New loop–loop tertiary interactions in self-splicing introns of subgroup IC and ID: a complete 3D model of the *Tetrahymena thermophila* ribozyme

Valérie Lehnert[#1], Luc Jaeger[#1], François Michel[2] and Eric Westhof[1]

Background: Group I introns self-splice via two consecutive trans-esterification reactions in the presence of guanosine cofactor and magnesium ions. Comparative sequence analysis has established that a catalytic core of about 120 nucleotides is conserved in all known group I introns. This core is generally not sufficient for activity, however, and most self-splicing group I introns require non-conserved peripheral elements to stabilize the complete three-dimensional (3D) structure. The physico-chemical properties of group I introns make them excellent systems for unraveling the structural basis of the RNA–RNA interactions responsible for promoting the self-assembly of complex RNAs.

Results: We present phylogenetic and experimental evidence for the existence of three additional tertiary base pairings between hairpin loops within peripheral components of subgroup IC1 and ID introns. Each of these new long range interactions, called P13, P14 and P16, involves a terminal loop located in domain 2. Although domains 2 of IC and ID introns share very strong sequence similarity, their terminal loops interact with domains 5 and 9 (subgroup IC1) and domain 6 (subgroup ID). Based on these tertiary contacts, comparative sequence analysis, and published experimental results such as Fe(II)–EDTA protection patterns, we propose 3D models for two entire group I introns, the subgroup IC1 intron in the large ribosomal precursor RNA of *Tetrahymena thermophila* and the SdCob.1 subgroup ID intron found in the cytochrome *b* gene of *Saccharomyces douglasii.*

Conclusions: Three-dimensional models of group I introns belonging to four different subgroups are now available. They all emphasize the modular and hierarchical organization of the architecture of group I introns and the widespread use of base-pairings between terminal hairpin loops for stabilizing the folded and active structures of large and complex RNA molecules.

Addresses: [1]Institut de Biologie Moléculaire et Cellulaire du CNRS, UPR9002,15 rue Descartes, 67084, Strasbourg, France and [2]Centre de Génétique Moléculaire du CNRS, 91198, Gif-sur-Yvette, France.

# indicates an equal contribution to this work

Correspondence: Eric Westhof
E-mail: westhof@ibmc.u-strasbg.fr

Keywords: group I intron, ribozyme, RNA folding, RNA modelling, RNA–RNA interactions

Received: **16 October 1996**
Revisions requested: **12 November 1996**
Revisions received: **20 November 1996**
Accepted: **20 November 1996**

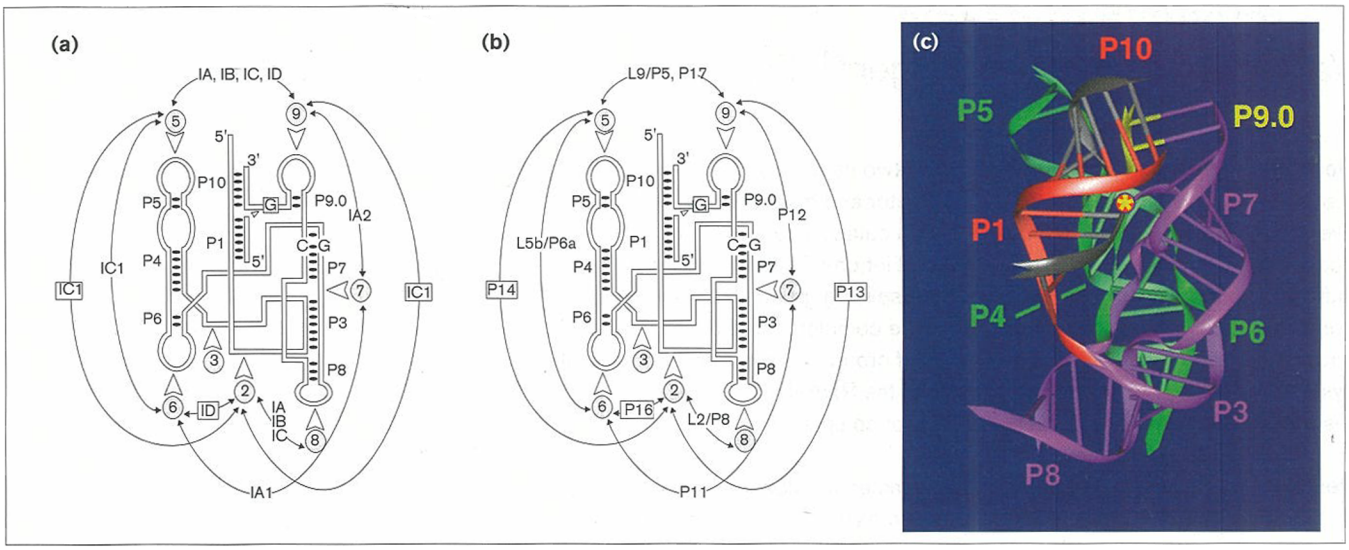
Chemistry & Biology
December 1996, **3**:993–1009

Introduction

Group I introns are catalytic RNA molecules that require magnesium ions for folding and catalysis [1,2]. They carry out self-splicing, generally in the absence of proteins, via two trans-esterification reactions, excising themselves from a precursor RNA and ligating together the flanking exon sequences. Catalysis is performed by a highly structured 'core' that is universally conserved among group I introns and consists of six to seven base-paired stems (P3–P9) with connecting segments (for reviews, see [1,3,4]). During the first step of splicing, a G-binding site located in stem P7 is responsible for the specific recognition of the free guanosine cofactor that attacks the 5′ intron–exon junction [5]. In the second step of splicing, a conformational change brings the 3′ terminal guanosine of the intron into the G-binding site by allowing the 3′ exon to be positioned next to the 3′ OH attacking group of the 5′ exon, so that the exons become ligated [5–7]. As illustrated by the three-dimensional (3D) model of the core proposed by Michel & Westhof [6] (see Fig. 1), the structure of the core is maintained by numerous tertiary interactions, some of which are indeed supported by experimental data (for review see [4]). The core itself, however, is generally not sufficient to maintain the active structure. The optimal activity of most, if not all, natural group I introns depends on a variety of non-core RNA components (e.g. [8–12]; see Fig. 1). These non-core components or peripheral domains are not conserved among all group I introns but are substantially conserved within intron subgroups [4,6]. For example, all introns of subgroup IA share one or two additional hairpins between core elements P3 and P7 but lack an extended P5 domain that is present in most subgroup IC and IB introns.

How do these different peripheral components contribute to the self-assembly of these large RNAs while maintaining a similar structure? In recent years, it has been shown that peripheral domains form specific tertiary interactions that help to fold and stabilize the ribozyme in a unique catalytically active structure (e.g. [10–15]). In two instances,

Figure 1



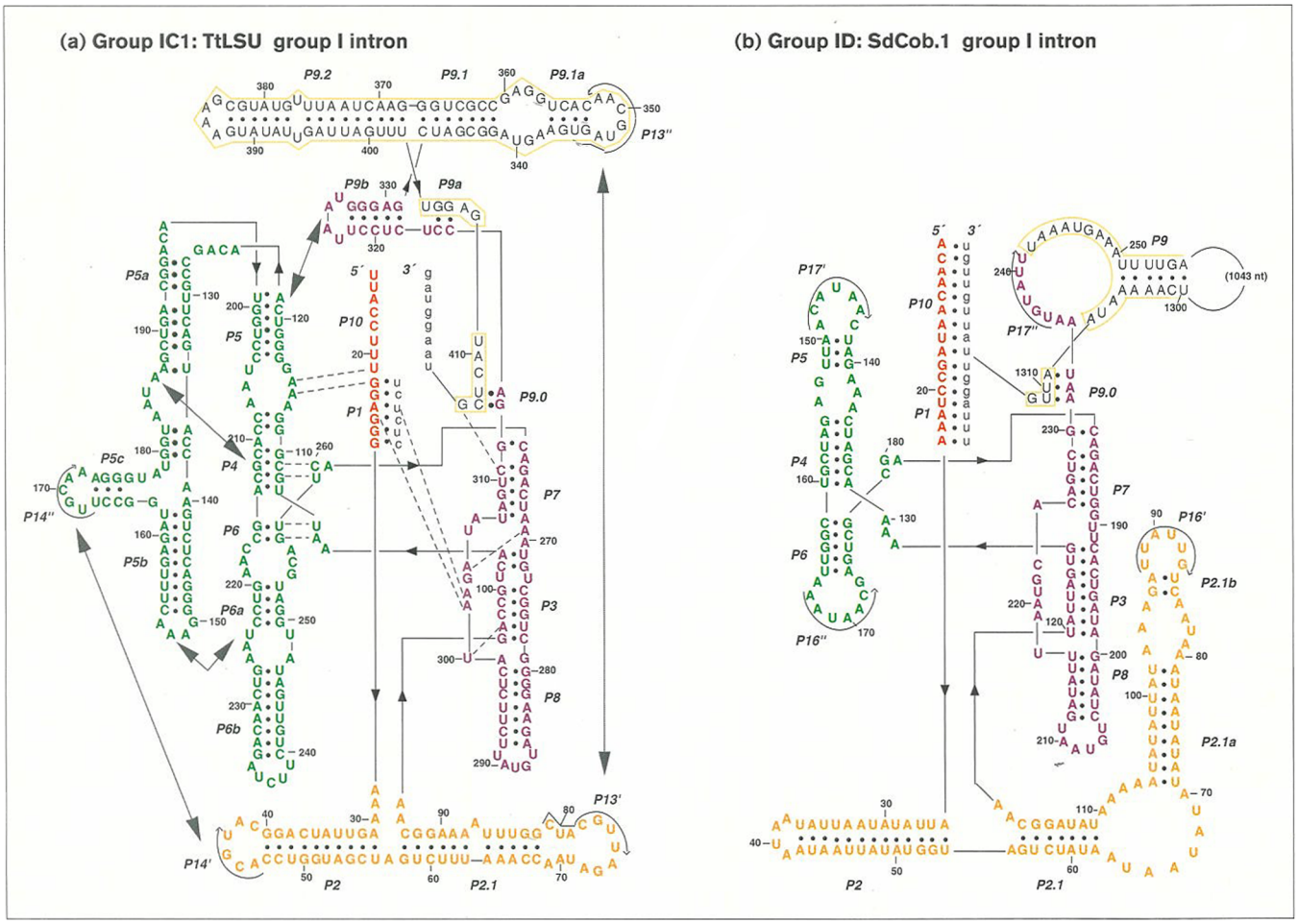
The architecture of the catalytic core of group I introns [6]. **(a)** Schematic secondary structure of the core of group I introns displaying the positions of insertion of the peripheral domains. The nomenclature is according to Cech *et al.* [72]. The G–C base pair constituting the G-binding site in helix P7 is indicated. The last G residue of the intron is within a square; the arrow head next to it represents the attack by the terminal O3′ atom of the 5′ exon on the junction between the last residue of the intron and the first one of the 3′ exon, occurring during the second trans-esterification of the self-splicing reactions. The positions of insertion of the peripheral components within single-stranded regions (2,3,5–9) are indicated in circles. The various long-range contacts occurring between those peripheral components in the different subgroups of group I introns [6], IA (sub divided into IA1 and IA2), IB, IC (and IC1), and ID are shown. **(b)** The nomenclature for all long-range contacts between the peripheral components established to date is given. Those starting with a P represent Watson–Crick base pairing between loops and those implying a contact between a GNRA apical loop and the shallow groove of a helix are noted by the name of the loop, L, and that of the helix, P. Notice that a loop–loop contact can be replaced by a GNRA/helix interaction (see L9/P5 and P17). The contacts within squares are those reported here. References for the other contacts are: P11 [10], P12 [11,13], P17 [33], L9/P5 [12], L5b/P6a [15], L2/P8 [31]. **(c)** Three-dimensional architecture of the catalytic core of group I introns [6]. The stage shown is prior to the second trans-esterification step (the ligation step). The site of catalysis is shown by an yellow asterisk within a red circle. Domain P4–P6 (green) and domain P3–P8 (purple) create a cleft into which is positioned the helical susbtrate formed by the stacked helices P1 and P10. Helix P1 is formed by the 5′-exon (black) base paired to the internal guide sequence (red) and in helix P10 the 3′-exon is base paired to the rest of the guide sequence. The nucleotide segment forming P9.0 is in yellow.

peripheral tertiary interactions were incorporated into the previous model of the core, leading to detailed and comprehensive 3D models for introns of subgroups IA1 and IA2 [10,11].

One of the ribozymes studied in most detail is certainly the self-splicing subgroup IC1 intron found in the large ribosomal precursor RNA of *Tetrahymena thermophila* [1,4,16]. Although the details of the catalytic reaction and the docking of the substrate to the ribozyme core are well understood, the global architecture of the ribozyme with all its peripheral components remained unclear. The domain P4–P5–P6 has been shown to fold independently from the rest of the ribozyme [14,15,17,18], but little was known about peripheral components that may stabilize the core domain P3–P7–P8.

