



RNA

A PUBLICATION OF THE RNA SOCIETY

Crystal structure of a group I intron splicing intermediate

PETER L. ADAMS, MARY R. STAHLEY, MICHELLE L. GILL, et al.

RNA 2004 10: 1867-1887

Access the most recent version at doi:[10.1261/rna.7140504](https://doi.org/10.1261/rna.7140504)

References

This article cites 83 articles, 27 of which can be accessed free at:
<http://rnajournal.cshlp.org/content/10/12/1867.full.html#ref-list-1>

Article cited in:

<http://rnajournal.cshlp.org/content/10/12/1867.full.html#related-urls>

Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#)

An advertisement for Thermo Scientific. On the left, there is a colorful molecular model with red, green, and blue spheres. The text reads: "When quality and value are equally important" in blue, "Trust the experts in custom RNA synthesis" in orange, and the Thermo Scientific logo in red and black. A "Learn More" button is located in the bottom right corner.

When quality and value are equally important
Trust the experts in custom RNA synthesis
Thermo SCIENTIFIC
Learn More

To subscribe to *RNA* go to:
<http://rnajournal.cshlp.org/subscriptions>

Crystal structure of a group I intron splicing intermediate

PETER L. ADAMS, MARY R. STAHLEY, MICHELLE L. GILL, ANNE B. KOSEK, JIMIN WANG, and SCOTT A. STROBEL

Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06520-8114, USA

ABSTRACT

A recently reported crystal structure of an intact bacterial group I self-splicing intron in complex with both its exons provided the first molecular view into the mechanism of RNA splicing. This intron structure, which was trapped in the state prior to the exon ligation reaction, also reveals the architecture of a complex RNA fold. The majority of the intron is contained within three internally stacked, but sequence discontinuous, helical domains. Here the tertiary hydrogen bonding and stacking interactions between the domains, and the single-stranded joiner segments that bridge between them, are fully described. Features of the structure include: (1) A pseudoknot belt that circumscribes the molecule at its longitudinal midpoint; (2) two tetraloop-tetraloop receptor motifs at the peripheral edges of the structure; (3) an extensive minor groove triplex between the paired and joiner segments, P6-J6/6a and P3-J3/4, which provides the major interaction interface between the intron's two primary domains (P4-P6 and P3-P9.0); (4) a six-nucleotide J8/7 single stranded element that adopts a μ -shaped structure and twists through the active site, making critical contacts to all three helical domains; and (5) an extensive base stacking architecture that realizes 90% of all possible stacking interactions. The intron structure was validated by hydroxyl radical footprinting, where strong correlation was observed between experimental and predicted solvent accessibility. Models of the pre-first and pre-second steps of intron splicing are proposed with full-sized tRNA exons. They suggest that the tRNA undergoes substantial angular motion relative to the intron between the two steps of splicing.

Keywords: RNA structure; ribozyme; self-splicing intron; X-ray crystallography; *Azoarcus*; tRNA

INTRODUCTION

We recently reported the 3.1 Å resolution crystal structure of an intact self-splicing group I intron in complex with both its 5' and 3' exons (Adams et al. 2004). This intron is found naturally within the pre-tRNA^{Ile} anticodon loop of the purple bacterium *Azoarcus* sp. *BH72* (Reinhold-Hurek and Shub 1992). The RNA in the crystal structure corresponds to the splicing intermediate prior to the exon ligation reaction, the second step of splicing (pre-2S) (Adams et al. 2004). In this complex, the 5'-exon has been cleaved, but remains bound via P1 base-pairing with the intron's internal guide sequence (IGS). The 3'-exon is covalently connected to the intron, and it base pairs with the IGS via helix P10 (Fig. 1A). The terminal nucleotide of the intron,

ΩG206, occupies the G binding site. In the ensuing reaction, the ΩG206 O3' would serve as the leaving group when the 5'-exon attacks the scissile phosphate to produce ligated exons. We targeted the pre-2S state because the subsequent reaction is chemically equivalent to that catalyzed during nuclear pre-mRNA splicing (Cech 1986).

The *Azoarcus* intron is of the same group IC subclass as the more extensively investigated *Tetrahymena* intron, and both introns appear to follow an equivalent reaction mechanism (Reinhold-Hurek and Shub 1992; Tanner and Cech 1996; Kuo et al. 1999; Kuo and Piccirilli 2001). The bacterial intron retains the same core secondary structural elements, i.e., helices P1 through P10 and the joiner segments (J) that connect the helices, yet it is the smallest known group I ribozyme due primarily to a lack of peripheral domains and the absence of an open reading frame. In addition to its small size, the intron is particularly well-suited for structural studies because it is unusually stable at high temperature ($T_m = 75^\circ\text{C}$), and it folds homogeneously into a native state by the addition of only moderate concentrations of

Reprint requests to: Scott A. Strobel, Department of Molecular Biophysics and Biochemistry, Yale University, 260 Whitney Ave., New Haven, CT 06520-8114, USA; e-mail: strobel@csb.yale.edu.

Article and publication are at <http://www.majournal.org/cgi/doi/10.1261/rna.7140504>.

