized in IndexHT-F12 and PEGRxHT-A10-C1-F3.

Diffraction data were collected at 100 K at the macromolecular crystallography beamlines X06DA at the Swiss Light Source and PROXIMA1 at Synchrotron SOLEIL. XDS was used for processing (38). Experimental

phase information of the CP LC ribozyme was obtained (Table S1) by a four-wavelength MAD experiment around the Iridium edge using the SHELX package (39). The wild-type LC ribozyme was solved by the molecular replacement method using Molrep (40) and the CP LC ribozyme crystal structure as a search model. The models were refined with Phenix (41) or Buster (42) and iteratively built by using Coot (43). Values of final R_{work}/R_{free} of 18.80%/23.57% and 20.99%/25.64% were obtained for the native and iridium derivative CP LC ribozymes, respectively. The wild-type form was refined to R_{work}/R_{free} of 24.55%/29.77%.

ACKNOWLEDGMENTS. We thank scientists at SOLEIL, Pierre Legrand, Andrew Thompson, and Jean Cavarelli for support in the synchrotron Block Allocation Group. X-ray data were collected at the Swiss Light Source—Paul Scherrer Institut and SOLEIL. We thank Marat Yusupov for the gift of osmium hexamine; Stéphane Bellemin-Laponaz for the gift of iridium hexamine; and Kelle Freel for English-language editing. This work was supported by the Centre National de la Recherche Scientifique. M.M. was supported by a University of Strasbourg PhD grant. This work was supported in part by L'Agence Nationale de la Recherche Project ANR-10-BLAN-1502-02 "GRP2CONF" (to E.V.), le Laboratoire d'Excellence Project ANR-11-LABX-0057-MITOCROSS, and the Danish Council for Independent Research in Natural Sciences (to H.N.).