In the present study, we have used two new tertiary base pairings found in subgroup IC1 introns to update the 3D model of the *Tetrahymena* ribozyme established by Michel and Westhof [6,19]. These two tertiary base pairings, called P13 and P14, involve terminal loops in domain 2 that interact with domains 9 and 5, respectively. Experimental evidence for the existence of these new tertiary interactions was obtained by site-directed mutagenesis. This model includes all peripheral extensions and helps to rationalize the patterns of protection from Fe(II)–EDTA cleavage [14,20–22]. We further propose a 3D model for subgroup ID introns which incorporates a new loop–loop interaction, called P16. Subgroups ID and IC1 share strong sequence similarity within the base of helices P2, P2.1, P3, P8 and their junction elements, suggesting that the local structure of this region is identical in both subgroups. Comparison of these two new models with the 3D structure of two other introns belonging to different subgroups indicates that peripheral elements are used in a modular way to stabilize the core of group I introns, and makes it plain that loop–loop base pairings are widely used 'anchors' in self-assembling RNA molecules.

Figure 2



Two-dimensional models of introns representative of the IC1 and ID subgroups, drawn according to the representation of Cech *et al.* [72]. The IC1 sequence **(a)** is that of the intron in the large ribosomal RNA precursor of *Tetrahymena thermophila*, abbreviated as TtLSU. The ID sequence **(b)** is that of the intron in the cytochrome *b* mitochondrial gene of *Saccharomyces douglasii*, abbreviated as SdCob.1 [32]. The structures are those expected to prevail just prior to exon ligation. Numbering of residues is specific to each intron. Exons (in lower case) are bound to the internal guide sequence (in red) at the 5′ end of the ribozyme. The P4–P6 domain is in green and the P3-P9 domain is in violet. Peripheral domain 2 is orange and domain 9 is boxed in yellow. Nucleotides involved in a tertiary base pairing are indicated by curved arrows: Pn′ and Pn″ correspond to the 5′ and 3′ strands of a tertiary base pairing, respectively. Heavy arrows indicate long-range interactions. Dashed lines indicate tertiary interactions within the catalytic core of group I introns.

Results

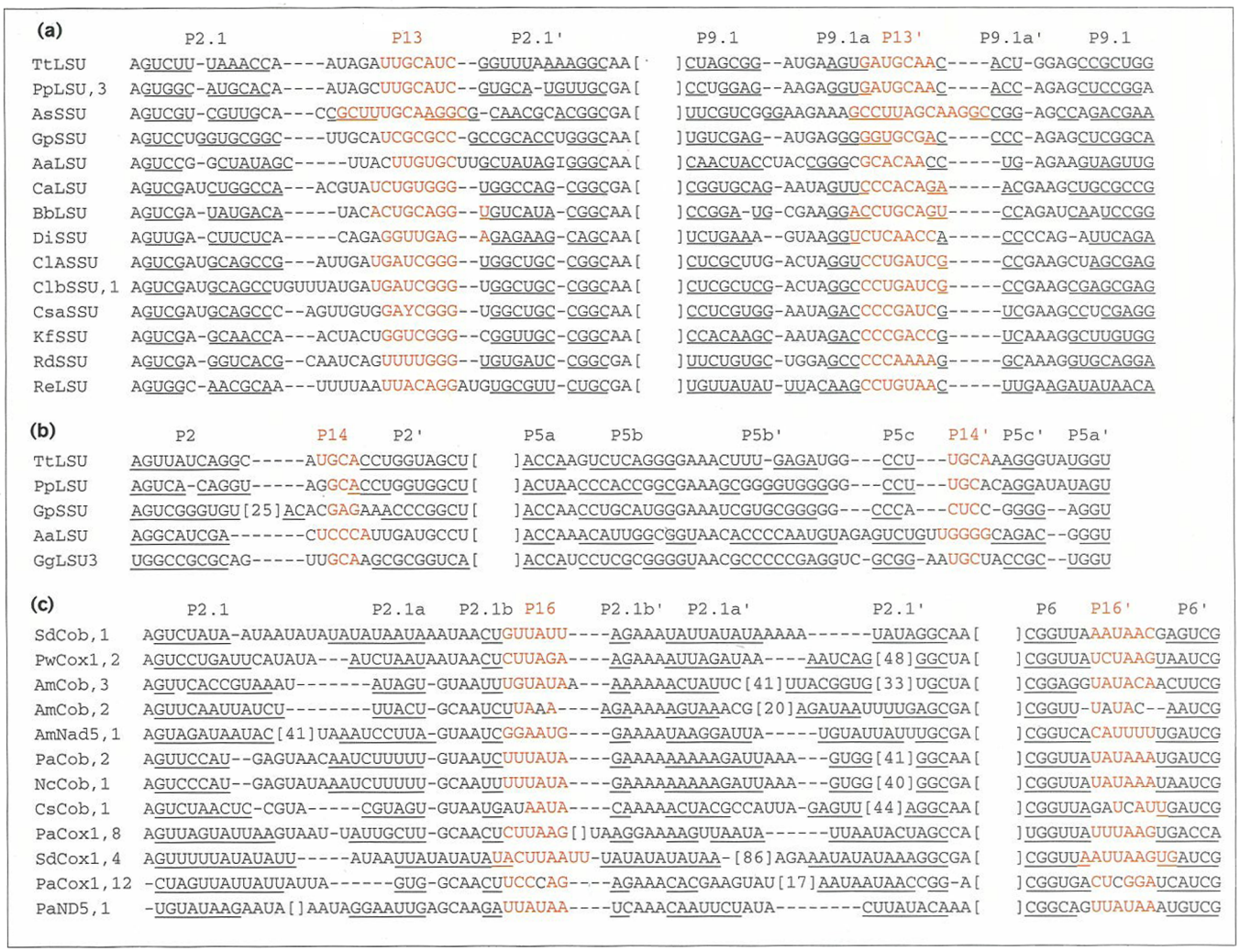
Long range tertiary interactions of the peripheral domain 2 of IC1

Several lines of experimental evidence identify peripheral domain 2 of the IC1 intron in the large ribosomal RNA precursor of *Tetrahymena thermophila* (TtLSU) (Fig. 2a) as an important, although not essential, domain for self-splicing activity. Removal of domain 2 does not abolish catalysis, but does result in defective molecules that typically require high salt and/or spermidine concentrations to function [8,23]. Moreover, nucleotide substitution or deletion in loop L2 and in connecting segments J1/2, J2/2.1 or J2.1/3 affects the efficiency and selectivity of substrate cleavage [24–27]. These data strongly suggest that domain 2 of the TtLSU intron is likely to be involved in multiple tertiary contacts and, perhaps, long range tertiary interactions similar to those found in subgroup IA1 and IA2 introns [10,11,13].

A tertiary interaction between domains 2 and 9

Been and Cech [28] originally proposed, based on examination of the sequence of the *Tetrahymena* ribozyme, that loops L2.1 and L9.1a might engage in base pairing. Subsequently, Banerjee *et al.* [29] investigated this further by using a number of group I sequences belonging to diverse subgroups for phylogenetic analysis. Most of the 27 pairings these authors suggested were inconsistent with the known secondary and tertiary structure of group I introns

Figure 3



Extracts from the alignments of sequences of subgroup IC1 and ID introns. **(a)** Phylogenetic evidence for the P13 interaction (subgroup IC1). **(b)** Phylogenetic evidence for the P14 interaction (subgroup IC1). **(c)** Phylogenetic evidence for the P16 interaction (subgroup ID). Names of introns are abbreviated as in Michel & Westhof [6]. References (other than these in Michel and Westhof [6]): AsSSU [73]; GpSSU [74]; AaLSU [75]; CaLSU [76]; BbLSU [77]; DiSSU [78]; ClASSU, V.A. Huss, unpublished, accession # X73997; ClBSSU,1, V.A. Huss, unpublished, accession # X73998; CsaSSU, V.A. Huss, unpublished, accession # X73991; KfSSU [79]; RdSSU [80]; ReLSU [81]; GgLSU3, J. Fogler, personal communication; SdCob.1 [32]; AmCob.3 [82]; AmCob.2 [82]; AmNad5.1 [82]; PwCox1.2 [83]; SdCox1.1 [84]. Underlined nucleotides participate in the secondary structure of group I introns. Names of secondary structure components are indicated on top of alignments (Pn designates the 5' branch of a pairing and Pn', its 3' branch). Tertiary pairings are in bold, red. Square brackets (or the letter I, or numbers) stand for those segments of sequence not included in the alignments.

that had already been established by phylogenetic and experimental analyses [6,10,11], however. Nevertheless, a careful analysis of 36 intron sequences from subgroup IC1 reveals that in 14 of these sequences, extended base pairing could exist between loops L2.1 and L9.1a, while preserving the consensus secondary structure of subgroup IC1. This tertiary pairing, henceforth called P13, is supported by numerous instances in which compensatory base changes occur to maintain base pairing (see the sequence alignments in Fig. 3a). Pairing P13 could involve 7–12 base pairs. Although the lengths of stems P2.1 and P9.1 tend to be well conserved among those subgroup IC1 introns that include a P13 pairing, formation of P13 requires, in some instances, the disruption of the apex of stems P2.1 and P9.1a. Our current model for the secondary structure of subgroup IC1, which is based on comparative sequence analysis, is shown in Figure 2a.

Further evidence for the existence of P13 was obtained experimentally by site-directed mutagenesis and *in vitro* self-splicing assays (see Materials and methods). To test whether the P13 pairing is indeed important for activity

we chose the TtLSU intron as an experimental system (Fig. 2a). We mutated loops L2.1 and L9.1a separately to destabilize P13 (mutants Tta and Tta′ respectively), and we expected to restore the postulated P13 pairing by combining these two mutations (mutant Ttaa′) (Fig. 4a). All TtLSU transcripts were found to be fully active at 10 mM magnesium (data not shown). At 2 mM magnesium and 30 °C, 90 % of Tta and Tta′ transcripts were trapped in an inactive state, however, whereas more than 70 % of wild type and Ttaa′ transcripts were spliced (Fig. 4b). These data therefore confirm the existence of the P13 pairing. We did not investigate the function of P13 further. It was shown by chemical probing and Fe(II)–EDTA cleavage, however, that deletion of the P9.1 and P9.2 secondary elements has a destabilizing effect on the TtLSU tertiary structure [22,29]. Furthermore, the 3′ terminal P9.1–P9.2 extension appears to guide the formation of the catalytic core [30]. Thus, these data support the idea that the P13 pairing may stabilize an active core conformation, in a manner similar to the stabilization of the introns of subgroups IA1 and IA2, by the P11 and P12 pairings respectively [10,11,13].

A tertiary contact between domains 2 and 5

About one third of known group I introns have, at the tip of stem P2, a GNRA tetraloop (where N is any nucleotide and R is a purine) that is at a defined distance from the conserved U–G pair in stem P1. The L2 tetraloop was shown to interact with specific pyrimidine–purine base pairs in helix P8 [6,31]. As suggested by comparative sequence analysis, stem–loop P2 from the *Tetrahymena* ribozyme is not homologous to GNRA-capped stem–loops and is unlikely to be stacked coaxially with P1 [27]. Even so, experimentally deduced constraints on the length of P2 and on the identity of nucleotides at both ends of the P2 element required for efficient substrate cleavage suggest that P2 may also have a function in positioning the substrate in the *Tetrahymena* ribozyme [27]. Moreover, the fact that nucleotides 44 and 45 of loop L2 are very sensitive to base substitutions [27], strongly suggests that L2 is directly involved in a long range interaction. Interestingly, we have found phylogenetic evidence for an interaction between loops L2 and L5c in five IC1 introns that have an extended domain P5abc, including the *Tetrahymena* intron (note the compensatory base changes between the TtLSU, GpSSU and AaLSU introns in Fig. 3b). This putative tertiary interaction, called P14, might be 3–5 base pairs long.

To test whether the L2-L5c pairing does occur, several mutant transcripts in which the putative P14 duplex is disrupted (Ttb and Ttb′) or restored (Ttbb′) (Fig. 4c), were generated and tested for self-splicing activity (see Materials and methods). Disruption of P14 had no significant effect on the TtLSU self-splicing reaction at magnesium concentrations ≥ 10 mM (data not shown). At lower magnesium concentrations, however, combinations Ttb and Ttb′ had much lower self-splicing activity than the wild type and matched double mutant. At 2 mM $MgCl_2$, 1mM GTP and 35 °C, compensatory Ttbb′ transcripts have ~50 % of the wild type activity whereas Ttb and Ttb′ transcripts are almost inactive (Fig. 4d). These data provide strong evidence for an interaction, the P14 base pairing, between the L2 and L5c loops. Further experiments are clearly needed to better understand the function of P14, but our experimental data are consistent with the idea that P14 is directly involved in stabilizing the active ribozyme conformation. The formation of P14 explains the fact that loop L5c was protected from dimethyl sulfate (DMS) modification in the TtLSU intron, but not when the P4–P6 domain is probed alone [14].