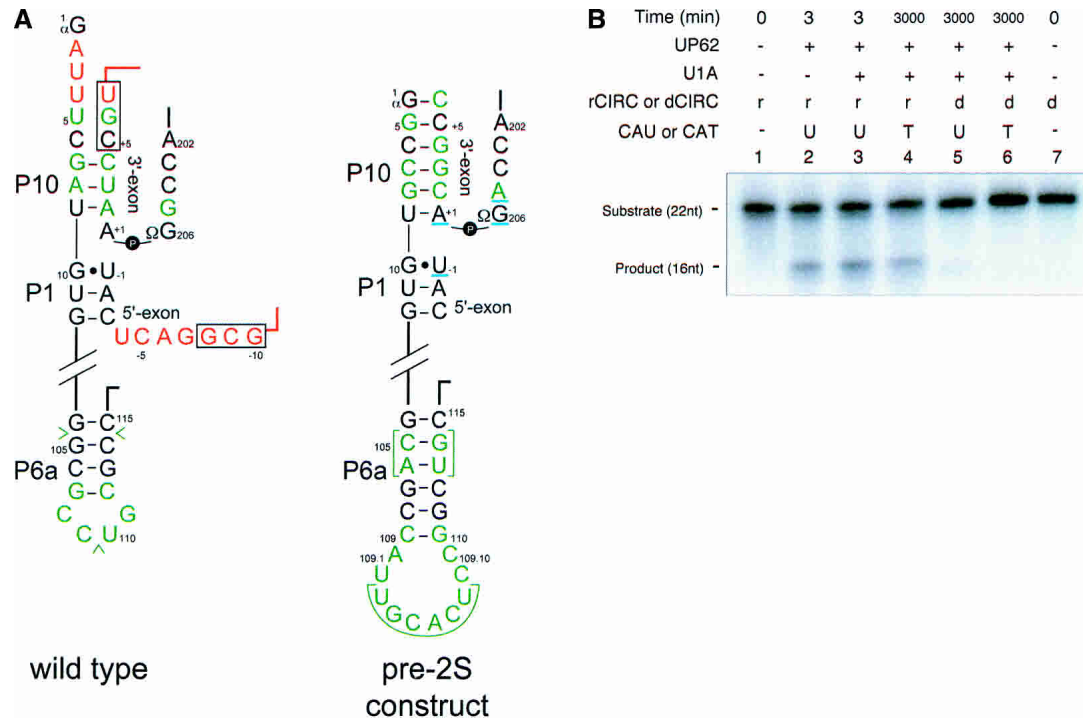


FIGURE 1. (A) Secondary structures for the wild type (left) and pre-2S crystallized (right) forms of the *Azoarcus* Ile-tRNA intron. Residues in the wild type sequence that were deleted within the pre-2S complex are shown in red. Residues mutated from wild type are shown in green. Note that there is a shift in the P10 base-pairing register between the wild type and crystallized complex. The four sites of 2'-deoxyribose substitution in the complex are underlined in blue (right). Exon residues that base pair to form the tRNA anticodon helix during the second step of splicing are indicated in black rectangles. The anticodon helix has been eliminated in the crystallized pre-2S complex. Nucleotide insertions within P6a and L6a are indicated with arrowheads (left) and brackets (right). The scissile phosphate is shown as a black circle. (B) Reactivity of the pre-2S complex containing ribose or 2'-deoxy at the ligation junction. The length of incubation time for each reaction is indicated, as well as the presence (+) or absence (-) of various components in the complex. In this experiment the 3'-exon oligonucleotide CIRC (d or r) oligonucleotide is 5'-end labeled. Reaction with the 5'-exon oligonucleotide (CAT or CAU) transfers the last six nucleotides onto the 5'-exon, which reduces the length of CIRC from 22 to 16 nucleotides. In all cases involving rCIRC the reaction proceeds to an equilibrium between the forward and the reverse ligation reactions with an endpoint of ~40% ligated (Mei and Herschlag 1996).

cations (Tanner and Cech 1996; Rangan et al. 2003; Rangan and Woodson 2003; Perez-Salas et al. 2004).

The initial report of this intron structure focused exclusively on residues within the immediate vicinity of the active site (Adams et al. 2004). Here, we describe features of the overall structure. These include: (1) a complete description of the intron construct used for crystallization; (2) the overall intron architecture and the tertiary interactions that stabilize the structure, including both tertiary hydrogen bonding and base stacking contacts; (3) the lattice contacts that facilitated crystal formation; (4) comparison of the *Azoarcus* pre-2S structure with the two previously reported structures of the *Tetrahymena* intron (the P4-P6 domain and the P3-P9 apoenzyme) and the biochemically based model of the *Azoarcus* intron (Cate et al. 1996a; Golden et al. 1998; Juneau et al. 2001; Rangan et al. 2003); (5) validation of the structure using hydroxyl radical footprinting; and (6) modeling of both splicing reactions in the context of full-length exons. A systematic description of the metal ions observed within the pre-2S complex will be the focus of a subsequent report.

RESULTS AND DISCUSSION

Construct design

Our crystallographic target was the pre-2S splicing intermediate complex formed between the 5'-exon and the covalently joined intron-3'-exon (Figs. 1A, 2A). This is an attractive complex for structural studies both for its biochemical relevance (Cech 1986), and because the covalent linkage between helices P9 and P10 might provide increased rigidity to the overall RNA architecture. The challenge in generating crystals of any complex that includes the intron and its exons is that the ribozyme is reactive and immediately produces ligated exons and a heterogeneous collection of intron circularization products (for review of intron reactions, see Cech 1990). As a result, the intron was engineered to promote formation of a stable homogeneous pre-2S complex that would remain unreactive during the multi-day incubations necessary to obtain crystals (Fig. 1A).

- (1) The 5'-exon was trimmed to just the three nucleotides (CAU) that are complementary to the IGS (residues