- Masquida B, Beckert B, Jossinet F (2010) Exploring RNA structure by integrative molecular modelling. *New Biotechnol* 27(3):170–183.
- Cheah MT, Wachter A, Sudarsan N, Breaker RR (2007) Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. *Nature* 447(7143):497–500.
- Lee ER, Baker JL, Weinberg Z, Sudarsan N, Breaker RR (2010) An allosteric self-splicing ribozyme triggered by a bacterial second messenger. *Science* 329(5993):845–848.
- Nielsen H, Beckert B, Masquida B, Johansen SD (2008) The GIR1 branching ribozyme. *Ribozymes and RNA catalysis*, eds Lilley DMJ, Eckstein F (Royal Society of Chemistry, London), pp 229–252.
- Nielsen H, Westhof E, Johansen S (2005) An mRNA is capped by a 2', 5' lariat catalyzed by a group I-like ribozyme. *Science* 309(5740):1584–1587.
- Johansen SD, Haugen P, Nielsen H (2007) Expression of protein-coding genes embedded in ribosomal DNA. *Biol Chem* 388(7):679–686.
- Vader A, Nielsen H, Johansen S (1999) In vivo expression of the nucleolar group I intron-encoded I-dir1 homing endonuclease involves the removal of a spliceosomal intron. *EMBO J* 18(4):1003–1013.
- Vader A, Johansen S, Nielsen H (2002) The group I-like ribozyme DiGIR1 mediates alternative processing of pre-rRNA transcripts in *Didymium iridis*. *Eur J Biochem* 269(23):5804–5812.
- Nielsen H, et al. (2003) The ability to form full-length intron RNA circles is a general property of nuclear group I introns. *RNA* 9(12):1464–1475.
- Johansen S, Vogt VM (1994) An intron in the nuclear ribosomal DNA of *Didymium iridis* codes for a group I ribozyme and a novel ribozyme that cooperate in self-splicing. *Cell* 76(4):725–734.
- Johansen S, Einvik C, Nielsen H (2002) DiGIR1 and NaGIR1: Naturally occurring group I-like ribozymes with unique core organization and evolved biological role. *Biochimie* 84(9):905–912.
- Beckert B, et al. (2008) Molecular modelling of the GIR1 branching ribozyme gives new insight into evolution of structurally related ribozymes. *EMBO J* 27(4):667–678.
- Nielsen H, Einvik C, Lentz TE, Hedegaard MM, Johansen SD (2009) A conformational switch in the DiGIR1 ribozyme involved in release and folding of the downstream I-Dir1 mRNA. *RNA* 15(5):958–967.
- Meyer M, Masquida B (2014) cis-Acting 5' hammerhead ribozyme optimization for in vitro transcription of highly structured RNAs. *Methods Mol Biol* 1086:21–40.
- Price SR, Ito N, Oubridge C, Avis JM, Nagai K (1995) Crystallization of RNA-protein complexes. I. Methods for the large-scale preparation of RNA suitable for crystallographic studies. *J Mol Biol* 249(2):398–408.
- Einvik C, Nielsen H, Westhof E, Michel F, Johansen S (1998) Group I-like ribozymes with a novel core organization perform obligate sequential hydrolytic cleavages at two processing sites. *RNA* 4(5):530–541.
- Tang Y, Nielsen H, Birgisdottir AB, Johansen S (2011) A natural fast-cleaving branching ribozyme from the amoebaflagellate *Naegleria pringsheimi*. *RNA Biol* 8(6):997–1004.
- Nissen P, Ippolito JA, Ban N, Moore PB, Steitz TA (2001) RNA tertiary interactions in the large ribosomal subunit: The A-minor motif. *Proc Natl Acad Sci USA* 98(9):4899–4903.
- Correll CC, et al. (1998) Crystal structure of the ribosomal RNA domain essential for binding elongation factors. *Proc Natl Acad Sci USA* 95(23):13436–13441.
- Adams PL, Stahley MR, Kosek AB, Wang J, Strobel SA (2004) Crystal structure of a self-splicing group I intron with both exons. *Nature* 430(6995):45–50.
- Lipchock SV, Strobel SA (2008) A relaxed active site after exon ligation by the group I intron. *Proc Natl Acad Sci USA* 105(15):5699–5704.
- Goddard MR, Burt A (1999) Recurrent invasion and extinction of a selfish gene. *Proc Natl Acad Sci USA* 96(24):13880–13885.
- Hafez M, et al. (2013) The mtDNA rns gene landscape in the Ophiostomatales and other fungal taxa: Twintrons, introns, and intron-encoded proteins. *Fungal Genet Biol* 53:71–83.
- Zaher HS, Shaw JJ, Strobel SA, Green R (2011) The 2'-OH group of the peptidyl-tRNA stabilizes an active conformation of the ribosomal PTC. *EMBO J* 30(12):2445–2453.
- Lee CS, Mui TP, Silverman SK (2011) Improved deoxyribozymes for synthesis of covalently branched DNA and RNA. *Nucleic Acids Res* 39(1):269–279.
- Carlomagno T, et al. (2013) Structural principles of RNA catalysis in a 2'-5' lariat-forming ribozyme. *J Am Chem Soc* 135(11):4403–4411.
- Anokhina M, et al. (2013) RNA structure analysis of human spliceosomes reveals a compact 3D arrangement of snRNAs at the catalytic core. *EMBO J* 32(21):2804–2818.
- Inoue T, Sullivan FX, Cech TR (1985) Intermolecular exon ligation of the rRNA precursor of *Tetrahymena*: Oligonucleotides can function as 5' exons. *Cell* 43(2 Pt 1):431–437.
- Vicens Q, Cech TR (2009) A natural ribozyme with 3',5' RNA ligase activity. *Nat Chem Biol* 5(2):97–99.
- Draper WE, Hayden EJ, Lehman N (2008) Mechanisms of covalent self-assembly of the *Azoarcus* ribozyme from four fragment oligonucleotides. *Nucleic Acids Res* 36(2):520–531.
- Vaidya N, et al. (2012) Spontaneous network formation among cooperative RNA replicators. *Nature* 491(7422):72–77.
- Adams PL, et al. (2004) Crystal structure of a group I intron splicing intermediate. *RNA* 10(12):1867–1887.
- Golden BL, Kim H, Chase E (2005) Crystal structure of a phage Twort group I ribozyme-product complex. *Nat Struct Mol Biol* 12(1):82–89.
- Guo F, Gooding AR, Cech TR (2004) Structure of the *Tetrahymena* ribozyme: Base triple sandwich and metal ion at the active site. *Mol Cell* 16(3):351–362.
- Toor N, Keating KS, Taylor SD, Pyle AM (2008) Crystal structure of a self-spliced group II intron. *Science* 320(5872):77–82.
- Reiter NJ, et al. (2010) Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA. *Nature* 468(7325):784–789.
- Ichetovkin IE, Abramochkin G, Shrader TE (1997) Substrate recognition by the leucyl/phenylalanyl-tRNA-protein transferase. Conservation within the enzyme family and localization to the trypsin-resistant domain. *J Biol Chem* 272(52):33009–33014.
- Kabsch W (2010) Xds. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):125–132.
- Sheldrick GM (2008) A short history of SHELX. *Acta Crystallogr A* 64(Pt 1):112–122.
- Vagin A, Teplyakov A (2010) Molecular replacement with MOLREP. *Acta Crystallogr D Biol Crystallogr* 66(Pt 1):22–25.
- Adams PD, et al. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213–221.
- Bricogne G, et al. (2011) BUSTER (Global Phasing Ltd, Cambridge, U.K.), Version 2.11.5.
- Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60(Pt 12 Pt 1):2126–2132.
- Einvik C, Decatur WA, Embley TM, Vogt VM, Johansen S (1997) *Naegleria nucleolar* introns contain two group I ribozymes with different functions in RNA splicing and processing. *RNA* 3(7):710–720.
- Jabri E, Cech TR (1998) In vitro selection of the *Naegleria* GIR1 ribozyme identifies three base changes that dramatically improve activity. *RNA* 4(12):1481–1492.
- Leontis NB, Westhof E (2001) Geometric nomenclature and classification of RNA base pairs. *RNA* 7(4):499–512.
- Kleywegt GJ, Jones TA (1994) A super position. *CCP4/ESF-EACBM Newslett Protein Crystallogr* 31:9–14.