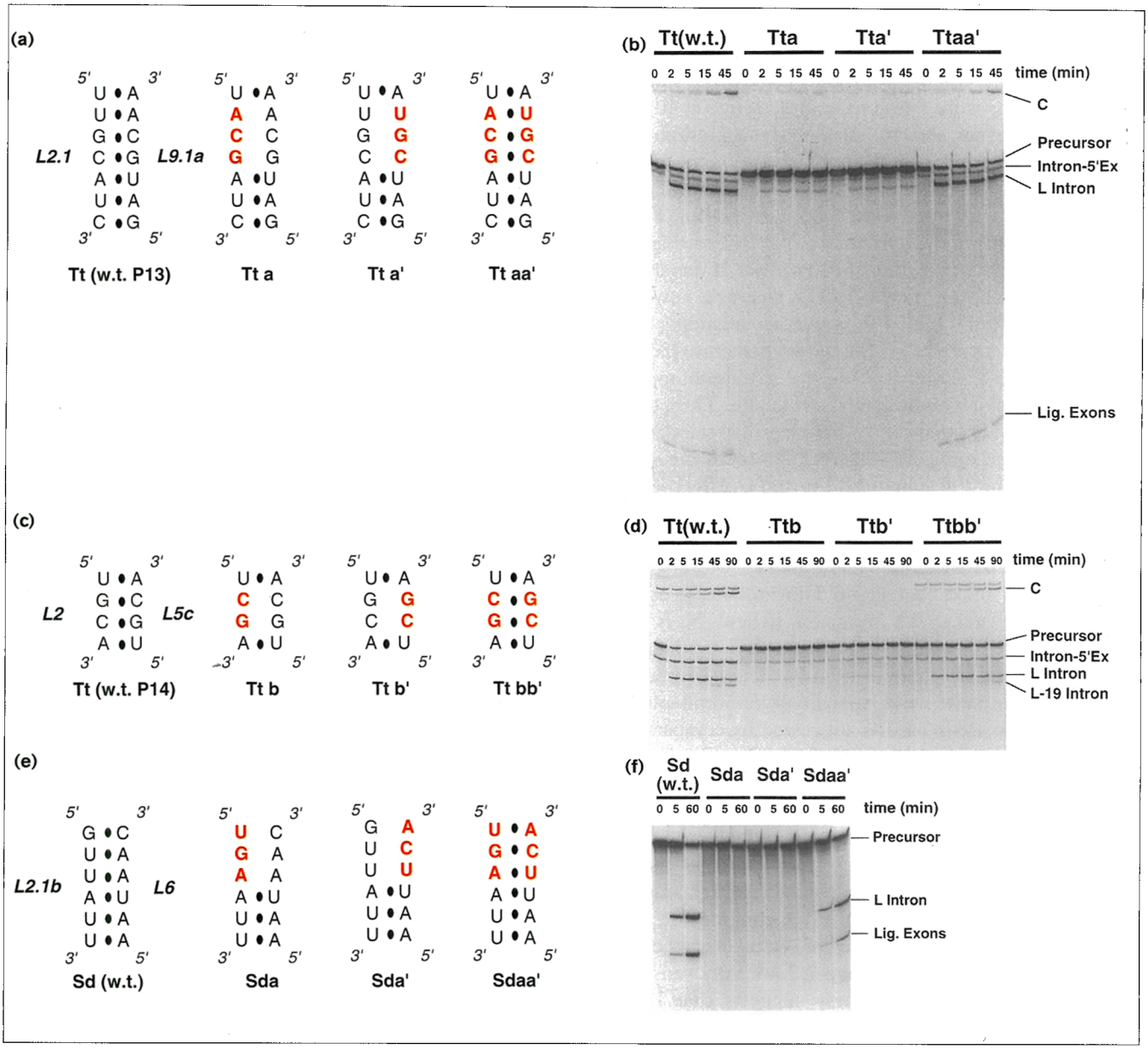
Interestingly, mutations that disrupt P14 seem to result in the same phenotype as those that disrupt P13. It remains to be determined whether P13 and P14 are mainly involved in stabilizing the entire catalytic core or whether they instead assist in the proper docking of domain 2 into the catalytic core, which leads in turn to the binding of P1.

Long range interaction between domain 2 of subgroup ID introns and domain 6

The secondary structure of a typical member of subgroup ID is shown in Figure 2b. Although subgroup ID is very different from the other subgroups based on patterns of nucleotide conservation within the catalytic core [6], it nevertheless shares a remarkable conservation of the first base pairs of P2, P2.1, P3, P8 and their connecting elements with subgroup IC1 (compare Fig. 2a and 2b, and see below). These similarities probably reflect the similarities in the structure of the P3–P8 helices at the base of P2 and P2.1. Twelve introns, including SdCob.1 in the cytochrome *b* gene of *Saccharomyces douglasii* [32], have a P2.1 stem typically formed by three helices (P2.1, P2.1a and P2.1b) separated by two A-rich internal loops (L2.1 and L2.1a) (see Fig. 3c). Interestingly, the terminal loop of stem P2.1b is involved in a phylogenetically conserved tertiary base pairing of 4–6 basepairs, which we term P16 (Figs 2b and 3c). The P2.1 stem–loop of subgroup ID introns interacts with loop L6, at the end of a 6-base pair stem, however, unlike its homolog in subgroup IC1 introns, P13, which contacts domain 9.

Despite the fact that intron SdCob.1 does not self-splice very efficiently at magnesium concentrations below 15 mM [33], we have attempted to probe the P16 pairing of this intron and chose to use optimal splicing conditions for this purpose. Mutant transcripts with a disrupted (Sda and Sda′) or restored (Sdaa′) P16 putative tertiary pairing Fig. 4e), were generated by site-directed mutagenesis and tested in self-splicing assays (see Materials and methods). At 45 °C and 50 mM magnesium, the Sda and Sda′ transcripts are completely inactive, whereas the Sdaa′ transcripts partially recover the activity of the wild-type

Figure 4



Evidence for the existence of P13, P14 and P16. **(a)** Wild type (w.t.) and mutant P13 combinations of the TtLSU intron. Mutations in L2.1 and L9.1a are shown in red **(b)** *In vitro* self-splicing activity of the wild type and P13 mutant transcripts of the Tt.LSU ribozyme generated from linearized plasmid DNA (see Materials and methods; 2 mM $MgCl_2$ and 1 mM GTP, 30 °C). RNAs that had been uniformly labelled with [α-^{32}P]UTP were separated by electrophoresis in a 5 % polyacrylamide/ 8M-urea gel. An autoradiogram is shown. Reaction products were identified using size markers (not shown). **(c)** Wild type and mutant P14 combinations of the TtLSU intron. **(d)** *In vitro* self-splicing activity of wild type and P14 mutant transcripts of the TtLSU intron generated from a PCR-created matrix (see Materials and methods; 2 mM $MgCl_2$ and 1 mM GTP at 35 °C). RNAs were uniformly labelled with [α-^{32}P]ATP. The ligated exons at the bottom of the gel are not shown. **(e)** Wild type and mutant P16 combinations of the SdCob.1 intron. **(f)** *In vitro* self-splicing activity of the wild type and P16 mutant transcripts of the SdCob.1 ribozyme (50 mM $MgCl_2$, 1 mM GTP, 45 °C). RNAs were uniformly labelled with [α-^{32}P]UTP. The 5′ exon at the bottom of the gel is not shown.

(Fig. 4f). In fact, we were unable to activate the RNA transcripts with a defective P16 pairing, even under conditions that enhance RNA stability (100 mM $MgCl_2$, 2 mM spermidine). Although these results clearly prove the existence of the P16 pairing, we cannot distinguish between a loss of activity resulting from the destabilization of the tertiary structure and a loss of activity resulting from the removal of a sequence directly involved in

catalysis. The former is more likely than the latter, however, given the peripheral location of this non-universal tertiary interaction.

Three-dimensional modelling of the group IC1 intron

A limited number of strong structural constraints involving peripheral domains may be enough to define the global architecture of an RNA molecule in a 3D model. The identification of base pairings P13 and P14 here, and of two other long-range interactions by Cech and his collaborators [15,34], has allowed us to extend our modelling of the TtLSU intron [6] to the entire 414-nucleotide molecule. In particular, pairings P13 and P14 proved essential for both positioning domain 2 with respect to the catalytic core, and simultaneously constraining the orientation of helices P9.1 and P5c. Only three terminal loops (L6b, L8 and L9.2) that are not known to be directly involved in a tertiary interaction remain. Even so, the orientation of helices P6b, P8 and P9.2 is strongly constrained by other secondary structure components.

We chose the revised version of our model for the catalytic core of TtLSU [19] as a starting point for a new model. The core of group I introns is organized in two domains: P4–P5–P6 (here abbreviated as P4–P6) and P3–P7–P8 (P3–P8) (Fig. 1). The P1–P10 helical substrate is positioned in a cleft formed by these two domains. In this respect, the revised version of the model does not differ significantly from the original one [6]. The current model takes into account the data of Pyle *et al.* [35], however, who showed that the base that interacts with the ribose at position –3 of the 5′ exon is A302, rather than G303, as initially proposed [6]. As a consequence, residue A301, previously proposed to interact with G26, is now proposed to contact the ribose of residue G25, consistent with the data of Strobel & Cech [36,37]. Data on the accessibility of riboses to cleavage by Fe(II)–EDTA proved invaluable for modelling the peripheral domains of the TtLSU molecule, and especially for orienting the helices towards each other [14,17,20–22].

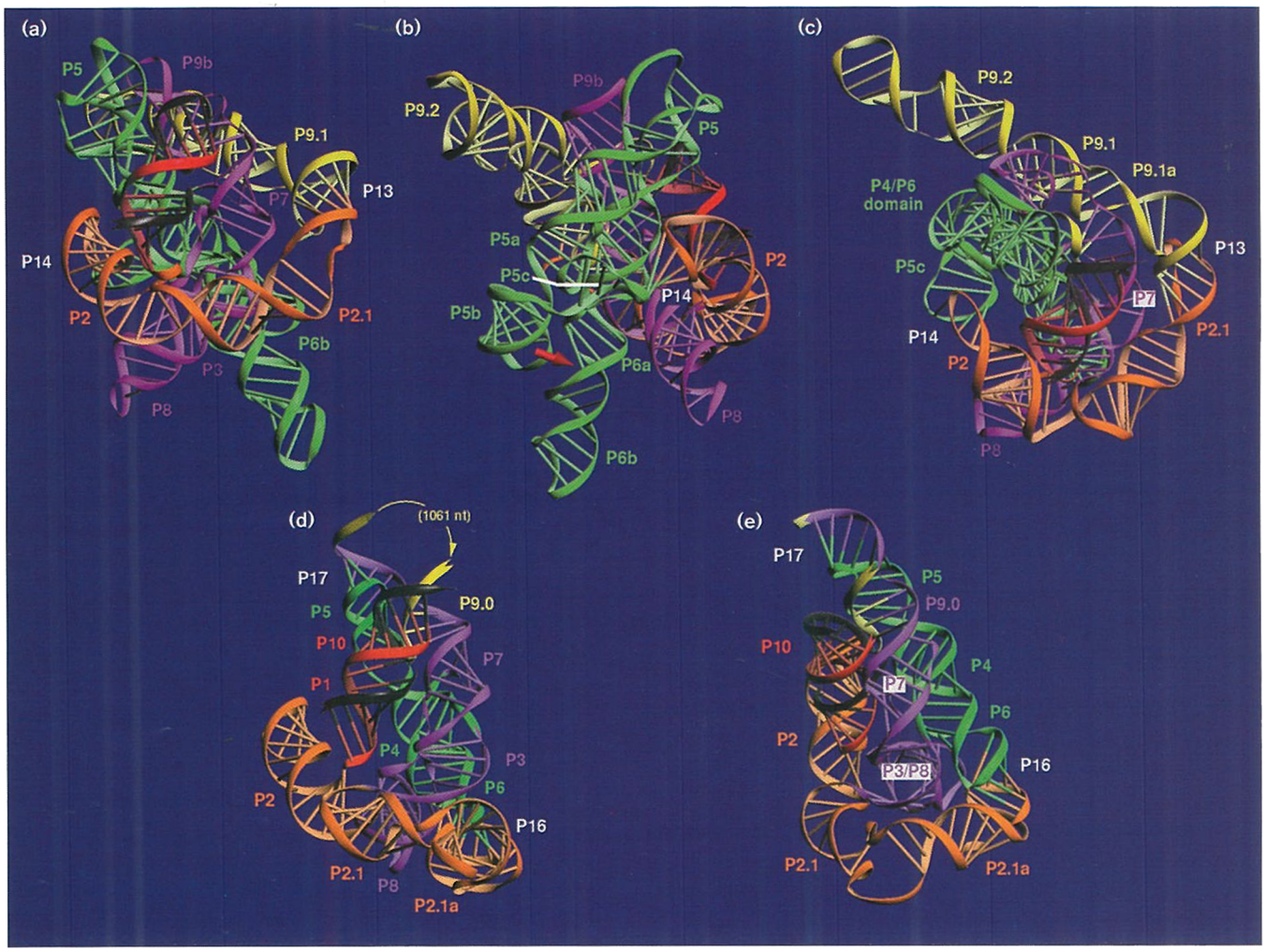
We began by modelling the P5abc extension of the P4–P6 domain, followed by domain 9. Once these two pieces had been grafted to the catalytic core, it became possible to orient helices P2 and P2.1, which interact with components P5c and P9.1, respectively, and to model their junctions with helices P3 and P1.

