

Molecular modelling of the GIR1 branching ribozyme gives new insight into evolution of structurally related ribozymes

This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits distribution, and reproduction in any medium, provided the original author and source are credited. This license does not permit commercial exploitation or the creation of derivative works without specific permission.

Bertrand Beckert¹, Henrik Nielsen^{2,3,*},
Christer Einvik⁴, Steinar D Johansen³,
Eric Westhof¹ and Benoît Masquida^{1,*}

¹Architecture et Réactivité de l'ARN, Université Louis Pasteur de Strasbourg, IBMC, CNRS, Strasbourg, France, ²Department of Cellular and Molecular Medicine, The Panum Institute, University of Copenhagen, Copenhagen, Denmark, ³Department of Molecular Biotechnology, Institute of Medical Biology, University of Tromsø, Tromsø, Norway and ⁴Department of Pediatrics, University Hospital of North Norway, Tromsø, Norway

Twin-ribozyme introns contain a branching ribozyme (GIR1) followed by a homing endonuclease (HE) encoding sequence embedded in a peripheral domain of a group I splicing ribozyme (GIR2). GIR1 catalyses the formation of a lariat with 3 nt in the loop, which caps the HE mRNA. GIR1 is structurally related to group I ribozymes raising the question about how two closely related ribozymes can carry out very different reactions. Modelling of GIR1 based on new biochemical and mutational data shows an extended substrate domain containing a GoU pair distinct from the nucleophilic residue that dock onto a catalytic core showing a different topology from that of group I ribozymes. The differences include a core J8/7 region that has been reduced and is complemented by residues from the pre-lariat fold. These findings provide the basis for an evolutionary mechanism that accounts for the change from group I splicing ribozyme to the branching GIR1 architecture. Such an evolutionary mechanism can be applied to other large RNAs such as the ribonuclease P.

The EMBO Journal (2008) 27, 667–678. doi:10.1038/emboj.2008.4; Published online 24 January 2008

Subject Categories: RNA

Keywords: GIR1 branching ribozyme; group I intron; RNA evolution; 3D modelling of RNA

Introduction

The list of naturally occurring ribozymes comprises a few that are fundamental for cellular life (the ribosome, RNase P, and possibly the spliceosome), two types of splicing

ribozymes that are abundant in organellar and microbial genomes (within group I and group II introns), and a number of cleavage ribozymes with a sporadic occurrence in viroids, plant satellite RNAs, bacteria and, more recently, within the human genome (hammerhead, hairpin, VS, HDV, glmS, and the CPEB3 ribozymes). Apart from the ribosome, all naturally occurring ribozymes catalyse phosphor transfer reactions (Ditzler *et al*, 2007; Scott, 2007; Serganov and Patel, 2007). A recent addition to the list is the GIR1 branching ribozyme. This ribozyme (Figure 1) catalyses cleavage of the RNA chain by transesterification resulting in the formation of a 2',5' phosphodiester bond between the first and the third nucleotide of the 3'-cleavage product. The downstream cleavage product is an mRNA encoding a homing endonuclease (HE) that is thereby capped with a lariat containing 3 nt in the loop (Nielsen *et al*, 2005). Both GIR1 and the downstream HE mRNA are inserted into a peripheral domain of a regular splicing ribozyme (GIR2) making up the characteristic configuration of a twin-ribozyme intron. Such introns have so far only been found in the SSU rDNA genes of a unique isolate of *Didymium iridis* and in several *Naegleria* strains where it has been vertically inherited from a common ancestor (Johansen *et al*, 2002; Wikmark *et al*, 2006; Nielsen *et al*, 2008). The biological function of GIR1 appears to be in the formation of the 5' end of the HE mRNA during processing from the spliced out intron and the resulting lariat cap seems to contribute by increasing the half-life of the HE mRNA (Vader *et al*, 1999; Nielsen *et al*, 2005), thus conferring an evolutionary advantage to the HE.

One of the interesting features of GIR1 is that the sequence and the secondary structure are very similar to that of eubacterial group IC3 introns at the second step of splicing (Figure 2), suggesting an evolutionary relationship with this specific subgroup of splicing ribozymes (Johansen *et al*, 2002). The secondary structure of GIR1 displays paired segments numbered P3–P10, similar to what is known in group I introns (Figure 2). The paired segments are generally shorter than those observed in group I introns consistent with the fact that the shortest form of DiGIR1 shown to catalyse branching *in vitro* is only 179 nt (Nielsen *et al*, 2005). Both ribozymes are organized as a compact bundle of three helical stacks (Figure 2; domains P3–P9, P4–P6, and P10–P2 (group I ribozyme) or P10–P15 (GIR1)). Group I intron classification is based on structural variation of peripheral elements organized around a very well-conserved catalytic core (Michel and Westhof, 1990). In contrast, the main distinctive features of GIR1 towards group I ribozymes occur within the catalytic core. Several characteristic single-stranded junctions tether the helices of a catalytic core containing a double pseudoknot in a way that leads to significant topological modifications

*Corresponding authors. B Masquida or H Nielsen, Architecture et Réactivité de l'ARN, Université Louis Pasteur de Strasbourg, IBMC, CNRS, 15 rue René Descartes, Strasbourg 67084, France. Tel.: +333 88 41 70 45; Fax: +333 88 60 22 18; E-mail: b.masquida@ibmc.u-strasbg.fr or Hamra@imbg.ku.dk

Received: 10 August 2007; accepted: 4 January 2008; published online 24 January 2008

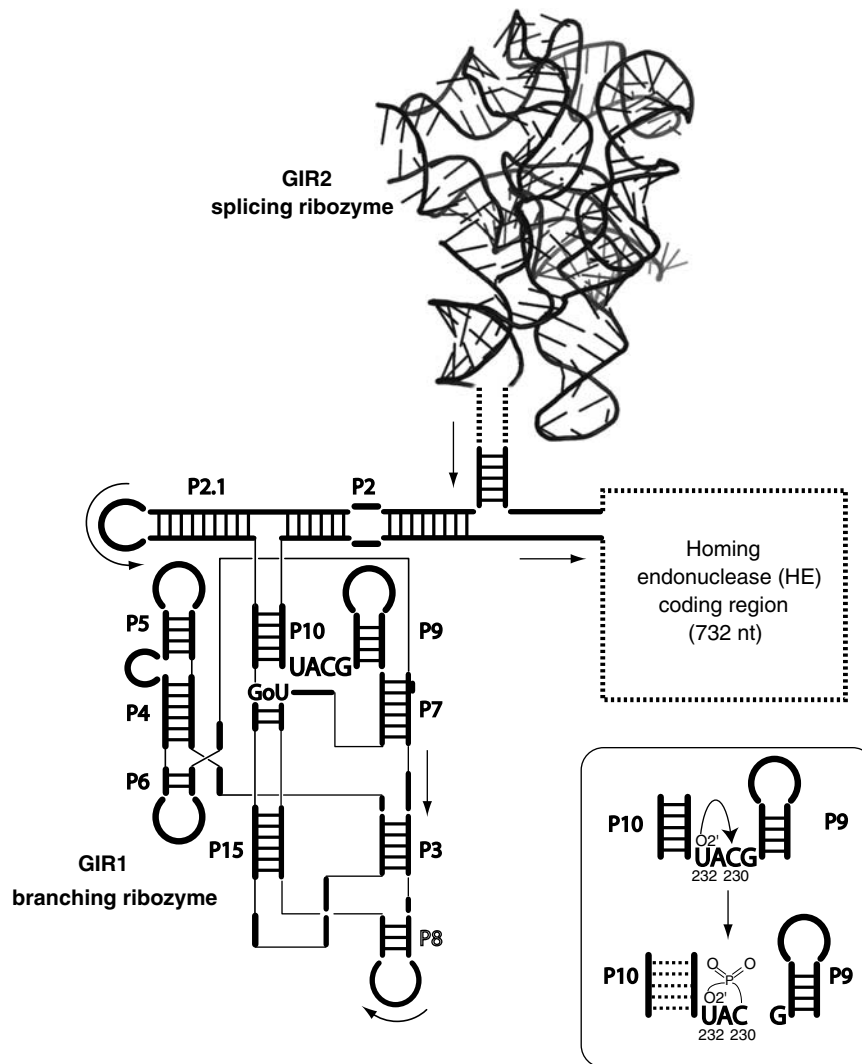


Figure 1 The RNA transcript from the *Dir.S956-1* locus is composed of a regular group I splicing ribozyme (GIR2) represented as a 3D ribbon generated from Guo *et al* (2004) containing an ORF encoding an homing endonuclease (HE, dashed line) preceded by a branching ribozyme (GIR1, schematic secondary structure). The branching reaction consists of the O^{2'}-hydroxyl group from U232 attacking the phosphate group from C230 leading to the formation of a 3-nt lariat (box). The branching reaction releases the cleaved ribozyme from the 5' end of the HE mRNA. In addition to stems P3–P15, DiGIR1 harbours a specific extension, namely P2/P2.1 (Einvik *et al*, 2000), which could not be included unambiguously into the present model.