Supporting Information

Meyer et al. 10.1073/pnas.1322248111

SI Experimental Procedures

Preparation of the Circularly Permuted Form of the LCrz Ribozyme.

RNA transcripts were obtained by in vitro transcription using a His-tagged version of the T7-RNA polymerase (1). A DNA fragment obtained by PCR corresponding to the circularly permuted form of the LCrz inserted between a 5' hammerhead ribozyme and a 3' H δ V ribozyme (2) or to the wild-type ribozyme followed by an H δ V ribozyme were inserted in an EcoRI/BamHI-digested pUC19 (NEB) by using the InFusion cloning kit (Clontech). The PCR introduced a U203C mutation, resulting in the presence of an A-C base pair in P15. This mutation has been reported to stabilize the *Naegleria* lariat-capping (LC) ribozyme (3). Fast cleavage in the course of the transcription reaction by the 5' hammerhead ribozyme was possible because the sequence of stem I was changed to that of the LC ribozyme while preserving the tertiary interactions between stems I and II (4). One-hour incubation at 60 °C was necessary to complete H δ V cleavage. Thus, the RNA of interest presents 5'OH and 2',3' cyclic phosphodiester ends, which do not correspond to the branching reaction products, in the sense that the absence of the phosphate group of C230 prevents tethering. Consequently, we observe that the lariat is folded in the crystal structure of the circular permutation (CP) form but not 2',5'-tethered. The transcripts were purified by 8% polyacrylamide gel electrophoresis and identified by UV shadowing. Bands containing the RNA were collected, and the RNA was recovered by passive elution overnight and, finally, filtrated on dispensable sterile 0.2- μ m filtration units (Nalgene). The RNA sample was concentrated by ethanol precipitation.

Mutational Studies of the P9 Stem. Mutations in P9 were introduced by using the QuikChange site-directed mutagenesis kit (Stratagene). The starting point for introducing the mutations was a published plasmid carrying a WT ribozyme (5). The oligos for deletion of a base pair were as follows: C274, GTTCAATGAACAGTCGTCGAAAGGAGCATCCGGTATCC; and C275, GGATACCGGATGCTCCTTTTCGGACGACTGTTTCATTGAAC. For insertion of a base pair, they were C276, GTTCAATGAACAGTCGTTACCGAAAGGTAAGCATCCGGTATCC; and C277, GGATACCGGATGCTTACCTTTTCGGTAACGACTGTTTCATTGAAC, respectively. Templates were made by PCR of the mutagenized plasmids using a 5' oligo including a T7 promoter (C287, AATTTAATACGACTCACTATAGGGTTGGGAAGTATCAT, together with C288, TCACCATGGTTGTTGAAGTGCACAGATTG). In vitro transcription and cleavage analyses were performed as described (5).