Modelling of the P4–P6 domain

Crystals are now available for the P4–P6 domain of the *Tetrahymena* intron [38] (nucleotides 104–261 in Fig. 2a), a stable domain that folds independently [14,18]. Although atomic coordinates for the P4–P6 domain were expected, we modelled this domain based on available experimental data (largely from Cech's laboratory) to build a complete 3D structure of the intron. Our model of the tertiary structure of the P4–P5–P6 domain does not differ significantly from the one proposed by Murphy & Cech [15] but includes all the junctions between helices and the P5c hairpin. The P4–P6 domain has a 180° bend between helices P5 and P5a [14] (see Fig. 2a and Fig. 5b). This bend allows the formation of a tertiary contact between the A-rich bulge within the P5a helix and the third base pair of helix P4 [34] and another tertiary contact between the L5b tetraloop and the internal loop between P6a and P6b [14,15]. Since Flor *et al.* [34] have shown that mutation of G212 in P4 affects the accessibility of A183 in the A rich-bulge to DMS modifications, A183 has been placed in the shallow groove of the third base pair of P4 (C109–G212). The other bases within the A-rich bulge are highly conserved in group I introns, so sequence comparisons do not provide much structural information. In our model, the presence of adenines can however, be justified by stacking interactions and/or the formation of multiple hydrogen bonds with the phosphate backbone or among themselves. Since L5b must reach the last base pair of P6a (C223–G250), helix P5b was positioned in continuity with helix P5a, consistent with Fe(II)–EDTA protection patterns (Fig. 6). Recently, GAAA tetraloops were shown to bind an 11-nucleotide RNA motif CCUAAG...UAUGG [31], unlike GURA tetraloops, which had previously been shown to interact with CU:AG or CC:GG helices [11]. When the last nucleotide of the tetraloop (A153) is modelled in interaction with the shallow groove of the last base pair of P6a (C223–G250), as proposed by Murphy and Cech [15], the two preceding residues, A151 and A152, can potentially interact with residues A248 and U249 respectively, within the internal loop between P6a and P6b. The position of helix P5c would have been impossible to predict without prior knowledge that L5c interacts with loop L2. This constraint imposes the helical axis of P5c to be almost perpendicular to helices P4 and P6 (Fig. 5b). Consequently, the 5′ strand of P5c is shielded from the solvent by P4 (Fig. 6). Helix P6b, which is not protected from OH radicals (Fig. 6), is positioned in continuity with P6a in the model (Fig. 5). Our previous modelling of the internal loop between P6 and P6a [6] had to be slightly modified to consider a covariation between the first base pair of P6a (G220–U253 in *Tetrahymena*) and the second base of the J6a/6 junction (C255 in *Tetrahymena*) [39]. When position 255 is A, there is usually a C–G basepair between positions 220 and 253, whereas when there is a C at position 255 the basepair at 220–253 is usually G–U. This observation can best be rationalized by proposing that nucleotide 255 interacts with the Hoogsteen positions of the conserved A219, this non-canonical base pair being stacked on the top of helix P6a. When the non-canonical base pair at 219–255 is A–C, a G–U as the first base pair of P6a provides better stacking than C–G. The opposite is true when the non-canonical base pair is A–A.

Some parts of the P4–P6 domain should become less accessible to the solvent in the context of the entire

Figure 5



Three-dimensional architectures of subgroup IC1 (TtLSU) and ID (SdCob.1) introns. The stage shown is just prior to the ligation step (see also Fig. 1). The colors and labeling are consistent with Figures 1 and 2. Domains P4–P6 (green) and P3–P8 (violet) create a cleft in which is positioned the helical substrate P1/P10 formed by the two exons (black) base paired to the internal guide sequence at the 5′ end of the intron (red). Domain 2 is orange and domain 9 is yellow. **(a)**, **(b)** and **(c)** Three-dimensional model of the TtLSU intron (subgroup IC1) (a) classical orientation. (b) 90° rotation to the left of view (a). (c). View from the top of (a). **(d)** and **(e)** Three-dimensional model of the SdCob.1 intron (subgroup ID). (d) Classical orientation. (e) 90° rotation to the right of view (a).

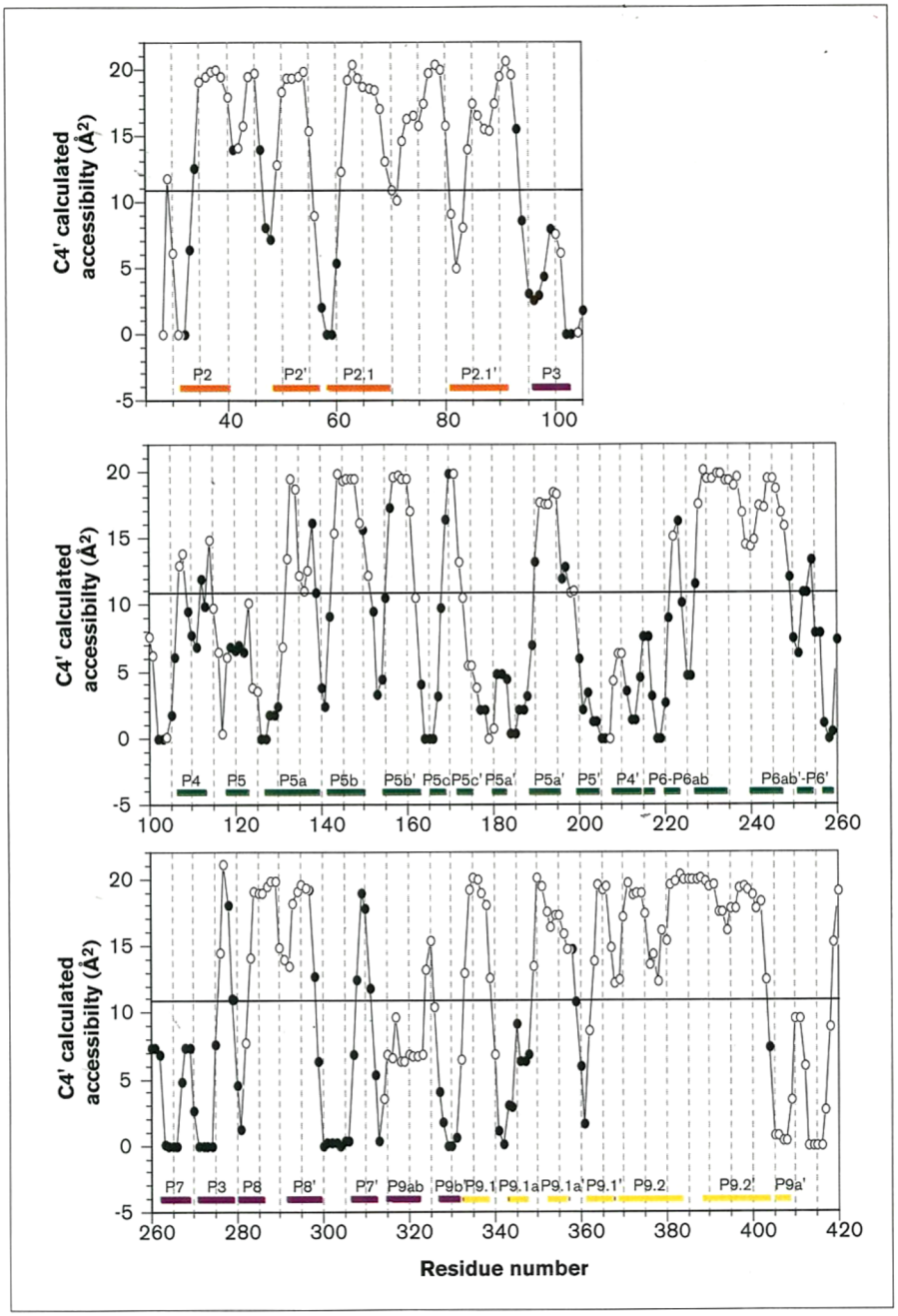
ribozyme. It is interesting to note that differences in the calculated accessibility of the C4′ atoms (data not shown) correlate well with experimentally observed differences in Fe(II)–EDTA protection patterns [14,21].

Modelling of domain 9

The second stage in the modelling involved grafting helices P9, P9.1 and P9.2 onto the core of the TtLSU intron. Helix P9 of the TtLSU intron can be regarded as a composite helix (Fig. 2). Subgroup IC1 introns that have the P13 interaction appear to include an additional base pairing P9a, which joins nucleotides –9 and –10 (counting upstream from the 3′ splice junction) to the first two nucleotides downstream of the 5′ strand of P9.0 (see Fig. 7 and Supplementary material available). In 7 of the 14 IC1 introns that have a potential P13 pairing, helix P9 is capped by a GNRA tetraloop, which is known to covary with the second and third base pairs of P5 [31]. Thus, P9b (i.e. P9 minus P9a) is part of a composite helix of seven base pairs that may be regarded as homologous to the P9 hairpin of the *sunY* intron of bacteriophage T4 [11,12] (Fig. 7). Accordingly, we have given the same orientation to the composite P9 stem of the TtLSU intron as that of helix P9 in our 3D models of the *sunY* intron and the related intron in the large ribosomal precursor of yeast mitochondria (abbreviated as ScLSU [6,11–13]). In the *Tetrahymena* model, the L9 tetraloop (UAAU) lies close to the distal base pairs of P5 and could potentially interact

Figure 6

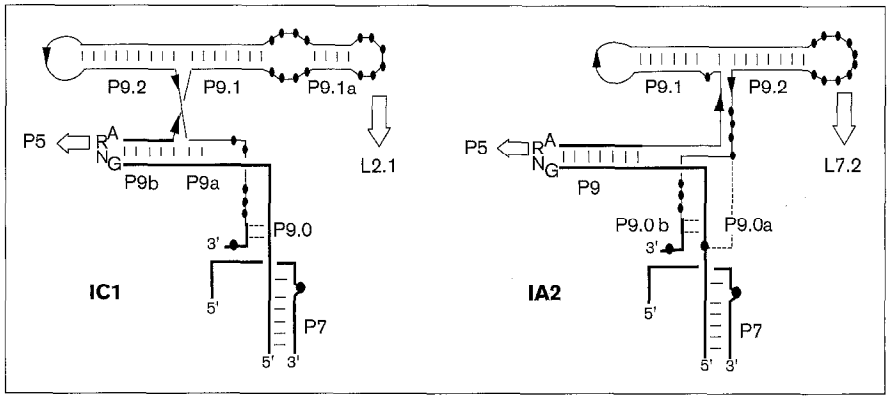
Theoretical and experimental accessibilities of the *Tetrahymena* ribozyme to Fe(II)–EDTA. The theoretical accessibility (line –) of the C4′ atoms is calculated for the model with the program ACCESS [85] using a 2.8 Å radius sphere. Open circles (○) and filled circles (•) indicate nucleotide positions accessible to, and protected from, cleavage by Fe(II)–EDTA, respectively [21,22]. Pn and Pn′ indicate the 5′ and 3′ strands of a pairing, respectively.



with P5 or the internal loop above P5 (Fig. 5). We lack experimental and phylogenetic data to propose a specific tertiary interaction since this loop is specific to the *Tetrahymena* intron. The protection towards Fe(II)–EDTA cleavage observed in the 3′ strand of P9b (Fig. 6) is also observed in the 3′ strand of helix P9 from the *sunY* intron [21], however, corroborating our view of a similar structural arrangement of P9 in both introns. The present model allows for the stacking of P9b on top of P9a, which leads in turn to the positioning of P9.1 in continuity with P9.2, despite a slight lateral gap between their helical axes (see Fig. 5c). Unlike P9, stems P9.1 and P9.2 of the TtLSU intron are not homologous to those of the *sunY* intron. To form P13 between L2.1 and L9.1a, P9.1 must be oriented towards P7, in the opposite direction of P9.1 in the *sunY* intron. Thus, P9.1 in the TtLSU intron corresponds to P9.2 in the *sunY* intron [11] (Fig. 7, and see below). Stem P9.1 consists of two base-paired segments separated by an internal loop that is rich in adenines (see the alignments in Fig. 3). Interestingly, this loop lies in the proximity of P7, perhaps allowing direct interaction of the 5′ strand with the sugar-phosphate backbone of P7

Figure 7

Two-dimensional diagrams of domains 9 from subgroup IC1 and IA2 introns. Large empty arrows indicate long range tertiary interactions. Heavier lines indicate those structures from IC1 and IA2 introns that may be regarded as homologous.