(Figure 2). GIR1 harbours a substrate domain different from the canonical P1 and P2 elements. However, the biochemical and mutational data presented in this study support the idea that they are replaced by a distinctive and unique 9-bp P15 stem starting with a GoU pair that should be able to dock onto the catalytic core in a way similar to that observed for group I introns. The close resemblance of GIR1 to a splicing ribozyme in an unrelated group of organisms and the structural organization of twin-ribozyme introns may be related to the propagation of group I introns by horizontal transfer. Group I introns are considered as mobile elements due to their sporadic occurrence in a wide variety of organisms, including protists, fungal mitochondria, bacteria, and phages (Haugen *et al*, 2005). Many lines of evidence point to reverse splicing and homing as mechanisms by which group I introns can transfer horizontally (Goddard *et al*, 2001; Bhattacharya *et al*, 2005; Haugen *et al*, 2005). The homing mechanism is well documented and appears to be particularly relevant to

GIR1 because its activity is intimately related to the expression of a HE mRNA (H Nielsen, in preparation).

The intriguing observation that GIR1 and the group I splicing ribozymes are structurally related, yet carry out different reactions (splicing versus branching) prompted us to revise our previous structural model of GIR1. This model (Einvik *et al*, 1998b; Johansen *et al*, 2002) was based on structure probing and mutational studies. It predated the discovery of the branching reaction (Nielsen *et al*, 2005) and could not account for this reaction. The model presented in this study is based on new mutational data and furthermore benefits from the recent crystal structures of various group I ribozymes (Guo *et al*, 2004; Adams *et al*, 2004a,b; Golden *et al*, 2005) in the sense that the GIR1 regions organized identically in group I introns could be modelled more accurately. In our new model, residues that are key to the branching reaction lie within a pocket formed at the interface of P10, P15, P7, and J5/4. All the distinctive features

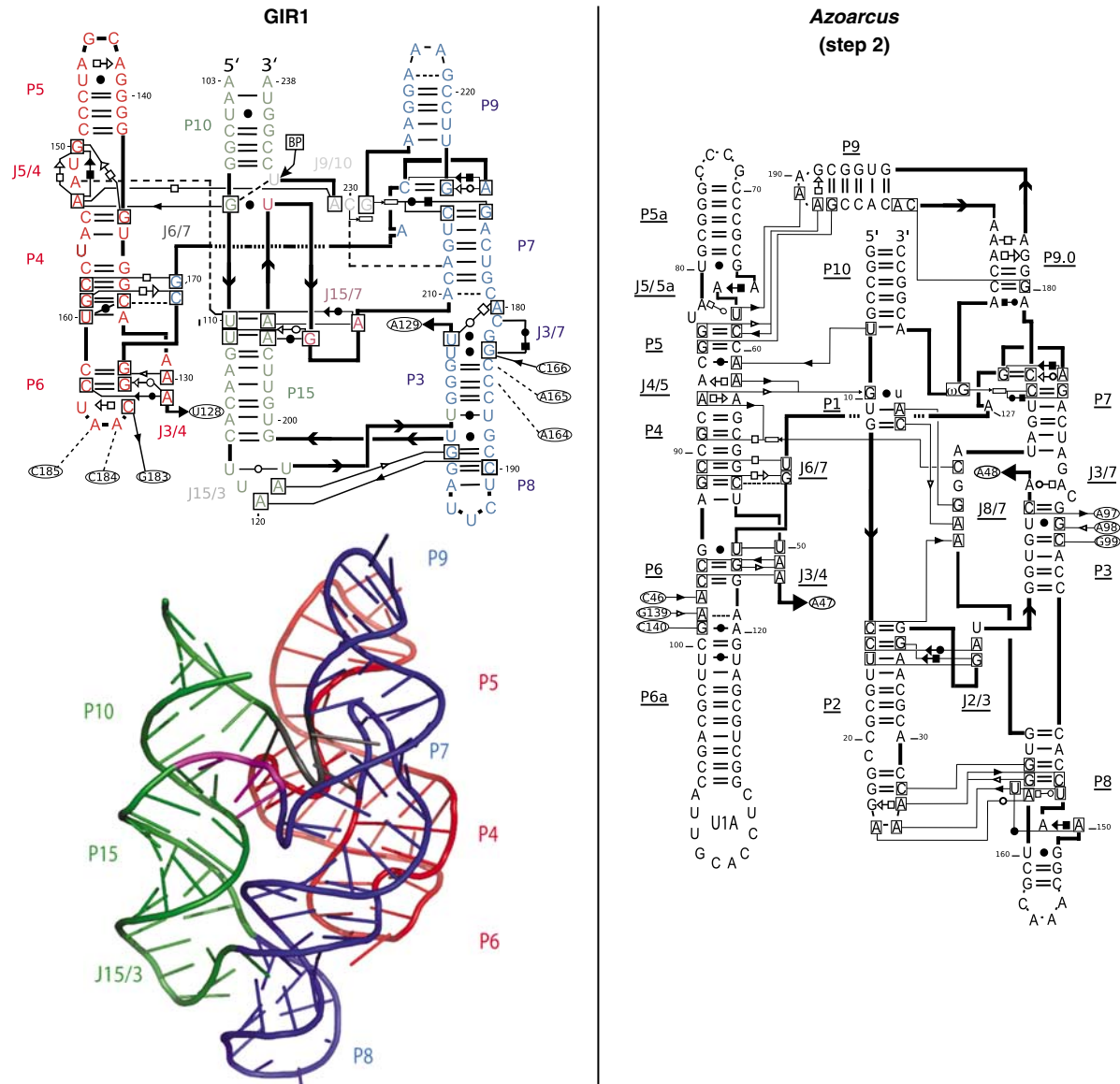


Figure 2 Overall representation of *Didymium* GIR1 secondary and 3D structures (left panel) and *Azoarcus* group I tRNA^{Ile} intron (*Azo*) secondary structure in (pre-) step 2 state (right panel). The secondary structure corresponding to the crystal structure model of Adams *et al* (2004a) harbours a U1A protein receptor motif. The non-canonical interactions are displayed on both secondary structures using the formalism elaborated by Leontis and Westhof (2001). The three helical domains of GIR1 (P10–P15 corresponding to *Azo* P10–P2, P4–P6, and P3–P9) are organized as a compact bundle at the centre of which lies the junction harbouring the residues involved in catalysis (J9/10, black ribbon). The overall architecture is stabilized by contacts recurrently observed in group I ribozymes, J15/3–P8 corresponding to L2–P8 in *Azo*, and L6–P3 corresponding to J6/6a–P8 in *Azo*. The contact between L9 and P5 is not observed in GIR1 despite the presence of the tetraloop sequence 5' GAAA in L9. Topological differences are due to the presence of the double pseudoknot involving P3, P7, and P15 in GIR1.

of GIR1 concentrate in this pocket and result in a topology very different from what is observed in group I ribozyme crystal structure models. The structure of the critical J8/7 segment of group I introns is dramatically changed and has been partly replaced by residues belonging to the GIR1 lariat fold J9/10. Other key features are the detachment of the nucleophile from the GoU pair at the catalytic site and a structural alteration of the GoU pair receptor. Taken together, these structural differences account for the different chemical reaction catalysed by GIR1. Comparison of the models of the *Azoarcus* tRNA^{Ile} intron at the second step of splicing and GIR1 suggests a relatively simple model for the conversion of the topology of one ribozyme to the other based on strand mispairing. Similar scenarios can apply to other RNAs,

for example, RNase P, and could constitute a general way of viewing the evolution of RNA molecules.

Results

In this section, the structure model of the *Didymium* GIR1 (DiGIR1) ribozyme is extensively compared with the *Azoarcus* group I ribozyme (*Azo*) crystal structure (Adams *et al*, 2004a). Hence, secondary structure elements and nucleotides corresponding to *Azo* are underlined throughout the text. The secondary structure of DiGIR1 and the similar ribozyme from *Naegleria* (NaGIR1) is generally supported by enzymatic and chemical probing (Einvik *et al*, 1998a; Jabri *et al*, 1997; Jabri and Cech, 1998). Furthermore, the *Naegleria*

structure is supported by covariations observed in most of the helical stems when NaGIR1 from different strains are compared (Johansen *et al*, 2002; Wikmark *et al*, 2006). NaGIR1 performs a branching reaction similar to that of DiGIR1 (H Nielsen, unpublished data) supporting the notion that the two GIR1 ribozymes adopt similar secondary structures. DiGIR1 harbours an additional domain P2/P2.1 (Einvik *et al*, 2000) not found in NaGIR1. This domain is excluded from the model because it is currently impossible to discriminate between several different models.