Cleavage Studies of the CP P8 LC Ribozyme. The CP form of the LC ribozyme presenting 5' and 3' ends in P8 was produced by PCR using oligonucleotides (Sigma) to generate the template by primer extension at 62 °C as follows: CP_P8_F_L, GTTAGG-TGTGTTCAATGAACAGTCGTTCCGAAAGGAAGCATCCGGTATCCCAAGACAATCTTCGGGTTGGGTTGGGAAGTATCATGGCTAATCACCATGATGC; and CP_P8_R_L, GCCGT-TAGGACGGATGTTACGGGATCGTCCCACCGTTTTAACCAATTAAGTGTTCACCCGATTGCATCATGGTGATTAGCCATGATACTTCC. The template was amplified by 25 cycles of PCR at 56 °C using the following oligonucleotides: CP_P8_F_C, AATTTAATACGACTCACTATAGGGTTAGG-TGTGTTCAATGAACAGTC; and CP_P8_F_C, GGGTTAGG-CAGGGCCGTGCAGTCTGTGCCGTTAGGACGGATGT-TACGG. The purified template was transcribed in vitro and further purified as other RNAs used in this study.

Kinetics experiments were carried out as described (6).

Crystallization of the LCrz. A 100 μ M RNA solution was heated for 1 min at 95 °C followed by 4 min on ice. MgCl₂ at a concentration of up to 4 mM was then added, and the solution was placed at 55 °C for 10 min and then at 4 °C overnight. Crystal growth was completed between a few weeks and few months at 20 °C in 400-nL droplets composed of 200 nL of a 100 μ M solution of purified LC RNA transcripts mixed with the same volume of crystallization solutions. Initial conditions were obtained from commercial high-throughput screening kits from Hampton Research (PEGRx, Index, Natrix) and Jena Biosciences (Nuc-Pro) set up with a mosquito robot (TTP Labtech). The CP LC ribozyme crystallized in IndexHT-D11-F11-F12-H6, NatrixHT-E11, and PEGRxHT-F1-F3, and the wild-type ribozyme crystallized in IndexHT-F12 and PEGRxHT-A10-C1-F3.

Data Collection, Structure Determination, and Model Refinement.

Crystals were flash-frozen by immersion in liquid nitrogen or ethane. Diffraction data were collected at 100 K at the macromolecular crystallography beamline X06DA at the Swiss Light Source (SLS) or PROXIMA1 at Synchrotron SOLEIL, and processed by using XDS (7). Phase information was determined with data obtained from crystal 1 (Table S1) by a four-wavelength multiwavelength anomalous diffraction (MAD) experiment around the iridium edge using the SHELX package (8). The wild-type LC ribozyme was solved by the molecular-replacement method using Molrep (9) with the structure of the CP form of the LC ribozyme solved by MAD as a search model.

The resulting experimental electron-density map allowed us to build an initial model that was later refined with Phenix (10) and iteratively built by using Coot (11) against higher-resolution data obtained from crystal 2. Values of final $R_{\text{work}}/R_{\text{free}}$ of 18.80%/23.57% and 19.81%/26.21% were obtained for the native and iridium derivative CP LC ribozymes, respectively. The wild-type form was refined with Buster (12) and Phenix (10) to $R_{\text{work}}/R_{\text{free}}$ of 24.55%/29.77% (Table S1). Density around J6/7 region as well as the first nucleotides of P7 was poorly defined but became visible when the structure was refined against the data from crystal.

Model statistics were obtained with MolProbity (13). Molecular visualizations and structures illustrations were performed by using Pymol (14). Data collection and refinement statistics are given in Table S1. Coordinates and structure factors have been deposited in the Protein Data Bank (PDB ID codes 4P9R, 4P95, and 4P8Z).

The Unique Architecture of the LC Ribozyme Is Supported by Small-Angle X-Ray Scattering on both Wild-Type and CP Ribozymes.