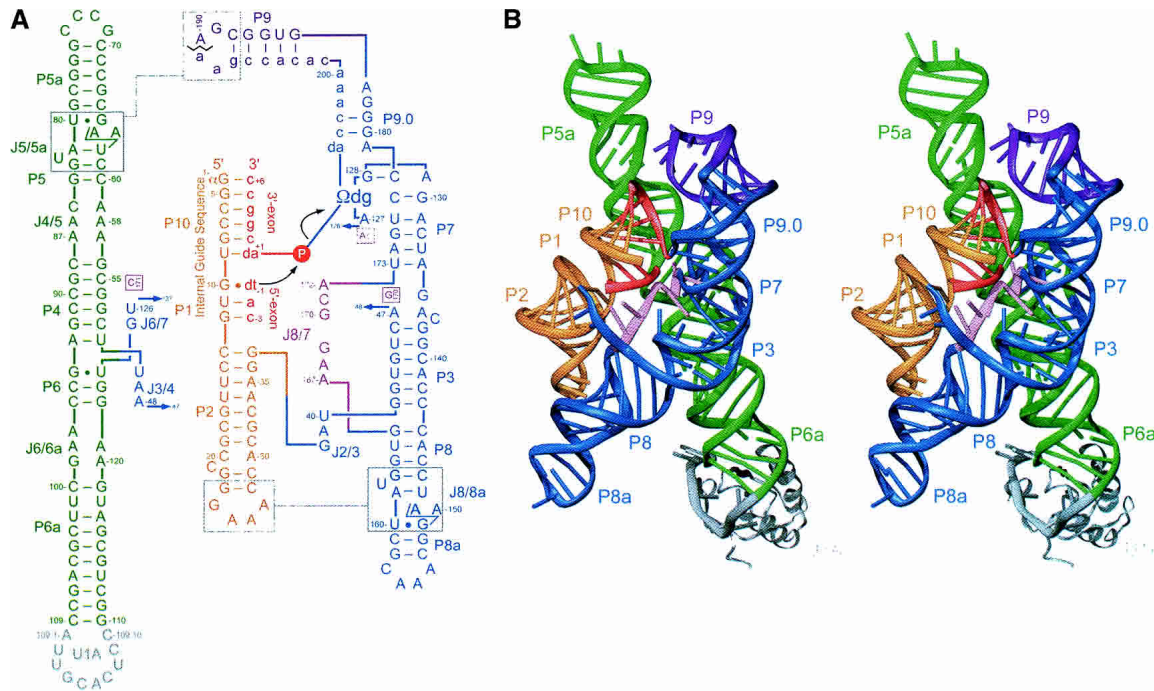


FIGURE 2. Overall secondary and tertiary structure of the *Azoarcus* Ile-tRNA intron pre-2S complex. (A) Secondary structure. The intron sequences, exon sequences, and structural elements (P and J elements) are depicted in the colors used for most of the subsequent figures. The secondary structure scheme follows that observed within the structure and differs slightly from that previously proposed for the intron, primarily in the P9.0 region (Reinhold-Hurek and Shub 1992; Tanner and Cech 1996). The ligation reaction catalyzed by this complex is shown by black arrows. The RNA transcript (UP62) comprising the majority of the intron is shown with capital letters, while residues derived from the two chimeric oligonucleotides dCIRC (intron/3'-exon segment) and CAT (5'-exon segment) are shown in lower case letters. The break between dCIRC and UP62 is indicated with a jagged line. The locations of the last three J8/7 residues are indicated as three labeled boxes positioned adjacent to the residues with which each nucleotide interacts. Intron numbering follows the canonical nomenclature established for group I introns, wherein the α G added in the first step of splicing is numbered 1 (Burke et al. 1987). This numbering differs by +1 from some reports on this intron (Reinhold-Hurek and Shub 1992; Tanner and Cech 1996; Rangan et al. 2003). (B) Tertiary structure in stereo representation. The backbone is depicted with a ribbon and individual bases depicted as cylinders. The color of the ribbon follows that shown in part A. This and many of the subsequent structure figures were prepared using RIBBONS (Carson 1991) and rendered using PovRay.

- 10–12). A short oligonucleotide of this sequence binds tightly to the intron as a result of P1 helix formation and tertiary interactions (Kuo et al. 1999; Strauss-Soukup and Strobel 2000).
- (2) The 3'-exon was truncated to six nucleotides and made fully complementary to the first six nucleotides of the intron. Normally, the P10 helix in the *Azoarcus* intron is only 3 bp long (Reinhold-Hurek and Shub 1992), but the P10 length varies from 3 to 7 bp among group I introns (Damberger and Gutell 1994; Cannone et al. 2002). An increased P10 length was chosen to enhance the stability of the complex. The six base-pair P10 helix was created by deleting three residues from the intron (A2-U4), and mutating three IGS residues and four residues in the 3'-exon to make them fully complementary. No U's were used within the 3'-exon to avoid competition with the U at the 5'-splice site. The first nucleotide of the 3'-exon (A+1) is unpaired within the *Azoarcus* P10 helix, but 3'-exon mispairing at this position is not a conserved group I intron feature (Cannone et al. 2002). This residue is more often mispaired or base-paired within the P10 helix. Consequently to
 - rigidify the structure, we made A+1 complementary to IGS residue U9.
 - (3) The penultimate residue of the intron was mutated from a G to an A. Previous work in the *Tetrahymena* intron has shown that Ω G mutation results in cryptic mis-splicing at neighboring G's within the sequence (Price and Cech 1988), which implies that group I introns may have a propensity for conformational heterogeneity. This was minimized by mutating G205, the only G within two nucleotides of Ω G.
 - (4) Three 2'-deoxy substitutions were introduced into the complex at positions A205, Ω G206, and A+1. 2'-Deoxyguanosine (dG) efficiently binds the intron, but it is not a substrate for the splicing reaction either in place of the exogenous G (first step of splicing) or when placed at the Ω G position (second step of splicing) (Bass and Cech 1984; Bass and Cech 1986; Moran et al. 1993). Although dG retains the O3', which serves as the leaving group in the second step of splicing, the nucleotide is inactive because the 2'-OH of G coordinates a catalytic metal ion that is necessary for O3' activation (Sjogren et al. 1997; Shan et al. 1999; Shan et al. 2001).

By introducing these three 2'-deoxy substitutions, we expected to eliminate splicing and 3'-exon hydrolysis at Ω G and both adjacent positions.

- (5) A single 2'-deoxy substitution was introduced at U-1 within the 5'-exon (Fig. 1A). The U-1 2'-OH contributes ~1000-fold to reactivity by forming a network of transition state specific interactions that activate the adjacent 3'-OH for nucleophilic attack (Herschlag et al. 1993; Strobel and Ortoleva-Donnelly 1999). Elimination of this 2'-OH group was designed to reduce the rate of mis-splicing at all cryptic positions that might be present within the intron. For synthetic simplicity, the 2'-deoxyribose substitution was achieved using thymidine (dT), which meant a 5-methyl substitution was also added at residue -1.
- (6) In order to introduce the three 2'-deoxy substitutions at A205, Ω G206 and A+1, it was necessary to make the 3'-splice junction as a synthetic oligonucleotide (Fig. 2A). A nick was designed into the P9 loop (L9) by preparing a 22-nucleotide oligonucleotide (dCIRC) that corresponded to the last 16 residues of the intron (A191- Ω G206) and the first six residues of the 3'-exon (A+1-C+6). Ligation schemes to make the full length chimeric RNA were envisioned and tested unsuccessfully, but intron reassembly was achieved efficiently by P10, P9, and P9.0 helix formation.
- (7) The majority of the intron, residues α G1-A190, was prepared as a T7 RNA polymerase transcript. The 3'-end of the transcript was made homogeneous by appending an antigenomic HDV ribozyme to the end of the intron and allowing it to cleave during the transcription reaction (Grosshans and Cech 1991; Been 1994; Ferre-D'Amare and Doudna 1996). Thus, the resulting transcript had a triphosphate at its 5'-end and a 2'-3' cyclic phosphate at its 3'-end.
- (8) Ferre D'Amare and colleagues reported success co-crystallizing ribozymes with the RNA binding protein U1A (Oubridge et al. 1994; Ferre-D'Amare and Doudna 2000; Rupert and Ferre-D'Amare 2001). Based upon this precedent, we introduced a U1A binding site at the ends of helices P5a, P6a, or P8a, and attempted to co-crystallize pre-2S as a ternary complex with the U1A protein. Several crystal forms were found for the P6a construct (UP6), though none of them diffracted to resolutions better than 6 Å.
- (9) Given the propensity of the UP6a complex to crystallize, we altered the register of the protein relative to the RNA by adding and subtracting base pairs from the P6a helix. A total of eight variants with P6a lengths from -1 to +6 bp were screened for crystallization. Each produced crystals under a range of conditions specific to each construct. The best diffraction (3.1 Å) was obtained from the construct with two additional base pairs in the P6a helix (UP62).