Extension of the P15 stem

Modelling of GIR1 is facilitated by the presence of a double pseudoknot at the core. In addition to the P3–P7 pseudoknot also found in all group I ribozymes (Michel and Westhof, 1990), a second pseudoknot, P3–P15, is found as a characteristic feature of GIR1 (Einvik *et al*, 1998b). P15 arises from base-pairing interactions between the 5' strand of P2 with residues that could be derived from the 3' strand of P8 and from J8/7, while the 3' strand of P2 has been shortened and now makes up the J15/3 segment (Figure 3). Thus, one can visualize P15 as replacing the shallow/minor groove interactions taking place between J8/7 and P2, which are conserved in group I ribozymes (Strauss-Soukup and Strobel, 2000; Soukup *et al*, 2002). Inspired by the comparison with *Azo*, we now propose an extension of P15 involving residues 205–207. Residues A205 and A206 appear to be equivalent to A residues in J8/7 responsible for recognition of the P1–P2 substrate (Figure 2). J8/7 is a highly conserved joining segment in group I ribozymes that is part of the active site and makes contacts with all of the three principal domains of the group I ribozyme. During the first and second steps of splicing, the two conserved adenosines at the 5' end of J8/7 are involved in recognition of the P1–P2 interface (Pyle *et al*, 1992; Tanner *et al*, 1997; Strauss-Soukup and Strobel, 2000; Adams *et al*, 2004a). In the original model of GIR1, a P1 was not included but could arise from a 3-bp extension of P15

resulting from the base complementarity between residues A205–U207 from J15/7 and U111–G109 separating P10 from P15, respectively. The existence of these three base pairs could originate from the interaction between a P1 having lost its 5' exon making the residues from the internal guide sequence (IGS) prone to base pair with J8/7. Indeed, the crystal structure of the *Tetrahymena* group I ribozyme (Guo *et al*, 2004) shows that residues in J8/7 are directed towards the solvent when the substrate domain P1–P2 is absent. It is therefore likely that some J8/7 residues could form Watson–Crick interactions with a substrate domain containing unpaired nucleotides as it occurs when the IGS is separated from the 5' exon. To confirm this possibility, disruptive and restoring mutations of the central base pair U110–A206 were tested by kinetic cleavage analyses (Supplementary Figure S1). *In vitro*, GIR1 catalyses (i) a forward branching reaction in equilibrium with (ii) a very efficient reversed reaction, and (iii) an inefficient hydrolytic cleavage reaction (Nielsen *et al*, 2005; Nielsen and Johansen, 2007). The outcome of the reaction can be analysed by primer extension with stop signals at the branch nucleotide or at the cleavage site representing branching and hydrolysis, respectively. All disruptive mutations resulted in reduced cleavage rates. The double mutations that restored base pairing (U110A–A206U and U110C–A206G) performed branching at a rate comparable to that of wild type. The possibility for nucleotides A205 and A206 to engage in base pairing with U110 and U111 additionally suggests that G109 base pairs with U207 to form a continuous helical stack at the junction between P10 and P15.

The GoU pair at the P10–P15 interface

The secondary structure of DiGIR1 allows for two different possibilities of forming a GoU pair at the catalytic site in analogy with the GoU pair in P1. In both cases, G109 is involved but the pairing partner could either be U207 or the branch nucleotide U232, as in the original model (Einvik

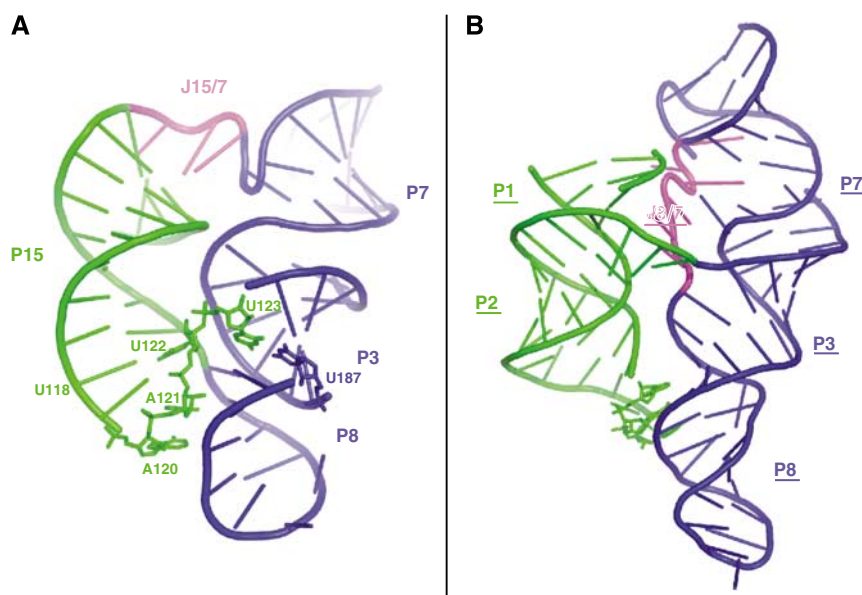


Figure 3 Comparison between the double-pseudoknot fold of GIR1 (A) and the corresponding region in *Azo* (B). (A) J15/7 forces P15 to be placed along P3 with which it forms a pseudoknot. Hence, the position of P15 corresponds to the position of P2 in *Azo* (B). In addition, this fold is confirmed by the conformation of the 3WJ since J15/3 (green stick residues) makes A-minor interactions in the shallow groove of P8, thus functionally replacing the interaction of L2 with the latter.

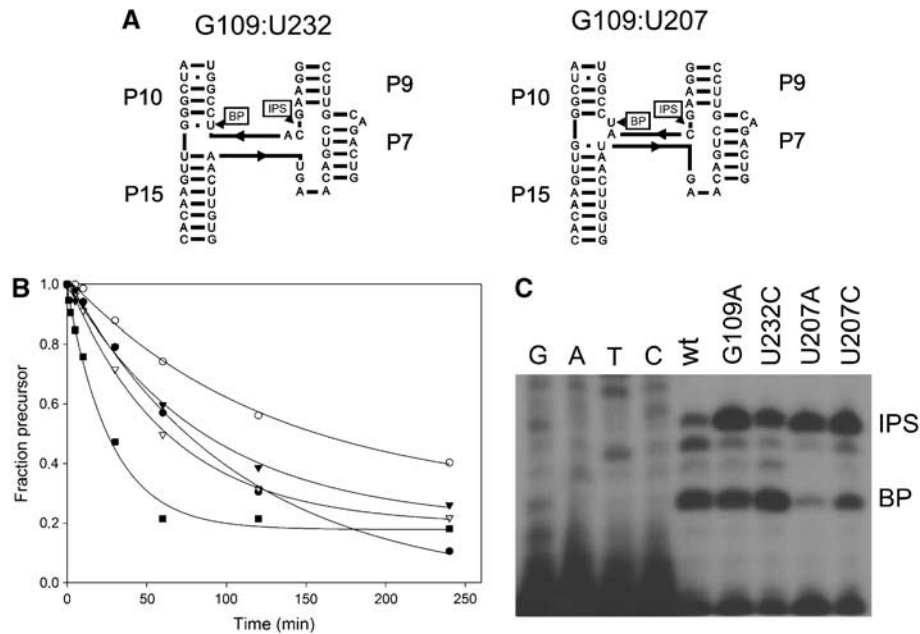


Figure 4 Mutational analysis of the GoU pair at the active site. (A) Two putative base-pairing schemes involve G109. In the left panel, G109 is base paired to the branch point (BP) nucleotide U232. In the right panel, this nucleotide is extra-helical and G109 is base paired to U207. (B) Kinetic cleavage analysis of mutations of the nucleotides involved. All of the mutations result in a decrease in the cleavage rate and mutations of the two potential base-pairing partners have a similar effect on the cleavage rate ((●) G109A, (○) U207A, (▼) U207C, (▽) U232C, and (■) wild-type GIR1). (C) Primer extension analysis of end point samples from the cleavage analysis. A primer extension stop signal at the BP indicates cleavage by branching and a stop at the internal processing site (IPS) indicates cleavage by hydrolysis. Mutations of G109 and U207 result in accumulation of more of the hydrolysis than the branching product in contrast to mutation of U232.