Small-angle X-ray scattering (SAXS) measurements were undertaken to compare the crystal structures to the structures of LC ribozyme in solution. Both LC ribozyme forms show very similar SAXS shapes (Fig. S6), with each ribozyme perfectly fitting its cognate SAXS data. The wild-type ribozyme is slightly more compact than the CP form due to the absence of the UUCG loop closing DP2. These observations indicate that the crystal environment does not perturb the conformations adopted by the ribozymes in solution. The values for R_g Guinier and D_{max} for the CP LC and for the WT LC ribozymes are 29.8, 100, 30.9, and 105 Å, respectively. Both forms have the same volume (88,000 Å³).

SAXS Data Collection and Analysis. X-ray scattering data were collected on beamline SWING at Synchrotron SOLEIL (Gif-sur-Yvette, France) with the 17 × 17-cm² low-noise Avix CCD

detector positioned at a distance of 1,800 mm from the sample. The beam wavelength was set to $\lambda = 1.033 \text{ \AA}$, and the resulting exploitable q range was $0.014\text{--}0.4 \text{ \AA}^{-1}$, where $q = 4\pi \sin \theta/\lambda$, and 2θ is the scattering angle. SAXS data from pure LC ribozyme samples were collected directly at the outlet of a size-exclusion chromatography HPLC system (15). To remove the aggregation contribution, concentrated solutions of CP or wild-type LC ribozyme (25 μL , 8 mg/mL) were injected into a Size Exclusion Column (Agilent Bio-sec3) by using an Agilent HPLC system and eluted directly into the SAXS flow-through capillary cell at a flow rate of 200 $\mu\text{L}/\text{min}$. SAXS data were collected online throughout the whole elution time, with frame duration of 1 s to prevent X-ray damage and a dead time between frames of 0.5 s. The frames corresponding to the main elution peak were inspected and averaged by using FOXTROT, a homemade program dedicated to data processing. In total, 20 frames were averaged to produce a curve with the same q range as previously, with no attractive interactions at low angles.

All data-processing, analysis, and modeling steps were carried out with PRIMUS (16) and other programs of the ATSAS suite (17). The radius of gyration R_g was evaluated by using the Guinier approximation (18), assuming that ($q < 1/3 R_g$), intensity may be represented as $I(q) = I(0) \exp(-(q \cdot R_g)^2/3)$ at very small angles. The same parameter was also calculated from the entire scattering pattern by using the indirect transform package

GNOM (19), which also provides the distance distribution function $p(r)$ of the particle and the maximal intramolecular distance (D_{max}). All parameters for the CP and wild-type LC ribozymes show no significant difference.

Low-resolution models of CP or wild-type LC ribozymes were generated ab initio, by using the dummy atom model method, as implemented in the program DAMMIF (20), which uses simulated annealing to search a compact model that fits the experimental data to minimize the discrepancy. The low-resolution models were compared with DAMAVER, a procedure to analyze and align ab initio models to select the most typical one (21). The superimposition of low-resolution molecular shapes and crystal structures were performed with SUPCOMB, a method to optimize the alignment of 3D objects (22).

Scattering from the crystal structures of the CP and wild-type forms of the LC ribozyme were calculated by using CRY SOL (23) and compared with the experimental curves obtained for each form of the LC ribozyme. The program fits the experimental intensity $I(q)$ with the atomic coordinates by adjusting the excluded volume, averaged atomic radius, and contrast of the hydration layer surrounding the particle in solution to minimize discrepancy as described by $I(q)$. Both forms of the ribozyme are highly similar, with χ^2 values close to 1, and confirm that the crystal environment does not affect the conformations of the ribozymes in solution.