(Fig. 5). This modelling of domain 9 of TtLSU is fully consistent with the strong protection from Fe(II)–EDTA cleavage observed in P9 and P9.1 (Fig. 6) [21,22]. The 3′ strand of stem P9 could thus be protected from OH radicals by the shallow groove of P5, while stem P9.1a would be protected from cleavage by the 5′ strand of P7 and the end of the 3′ strand of P7. Full Fe(II)–EDTA cleavage of P9.2 can be explained by the fact that P9.2 is pointing outwards from the core (Fig. 5).

Modelling of domain 2

The distance between loops L5c and L9.1a is such that domain 2 must span almost 90 Å to form the long range interactions P13 (between L2.1 and L9.1a) and P14 (between L2 and L5c). These interactions can be made by orienting stems P2 and P2.1 perpendicular to P1, towards L5c and L9.1, respectively (Fig. 5a), creating an angle of almost 130° between the helical axes of P2 and P2.1 (Fig. 5c). Comparative sequence analysis assigns a length of seven base pairs to the P13 pairing (see Fig. 3). To generate a P13 helix of that length, we had to disrupt the two apical base pairs of P2.1 and P9.1a (Fig. 2). In our model, P13 is stacked on P2.1 and lies in the proximity of core element P7 (Fig. 5). The L2–L5c interaction was modelled with four base pairs (Fig. 5b). It is possible, however, that this interaction includes fewer Watson– Crick base pairs, since some of the other introns with a putative P14 pairing have only three base pairs (see Fig. 3). We cannot rule out the possibility that some nucleotide positions within loop L2 are important only for the structure of the loop itself. The effects of mutations within this loop nevertheless suggest that the last four bases of L2 (nucleotides 43–46) could be part of the P14 interaction [27] although the mutations at positions 44 and 45 are much more detrimental for substrate cleavage than those at position 43 and 46. To insert stems P2 and P2.1 between core elements P1 and P3 we considered the structural constraints suggested by phylogenetic analysis of IC1 introns (Fig. 8), site-directed mutagenesis and cross-linking data [26,40]. The sequence at the base of domain 2 and surrounding elements is highly conserved among IC1 introns (Fig. 8). On the other hand, the length of the junction connecting P1 and P2 (J1/2) covaries with the length of stem P1; J1/2 segments of 0 and 3 nucleotide correlate with P1 helices of 5 and 6 basepairs, respectively. This suggests that the 3-nucleotide J1/2 of TtLSU bulges out instead of being stacked under P2 or P1. P2 was therefore oriented so that its first nucleotide (A31) might be connected to residue G26 from the fifth base pair of P1. As a result, the J1/2 junction of the *Tetrahymena* ribozyme is very compact and could potentially interact with P1, J7/3, P3 or P2.1 to lock the position of P1 (Fig. 8b), as suggested by experimental data [24,25]. A UV cross-link was identified between nucleotides A57 and A95 of the *Tetrahymena* ribozyme [40], two residues phylogenetically conserved among subgroup IC1 and ID introns (Fig. 8a). This cross-link, which occurs in the active form of the ribozyme [40], imparts a characteristic structure to the conserved base of stems P2 and P2.1. Thus, in our model, A57 and A95 are stacked and could restrain the local structure of junctions J2/2.1 and J2/3 by making several contacts with G279 (at the P3–P8 junction) and the first two conserved base pairs of P2.1 (see Fig. 8). In particular, the presence of a conserved G–U pair inP2.1 leads to the formation of a pocket in the shallow groove of that helix that could bind the ribose of A95. Although the interactions we propose within this local region of the *Tetrahymena* ribozyme are consistent with the site-directed mutagenesis data of Downs and Cech [26], it must be kept in mind that this is not a crystallographic structure at atomic resolution. Multiple contacts mediated through water molecules or magnesium ions are likely to contribute to the stability of domain 2. Despite this, our modelling of domain 2 is consistent with Fe(II)–EDTA protection data (Fig. 6), with the 5′ strand of P2 and the 5′ strand of P2.1 protecting the 5′ strand of P8 (Fig. 5). Another reason that our model is plausible is that those helices inserted within stems P2 or P2.1 in some IC1 introns would always point towards the solvent (Fig. 3).

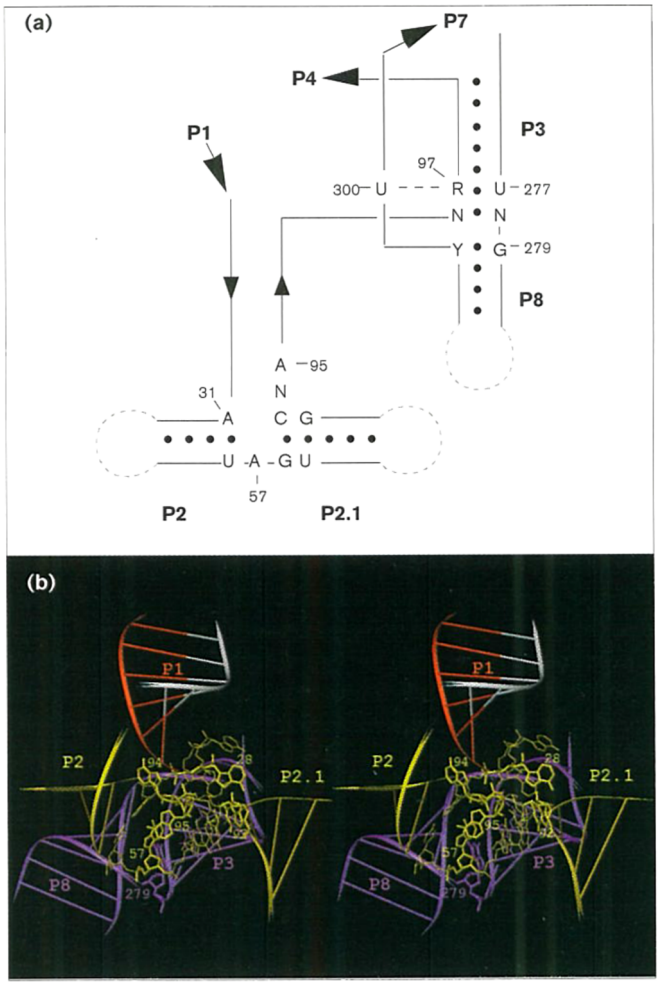
This model of the TtLSU intron enables the strong protections from Fe(II)–EDTA cleavage of core elements P3, P7 and P8 and their connections [21,22] to be rationalized (see Fig. 6). According to the 3D model shown in Figure 5, the *Tetrahymena* ribozyme has a globular shape, with the majority of its peripheral components being involved in long range contacts. Only P6b, P8 and P9.2 point outside from the core towards the solvent.

Three-dimensional model of SdCob.1, a group ID intron

We have already assumed that the sequence conservation shared by subgroups ID and IC1 in domain 2 reflects a common local structure of the base of domain 2, which is essential for the positioning of stems P2 and P2.1 relative to the group I core structure (Fig. 8). The constraints that result for domain 2 have allowed us to model those peripheral regions that surround the P16 tertiary base pairing in the SdCob.1 subgroup ID intron. Thus, the local structure of the base of domain P2 and its orientation towards P3 and P8 were assumed to be exactly the same as in the TtLSU intron (Fig. 8). Still, introns of subgroup ID can be distinguished by the following features. First, the J1/2 junction of SdCob.1 is shorter than that of the TtLSU intron: the first base pair of P2 and the U–G base pair of P1 are usually separated by seven nucleotides in ID introns instead of nine nucleotides in most IC1 introns (see Supplementary material and Fig. 2). Second, in most ID introns, helix P3 can be extended by one base pair with respect to TtLSU intron (Fig. 2). And third, there is one nucleotide missing between residue G222 (G303 in TtLSU) and the 3′ strand of P7 in the J8/7 segment of ID introns. We have taken into account these constraints by slightly reorienting helices P3 and P8 with respect to the other conserved core elements. Residue G222 in the J8/7 segment of SdCob.1 was kept at the same position as residue G303 in the TtLSU model because this base is almost universally conserved among group I introns (see appendix in [6] and Fig. 2). We then slightly rotated the entire section formed by P2, P2.1, P3 and P8 to bring the base of P1 closer to that of P2. Finally, the various elements of P2.1 (L2.1, P2.1a, L2.1a and P2.1b) were assembled to generate the P16 pairing between L2.1b and L6. We were able to form a P16 interaction of six base pairs, nucleotides 87–92 of L2.1b being paired with nucleotides 167–172 of L9.1, without disrupting the secondary structure.

The model (Fig. 5d,e), although not yet optimal, does show both the similarities (Fig. 8) and the differences (J1/2, J8/7, J7/3) within the consensus sequences of introns of subgroups ID and IC1 and it also suggests additional potential tertiary contacts. Thus, in most subgroup ID introns with a potential P16 pairing, stem P8 comprises six base pairs and ends with a UGNAAU terminal loop. In our model, the two adenines of this loop could interact with the 5′ strand of the conserved internal loop between P2.1a and P2.1b (Fig. 3c). Junction segments in P2.1 have been

Figure 8



The peripheral domain 2 of subgroup IC1 and ID introns. **(a)** Consensus sequence of IC1 and ID domains 2 and surrounding elements. **(b)** Detailed stereoview of the base of helices P2, P2.1, P3 and P8 and their junctions. The sequence and numbering is that of TtLSU. The postulated contact between U300 and the second base pair of P3 was first proposed by Michel & Westhof [6] for subgroups IC1 and IC2. Position 299 is always a pyrimidine (Y) except in TtLSU and PpLSU. The RMS deviation between the 3D models of TtLSU and SdCob.1 is no greater than 1.5 Å in this region.

modelled to be consistent with the shortest existing junctions in those ID introns with a P16 interaction. One phylogenetic argument suggesting that P2.1 should have the same orientation in the SdCob.1 and TtLSU introns is that five introns (PaCob.2, NcCob.1, CsCob.1, AmCob.3, PwOx1,2) have two helices inserted in the 3′ strand of P2.1, which is the strand that lies on the outside of the core (Fig. 3c). It is also interesting to note that the SdCob.1 intron includes an additional tertiary base pairing, called P17, between L5 and the 5′ strand of the L9.0 internal loop [33] (Fig. 2). This long range interaction, which is important for the folding of the intron into an active conformation [33], is in the same place as the proved L9/P5 tertiary

interaction, which involves the GNRA terminal loop that exists at the tip of stem P9 in more than half of group I introns [9]. The fact that we were able to compensate for the removal of the L9/P5 interaction from the *td* intron of bacteriophage T4 by engineering a pseudoknot between L5 and a sequence downstream of P7 [12] suggests that the two interactions may be regarded as equivalent.

Discussion

The identification of long-range interactions within a large RNA molecule contributes to the understanding of its 3D architecture (reviews [4,41–43]). Here, we provide evidence for the existence of three additional long range interactions involving the peripheral domains of IC1 and ID introns. We have used these additional structural constraints to construct complete 3D models of two members of these subgroups, based on the pre-existing model of the catalytic core [6,19]. The addition of peripheral components did not require any extensive revision of the core structure, thus supporting the earlier model. Considering the large size of group I intron RNA molecules and the lack of detailed structural information on them, our models should obviously not be regarded as having atomic resolution. But as the relative positions of helices are dictated by long-range interactions, we expect their overall arrangement to be fairly accurate. The 3D architecture we propose here for the IC1 intron of *Tetrahymena thermophila*, one of the most studied ribozymes, is generally in very good agreement with the available experimental data (see Results and the discussion below). Moreover, this model provides some insights into the function of domain 2 in the stabilization of the P3–P8 core domain and the positioning of the P1 substrate.