et al, 1998b) (Figure 4A). To distinguish between these two possibilities, the effect of mutations of the involved nucleotides on the cleavage rate (Figure 4B) and on the type of reaction (branching versus hydrolytic cleavage; Figure 4C) was assessed. The wild type is characterized by predominant cleavage by branching with only a small fraction (<10%) of stop signal, indicating hydrolytic cleavage after a 4 h incubation. Mutations that disrupt the catalytic pocket would be expected to affect both the branching and the hydrolysis rates. Such an activity loss has been previously observed following mutations of the ωG nucleotide (G229; Johansen *et al*, 2002) and disruption of the ωG-binding site in P7 (G174C; Decatur *et al*, 1995). Conversely, mutations that affect the positioning of the U232 relative to the ωG without disrupting the catalytic pocket would be expected to shift the reaction from branching to hydrolysis. The mutation G109A maintains the base-pairing potential with U232 or U207. The effect of the mutation is a moderate reduction in cleavage rate (Johansen *et al*, 2002). However, the reaction results in more hydrolysis than branching products. The mutations U232C and U207C similarly maintain the ability of these residues to base pair with G109. The effect on the cleavage rate is a comparable reduction in cleavage rate to that of G109A. However, cleavage in the U232C mutant results in more branching than hydrolysis product as in wild type, whereas cleavage in the U207C mutant results in more hydrolysis than branching product. The disruptive mutant U207A displays an even more reduced cleavage rate and cleaves almost exclusively by hydrolysis. The accumulation of more hydrolysis than branching product as seen in G109A and U207 mutants is an unusual phenotype as judged from our analysis of over 50 GIR1 mutants. Furthermore, mutants U232A and U232G

that would likely disrupt base pairing involving this nucleotide cleave more by branching than by hydrolysis similar to U232C (H Nielsen, in preparation). These observations are in favour of base pairing of G109oU207 instead of G109oU232 as in the original model (Einvik *et al*, 1998b). In this way, the critical GoU pair at the active site belongs to a P1-like helix (the extended P15) as in splicing group I ribozymes and not to P10 as in the original model (Einvik *et al*, 1998b). A further implication is that U207 forming the GoU pair does not provide the nucleophile for the branching reaction as it occurs in group I ribozymes. Rather, U232 lies in the shallow/minor groove of the G109oU207 pair, where it potentially interacts with the amino group of G109 to drive the branching reaction (Figure 5A). These mutational data furthermore validate the 3-bp extension of P15, which contributes significantly to the re-design of the catalytic core by forming a continuous helical stack between P10 and P15.

Recognition of the substrate domain P10–P15 by J5/4

In group I ribozymes, the GoU pair in P1 is recognized by a wobble receptor located at the interface of P4 and P5 (Michel and Westhof, 1990; Wang and Cech, 1992; Strobel and Cech, 1994; Strauss-Soukup and Strobel, 2000). When viewed in secondary structure diagrams, the structure of this interface is a 3-nt symmetrical internal loop. In DiGIR1, the interface between P4 and P5 is asymmetrical with a 4-nt junction, 5'-GUAA, as J5/4 and no intervening nucleotides at the 5' strand (Supplementary Figure S2A). Furthermore, J5/4 is one of the most variable segments as deduced from the *Naegleria* GIR1 sequence alignment albeit with conserved features (Wikmark *et al*, 2006). To assess the importance of J5/4 in DiGIR1, systematic mutational analysis of J5/4 residues was

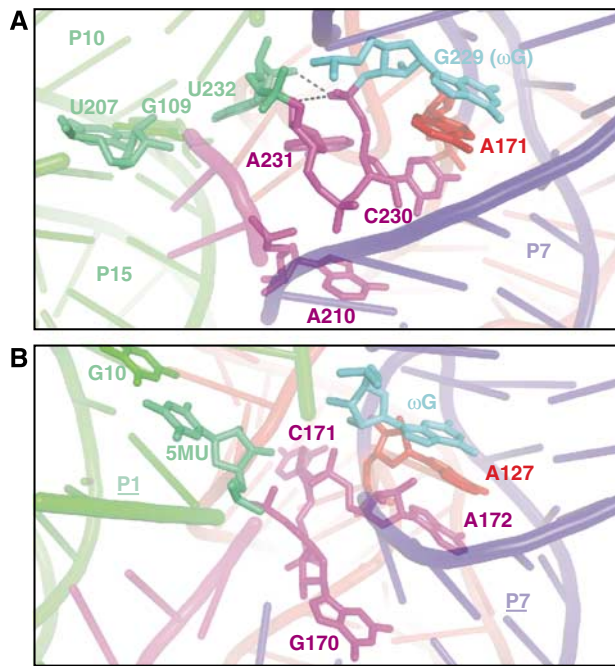


Figure 5 The catalytic pocket of the GIR1 ribozyme (**A**) spans from the ω G (G229, cyan residue) binding pocket to the nucleophile U232 (green residue); the GoU pair in the substrate domain lies at the P10–P15 interface (left side). G229 interacts in the deep groove side of the second base pair of P7 (blue ribbons). The lariat fold residues C230 and A231 interact with P7 and J5/4, respectively. The oxygen atoms of the scissile phosphate group are in contact with the 2'-hydroxyl groups of A231 and U232 (dashed lines). The 2'-hydroxyl and phosphate groups in the vicinity of the scissile phosphate potentially provide ligands for the catalytic magnesium ions. (**B**) In the same orientation and using the same colouring scheme (ω G in cyan, nucleophilic 5-methyl uracil in green (5MU)), the *Azoarcus* ribozyme shows that the GIR1 lariat fold residues functionally replace nucleotides from J8/7. The lariat fold together with J15/7 thus constitute a composite J8/7.

performed. Major alterations of the structure, such as deletion of J5/4, substitution of J5/4 with 5'-UUCG, or deletion of the bulged U156 all resulted in a complete loss of activity (data not shown). Substitution of the individual nucleotides resulted in decreased cleavage rates in all cases (Supplementary Figure S2B). The effect of mutating G150, U151, and A152 was moderate but the effect of the A153G mutant was dramatic pointing to this nucleotide as a key nucleotide for reactivity. Taken together, these results demonstrate an important function of J5/4 in GIR1 consistent with a preserved role of this structure in GoU recognition at the active site.

Alterations in the catalytic core do not affect the overall structure

The double pseudoknot provides a high level of constraint that guarantees confident model building of this region. The three stems P15, P3, and P8 together form a three-way junction (3WJ) already constrained in the P3–P15 pseudoknot. In the present model, the extended P15 is docked along P3 and adopts a parallel orientation with the co-axial stack occurring between stems P3 and P8. This conformation is promoted by the presence of the fairly long J15/3 stretch that forms a loop capping P15 and is able to interact in the shallow groove of P8 (Figure 3). A recent survey of 3WJ structures shows that J15/3 is part of a kind of 3WJ that

occurs at 10 ribosomal RNA locations and in several other RNA crystal structures (Lescoute and Westhof, 2006). Moreover, the above-mentioned survey shows that, when present in the longest loop, adenine residues are instrumental in stabilizing the junction architecture through the formation of A-minor interactions in the narrow groove of the facing stem, hence mimicking the GNRA/tetraloop receptor inter-domain interactions between P2 and P8 observed in group I introns crystal structures (see below).

Consequently, the single strands connecting P15 to the neighbouring helices can be considered as characteristic features distinguishing GIR1 from group I ribozymes. The constraints due to the 3WJ and to the double pseudoknot result in P15 occupying the same place as P2 (Figure 3). Furthermore, P15 directly stacks onto P10 by taking advantage of the 3-bp extension of P15 that was not considered in the original model (Einvik *et al*, 1998b). Hence, P10 and P7 adopt a relative position to the 3-bp P15 extension similar to what is observed for stem P1 in the crystal structures of group I introns (Adams *et al*, 2004b; Golden *et al*, 2005). This conformation is also supported by the fact that it leads to the formation of a pocket where all the structural elements necessary to form the catalytic site, namely ω G, U232, and the G109oU207 pair from P15 are gathered, a condition not satisfied by other tested models of 3WJ.

P8 and P9 were then directly connected to the double pseudoknot to form the catalytic domain. The connections between the P3–P9 catalytic core and the P4–P6 domain of GIR1 are similar to what is observed in *Azo* crystal structure (Adams *et al*, 2004a). In other words, J3/4 and J6/7 are modelled so as to weave the same contacts as those observed in *Azo* with the shallow groove of P6 and the narrow groove of P4, respectively (Figure 2). The last residue from J6/7 (A171) plays the same role as in *Azo* by providing stacking continuity between G229 (ω G) and the closest residue from J9/10 (C230), which corresponds to the last residue of J8/7 (A172) in *Azo* (Figure 5B). Since the P4–P6 domain is connected to the core as in group I introns, P6 consequently resides in the vicinity of P3, and the P4–P5 interface is able to contact the P10–P15 interface. Regarding the P7–P9 interface, a very discrete difference occurs. P7 is tethered to P9 without the intervening A residue frequently observed in group I ribozymes. This observation is important because J7/9 has been proposed to sequester ω G during the first step of splicing (Rangan *et al*, 2004), a condition not necessary in GIR1.

Apart from tertiary interactions specifically found in the core of the ribozyme, the group I intron architecture is stabilized by three sets of tertiary interactions (Figure 2). The first two interlock elements P2 and P6 with P8 and P3, respectively, and the third one allows L9 to contact P5 (Jaeger *et al*, 1996). The structural homology between group I introns and GIR1 ribozymes would plead for the existence of similar inter-domain interactions. However, these interactions are necessarily affected by the fact that the secondary structure elements from GIR1 are different or shorter than in group I introns.