1. Ichetovkin IE, Abramochkin G, Shrader TE (1997) Substrate recognition by the leucyl/phenylalanyl-tRNA-protein transferase. Conservation within the enzyme family and localization to the trypsin-resistant domain. *J Biol Chem* 272(52):33009–33014.
2. Price SR, Ito N, Oubridge C, Avis JM, Nagai K (1995) Crystallization of RNA-protein complexes. I. Methods for the large-scale preparation of RNA suitable for crystallographic studies. *J Mol Biol* 249(2):398–408.
3. Jabri E, Cech TR (1998) In vitro selection of the Naegleria GIR1 ribozyme identifies three base changes that dramatically improve activity. *RNA* 4(12):1481–1492.
4. Martick M, Scott WG (2006) Tertiary contacts distant from the active site prime a ribozyme for catalysis. *Cell* 126(2):309–320.
5. Einvik C, Nielsen H, Westhof E, Michel F, Johansen S (1998) Group I-like ribozymes with a novel core organization perform obligate sequential hydrolytic cleavages at two processing sites. *RNA* 4(5):530–541.
6. Nielsen H, Westhof E, Johansen S (2005) An mRNA is capped by a 2', 5' lariat catalyzed by a group I-like ribozyme. *Science* 309(5740):1584–1587.
7. Kabsch W (2010) Xds. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):125–132.
8. Sheldrick GM (2008) A short history of SHELX. *Acta Crystallogr A* 64(Pt 1):112–122.
9. Vagin A, Teplyakov A (2010) Molecular replacement with MOLREP. *Acta Crystallogr D Biol Crystallogr* 66(Pt 1):22–25.
10. Adams PD, et al. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213–221.
11. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60(Pt 12 Pt 1):2126–2132.
12. Bricogne G, et al. (2011) BUSTER (Global Phasing Ltd., Cambridge, U.K.), Version 2.11.5.
13. Chen VB, et al. (2010) MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* 66(Pt 1):12–21.
14. Schrodinger, LLC (2010) The PyMOL Molecular Graphics System, Version 1.3r1.
15. David G, Perez J (2009) Combined sampler robot and high-performance liquid chromatography: A fully automated system for biological small-angle X-ray scattering experiments at the Synchrotron SOLEIL SWING beamline. *J Appl Cryst* 42(5):892–900.
16. Konarev PV, Volkov VV, Sokolova AV, Koch MHJ, Svergun DI (2003) PRIMUS: A Windows PC-based system for small-angle scattering data analysis. *J Appl Cryst* 36(5):1277–1282.
17. Konarev PV, Petoukhov MV, Volkov VV, Svergun DI (2006) ATSAS 2.1, a program package for small-angle scattering data analysis. *J Appl Cryst* 39(2):277–286.
18. Guinier A (1939) La diffraction des rayons X aux très petits angles: Application à l'étude de phénomènes ultra microscopiques. *Ann Phys* 12:161–237.
19. Svergun D (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J Appl Cryst* 25(4):495–503.
20. Franke D, Svergun DI (2009) DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J Appl Cryst* 42(2):342–346.
21. Volkov VV, Svergun DI (2003) Uniqueness of ab initio shape determination in small-angle scattering. *J Appl Cryst* 36(3 Part 1):860–864.
22. Kozin MB, Svergun DI (2001) Automated matching of high- and low-resolution structural models. *J Appl Cryst* 34(1):33–41.
23. Svergun D, Barberato C, Koch MHJ (1995) CRY SOL - A program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. *J Appl Cryst* 28(6):768–773.

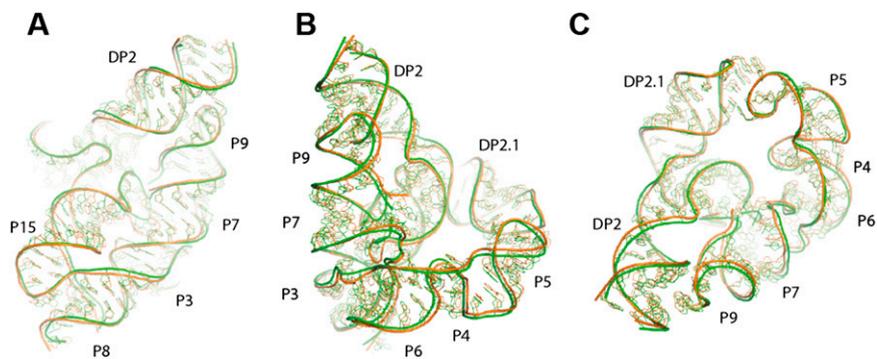


Fig. S1. Different views of the superimposition between the CP (orange) and wild-type (green) forms of the LC ribozyme crystal structures. The rmsd value of 3.63 Å between the 4,013 atoms common to both structures was calculated in Chimera (1). The presented images show a very good superimposition between the backbone traces, which indicates that the overall folds are similar. The fairly high rmsd value is due to the fact that entire domains undergo slight translational motions and that the junctions within the core adopt slightly different paths. (A) The ribozyme is depicted with the double-pseudoknot P3–P7–P15 in the foreground. (B) A clockwise rotation of 90° on a vertical axis reveals the original shape of the P4–P6 domain and the tertiary interaction between L5 and L2.1. (C) A 120° clockwise rotation of A shows the hollow resulting from the curvature of the P4–P6 and DP2.1 domains.