The architecture and folding of the *Tetrahymena* ribozyme

As first suggested by Fe(II)–EDTA probing studies [14, 15,20–22] and later corroborated by electron microscopy [44], the *Tetrahymena* ribozyme is extremely compact. Almost all of the core elements are protected from cleavage by OH radicals [21], a situation that could not be accounted for without invoking the contributions of the intron peripheral elements. The compact and globular shape of the full-length *Tetrahymena* ribozyme, as revealed by electron microscopy [44], is fully consistent with our 3D model. The distances measured between sites in the modelled molecule range from 85 to 135 Å (the distances between J1/2 and P5b, P13 and L8, J5/5a and L6b, J1/2 and L9.2 are 85, 95, 128 and 135 Å, respectively), consistent with the average diameter of 116 (±14) Å obtained by electron microscopy for the TtLSU intron [44]. The calculated radius of gyration of our model is 36.7 Å, only 14.3 Å larger than the one calculated for the X-ray structure of $tRNA^{Asp}$, and thus slightly smaller than expected from the difference in size between $tRNA^{Asp}$ and the *Tetrahymena* ribozyme (415 nucleotides against 73 nucleotides). While the proven stabilization of the P4–P6 core domain by the P5abc extension might account for the extensive protection from Fe(II)–EDTA cleavage within this region of the molecule, it would have been difficult to interpret the pattern of protection observed in helices P3 and P7 without knowing that domains 2 and 9 interact via the P13 loop–loop interaction. In our model, most of the residues in P7 and P3 have their C4′ atoms screened by peripheral elements P9.1, P2.1 and P13 (Fig. 5). Only the 3′ strand of P7 is somewhat more accessible to the solvent than expected from chemical probing data (Fig. 6). In fact, the topology of domains 2 and 9 after formation of the P13 interaction readily suggests how the P3–P8 core domain is stabilized in the *Tetrahymena* ribozyme [30]. Domain 2 of the *Tetrahymena* intron may be regarded as forming the two horizontal arms of an anchor, in which the P1 substrate forms the vertical axis. By interacting simultaneously with domains 5 and 9, the terminal loops of domain 2 appear to assist in the docking of P1 into the cleft between the P4–P6 and P3–P8 core domains. In fact, available experimental data (see Results and [26,27]) do suggest that domain 2 is involved in the proper docking of P1 on the catalytic core.

Additional experiments are clearly required to assess the function of P13 and P14, but we may already anticipate the role of domain 2 in the folding process of the *Tetrahymena* ribozyme. Using time-resolved oligonucleotide probe hybridization, Zarrinkar and Williamson [45,46] showed that the rate-limiting step in the folding of the molecule is the formation of core elements P3 and P7, while folding of the P4–P6 domain comes first. Interestingly, peripheral domain 9 was shown to guide the formation of domain P3–P8 [30]. We now have experimental evidence showing that, as previously suggested [22,29,30], P13 participates in the stabilization of the catalytic core, presumably by locking the P3–P8 core domain into place, functioning in a similar way, to the P12 tertiary interaction in the *sunY* intron [11]. In fact, it is likely that domain 2 of the *Tetrahymena* intron behaves somewhat like domain 2 of the yeast mitochondrial group I intron ScCob.5, except that the latter requires the presence of the ribonucleoprotein CBP2 [47–49]. It is, therefore, plausible that the folding process of the TtLSU intron would include, as a final step after core assembly, the docking of domain 2 to the catalytic core, locking the core in place and allowing binding of P1.

While our model accounts for the positioning of P1 for the splicing reactions, it does not explain how the cyclization reaction occurs. In the *Tetrahymena* ribozyme, cyclization requires a translocation of the IGS by three base pairs. During cyclization, the U residue preceding the cyclization site appears to be paired to G25 instead of G22 [28]. Thus, as suggested by Downs and Cech [26], it is likely that a local rearrangement of domain 2 precedes the cyclization reaction. This rearrangement probably consists of disrupting the P14 interaction and/or rearranging junctions J1/2 and J2/2.1.

In our 3D model of the TtLSU intron, the orientation of RNA helices was essentially constrained by long-range contacts that were assumed to co-exist with an active core in the final folded form of the molecule. Recently, the relative orientation of RNA helices in the TtLSU ribozyme was determined by electron microscopy using molecules with engineered helical extensions [50]. Interestingly, the angle measured experimentally between P8 and P6b is similar to the one in our model. The positioning of stem P2.1 relative to helices P6b and P8, as inferred from helix extension electron microscopy, is incompatible with the formation of the P13 tertiary interaction, however. In the model proposed by Nakamura *et al.* [50], stem P2.1 is pointing in the opposite direction to the core. Although it is possible that the tertiary interactions P13 and P14 are only formed transiently by the ribozyme, disruption of P13 or P14 has detectable effects only at magnesium concentrations below 3 mM (see Results). By contrast, versions of the *Tetrahymena* ribozyme with an extended P2.1 helix were analyzed at magnesium concentrations above 10 mM (i.e. in conditions under which P13 and P14 are dispensable in solution). Thus, it is possible that the electron micrographs show molecules in which domain 2 is not properly docked on the catalytic core, with magnesium ions compensating for the lack of P13 and P14. Note also that protection of core elements P3 and P7 from the solvent is difficult to account for in the model of Nakamura *et al.* [50].

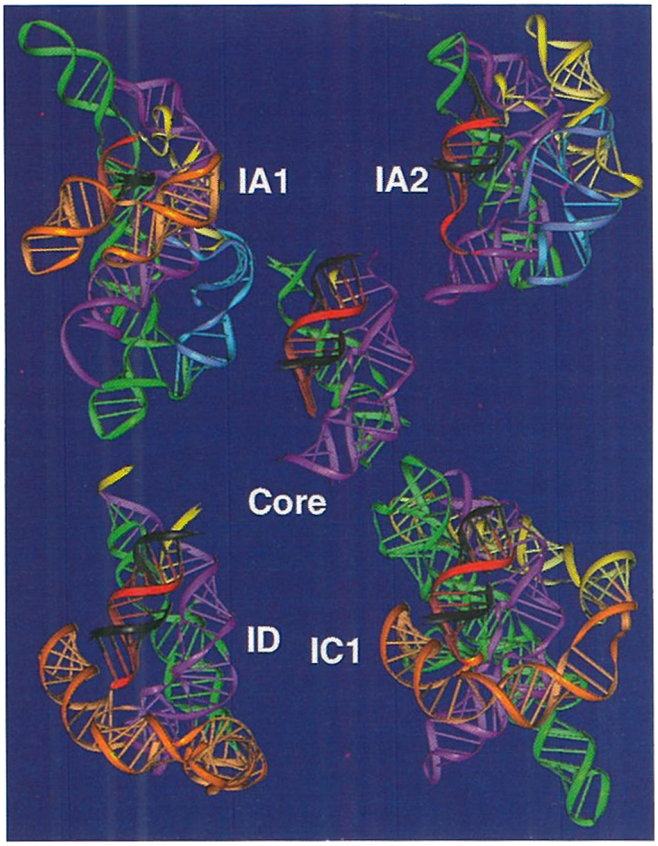
The modular organization of group I introns

Given the diversity of peripheral components in group I introns, it is remarkable that they should coexist with a constant architecture of the group I catalytic core. In this respect, it is particularly revealing to compare the architecture of group I introns from different subgroups. There are now full 3D models available for introns belonging to subgroups IA1, IA2, IC1 and ID ([10,11] and this work; Fig. 9) and the variety of shapes and sizes that mask a common catalytic core is striking. But the underlying pattern is more similar than is apparent from a casual inspection of these models. Although the peripheral appendices are very different in different introns, they all participate in the formation of similar long-range tertiary interactions and all contribute to solving the problem of folding a common core essentially by stabilizing elements P3 or P7.

A closer look at the models shows that there are similar structural features within peripheral domains, such as similar types of long-range interactions acting as RNA–RNA anchors. Two main classes of long-range RNA motifs may be distinguished; loop–helix interactions and loop–loop interactions (see [10,11,51]). Loop–helix interactions involve GNRA tetraloops that are able to recognize base pairs in the shallow groove of a helix [6,12,52] or, like GAAA loops, a conserved internal loop motif [15,31]. This type of interaction exists between loop L9 and helix P5 in introns from subgroups IA1, IA2 and IC1, but also between L2 and P8 of IA1 or IA2 introns, or L5b and P6 of IC1 introns. Similarly, there is a modular use of loop–loop anchors within peripheral domains of group I introns. Interactions P11 in subgroup IA1 [10], P12 in subgroup IA2 [11,13], P13 and P14 in subgroup IC1 (this work) and P16 and P17 in subgroup ID (this work and [33]), all belong to the loop–loop anchor class. They were modelled as Watson–Crick base pairings, resulting in pseudoknots [53], although P12 and P14 could well be a mixture of base pairing and intercalation like the intramolecular association between the thymine and dihydrouridine loops of tRNAs [54]. Interestingly, most of the loops involved in these tertiary base pairings contain six to eight nucleotides. These loops can potentially adopt conformations very similar to that of the tRNA anticodon loop of seven nucleotides, which would favor the stacking of nucleotides on the 3′ side of the loop in continuity with the 3′ strand of the helical stem. Such an arrangement would in turn facilitate the formation of tertiary base pairings during the folding process, as occurs for the closely related intermolecular interactions of anticodon loops [55,56] and for the interactions that result in the initiation or 'kissing' complexes of natural antisense RNAs with their targets [57–62]. Clearly, the stability of such base pairings must be highly dependent on the context in which they are found when they consist of intramolecular associations between loops such as those found in group I introns (shown here) or RNase P [63]. The pairing of two distant regions leads to the cooperative formation of a chain of additional interactions in the immediate environment of that pairing, and such interactions are unlikely to be shared by other RNA molecules. The importance of tertiary anchor motifs in a global molecular context, therefore, lies more in their ability to ensure specific recognition between distant sites and to restrain the number of alternative conformations than in the intrinsic stability of the interaction itself. Also, similar 3D motifs may be used to bring distant peripheral domains closer together, even though the peripheral structures they join may differ from subgroup to subgroup.

Peripheral domains from different subgroups could also have similar structurally based functions, such as the positioning of P1 or the stabilization of core elements within domain P3–P8. Group I introns have evolved at least two different structural mechanisms to help the P1 substrate to dock into the catalytic core, typified by the molecular devices found in the ScLSU and *Tetrahymena* introns. Both involve the formation of long-range tertiary interactions, but these depend on the position of P2 relative to P1 (Fig. 9). Although P2 is stacked under P1 in the ScLSU intron, P2 is almost perpendicular to P1 in the TtLSU intron. Most group I introns also appear to have developed strategies to stabilize either P3 or P7 within the P3–P8 core domain. In the TtLSU intron, domain 2 interacts with domain 9 to form P13 and tethers P7, whereas the related domain 2 from the SdCob.1

Figure 9



Comparison of subgroup IC1 and ID introns with subgroup IA1 and IA2 introns. Center: the common architecture of the group I intron catalytic core [6,19]. Clockwise from the top left corner: the ScLSU intron, from the small ribosomal RNA precursor of *Sacharomyces cerevisiae* mitochondria, a representative of subgroup IA1 [10]; the *sunY* intron from bacteriophage T4, a representative of subgroup IA2 [11,13]; the TtLSU intron, a representative of subgroup IC (this study); the SdCob.1 intron, a representative of subgroup ID (this study, [33]). The colors are consistent with Figures 1 and 5. Exons (lower case) are bound to the internal guide sequence (red) at the 5′ end of the ribozyme. The P4–P6 domain is green and the P3-P9 domain is violet. Peripheral domain 2 is orange and domain 9 is yellow. For the ScLSU and *sunY* introns, the peripheral domains inserted between P3 and P7 are light blue.

intron interacts with domain 6 to form P16, tethering P3 (Fig. 9). Similarly, the peripheral domain 7 of introns ScLSU and *sunY*, (inserted between P3 and P7) interacts either with domain P6 (subgroup IA1 situation) or with domain 9 (subgroup IA2 situation) (Fig. 1). In the ScLSU intron, the P11 long-range pairing between loops L7.1a and L6a tethers P3, whereas in the *sunY* intron, interaction P12 between loops L7.2 and L9.2 tethers P7. Structural elements located in similar spatial positions in related molecules can therefore interact differently, yet for the same purpose, namely the stabilization of the catalytic core, obtained by locking either P3 or P7 into position. That stabilizing either P7 or P3 should lead to the stabilization of the entire P3–P8 domain is fully consistent with the interdependent formation of P3 and P7 observed by Zarrinkar and Williamson [46] using the TtLSU intron.

There are only a few examples of group I introns that have no peripheral appendices but can still stabilize core elements P3 or P7. Such stabilization need not always involve the direct tethering of P3 or P7 and in fact most known group I introns possess a domain 9 which participates in the folding of the core. For example, the P9 hairpin interacts with P5 in the GC-rich group I intron from the bacterium *Azoarcus*, the smallest natural self-splicing group I intron identified to date [64]. The same interaction was shown to contribute significantly to the overall stability of the *sunY* intron [12]. In some cases, proteins assist in the proper folding of group I ribozymes. Interestingly, the stabilizing effect of the protein seems to result from the stabilization of those peripheral domains that are supposed to direct the proper folding of P3–P8 core. Thus, the *Neurospora crassa* mitochondrial tyrosyl-tRNA synthetase (CYT-18 protein), which binds to the P4–P6 domain of the *N. crassa* intron NcLSU, a close relative of ScLSU, also recognizes domain 9 [65]. Similarly, the CBP2 protein, which is involved in the stabilization of another relative of the ScLSU intron (intron ScCob.5 from yeast), recognizes domain 7, which is inserted between P3 and P7 [48] and is known to interact with domain 6 via the P11 tertiary pairing [10].