The double pseudoknot corresponds to a motif swap for the known interaction between P2 and P8 (Michel and Westhof, 1990; Salvo and Belfort, 1992; Costa and Michel, 1995). The resulting model suggests that J15/3 replaces the tetraloop located at the tip of P2 and interacts in the shallow groove of P8 (Figure 3). As in the *Tetrahymena* ribozyme

(Guo *et al*, 2004), the loop receptor in GIR1 P8 consists of two consecutive G=C pairs instead of an 11-nt receptor motif as observed in *Azo* (Adams *et al*, 2004b). J15/3 loops over itself to enter the 5' strand of P3. U123 interacts with U187 and the resulting base pair intercalates at the P3–P8 interface. U122 base pairs with U118 and provides stacking continuity between P15 and J15/3. A120 and A121 form A-minor interactions in the minor groove of P8. This conformation is supported by mutants of U residues to C that affect the structure of the 3WJ (data not shown).

In *Didymium*, P6a does not exist and J6/6a thus becomes the tetraloop L6 capping P6. Hence, L6 is perfectly located to form the recurrent interaction between J6/6a and P3 (Waldsich *et al*, 2002; Adams *et al*, 2004a; Golden *et al*, 2005) using A-minor interactions (Doherty *et al*, 2001; Nissen *et al*, 2001) between A residues from L6 and two consecutive G=C pairs from P3 in the shallow/minor groove. Moreover, this tertiary contact is supported by mutational analysis of the two consecutive A residues from L6 (data not shown) and by the secondary structure of the NaGIR1 which displays a P6 element longer than in DiGIR1, albeit interrupted by an A-rich internal loop that could presumably function as J6/6a (Einvik *et al*, 1998b). It is noteworthy that the two tertiary interactions described above are also supported by chemical probing experiments showing that A residues important for the described contacts are protected from DMS and DEPC (Einvik *et al*, 1998b).

In contrast to the previous interactions, the characteristic interaction formed in group I between L9 and P5 is not conserved in GIR1 ribozymes (Figure 2). In DiGIR1, P9 is short (4 bp) and does not contain any hinge point that allows it to bend towards P5. However, its 5'-GAAA tetraloop could interact with a receptor embedded in the P2/P2.1 extension (Einvik *et al*, 2000; Nielsen *et al*, 2005) that was not included in the model (Figure 1).

Organization of the catalytic core

The next step consisted in understanding the architecture of the catalytic core in this unforeseen structural context. Around the catalytic region, the only fully identical feature shared by group I and GIR1 ribozymes resides in the organization of the ω G (G229) binding pocket (Figures 2 and 5). The Watson–Crick edge of G229 H-bonds with the Hoogsteen edge of the second G=C pair of P7, and is stabilized by stacking interactions between the first C=G pair of P7 and A171 from J6/7.

On the opposite side of the ribozyme, the J5/4 junction on which the substrate domain docks is organized quite differently in GIR1 compared to *Azo*. The dramatic loss of activity in the A153G mutant is consistent with its protection from chemical modification by DMS (Einvik *et al*, 1998b), and justifies orienting this key residue towards the core of the ribozyme. To achieve this, the two 5' nucleotides from J5/4 (G150 and U151) are placed so as to lie in the deep/major groove of P4 to improve the stacking continuity between P4 and P5. A kink performed around the phosphate group of A152 allows A153 to loop back into P5 ejecting A152 and A153 towards P10–P15. Thus, J5/4 becomes part of the catalytic pocket and shields P4 (Supplementary Figure S3). In such a situation, A153 can bind the GoU base pair from P15 as does A58 in J4/5 (Adams *et al*, 2004b). Moreover, A152 interacts with U110 to provide a tandem of A-minor

interactions. These A-minor interactions account for the observed loss of activity in the A153G mutant since G residues are rarely observed in contact with G=C base pairs in this motif (Doherty *et al*, 2001; Nissen *et al*, 2001). We propose that the interaction between P1 and J4/5 is replaced by A-minor tandem interactions involving the G109oU207 and U110–A206 pairs resulting from the extension of P15 with A153 and A152 from J5/4, respectively.

The lariat residues C230 and A231 replace important residues from the J8/7 junction in group I ribozyme

To suggest a relevant position for the residues involved in the lariat, a best-fitting lariat model with a 3-nt loop, obtained by an NMR study of an A2'-pG branched RNA (Agback *et al*, 1993) was accommodated in the catalytic pocket between P7, P5, and P10. The NMR models of these lariat RNAs provide a starting model from which several structural features can be characterized. The short length of the lariat loop forces the ribose–phosphate backbone to form the inner ring of the loop while ejecting the base moieties on the outside. As a result, the base rings cannot stack together but occupy distinct volumes in which they could interact with other chemical moieties. A lariat harbouring the DiGIR1 5'-CAU sequence while keeping the conformation of the RNA studied by NMR (Agback *et al*, 1993) was docked in the active site in a search for the best orientation. In the course of the refinement, the lariat was debranched to model a conformation corresponding to the pre-cleavage state. The ring formed by the lariat is short and tight with a kink around the phosphate group of A231 forcing base moieties of residues C230 and A231 to point towards structural elements forming the catalytic pocket with which they can interact (Figure 4A).

Interestingly, the lariat residues could be placed within the pocket left free following the relocation of residues 207–209 from J15/7 that extend P15. In this position, A231 and C230 take over the role of residues C171 and A172 from J8/7 in their ability to interact towards P4 and P7, respectively. C230 stacks with A171 (J6/7) strengthening the deep groove 4-nt stack including G229, and taking the place occupied by the 3' A residue from J8/7 in *Azo* (Figure 5). A231 points towards J5/4 as a consequence of the lariat sharp turn and places this nucleotide at hydrogen bonding distance of A153. Although a base-pairing interaction is implicated, a geometry explaining the deep effect of the A231G mutant could not be clearly deduced based on chemical footprinting data (H Nielsen, in preparation). In the course of the catalytic formation of the GIR1 U2'-pC lariat, C230 and U232 are covalently attached following the nucleophilic attack of the O2' group of U232 onto the phosphorus atom of C230. It is thus reasonable to place these chemical groups at H-bonding distance (2.8 Å) by taking advantage of the closest oxygen atom of the phosphate group. In the conformation proposed, the 2'-hydroxyl group of A231 interacts with the oxygen atom of the phosphate group of C230, which is not already in contact with the O2' atom of U232 (Figure 4A). A deoxy substitution scan experiment (Nielsen *et al*, 2005) pointing out the important role of the 2'-hydroxyl group of A231 comes to support the proposed architecture since nucleotides at the 3' end of J8/7 are involved in coordinating the catalytic magnesium ions using phosphate oxygen atoms and/or hydroxyl groups. Hence, the model strongly suggests that magnesium ions are relocated in J9/10 in the vicinity of C230 and A231.

J15/7 and J9/10 complement each other to stabilize the ribozyme catalytic core by forming a composite J8/7 junction that coordinates magnesium ions, and places the nucleophilic U residue in close vicinity of the targeted phosphate group.

Discussion

Building a new structural model for the GIR1 ribozyme was prompted by the recent finding that the ribozyme catalyses the formation of a short lariat containing a 3-nt loop by transesterification (Nielsen *et al*, 2005). The new model is consistent with previous biochemical and mutational data and incorporates new mutational data presented in this study. The model shows that the group I ribozyme substrate stems P1 and P2 are replaced in GIR1 by a distinctive and unique 9-bp P15 stem starting with a GoU pair. The modelling strategy relied on the existence of a double pseudoknot involving stems P3 and P7 on the one hand (a mandatory feature of group I intron catalytic core structure), and stems P3 and P15 on the other hand (Figure 3). The proposed architecture of this highly constrained double pseudoknot is consistent with the conformation of the 3WJ additionally encompassing P8 (Lescoute and Westhof, 2006). Apart from the characteristic P15, the secondary structure of GIR1 is similar enough to canonical group I introns to unambiguously claim their phylogenetic relationship. Surprisingly, the elements distinguishing GIR1 from the group I splicing ribozymes lie within the usually very well-conserved catalytic core (Michel and Westhof, 1990). The different topology results in a core that despite the marked similarity in base-pairing scheme between GIR1 and the group I ribozymes at the second step does not carry out splicing. Rather, the position of the nucleophile is shifted from the last base pair in P1 to the interface between the analogous P15 extension and P10, thereby allowing for the branching reaction. Thus, the function of carrying the nucleophile is detached from the GoU pair and the catalytic reaction occurs in *cis* rather than in *trans*.

The topological differences between the catalytic cores of group I and GIR1 ribozymes resulting from the presence of the double pseudoknot heavily impacts the architecture of the catalytic pocket. GIR1 harbours a G-binding pocket in P7 identical to the pocket observed in *Azoarcus* pre-tRNA^{le} intron. The extended substrate helix P15 (analogous to P1–P2) is recognized by the protruding J5/4 using two consecutive A-minor interactions (Doherty *et al*, 2001; Nissen *et al*, 2001) instead of the base-pair tandem formed between the sugar and Hoogsteen edges (Leontis and Westhof, 2001) of A residues in J4/5. Residues C230 and A231 in the loop of the lariat fold replace key residues from *Azo* J8/7 (A172 and C171) in their ability to interact with P7 and P4, respectively. Since in *Azo*, C171 and A172 are involved in binding the two magnesium ions that are required for catalysis (Adams *et al*, 2004b), it is tempting to suggest that residues constituting the lariat fold provide some of the ligands for binding the active site metal ions.