Splicing activity

The pre-2S complex was tested for splicing activity with and without the 2'-deoxy substitutions and the U1A protein (Fig. 1B; Adams et al. 2004). The complete pre-2S complex was assembled from the UP62 RNA and two oligonucleotides. Activity was tested using both the ribose (rCIRC and CAU) and the chimeric ribose/deoxyribose (dCIRC and CAT) variants of the oligonucleotides. The intron catalyzed the reversible exon ligation reaction of the all-ribose substrates rCIRC and CAU at a rate of 2 min⁻¹ with an endpoint at equilibrium between the forward and reverse exon ligation reaction of ~40% ligated. The 2'-deoxy substitution at U-1 (reaction of CAT with rCIRC) reduced the rate 3000-fold (Fig. 1B, lane 4), which is consistent with the ~10³-fold effect reported for this substitution (Herschlag et al. 1993). The dG substitution at Ω G (reaction of CAU with dCIRC) reduced the rate by ~500,000-fold (Bass and Cech 1984; Bass and Cech 1986; Moran et al. 1993). The doubly substituted complex (reaction of CAT with dCIRC) reacted at the almost imperceptible rate of 8 × 10⁻⁷ min⁻¹, approximately 2-million-fold slower than the wild type rate. Less than 1% of the substrates reacted after incubation for a week. Failure to react was not due to lack of binding. dCIRC binding was demonstrated by gel shift analysis (Adams et al. 2004). Thus, the relatively conservative removal of four 2'-OH groups from the pre-2S complex left it structurally intact, but catalytically impaired. Furthermore, activity of the pre-2S complex was unaffected by the additional base pairs in P6a or by the addition of the U1A protein (Fig. 1B).

Model building and structural validation

The backbone of all 222 nucleotides in the complex and the vast majority of the bases could be traced through the experimental electron density (Fig. 2B). The pre-2S intron U1A complex has dimensions of 136 × 65 × 50 Å, which is relatively compact compared to other ribozyme structures (Ferre-D'Amare and Doudna 1999), with an average surface accessible area of 145 Å²/nucleotide. The solvent content of the crystal is 72%, which would be high for a protein crystal, but consistent with that of other RNA crystals (Ferre-D'Amare and Doudna 1999). The temperature factors (B factors) throughout the structure are also high, consistent with values observed in other large RNA structures (Fig. 3; Ferre-D'Amare and Doudna 2000). In general, the RNA is well ordered, particularly around the active site. The RNA is most disordered within the P6a helix, particularly below the U101·U118 mispair. The mispair may lead to greater mobility for the helix and the associated U1A protein. Outside of P6a, the bases of only three residues are disordered (C21, U82, and U162), all three of which are bulged out of the helical stack. The U1A protein is the most disordered portion of the complex. This is opposite of the trend observed

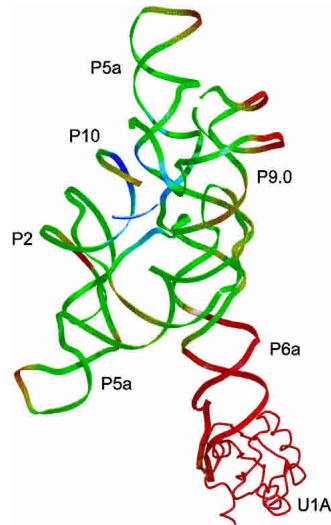


FIGURE 3. Temperature (B) factors superimposed on a ribbon of the pre-2S tertiary structure. B factors between 0–50 Å² are shown in blue, values between 50–100 Å² in green, and values between 100–150 Å² in red. The region of greatest disorder within the structure is the P6a helix and the U1A protein bound at its end. The lowest B factors are found within the intron active site.

within the hairpin ribozyme, HDV ribozyme, and stem loop RNA cocrystal structures with U1A (Oubridge et al. 1994; Ferre-D'Amare and Doudna 2000; Rupert and Ferre-D'Amare 2001). As a result, the U1A was built into the density as a rigid body (Oubridge et al. 1994), and the protein structure held rigid during refinement.

The orientation of the RNA and U1A within the pre-2S complex was established using site-specific heavy atom substitutions. The orientation of the U1A was validated using seleno-methionine substituted U1A and by observing that the anomalous electron density for selenium superimposed on the methionines within the model. Similarly, the register and orientation of the RNA was established by the anomalous difference density of bromine substituted at the 5'-splice site (5-bromo-2'-deoxy uridine substitution at the –1 position) and selenium substitution at C194 within the 3'-exon oligonucleotide (dCIRC) (Höbartner and Micura 2004). The bromine substitution established the location of the active site, while the Se substitution established the register of dCIRC, which by direct extension identified the boundary between the intron and the 3'-exon (Adams et al. 2004).