1. Yang Z, et al. (2012) UCSF Chimera, MODELLER, and IMP: An integrated modeling system. *J Struct Biol* 179(3):269–278.

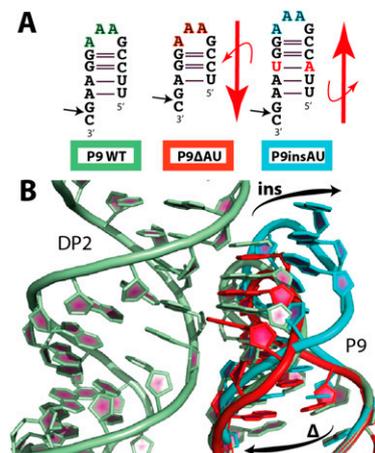


Fig. 54. P9 mutants demonstrate the role of the P2/L9 interaction. (A and B) Deleting or adding one base pair to P9 translates and rotates L9 and disturbs its interaction with DP2. (A) Black arrows point at the cleavage site. (B) Insertion of one base pair orients the loop adenine residues away from the receptor, whereas deletion of one base pair apparently maintains the interaction of two adenine residues. (C) These effects are suggested by the catalytic tests determined under single turnover conditions. Molecular models from the P9 mutants were obtained with the Assemble program suite (1) coupled to Chimera (2).

- Jossinet F, Ludwig TE, Westhof E (2010) Assemble: An interactive graphical tool to analyze and build RNA architectures at the 2D and 3D levels. *Bioinformatics* 26(16):2057–2059.
- Yang Z, et al. (2012) UCSF Chimera, MODELLER, and IMP: An integrated modeling system. *J Struct Biol* 179(3):269–278.