Shortening of J8/7 may account for the appearance of the branching reaction

The most dramatic feature of the topological change in GIR1 is that the joining segment that connects to the 5' strand of P7 comes from P15 (J15/7) rather than from P8 (J8/7) and that it has been shortened down to 3 nt as a consequence of the extension of P15. Hence, J15/7 is stretched and adopts a very

different path compared to J8/7 in *Azo*. As a consequence, the two 5' residues are excluded from residing inside the pocket and are located towards the outer shell of the molecule and the branching residue U232 is allowed to dock in the shallow groove of P15 and be accommodated into the catalytic pocket. The 3' residue of J15/7 (A210) functionally replaces the fourth nucleotide of J8/7 (G170) in *Azo* by interacting with J7/3. Hence, the space left free by the absence of the two 3' nucleotides of J8/7 can be occupied by the two first residues (C230 and A231) from the lariat fold. These nucleotides are followed by the branching U232, with a 3'-hydroxyl group already tethered to the downstream RNA chain. U232 is positioned similarly with respect to the cleavage site as the 3' U of the 5' exon in *Azo*. The absence of a free 3'-hydroxyl group in an environment prone to bind magnesium ions and generate nucleophilic oxygen atoms from hydroxyl groups may have driven the O2' of U232 to attack the facing phosphate group of C230 and form the 3-nt lariat characteristic of GIR1 ribozymes.

GIR1 topology may have arisen by drift of the 3' strand of P2, the 5' strand of P3, and J8/7 sequences of the ancestor intron

The three most notable features of GIR1 are that (i) it has so far only been found in the setting of twin-ribozyme introns, (ii) it closely resembles the eubacterial IC3 introns, and (iii) despite this close similarity, GIR1 catalyses a branching reaction rather than splicing. Although it is generally difficult to demonstrate any evolutionary path, only a few discrete events would be required to account for the emergence of the GIR1 branching ribozyme from group I introns. For the emergence of the twin-ribozyme configuration, it is plausible that a bacterial intron invaded a group I intron containing a HEG insertion. Myxomycetes are rich in nuclear rDNA introns with a relatively large proportion containing HEGs and the possibility of an invading bacterial intron is supported by the recent observation of a sister intron to Dir.S956-1 in the myxomycete *Diderma* (SD Johansen, unpublished data). The *Diderma* intron is located at the exact same rDNA position as the *Didymium* intron and has an almost identical group I splicing ribozyme with a very similar HEG inserted into the P2 segment. However, the *Diderma* intron lacks a GIR1 ribozyme and thus may represent the pre-existing receptor intron in the model. The configuration of an intron within an intron is reminiscent of the case of group II/III twintrons (Copertino and Hallick, 1993) and renders the invading intron in a situation with no stringent requirement to preserve the splicing activity. This could have set the stage for the subsequent transition of the invading intron into a branching ribozyme.

In the present study, we have shown that the difference between the *Azoarcus* intron representing the IC3 group I introns and the GIR1 branching ribozyme consists mainly of topological changes in the core. At the sequence level, we noticed that a single transposition event of the 5'-GUGUUC stretch from the 3' strand of P15 of the wild-type GIR1 to A120 in J15/3 would restore the topology and the base-pairing scheme found in *Azo* (Supplementary Figure S4). Further, at the 3D level, exchanges of phosphate bonds at positions that come in fairly close distance on the *Azo* crystal structure would result in the GIR1 topology (Figures 2 and 6A). From a mechanistic point of view, a model based

on sequence drift and strand exchange promoted by the absence of a selection pressure for splicing can be envisaged. In this scenario, a gradual sequence change leads to sequence similarities between the 5' strand of P3, the 3' strand of P2 and J8/7 resulting in a strand pairing switch within the core (Figure 6B). This evolutionary pathway via alternative pairing is similar to the mutational drift experimentally demonstrated between the HδV ribozyme and the artificial class III ligase (Schultes and Bartel, 2000). In the evolutionary model of GIR1, misfolding has been promoted by the loss of the 5' exon that has driven the J8/7 from GIR1 to form a pseudo P1 that expanded to form P15. It is noteworthy that this process gives rise to the double pseudoknot (P3–P7 and P3–P15) that contributes to the energetic stabilization of the core of the ribozyme. The higher stability conferred to the ribozyme core

by the appearance of the double pseudoknot may have been important to allow the peripheral domains to evolve with only minor implications on the core structure explaining why they are reduced to short appendices. In a context with low selection pressure towards splicing, since GIR1 was already embedded in a self-splicing intron, this process relied on sequence similarities between the segments of the ribozyme involved in mispairing (P1, P2, P3, and J8/7). Misfolding of a group I ribozyme around J8/7 and P3 has been experimentally observed (Pan and Woodson, 1998). The misfolding generating GIR1 may have been positively selected by allowing the branching reaction to occur, which conferred a selective advantage by increasing the half-life of the HE mRNA (H Nielsen, in preparation; Johansen *et al*, 2007). The topological shuffle described here fully accounts for the topological

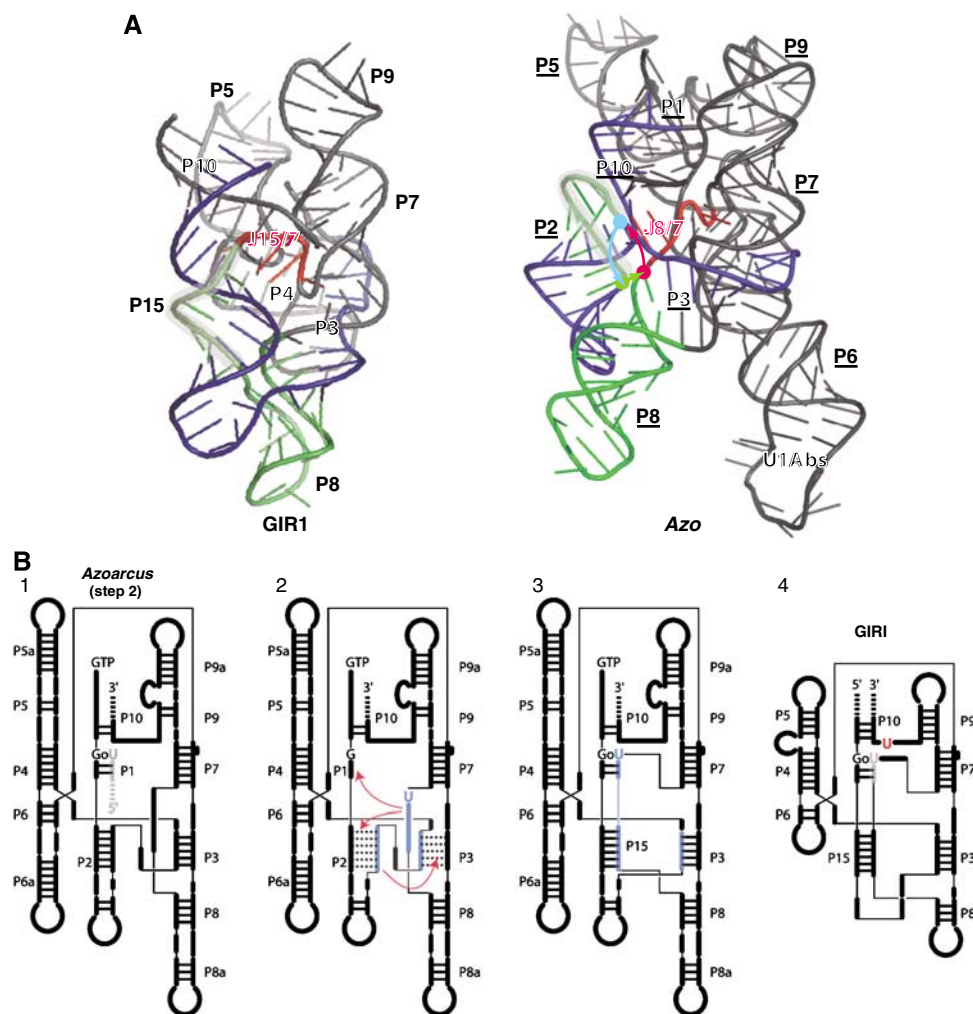


Figure 6 The different topologies of GIR1 and group I ribozymes. (A) At the 3D level, shuffling the backbone in regions where RNA backbones are in close vicinity (coloured arrows on the right panel displaying *Azo* crystal structure) leads to changing the topology of the *Azo* ribozyme to the GIR1 ribozyme. The first shuffling point intervenes at A27 in P2, the second involves G37 in P2, and the third one takes place at G166 in P8. Once *Azo* is cleaved at these positions, the intron can be religated. C28 is attached to G38 to connect the loop of P2 to P3 becoming J15/3. G37 is then attached to A167 to make J15/7, and finally G166 is linked to C28 to complete the shuffling. Equivalent backbone portions are coloured identically. Note the conserved position of the outlined segment which corresponds either to the 3' strand of P15 in GIR1 or to the 3' strand of P2 in *Azo* (U1Abs: U1A protein binding site). (B) A model for the evolution of a group I intron into the GIR1 ribozyme. (1) The *Azoarcus* group I ribozyme at a stage prior to the second catalytic step undertakes mutations in J8/7, P2, and P3 that lead to partial alternative pairing in folding. (2) The loss of the 5' exon favours misfolded species by drifting of neighbouring sequence stretches altering the overall secondary structure. (3) The GIR1 fold is selected due to the appearance of a new chemical reaction allowing the formation of the lariat and conferring an increased half-life to the homing endonuclease mRNA. (4) The gain in energetic stabilization due to the presence of the new pseudoknot P3–P15 allows for a shortening of some peripheral elements leading to the final version of GIR1 ribozyme that is always shorter than group I ribozymes.