Structural overview

The majority of the pre-2S intron structure is contained within three helical domains: P1-P10 (helices P1, P2, P10, orange/red), P4-P6 (helices P4, P5, P5a, P6, P6a, green), and P3-P9.0 (helices P3, P7, P8, P8a, P9.0, blue) (Fig. 2B). A fourth, short helix, P9 (purple), is arranged perpendicular to P9.0. Helices P1-P10 and P4-P6 are effectively straight rods, while P3-P9.0 makes an ~30° bend in the course of the

helix. This bend occurs primarily at the junction between the P3 and P7 helices. A comprehensive diagram depicting the collection of non-canonical base pairs and tertiary interactions that occur throughout the structure using geometric nomenclature is shown in Figure 4 (Leontis and Westhof 2001).

Surprisingly, the three main helical domains directly contact each other in only three places (Fig. 4). In each case, the interaction occurs between a mispaired (or joiner) region in one domain and a helical region of the other. (1) Helix P1 docks against the J4/5–J5/4 segment of P4–P6. This contact defines the 5'-splice site and has been discussed in previous biochemical and structural reports (Strobel et al. 1998; Adams et al. 2004). (2) The tetraloop at the end of P2 contacts a tetraloop receptor in J8/8a (see below for discussion; Tanner and Cech 1996). (3) The minor groove of P3 interacts

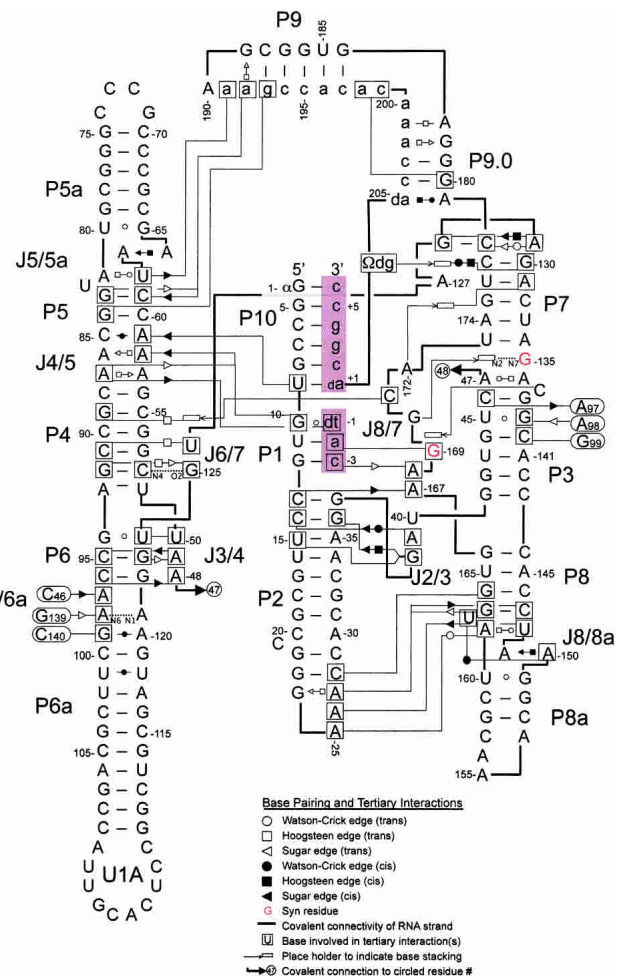


FIGURE 4. Schematic tertiary structural view of the intron using the symbolism described in Leontis and Westhof (2001). The symbol key is included at the bottom of the figure. Individual nucleotide numbers, helical elements, and joiner regions are indicated. The 5' and 3' exons are indicated within shaded red rectangles. Nucleotides within the UP62 transcript are in capital letters, while those in the two oligonucleotides are depicted in lowercase letters.

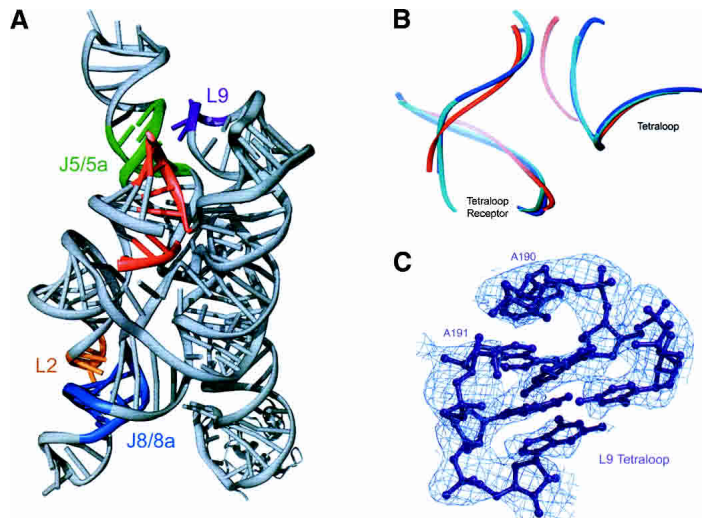


FIGURE 5. Two tetraloop-tetraloop receptor (TL/TLR) interactions on opposite ends of the *Azoarcus* intron. (A) Depiction of the pre-2S complex with emphasis on the two TL/TLR interactions. The L2 (orange) interaction with J8/8a (blue) and the L9 (purple) interaction with J5/5a (green) are both highlighted. The rest of the intron and the U1A protein are in gray. The 5' and 3' exons are in red. (B) Superposition of the backbone trace of the *Tetrahymena* L5b-J6a/6b TL/TLR motif (dark blue) with the two examples found in the *Azoarcus* intron. The L2-J8/8a interaction is shown in light blue, while the L9-J5/5a interaction is in red. (C) Electron density for the L9-J5/5a TL/TLR showing that the RNA is well ordered despite the nick in the L9 tetraloop. The experimental electron density contoured at 2σ is shown as a blue web. Note the break in electron density between residues A190 and A191, which marks the covalent break in the RNA.

with unpaired and mismatched residues in J6/6a. These contacts as well as the interactions between J3/4 and P6 created an interface that encompasses six continuous base pairs in the P6-J6/6a minor groove.

The alignment of the three main helical domains primarily involves tertiary interactions with the four joiner segments that bridge between the helical domains (J2/3, J3/4, J6/7, and J8/7) and the coordination of 18 metal ions (M.R. Stahley, P.A. Adams, J. Wang, S.A. Strobel, in prep.; Fig. 4). The most structurally dramatic of these segments is J8/7, which adopts a μ shaped conformation as it flips and turns its way through the active site (Fig. 2B, pink). Despite its short six residue length, J8/7 makes extensive and functionally critical tertiary contacts to all three helical domains (Fig. 4; see J8/7 section for further discussion). In comparison, J2/3 contacts only P1-P10, and J3/4 and J6/7 contact only P4-P6 (Fig. 4).