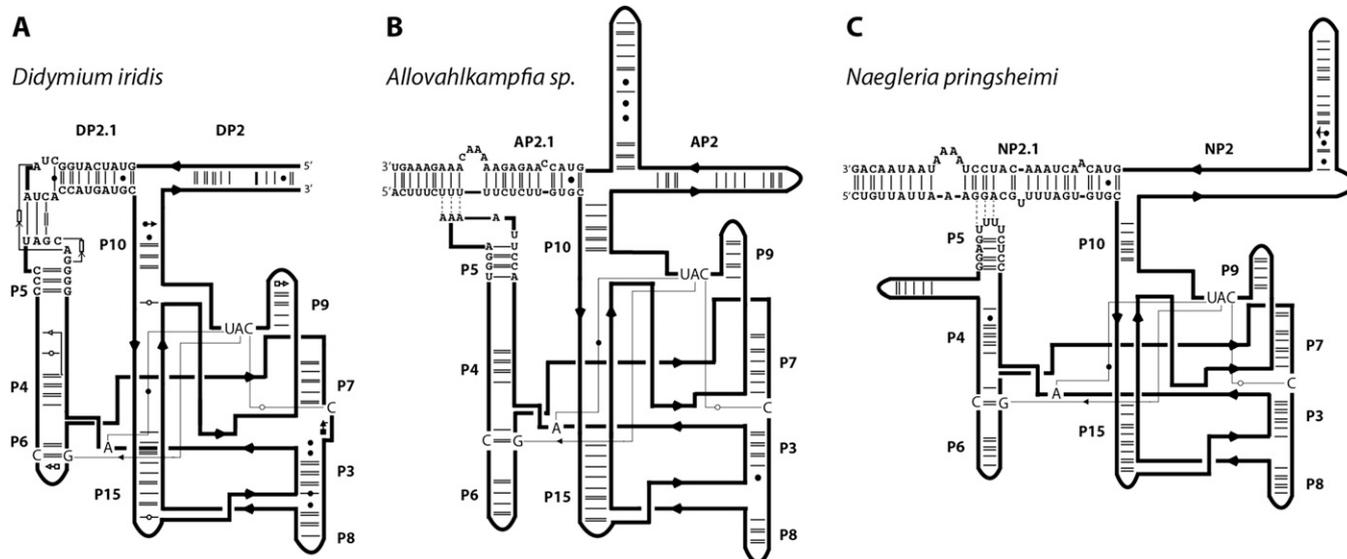


Fig. 55. Phylogenetic conservation of the L2.1/L5 interaction in the class LC ribozymes. (A) Secondary structure with tertiary interactions observed between L2.1 and L5 in the crystal structure of the *Didymium iridis* LC ribozyme. (B) Secondary structure of the *Allovahlkampfia sp.* LC ribozyme with the proposed tertiary interaction between AP2.1 and L5 (dashed lines). (C) Secondary structure of the *Naegleria sp.* NG393 LC ribozyme. The stem P2.1 region involved in the tertiary interaction with the loop of P5 intervenes after one helical turn or beyond (10–12 bp) from the three-way junction formed between P2, P2.1, and P10. The tertiary interactions fully conserved in all LC ribozyme sequences known to date, mediated between the residues from the lariat fold and the core of the ribozyme, are indicated using the Leontis–Westhof nomenclature (1).

- Leontis NB, Westhof E (2001) Geometric nomenclature and classification of RNA base pairs. *RNA* 7(4):499–512.

Table S1. Statistics on crystallographic data collection and refinement

	Crystal 1				Crystal 2	Crystal 3
	Peak	Inflection	High remote	Low remote		
Data collection						
Space group	P2 ₁ 2 ₁ 2 ₁					
Unit cell a, Å	59.44	59.44	59.39	59.47	57.63	52.56
b, Å	86.26	86.32	86.10	86.38	85.71	88.86
c, Å	109.78	109.82	109.65	109.88	108.53	105.70
Beamline	X06DA	X06DA	X06DA	X06DA	X06DA	PROXIMA1
	SLS	SLS	SLS	SLS	SLS	SOLEIL
Wavelength, Å	1.10530	1.10564	1.09553	1.11525	1.6	0.98011
Resolution limits, high-resolution shell, Å	48.9–2.9	48.9–2.9	48.9–2.7	48.9–2.9	47.8–2.5	28.55–3.85
	3.0–2.9	3.0–2.9	2.8–2.7	3.0–2.9	2.6–2.5	3.95–3.85
Reflections: measured unique	82,329	82,184	207,712	85,513	494,356	34,494
	23,432	23,367	29,825	24,203	34,693	5,021
Completeness, * %	96.8 (69.3)	96.4 (65.9)	99.9 (99.9)	99.6 (99.7)	96.4 (94.5)	99.6 (100)
R_{meas} , * %	15.1 (92.1)	16.7 (114.6)	18.5 (159.7)	16.9 (124.9)	10.4 (110.9)	14.7 (247.6)
I/σ *	9.81 (1.49)	9.27 (1.19)	12.38 (1.32)	9.33 (1.19)	24.88 (2.49)	9.63 (0.88)
CC(1/2)*	99.3 (57.4)	99.2 (42.5)	99.6 (45.6)	99.3 (40.9)	99.9 (76.5)	99.9 (70.8)
Phasing	Four-wavelengths Ir MAD					MR
Resolution range used, Å	48.9–3.4					
Heavy atom sites [†]	4 Ir					
Correlation coefficient (all/weak)	47.0/32.2					
FOM [‡]	0.7					
Pseudofree CC [‡]	72.23					
Refinement						
Reflections $R_{\text{cryst}}/R_{\text{free}}$			29,782/1,470		34,663/1,968	4,759/251
R_{cryst} , %			20.99		18.80	24.55
R_{free} , % [§]			25.64		23.57	29.77
RNA atoms heterogen atoms			4076		4076	4014
			90		33	0
Water molecules			21		88	0
Mean B factor RNA, Å ²			45.04		18.79	174.95
Mean B factor hetero, Å ²			68.69		17.36	—
Mean B factor solvent, Å ²			38.70		8.98	—
Coordinate error, ML-based			0.40		0.38	0.61
rmsd bond length, Å			0.002		0.002	0.001
rmsd angles length, °			0.504		0.629	0.842
PDB ID code			4P9R		4P95	4P8Z

ML, maximum likelihood.

*Values in parentheses correspond to high-resolution shell in data collections.

[†]Shelxd output.

[‡]Shelxe output.

[§]5% of the reflections were set aside for an R_{free} test before initiating any refinement.