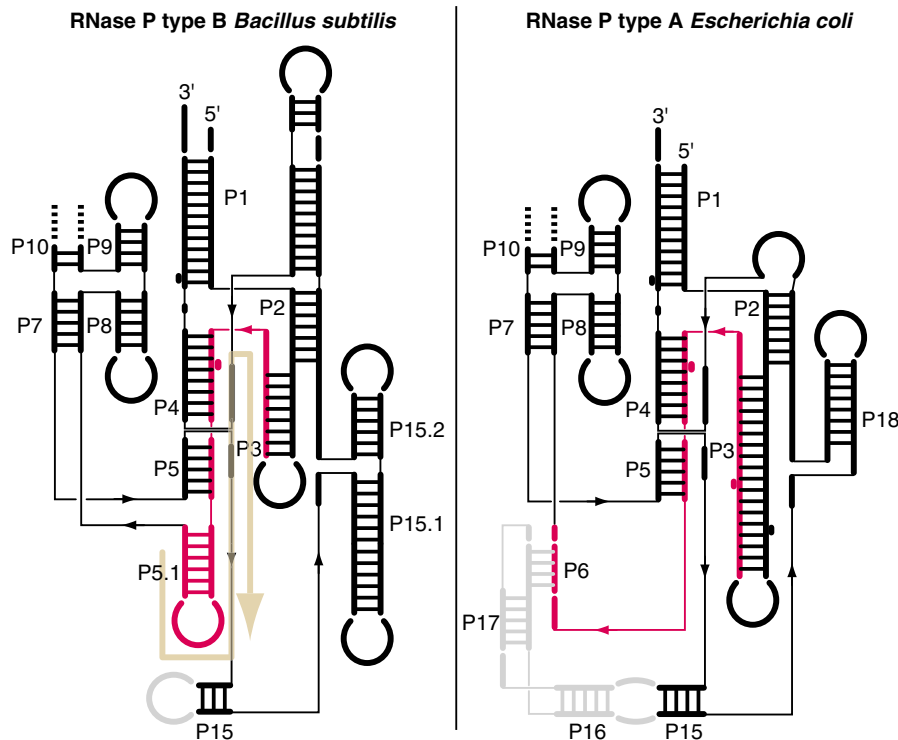


Figure 7 The drift model can be applied to the bacterial RNase P ribozymes of subtypes A and B. Starting from the type B ribozyme (left panel), pulling the 5' end of the red strand, as suggested by the orange arrow, lengthens P3 as in type A ribozyme (right panel), and shortens P5.1. P5.1 does not fold anymore as a hairpin and finds a new pairing partner in its vicinity (region depicted in grey) leading to the formation of the P6 pseudoknot. This scheme is made possible by the existence of an extended L15 loop in type A ribozyme compared to the short L15 in type B ribozyme.

changes observed between the core of group I ribozymes and GIR1, the appearance of the double pseudoknotted structure with the extended P15, and the redefinition of the role of J8/7 (now J15/7).

The proposed mechanism for evolution of new RNA molecules may apply to other RNAs. The two main families of ribonuclease P ribozymes (Darr *et al*, 1992a,b) can be distinguished by secondary structure changes occurring in a single contiguous region: the path from family B to family A involves lengthening of stem P3, disruption of stem P5.1 with formation of a new pseudoknot P6 (Figure 7). In contrast, in the H δ V/ligase case (Schultes and Bartel, 2000), all pairing stems are involved in strand exchange. Even though the above scenario leading to GIR1 evolution seems to be the most relevant, we cannot rule out that some unknown transposition events could have taken place at the RNA level with subsequent transfer to the DNA level by reverse transcription and integration into the genome or by the recently described RNA-directed DNA repair mechanism (Storici *et al*, 2007).

In conclusion, we have provided a model that correlates the branching activity of GIR1 with its topological difference compared to that of group I splicing ribozymes. We suggest also an evolutionary mechanism for the emergence of GIR1 based on the shuffling between functional motifs promoted by sequence shift and alternative pairings. Additional proofs will be needed and could be inspired from studies of other GIR1 ribozymes, such as those found in *Naegleria* (Einvik *et al*, 1997; Jabri *et al*, 1997; Johansen *et al*, 2002; Wikmark *et al*, 2006).

Materials and methods

In vitro mutagenesis

Extension of P15. Mutations at U110 were introduced by PCR using *Pfu* DNA polymerase of a wild-type GIR1 template (pDi162G1 Decatur *et al* (1995)) and oligos C377 (see Supplementary data for details) or C378 as the 5'-oligo and OP12 as the 3'-oligo. The PCR product was re-amplified to make templates for *in vitro* transcription (see below). Mutations of A206 were introduced by *in vitro* mutagenesis using the Quick Change site-directed mutagenesis kit (Stratagene) and oligos C405/C406 and C407/C408. To make double mutants, the U110 mutations were introduced into the A206 mutated templates.

GU pair. Construction of G109A and U232C were previously published (Johansen *et al*, 2002). Mutations at U207 were made as described above using oligos C477/C478 or C479/C480.

J5/4. Mutations were made in a wild-type GIR1 template as described above and oligos C415/C416, C424/C425, C270/C271, and C417.

Cleavage analyses and primer extension

Templates for *in vitro* transcription were made from wild-type and mutant templates by PCR using *Pfu* DNA polymerase (Fermentas) and oligos C287: 5'-AATTTAATACGACTCACTATAGGGTTGGGAAGTATCAT and C288: 5'-TCACCATGTTGTTGAAGTGCACAGATTG. C287 carries a T7 RNA polymerase promoter. The run-off transcript from the PCR template includes 162 nt upstream and 65 nt downstream of the cleavage site. All templates were transcribed *in vitro* using T7 RNA polymerase (Fermentas) with trace amounts of [α - 32 P]UTP. Cleavage analysis was performed as described in Einvik *et al* (2000). Briefly, radioactively labelled *in vitro* transcripts were renatured in 1 M KCl, 25 mM MgCl $_2$ at pH 5.5 for 10 min at 45°C. Then the reaction was jump started by increasing the pH to 7.5 by addition of Hepes-KOH. Time samples were withdrawn and run on

6% denaturing (urea) polyacrylamide gels. The gels were analysed on storage phosphor screens and the data fitted to a nonlinear first-order decay equation. The experiments shown are representative results of 3–5 independent experiments. Primer extension analysis was carried out as described (Einvik *et al*, 1998b) using end-labelled oligo C291: 5'-GATTGCTCTGGGAT. Sequencing ladders were made using the same primer and the plasmid pDi162G1 (Einvik *et al*, 1998b) as template. The reactions were analysed on 8% denaturing (urea) polyacrylamide gels.

Molecular modelling

Molecular modelling was performed as described in Masquida and Westhof (2005). The lariat model taken from Agback *et al* (1993) corresponds to the RNA lariat with a 3-nt loop in which all residues presents a C2'-endo conformation taken from http://www.boc.uu.se/boc14www/res_proj/final_struct/pictures/Welcome.html (file cGUAC_md25_A.pdb). The lariat was debranched to allow the phosphate group of the 5' residue to be tethered to ωG. The sequence of J9/10 was applied to the lariat fold using the program fragment embedded in the manip software (Massire and Westhof, 1998). This program was also used to build in three dimensions (pdb file format) all GIR1 pieces similar to the corresponding group

I intron regions. All the 3D elements were assembled interactively on a SGI Octane graphical workstation (IRIX64 v6.5, IP30) using the manip software. Each step of manual modelling was followed by several least-square refinement step (Westhof *et al*, 1985). The modelling/refinement cycles were iterated until a model satisfying all the constraints was obtained. Figures were prepared using the PYMOL program (DeLano WL, The PyMOL Molecular Graphics System (2002) <http://www.pymol.org>). Secondary structure diagrams in Figure 2 were directly generated from the PDB files using the program S2S (Jossinet and Westhof, 2005).

Supplementary data

Supplementary data are available at *The EMBO Journal* Online (<http://www.embojournal.org>).