Two tetraloop-tetraloop receptors

The outermost ends of the tertiary interface are stabilized by two GAAA tetraloop-tetraloop receptor contacts (TL-TLR), one between the L2 tetraloop and a receptor in J8/8a, and a second between L9 and J5/5a (Fig. 5A; Costa and Michel 1995; Cate *et al.* 1996a,b; Tanner and Cech 1996). These complex tertiary interaction motifs are located on opposite ends of the intron, and each is consistent with phylogenetic and biochemical predications (Costa and Michel 1995; Tan-

ner and Cech 1996; Basu *et al.* 1998; Strauss-Soukup and Strobel 2000). Both interactions have essentially the same architecture as originally observed for the L5b tetraloop and its J6a/6b receptor in the *Tetrahymena* intron P4-P6 domain (Fig. 5B; Cate *et al.* 1996a). The L2-J8/8a interaction superimposes almost exactly on L5b-J6a/6b (rmsd of only 0.6 Å) and includes the presence of a monovalent metal ion coordinated underneath the AA platform (Basu *et al.* 1998). The L9-J5/5a contact shows slight (~1 Å) displacement of the tetraloop away from the receptor (rmsd 1.4 Å), relative to the other two examples. The A-minor interactions (A191 and A192) are still in place, but the AA *trans*-Watson-Crick pair, which is also characteristic of this TL-TLR motif (Costa and Michel 1995; Cate *et al.* 1996a), does not form between A81 and A190 (3.8 Å apart) as a result of the displacement. This slight displacement might result from the extra charge on the 2'-3' cyclic phosphate of A190 or structural restraints coming from the size and con-

nectivity of P9. The nick we introduced into L9 does not appear to cause structural disorder. The strand termini at the nick occupy contiguous electron density within the experimental map (Fig. 5C). The interaction is sufficiently stabilizing to hold both strands of the tetraloop in place despite the dynamic breathing expected for mismatched nucleotides at the end of a helix. These flanking TL-TLR contacts and the high G-C content are likely to be primary factors in the thermostability of the intron (Tanner and Cech 1996).

P3 pseudoknot belt

A single stretch of ~25 nucleotides (A34-A59) is included within all three of the intron's main helical domains (Fig. 2A). These residues form a pseudoknot belt that wraps completely around the circumference of the molecule (Fig. 6). Residues A34-G37 base pair within P2, residues G41-A47 form base pairs that include the full length of P3, and residues U51-A59 extend into helix P4. There is no other consecutive sequence that is part of all three helical domains. These residues track around the outer edge of the intron at its longitudinal midpoint. The radius of the belt is ~20 Å, and the circumscription is sufficiently complete that the phosphates of P2 residue A34 and J4/5 residue A59 approach to within 15 Å of each other on the side opposite of the P3 helix (Fig. 6B).

The topology of the pseudoknot belt is consistent with

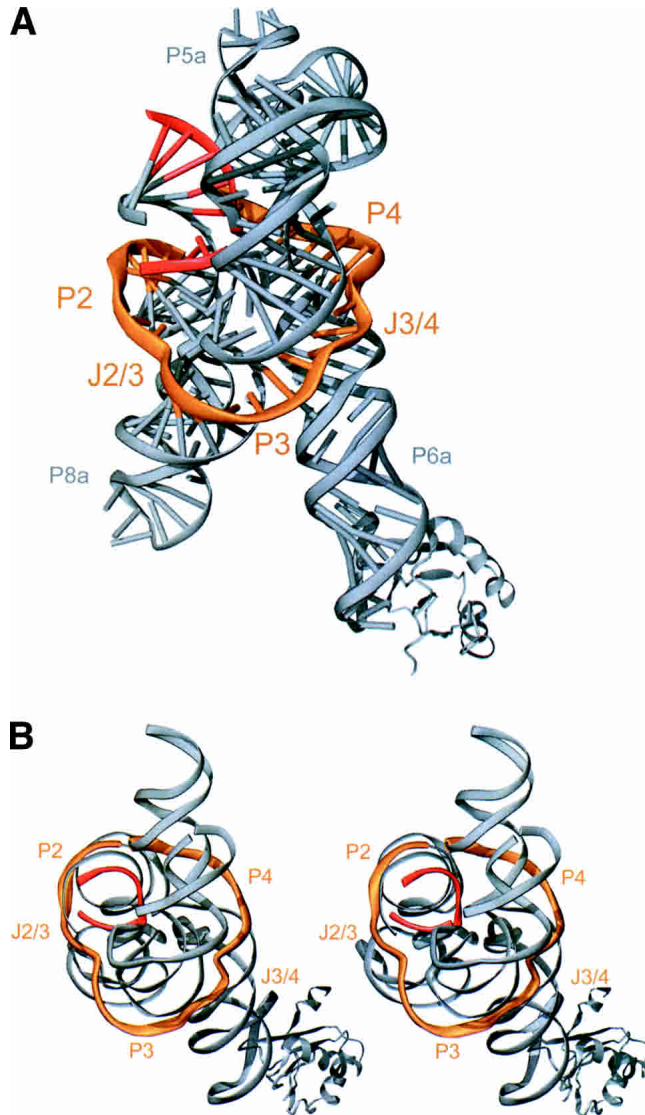


FIGURE 6. The pseudoknot belt that reaches around the circumference of the pre-2S complex at the intron's midpoint. Residues A34–A59 are shown as an orange ribbon with orange cylinders for the bases. The U1A protein and the rest of the intron are shown in gray with gray cylinders for the bases. The 5' and 3' exons are shown in red. (A) Front view. (B) Top view (in stereo representation) looking down the P1-P10 helical axis (upper left).