Some of the peripheral domains of group I introns are mutually exclusive because there is a spatial overlap of some of their long-range interactions (see Fig. 1a,b). Thus, domain P5abc from IC1 introns, domain 7 from IA1 introns and domain 2 from ID introns cannot coexist because they are all using domain 6 as a tethering location. Similarly, domain 2 from the IC and ID introns cannot coexist with domain 7 from IA introns because it interacts with domains 6 and 9, which are the targets of domain 7. Some other peripheral appendices are shared by different subgroups, however. Thus, the interaction between loop L9 and P5 exists in different subgroups (e.g. IC1, IA1 and IA2), while the domain 2 from ScLSU (subgroup IA1) can potentially coexist with a P5abc extension, as it does in some IB introns. Such a modular view of group I introns could lead to the design and engineering of novel combinations of peripheral components [51]. Thus, based on our current understanding of the structure of group I introns, it appears feasible to construct a molecule in which a P5abc domain interacting with P6 would coexist with domains 7 and 9 of the *sunY* intron. Since this particular arrangement has never yet been observed, this would allow one to ask whether peripheral elements govern the choice of bases within the catalytic core. While the answer is probably yes, this question would be best addressed by *in vitro* selection techniques [66] which are especially suited to reveal subtle tertiary contacts.

Comparisons with the P4–P6 domain crystal structure

The details of the crystal structure of the P4–P6 domain at 2.8 Å resolution have been published very recently [67,68]. The global architecture of the X-ray structure and of our model of P4–P5–P6 compares favorably. Indeed, the overall root mean square (rms) deviation between equivalent phosphorus atoms (157 atoms) after automatic superposition of the two molecules is 9.8 Å. We also compared all intra-molecular distances between phosphorus atoms in the two coordinate sets. The resulting intramolecular rms is 9.6 Å. (Recall that in a RNA helix, a good yardstick is the distance between successive phosphorus atoms, which is 6 Å). Four regions, in terminal and internal loops, deviate by distances between 15 and 20 Å while, in the most comparable regions, which consist of all helices, the deviations are around 5 Å. The radius of gyration of the X-ray and modelled structures is 30 Å and 33.8 Å respectively. The modelled domain is, therefore, slightly less compact than the X-ray structure. The major aim of our modelling was not to predict the conformation of the P4–P6 domain but instead to derive a global model of a full intron (429 nucleotides) belonging to subgroup IC so as to verify some proposed long-range contacts and discover new ones. We used atomic resolution during the modelling because the geometry and stereochemistry of a model must be defined to this level for the possibility that long-range contacts may form to be assessed. The replacement of our modelled structure of P4–P5–P6 by the crystallographic structure does not alter our conclusions regarding the global architecture and the underlying long-range contacts of the *Tetrahymena* intron that were derived here.

Significance

Self-splicing group I introns are remarkable systems for approaching the RNA-folding problems experimentally both in terms of physical chemistry and in an evolutionary perspective. Understanding how these complex molecules fold into a catalytically active conformation should give insight into the plasticity in the design of self-splicing group I introns.

In this work, we have identified three novel long-range interactions between peripheral elements of subgroup IC1 and ID introns and used them to produce complete 3D models of representative molecules of each of these two subgroups of group I introns. In the case of the *Tetrahymena* group I ribozyme, anchoring of domain 2 to domains 5 and 9 led to a model that rationalizes much of the available information on this extensively studied RNA molecule. Six out of the nine terminal loops of this ribozyme are involved in RNA–RNA long-range interactions that contribute either to the stabilization of the catalytic core or the positioning of the helical substrate P1.

Long range loop–loop interactions are important for the self-assembly of these structured RNA molecules into their active conformation. The strong distance constraints they impose on folding are very useful for predicting the 3D architectures of large RNAs. We show that the molecular devices evolved by group I introns exploit different structural elements in related molecules located in similar spatial positions to promote different long-range tertiary contacts that have identical functional purposes. No similar and transposable study exists in the protein universe and it is possible that the folding rules derived here are more prevalent in the RNA world, most probably because of the well-defined nature and high-energy content of RNA secondary structure motifs.

At present, 3D models are available for molecules belonging to four different subgroups of group I introns. These models emphasize the modular and hierarchical organization of the architecture of group I ribozymes. A general picture emerges in which non-homologous peripheral appendices engage in similar RNA–RNA long range tertiary interactions and, in doing so, contribute to the folding of a common catalytic core, essentially through the stabilization of either one of the two helical stems P3 and P7 within the core.

Materials and methods

DNA constructs

The *Eco*RI–*Bam*HI fragment of plasmid PSZ241 [5], which codes for the intron in the large ribosomal precursor of *Tetrahymena thermophila* (TtLSU) and surrounding sequences, was transferred into plasmid pTZ18U downstream from the T7 RNA polymerase promoter, resulting in plasmid pTZ18U/Tt. The complete TtLSU intron (414 nucleotides) is flanked by a 5′ exon of 49 nucleotides. The SdCob.1 intron is encoded by plasmid pSd2, which lacks intron nucleotides 293–1251 and includes mutations U98G/A99C and U77G/A78C in the P2.1a pairing [33]. This 352-nucleotide intron is flanked by a 5′ exon of 125 nucleotides. Mutations were introduced into the P13 and P16 pairings of constructs pTZ18U/Tt and pSd2 as described by Kunkel *et al.* [69]. Mutated versions of P14 were generated by PCR using the *Pfu* thermostable polymerase from Stratagene. All constructs were verified by DNA sequence analysis of the intron and flanking sequences.

RNA preparation, purification and renaturation

TtLSU P13 transcripts with a 44 nucleotide 3′ exon and SdCob.1 transcripts with a 164-nt 3′ exon were generated from *Eco*RI-cleaved plasmid DNA. TtLSU P14 transcripts were generated from a PCR-created matrix, cleaved by *Eco*RI. RNA synthesis and purification were carried out as described in Michel *et al.* [13]. After purification from denaturing acrylamide gels and resuspension in water, RNA samples were denatured at 90 °C (TtLSU transcripts) or 85 °C (SdCob.1 transcripts) for 2 min. Renaturation of TtLSU samples was achieved by adding 5×-concentrated splicing buffer at 60 °C (P13-matrix transcripts) or 50 °C (P14-matrix transcripts), quick-cooling to the reaction temperature and further incubation at this temperature for 3–10 min. These renaturation protocols lead to highly reproducible data. Renaturation of SdCob.1 was achieved by adding 5×-concentrated splicing buffer at 45 °C followed by an incubation of 15 min at 45 °C.

In vitro *splicing assays*

The concentration of precursor RNA was routinely set at 30 nM. For the TtLSU intron, splicing was carried out at 30 °C or 35 °C in 50 mM Tris (pH 7.5 at 25 °C), 10 mM NaCl, 2 mM $MgCl_2$, 0.02 % SDS. Splicing of

SdCob.1 transcripts was in 50mM Tris (pH 7.5 at 25 °C), 10 mM NH_4Cl, 50 mM $MgCl_2$, 0.02 % SDS. Splicing reactions were started by addition of GTP at a final concentration of 1 mM and stopped by addition of urea loading buffer and Na_2EDTA to a final concentration of 25 mM (TtLSU samples) or 100 mM (SdCob.1 samples). Samples were analyzed by loading on a 5 % polyacrylamide/8M-urea gel.

Computer modelling

Molecular modelling was performed as described by Westhof [70] and Michel & Westhof [6]. The 3D structures were geometrically and stereochemically refined with the restrained least-square program NUCLIN-NUCLSQ. Drawings were produced by DRAWNA [71] on a Silicon Graphics workstation. The coordinates are available by anonymous ftp (130.79.17.244) in directory (cd/pub).

Supplementary material available

Alignments of 35 sequences of subgroup IC1 introns and 15 sequences of subgroup ID introns.

Acknowledgements

The authors thank Annie Hoeft (IBMC, Strasbourg) and Denise Menay (CGM, Gif-sur-Yvette) for synthesizing DNA oligonucleotides. We are very grateful to Jamie Cate and Jennifer Doudna for sending us the coordinates of the crystal structure of the P4–P5–P6 domain. This work was partially supported by a grant from the CNES (96/CNES/0247).

References

1. Cech, T.R. (1993). Structure and mechanism of the large catalytic RNAs: group I and group II introns and ribonuclease P. In *The RNA world* (Gesteland, R.F. & Atkins, J.F., ed.), pp. 239–269, Cold Spring Harbor Laboratory, Cold Spring Harbor.
2. Streicher, B., Westhof, E. & Schroeder, R. (1996). The environment of two metal ions surrounding the splice site of a group I intron. *EMBO J.* **15**, 2556–2564.
3. Cech, T.R. (1990). Self-splicing of group I introns. *Annu. Rev. Biochem.* **59**, 543–568.
4. Jaeger, L., Michel, F. & Westhof, E. (1996). The structure of group I ribozymes. In *Catalytic RNA*. Vol. 10, (Eckstein, F. & Lilley, D.M.J., eds.), pp. 33–51, Springer-Verlag, Berlin Heidelberg.
5. Michel, F., Hanna, M., Green, R., Bartel, D.P. & Szostak, J.W. (1989). The guanosine binding site of the *Tetrahymena* ribozyme. *Nature* **342**, 391–395.
6. 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.
7. Been, M.D. & Perrotta, A.T. (1991). Group I intron self splicing with adenosine: evidence for a single nucleoside binding site. *Science* **252**, 434–437.
8. Beaudry, A.A. & Joyce, G.F. (1990). Minimum secondary structure requirements for catalytic activity of self-splicing group I intron. *Biochemistry* **29**, 6534–6539.
9. Van Der Horst, G., Christian, A. & Inoue, T. (1991). Reconstitution of a group I intron self-splicing reaction with an activator RNA. *Proc. Natl. Acad. Sci.* **88**, 184–188.
10. Jaeger, L., Westhof, E. & Michel, F. (1991) . Function of P11, a tertiary base pairing in self-splicing introns of subgroup IA. *J. Mol. Biol.* **221**, 1153–1164.
11. Jaeger, L., Westhof, E. & Michel, F. (1993). Monitoring of the cooperative unfolding of the *sunY* group I intron of the bacteriophage T4. The active form of the sunY ribozyme is stabilized by multiple interactions with 3′ terminal intron components. *J. Mol. Biol.* **234**, 331–346.
12. Jaeger, L., Michel, F. & Westhof, E. (1994). Involvement of a GNRA tetraloop in long-range RNA tertiary interactions. *J. Mol. Biol.* **236**, 1271–1276.
13. Michel, F., *et al.*, & Shub, D.A. (1992). Activation of the catalytic core of a group I intron by a remote 3′ splice junction. *Genes & Dev.* **6**, 1373–1385.
14. Murphy, F.L. & Cech, T.R. (1993). An independently folding domain of RNA tertiary structure within the *Tetrahymena* ribozyme. *Biochemistry* **32**, 5291–5300.
15. Murphy, F.L. & Cech, T.R. (1994). GAAA tetraloop and conserved bulge stabilize tertiary structure of a group I intron domain. *J. Mol. Biol.* **236**, 49–63.
16. Narlikar, G.J. & Herschlag, D. (1996). Isolation of a local tertiary folding transition in the context of a globally folded RNA. *Nat. Struct. Biol.* **3**, 701–710.
17. Celander, D.W. & Cech, T.R. (1991). Visualizing the higher folding of a catalytic RNA molecule. *Science* **251**, 401–407.
18. Murphy, F.L., Wang, Y., Griffith, J.D. & Cech, T.R. (1994). Coaxially stacked RNA helices in the catalytic center of the *Tetrahymena* ribozyme. *Science* **265**, 1709–1712.
19. Michel, F. & Westhof, E. (1994). Slippery substrate. *Nat. Struct. Biol.* **1**, 5–7.
20. Latham, J.A. & Cech, T.R. (1989). Defining the inside and outside of a catalytic RNA molecule. *Science* **245**, 276–282.
21. Heuer, T.S., Chandry, R.S., Belfort, M., Celander, D.W. & Cech, T.R. (1991). Folding of group I introns from bacteriophage T4 involves internalization of the catalytic core. *Proc. Natl. Acad. Sci.* **88**, 11105–11109.
22. Laggerbauer, B., Murphy, F.L. & Cech, T.R. (1994). Two major tertiary folding transitions of the *Tetrahymena* catalytic RNA. *EMBO J.* **13**, 2669–2676.
23. Joyce, G.F., Van der Horst, G. & Inoue, T. (1989). Catalytic activity is retained in the *Tetrahymena* group I intron despite removal of the large extension of element P5. *Nucleic Acids Res.* **17**, 7879–7889.
24. Young, B., Herschlag, D. & Cech, T.R. (1991). Mutations in a non-conserved sequence of the *Tetrahymena* ribozyme increase activity and specificity. *Cell* **67**, 1007–1019.
25. Herschlag, D. (1992). Evidence for processivity and two-step binding of the RNA substrate from studies of J1/2 mutants of the *Tetrahymena* ribozyme. *Biochemistry* **31**, 1386–1399.
26. Downs, W.D. & Cech, T.R. (1994). A tertiary interaction in the *Tetrahymena* intron contributes to selection of the 5′ splice site. *Genes & Dev.* **8**, 1198–1211.
27. Peyman, A. (1994). P2 functions as a spacer in the *Tetrahymena* ribozyme. *Nucl. Acids Res.* **22**, 1383–1388.
28. Been, M.D. & Cech, T.R. (1987). Selection of cyclization sites in group I IVS RNA requires multiple alignments of an internal template like sequence. *Cell* **50**, 951–961.
29. Banerjee, A.R., Jaeger, J.A. & Turner, D.H. (1993). Thermal unfolding of a group I ribozyme: the low temperature transition is primarily disruption of tertiary structure. *Biochemistry* **32**, 153–163.
30. Zarrinkar, P.P. & Williamson, J.R. (1996). The P9.1–P9.2 peripheral extension helps guide folding of the *Tetrahymena* ribozyme. *Nucl. Acids Res.* **24**, 854–858.
31. Costa, M. & Michel, F. (1995). Frequent use of the same tertiary motif by self-folding RNAs. *EMBO J.* **14**, 1276–1285.
32. Tian, G.-L., Michel, F., Macadre, C., Slonimski, P.P. & Lazowska, J. (1991). Incipient mitochondrial evolution in yeast. II. The complete sequence of the gene coding for cytochrome *b* in *Saccharomyces douglasii* reveals the presence of both new and conserved introns and discloses major differences in the fixation of mutations in evolution. *J. Mol. Biol.* **218**, 747–760.
33. Jaeger, L., Westhof, E. & Michel, F. (1996). Function of a pseudoknot in the suppression of an alternative splicing event in group I intron. *Biochimie* **78**, 466–473.
34. Flor, P.J., Flanegan, J.B. & Cech, T.R. (1989). A conserved base pair within helix P4 of the *Tetrahymena* ribozyme helps to form the tertiary structure required for self-splicing. *EMBO J.* **8**, 3391–3399.
35. Pyle, A.M., Murphy, F.L. & Cech, T.R. (1992). RNA substrate binding site in the catalytic core of the *Tetrahymena* ribozyme. *Nature* **358**, 123–128.
36. Strobel, S.A. & Cech, T.R. (1993). Tertiary interactions with the internal guide sequence mediate docking of the P1 helix into the catalytic core of the *Tetrahymena* ribozyme. *Biochemistry* **32**, 13593–13604.
37. Strobel, S.A. & Cech, T.R. (1994). Translocation of an RNA duplex on a ribozyme. *Nat. Struct. Biol.* **1**, 13–17.
38. Doudna, J.A., Grosshans, C., Gooding, A. & Kundrot, C.E. (1993). Crystallization of ribozymes and small RNA motifs by sparse matrix approach. *Proc. Natl. Acad. Sci.* **90**, 7829–7833.
39. Gautheret, D., Damberger, S.H. & Gutell, R.R. (1995). Identification of base-triples in RNA using comparative sequence analysis. *J. Mol. Biol.* **248**, 27–43.
40. Downs, W.D. & Cech, T.R. (1990). An ultraviolet-inducible adenosine-adenosine cross-link reflects the catalytic structure of the *Tetrahymena* ribozyme. *Biochemistry* **29**, 5605–5613.
41. Harris, M.E., Nolan, J.M., Malhotra, A., Brown, J.W., Harvey, S.C. & Pace, N.R. (1994). Use of photoaffinity crosslink and molecular modelling to analyze the global architecture of ribonuclease P RNA. *EMBO J.* **13**, 3953–3963.
42. Brown, J.W., Nolan, J.M., Haas, E.S., Rubio, M.A., Major, F. & Pace,