Acknowledgements

This project was supported the Danish Natural Science Research Council. BB was supported by CEE BAC RNA program (LSHG-CT-2005-018618) and the Lundbeck Foundation. We thank Pascale Romby for critical reading of the manuscript.

References

- Adams PL, Stahley MR, Gill ML, Kosek AB, Wang J, Strobel SA (2004a) Crystal structure of a group I intron splicing intermediate. *RNA* **10**: 1867–1887
- Adams PL, Stahley MR, Kosek AB, Wang J, Strobel SA (2004b) Crystal structure of a self-splicing group I intron with both exons. *Nature* **430**: 45–50
- Agback P, Sandstrom A, Yamakage S, Sund C, Glemarec C, Chattopadhyaya J (1993) Solution structure of lariat RNA by 500 MHz NMR spectroscopy and molecular dynamics studies in water. *J Biochem Biophys Methods* **27**: 229–259
- Bhattacharya D, Reeb V, Simon DM, Lutzoni F (2005) Phylogenetic analyses suggest reverse splicing spread of group I introns in fungal ribosomal DNA. *BMC Evol Biol* **5**: 68–78
- Copertino DW, Hallick RB (1993) Group II and group III introns of twintrons: potential relationships with nuclear pre-mRNA introns. *Trends Biochem Sci* **18**: 467
- Costa M, Michel F (1995) Frequent use of the same tertiary motif by self-folding RNAs. *EMBO J* **14**: 1276–1285
- Darr SC, Brown JW, Pace NR (1992a) The varieties of ribonuclease P. *Trends Biochem Sci* **17**: 178–182
- Darr SC, Zito K, Smith D, Pace NR (1992b) Contributions of phylogenetically variable structural elements to the function of the ribozyme ribonuclease P. *Biochemistry* **31**: 328–333
- Decatur WA, Einvik C, Johansen S, Vogt VM (1995) Two group I ribozymes with different functions in a nuclear rDNA intron. *EMBO J* **14**: 4558–4568
- Ditzler MA, Aleman EA, Rueda D, Walter NG (2007) Focus on function: single molecule RNA enzymology. *Biopolymers* **87**: 302–316
- Doherty EA, Batey RT, Masquida B, Doudna JA (2001) A universal mode of helix packing in RNA. *Nat Struct Biol* **8**: 339–343
- 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**: 710–720
- Einvik C, Elde M, Johansen S (1998a) Group I twintrons: genetic elements in myxomycete and schizopyrenid amoebiflagellate ribosomal DNAs. *J Biotechnol* **64**: 63–74
- Einvik C, Nielsen H, Nour R, Johansen S (2000) Flanking sequences with an essential role in hydrolysis of a self-cleaving group I-like ribozyme. *Nucleic Acids Res* **28**: 2194–2200
- Einvik C, Nielsen H, Westhof E, Michel F, Johansen S (1998b) Group I-like ribozymes with a novel core organization perform obligate sequential hydrolytic cleavages at two processing sites. *RNA* **4**: 530–541
- Goddard MR, Greig D, Burt A (2001) Outcrossed sex allows a selfish gene to invade yeast populations. *Proc Biol Sci* **268**: 2537–2542
- Golden BL, Kim H, Chase E (2005) Crystal structure of a phage Twort group I ribozyme–product complex. *Nat Struct Mol Biol* **12**: 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**: 351–362
- Haugen P, Simon DM, Bhattacharya D (2005) The natural history of group I introns. *Trends Genet* **21**: 111–119
- Jabri E, Aigner S, Cech TR (1997) Kinetic and secondary structure analysis of *Naegleria andersoni* GIR1, a Group I ribozyme whose putative biological function is site-specific hydrolysis. *Biochemistry* **36**: 16345–16354
- Jabri E, Cech TR (1998) *In vitro* selection of the *Naegleria* GIR1 ribozyme identifies three base changes that dramatically improve activity. *RNA* **4**: 1481–1492
- Jaeger L, Michel F, Westhof E (1996) The structure of group I ribozymes. In *Nucleic Acids and Molecular Biology*, Eckstein F, Lilley DMJ (eds), Vol. 10, pp 33–51. Berlin: Springer Verlag
- 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**: 905–912
- Johansen SD, Haugen P, Nielsen H (2007) Expression of protein-coding genes embedded in ribosomal DNA. *Biol Chem* **388**: 679–686
- Jossinet F, Westhof E (2005) Sequence to structure (S2S): display, manipulate and interconnect RNA data from sequence to structure. *Bioinformatics* **21**: 3320–3321
- Leontis NB, Westhof E (2001) Geometric nomenclature and classification of RNA base pairs. *RNA* **7**: 499–512
- Lescoute A, Westhof E (2006) Topology of three-way junctions in folded RNAs. *RNA* **12**: 83–93
- Masquida B, Westhof E (2005) Modeling the architecture of structured RNAs within a modular and hierarchical framework. In *Handbook of RNA Biochemistry*, Hartmann RK, Bindereif A, Schön A, Westhof E (eds), pp 536–545. Weinheim, Germany: Wiley VCH Verlag GmbH & Co.
- Massire C, Westhof E (1998) MANIP: an interactive tool for modelling RNA. *J Mol Graph Model* **16**: 197–205, 255–197
- Michel F, Westhof E (1990) Modelling of the three-dimensional architecture of group-I catalytic introns based on comparative sequence analysis. *J Mol Biol* **216**: 585–610
- Nielsen H, Beckert B, Masquida B, Johansen SD (2008) The GIR1 branching ribozyme. In *Ribozymes and RNA Catalysis*, Lilley DMJ, Eckstein F (eds), pp 229–252. London: Royal Society of Chemistry
- Nielsen H, Johansen SD (2007) A new RNA branching activity: the GIR1 ribozyme. *Blood Cells Mol Dis* **38**: 102–109
- 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**: 1584–1587
- 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**: 4899–4903

- Pan J, Woodson SA (1998) Folding intermediates of a self-splicing RNA: mispairing of the catalytic core. *J Mol Biol* **280**: 597–609
- Pyle AM, Murphy FL, Cech TR (1992) RNA substrate binding site in the catalytic core of the *Tetrahymena* ribozyme. *Nature* **358**: 123–128
- Rangan P, Masquida B, Westhof E, Woodson SA (2004) Architecture and folding mechanism of the *Azoarcus* group I pre-tRNA. *J Mol Biol* **339**: 41–51
- Salvo JL, Belfort M (1992) The P2 element of the td intron is dispensable despite its normal role in splicing. *J Biol Chem* **267**: 2845–2848
- Schultes EA, Bartel DP (2000) One sequence, two ribozymes: implications for the emergence of new ribozyme folds. *Science* **289**: 448–452
- Scott WG (2007) Ribozymes. *Curr Opin Struct Biol* **17**: 280–286
- Serganov A, Patel DJ (2007) Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nat Rev Genet* **8**: 776–790
- Soukup JK, Minakawa N, Matsuda A, Strobel SA (2002) Identification of A-minor tertiary interactions within a bacterial group I intron active site by 3-deazaadenosine interference mapping. *Biochemistry* **41**: 10426–10438
- Storici F, Bebenek K, Kunkel TA, Gordenin DA, Resnick MA (2007) RNA-templated DNA repair. *Nature* **447**: 338–341
- Strauss-Soukup JK, Strobel SA (2000) A chemical phylogeny of group I introns based upon interference mapping of a bacterial ribozyme. *J Mol Biol* **302**: 339–358
- Strobel SA, Cech TR (1994) Translocation of an RNA duplex on a ribozyme. *Nat Struct Biol* **1**: 13–17
- Tanner MA, Anderson EM, Gutell RR, Cech TR (1997) Mutagenesis and comparative sequence analysis of a base triple joining the two domains of group I ribozymes. *RNA* **3**: 1037–1051
- Vader A, Nielsen H, Johansen S (1999) *In vivo* expression of the nucleolar group I intron-encoded I-DirI homing endonuclease involves the removal of a spliceosomal intron. *EMBO J* **18**: 1003–1013
- Waldsich C, Masquida B, Westhof E, Schroeder R (2002) Monitoring intermediate folding states of the td group I intron *in vivo*. *EMBO J* **21**: 5281–5291
- Wang JF, Cech TR (1992) Tertiary structure of around the guanosine-binding site of *Tetrahymena* ribozyme. *Science* **256**: 526–529
- Westhof E, Dumas P, Moras D (1985) Crystallographic refinement of yeast aspartic acid transfer RNA. *J Mol Biol* **184**: 119–145
- Wikmark O-G, Einvik C, De Jonckheere J, Johansen S (2006) Short-term sequence evolution and vertical inheritance of the *Naegleria* twin-ribozyme group I intron. *BMC Evol Biol* **6**: 39–50



The EMBO Journal is published by Nature Publishing Group on behalf of European Molecular Biology Organization. This article is licensed under a Creative Commons Attribution License <<http://creativecommons.org/licenses/by/2.5/>>