kinetic investigations of intron folding (Zarrinkar and Williamson 1994; Sclavi et al. 1998). In the L-21 ScaI ribozyme form of the *Tetrahymena* intron, the P3 helix is the last secondary structural element to form (Pan and Woodson 1998). This agrees with the structural observation that the pseudoknot belt establishes the outer circumference of the structure. Given the RNA topology, a late stage folding intermediate may include the formation of all core helices except P3, including P1, P2, and P4. In this intermediate, P1/P2 would be tethered to P4 and the rest of the intron via a single stranded segment comprised of residues in the pseudoknot belt (G38–A47). P3 formation, which must in-

volve the wrapping of this tether around the rest of the intron, would bring P1/P2 into the active site cleft where the P1 helix and its 5'-exon could assume a docked conformation (Herschlag 1992). Topological constraints dictate that the P10 helix could only form after P3 formation. Although P10 was not explicitly included in the folding constructs studied to date, there is experimental evidence to support this prediction. The P10 helix is specific to the second step of splicing. It only forms after the 5'-exon is covalently cleaved off the intron, which makes the 5'-end of the intron accessible for P10 base-pairing to the 3'-exon. Thus, in the context of exon splicing, the P10 helix must form last, well after formation of P3.

Tertiary contacts to each helical domain

P4-P6

The pre-2S crystal structure strongly supports the hypothesis that the P4-P6 domain serves as an architectural scaffold upon which the rest of the intron is built (Fig. 7; Murphy and Cech 1993; Murphy et al. 1994). The highest density of tertiary contacts to helical regions occurs within helices P4, P5, and P6 (Fig. 4). Fourteen of the 17 bp between G99–A120 in J6/6a and U62–U81 in P5a make at least one tertiary contact in the structure. These interactions oc-

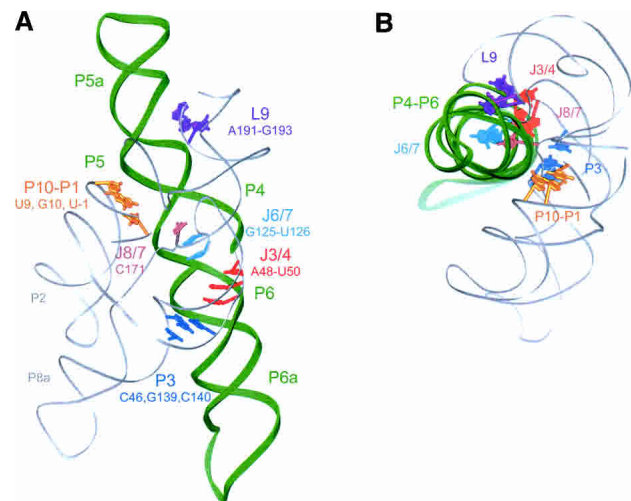


FIGURE 7. Tertiary interactions with the P4-P6 helical domain. The P4-P6 helix is shown as a green ribbon. The rest of the intron and both exons are in a thin light gray ribbon. All of the residues within the complex that make tertiary contact with the domain are shown as bases with a stick connection to the backbone trace. The identity of each of these is labeled. The color scheme follows that established in Figure 2A. Alternative colors are used in the joiner regions to emphasize the large number of structural elements that are organized around the P4-P6 helix. (A) Side view. (B) Top view. The tertiary contacts made with the P4-P6 minor groove are located outside the green helical ribbon, while the major groove contacts made by J6/7 and J8/7 (Fig. 12) are positioned inside the cylinder created by the helical ribbon. A complete list of the tertiary interactions to the P4-P6 domain is compiled in Table 1.

cur in both helical grooves and include contacts with helices P1, P3, P9, and P10, and three of the four joiner regions (J3/4, J6/7, and J8/7) that bridge between the helical domains (Fig. 7). A detailed list of the tertiary hydrogen contacts is compiled in Table 1. Of the three P4-P6 base pairs that do not make direct tertiary interactions, one of them, G54-C90 (P4 base pair 4), interacts with the A-rich bulge among those introns that have a P5abc extension (Cate et al. 1996a). A second, G56-C88, coordinates both catalytic metal ions using the phosphates 5' and 3' of C88 (Szewczak et al. 2002; Adams et al. 2004). Thus, only the central A93-U51 pair (P4 base pair 1) at the junction between helices P4 and P6 is without a known tertiary contact.

P3-P9.0

In comparison to the other elements, P3-P9.0 has relatively few tertiary interactions (Figs. 4, 8A). This deficit is surprising given its irregularly bent shape and the way it appears to wrap around the P4-P6 helix (Fig. 4). There are two primary sites of tertiary interaction and a few isolated contacts (Fig. 8A; Table 2). The only direct contact between the P3-P9.0 and the P4-P6 domains occurs between the top of

TABLE 1. Pairwise tertiary hydrogen bonding interactions to the P4-P6 domain

Residue	Chemical Group	Bonding	Region	Residue	Chemical Group	Bonding
U62	O2'	d	P9	A191	N3	a
	O2'	a		A191	O2'	d
C61	O2'	d		A192	O2'	a
	O2	a		A192	O2'	d
G84	O2'	a/d		G193	O2'	a/d
A59	O2'	a	P10	U9	O2'	d
	O2'	d		U9	O2	a
A58	O2'	d	P1	G10	O2'	a
	N3	a		G10	O2'	d
A87	N3	a		G10	N2	d
	O2'	a		G10	N2	d
	*O2'	a/d		dT-1	*O2'	a/d
G89	N7	a	J8/7	C171	N4	d
C91	N4	d	J6/7	U126	O4	a
G92	O6	a		G125	N2	d
C52	N4	d		C125	O2'	a
U124	O2'	a/d	J3/4	U50	O2'	a/d
G123	O2'	d		A49	N3	a
	N2	d		A49	N1	a
C95	O2'	a/d		A49	O2'	a/d
C96	O2'	a/d		A48	O2'	a/d
A97	N3	a	P3	C46	O2'	d
	O2'	d		C46	O2'	a
A98	N3	a		G139	N2	d
	N1	a		G139	O2'	d
G99	O4'	a		C140	O2'	d

Note: a and d for hydrogen bond donor acceptor and donor, respectively; a/d for interactions that are ambiguous.

*Predicted interaction based upon modeling of O2' at U-1.