N.R. (1996). Comparative analysis of ribonuclease P RNA using gene sequences from natural microbial populations reveals tertiary structural elements. *Proc. Natl. Acad. Sci.* **93**, 3001–3006.
43. Westhof, E. & Altmann, S. (1994). Three-dimensional working model of M1 RNA, the catalytic RNA subunit of ribonuclease P from *Escherichia coli. Proc. Natl. Acad. Sci.* **91**, 5133–5137.
44. Wang, Y.-H., Murphy, F.L., Cech, T.R. & Griffith, J.D. (1994). Visualization of a tertiary structural domain of the *Tetrahymena* group I intron by electron microscopy. *J. Mol. Biol.* **236**, 64–71.
45. Zarrinkar, P.P. & Williamson, J.R. (1994). Kinetic intermediates in RNA folding. *Science* **265**, 918–924.
46. Zarrinkar, P.P. & Williamson, J.R. (1996). The kinetic folding pathway of the *Tetrahymena* ribozyme reveals possible similarities between RNA and protein folding. *Nat. Struct. Biol.* **3**, 432–438.
47. Weeks, K.M. & Cech, T.R. (1995). Protein facilitation of group I intron splicing by assembly of the catalytic core and the 5′ splice site domain. *Cell* **82**, 221–230.
48. Weeks, K.M. & Cech, T.R. (1995). Efficient protein-facilitated splicing of the yeast mitochondrial bI5 intron. *Biochemistry* **34**, 7728–7738.
49. Weeks, K.M. & Cech, T.R. (1996). Assembly of a ribonucleoprotein catalyst by tertiary structure capture. *Science* **271**, 345–348.
50. Nakamura, T.M., Wang, Y., Zaug, A.J., Griffith, J.D. & Cech, T.R. (1995). Relative orientation of RNA helices in a group I ribozyme determined by helix extension electron microscopy. *EMBO J.* **14**, 4849–4859.
51. Westhof, E., Masquida, B. & Jaeger, L. (1996). RNA tectonics: towards RNA design. *Fold. Des.* **1**, 78–88.
52. Pley, H.W., Flaherty, K.M. & Mckay, D.B. (1994). Three-dimensional structure of a hammerhead ribozyme. *Nature* **372**, 68–74.
53. Westhof, E. & Jaeger, L. (1992). RNA pseudoknots. *Curr. Opin. Struct. Biol.* **2**, 334–337.
54. Saenger, W. (1984). *Principles of Nucleic Acid Structure*. Springer Verlag, New York.
55. Grosjean, H. & Chantrenne, H. (1980). On codon-anticodon interactions. In *Molecular Biology Biochemistry and Biophysics, Chemical Recognition in Biology, Vol. 32*, (Chapeville, F. & Haenni, A.-L. eds), pp. 347-367, Springer-Verlag, Berlin, Heidelberg.
56. Westhof, E., Dumas, P. & Moras, D. (1983). Loop stereochemistry and dynamics in transfer RNA. *J. Biomol. Struct. Dyn.* **1**, 337–355.
57. Tomizawa, J. (1984). Control of ColE1 plasmid replication: the process of binding of RNA I to the primer trascript. *Cell* **38**, 861–870.
58. Tomizawa, J. (1985). Control of ColE1 plasmid replication : initial interaction of RNA I and the primer transcript is reversible. *Cell* **40**, 527–535.
59. Persson, C., Wagner, E.G.H. & Nordström, K. (1990). Control of replication of plasmid R1: structures and sequences of the antisense RNA, CopA, required for its binding to the target RNA, CopT. *EMBO J.* **9**, 3767–3775.
60. Persson, C., Wagner, E.G.H. & Nordström, K. (1990). Control of replication of plasmid R1: formation of an initial transient complex is rate-limiting for anti-sense RNA/target RNA pairing. *EMBO J.* **9**, 3777–3785.
61. Gregorian, R.S.J. & Crothers, M.D. (1995). Determinants of RNA hairpin loop–loop complex stability. *J. Mol. Biol.* **248**, 968–984.
62. Marino, J.P., Gregorian, R.S.J., Gsankovszki, G. & Crothers, M.D. (1995). Bent helix formation between RNA hairpins with complementary loops. *Science* **268**, 1148–1454.
63. Haas, E.S., Morse, D.P., Brown, J.W., Schmidt, F.J. & Pace, N.R. (1991). Long-range structure in ribonuclease P RNA. *Science* **254**, 853–856.
64. Tanner, M.A. & Cech, T.R. (1996). Activity and thermostability of the small self-splicing group I intron in the pre-tRNA Ile of the purple bacterium *Azoarcus. RNA* **2**, 74–83.
65. Caprara, M.G., Mohr, G. & Lambowitz, A.M. (1996). A tyrosyl-tRNA synthetase protein induces tertiary folding of the group I intron catalytic core. *J. Mol. Biol.* **257**, 512–531.
66. Breaker, R.R. & Joyce, G.F. (1994). Inventing and improving ribozyme function: rational design versus iterative selection methods. *Trends Biotechnol.* **12**, 268–275.
67. Cate, J.H., *et al.*, & Doudna, J.A. (1996). RNA tertiary structure mediation by adenosine platforms. *Science* **273**, 1696–1699.
68. Cate, J.H., *et al.*, & Doudna, J.A. (1996). Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science* **273**, 1678–1684.
69. Kunkel, T.A., Roberts, J.D. & Zakour, R.A. (1987). Rapid and efficient site-specific mutagenesis without phenotypic selection. *Meth. Enzymol.* **154**, 367–382.
70. Westhof, E. (1993). Modelling the three-dimensional structure of ribonucleic acids. *J. Mol. Struct. (Theochem)* **286**, 203–210.
71. Massire, C., Gaspin, C. & Westhof, E. (1994). DRAWNA: a program for drawing schematic views of nucleic acids. *J. Mol. Graphics* **12**, 201–206.
72. Cech, T.R., Damberger, S.H. & Gutell, R.R. (1994). Representation of the secondary and tertiary structure of group I introns. *Nat. Struct. Biol.* **1**, 273–280.
73. Davila-Aponte, J.A., Huss, V.A.R., Sogin, M.L. & Cech, T.R. (1991). A self-splicing group I intron in the nuclear pre-rRNA of the green alga, *Ankistrodesmus stipitatus. Nucleic Acids Res.* **19**, 4429–4436.
74. Bhattacharya, D., Damberger, S., Srek, B. & Melkonian, M. (1996). Primary and secondary structure analyses of the rDNA group I introns of the *Zygnematales (Charophyta). Curr. Genet.* **29**, 282–286.
75. Kunze, G. & Kunze, I. (1994). Characterization of *Arxula adeninivorans* strains from different habitats. *Antonie Van Leeuwenhoek* **65**, 29–34.
76. Mercure, S., Montplaisir, S. & Lemay, G. (1993). Correlation between the presence of a self-splicing intron in the 25S rDNA of *Candida albicans* and strain susceptibility to 5-fluorocytosine. *Nucleic Acids Res.* **21**, 6020–6027.
77. Neuvéglise, C. & Brygoo, Y. (1994). Identification of group I introns in the 28S rDNA of the entomopathogenic fungus *Beauveria brongniartii. Curr. Genet.* **27**, 38–45.
78. Johansen, S. & Vogt, V.M. (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**, 725–734.
79. Kranz, H.D., Miks, D., Siegler, M.L., Capesius, I., Sensen, C.W. & Huss, V.A. (1995). The origin of land plants: phylogenetic relationships among charophytes, bryophytes and vascular plants inferred from complete small-subunit ribosomal RNA gene sequences. *J. Mol. Evol.* **41**, 74–84.
80. Suh, S. & Sugiyama, J. (1994). Phylogenetic placement of the basidiomycete yeasts *Kondoa mavinella* and *Rhodoporidium dacryoidum*, and the anamorphic yeast *Sympodiomycepsis paphiopedili* by means of 18S rRNA gene sequence analysis. *Mycoscience* **35**, 367–375.
81. Pawlowski, J., Bolivar, I., Fahrni, J. & Zaninetti, L. (1994). Taxonomic identification of foraminifera using ribosomal RNA sequences. *Micropaleontol.* **40**, 373–377.
82. Paquin, B. & Lang, B.F. (1996). The mitochondrial DNA of *Allomyces macrogynus*: the complete genomic sequence from an ancestral fungus. *J. Mol. Biol.* **255**, 373–377.
83. Wolff, G., Burger, G., Lang, B. & Kueck, U. (1993). *Mitochondria* genes in the colourless alga *Prototheca wickerhamii* resemble plant genes in their exons but fungal genes in their introns. *Nucleic Acids Res.* **21**, 719–726.
84. Tian, G.-L., Michel, F., Macadre, C. & Lazowska, J. (1993). Sequence of the mitochondrial gene encoding subunit I of cytochrome oxidase in *Saccharomyces douglasii. Gene* **124**, 153–163.
85. Richmond, T.J. (1984). Solvent accessible surface area and excluded volume in proteins. *J. Mol. Biol.* **178**, 63–89.