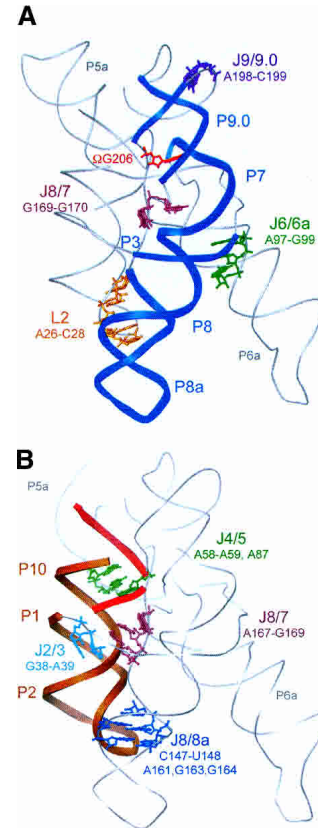


FIGURE 8. Tertiary interactions to the P3-P9.0 helical domain (A) and P1-P10 domain (B). The figures are organized as in Figure 7 with the P3-P9.0 helix shown in blue ribbons (A) and the P1-P10 helix shown in orange and red ribbons (B). Each nucleotide that contacts the helix is shown in full. Lists of the tertiary interactions to the P3-P9.0 domain and P1-P10 domain are compiled in Tables 2 and 3, respectively.

P3 and residues in J6/6a. This interface utilizes consecutive type II and type I A-minor tertiary contacts (Nissen et al. 2001). The only direct contact between P3-P9.0 and P1-P10 is the TL-TLR between J8/8a and L2, which also involves A minor and stacking interactions (see above). Isolated tertiary contacts include stacking of J8/7 residue G170 at the interface between P3 and P7 and the previously described G binding site with the G130-C177 pair in P7 (Adams et al. 2004). The only contacts to the P9.0 helix come from the exocyclic amines of A198 and C199 in J9/9.0, both of which hydrogen bond to the phosphate of G180. Thus, among the 27 bp from the bottom of P8 to the top of P9.0, only nine are involved in tertiary hydrogen bonding within this structure.

P1-P10

The 5' and 3' exons both base pair to the intron's IGS resulting in P1 and P10 helix formation, respectively (Fig. 2A; Waring et al. 1986). Consistent with extensive biochemical investigations, the tertiary contacts to these helices

TABLE 2. Pairwise tertiary hydrogen bonding interactions to the P3-P9 domain

Residue	Chemical		Region	Residue	Chemical	
	Group	Bonding			Group	Bonding
G180	O2P	a	P9	A198	N6	d
	O1P	a		C199	N4	d
C177	N4	d	ΩG	dG206	O6	a
G130	O6	a		dG206	N1	d
	N7	a		dG206	N2	d
A127	N6	d	P7	A131	O2P	a
G135	N7	a	J8/7	G170	N2	d
C46	O2'	d	J6/6a	A97	N3	a
	O2'	a		A97	O2'	d
G139	N2	d		A98	N3	a
	O2'	d		A98	N1	a
C140	O2'	d		G99	O4'	a
G164	O2'	a	P2	C28	O2'	d
	O2'	d		C28	O4'	a
G163	N2	d	L2	A27	N3	a
	O2'	d		A27	N1	a
C147	O2'	d		A27	O2'	a
	O2	a		A27	O2'	d
A161	N6	d		A25	N1	a
U148	O2'	d		A25	N3	a
	O2'	a		A26	O2'	d

Note: a and d for hydrogen bond donor acceptor and donor, respectively.

are focused on P1 and the 5'-exon and the base pairs immediately adjacent to it in P10 and P2 (Figs. 4, 8B; Table 3; Herschlag and Cech 1990; Bevilacqua and Turner 1991; Pyle and Cech 1991; Strobel and Cech 1993; Strobel and Cech 1995). No tertiary contacts are made to the 3'-exon. All six of the base pairs between U9-A+1 and C14-G36 form tertiary hydrogen bonds, including contacts to all three nucleotides within the 5'-exon (Fig. 8B). These interactions occur exclusively in the minor groove and involve joiner regions J2/3, J4/5-5/4, and J8/7. With the exception of the contact to the G·U pair at the 5'-splice site, the interactions are not sequence specific (Doudna et al. 1989). Minimal sequence constraints for exon placement into the active site may serve an important biological role because they maximize the potential sequences into which the intron can be inserted while retaining splicing function.

P9

P9 is a short helix that makes tertiary contacts only at its ends. The L9 loop forms a TL/TLR interaction with J5/5a (discussed above, Fig. 5). Interactions at the other end only involve the unpaired nucleotides that are coaxial with P9 in the joiner region between the P9 and P9.0 helices (Fig. 8A; Strobel et al. 2004).

Joiner regions between helical domains

The sequences that span the junctions between the helical domains form the majority of the tertiary interactions in the

pre-2S structure. These joiner (J) regions make bridging contacts within and between the helical domains. Sequences in the J segments are often highly conserved or show strong covariation with residues within the helices (Michel et al. 1990; Tanner and Cech 1997; Cannone et al. 2002). There are five joiner regions that span between domains, including J2/3, J3/4, J6/7, J8/7, and J9/9.0. Every residue in the first four of these joiner regions makes a tertiary hydrogen bonding or tertiary stacking interaction within the intron core.

J2/3

The three nucleotide joiner segment J2/3 is folded back into the P2 minor groove (Fig. 9). There, it makes several functionally critical tertiary contacts with the second and third base pairs of the P2 helix (Table 3; Strauss-Soukup and Strobel 2000). G38 and A39, the first two residues of J2/3, both occupy the minor groove of the C14-G36 pair, the penultimate base pair in the P2 helix. A39 contacts C14, while G38 hydrogen bonds to G36 and reaches even further down the side of the P2 helix to contact U15 in the preceding U15-A35 pair. The residues in J2/3 initiate a single stranded stacking interface between P1-P10 and P3-P9.0 that extends into the first two nucleotides of J8/7 (A167 and A168, see below).

TABLE 3. Pairwise tertiary hydrogen bonding interactions to the P1-P10 domain

Residue	Chemical		Region	Residue	Chemical	
	Group	Bonding			Group	Bonding
U9	O2	a	P5	A59	O2'	d
	O2'	d		A59	O2'	a
*dT-1	O2'	d	J4/5	A87	*O2'	a
G10	N2	d		A87	N3	a
	N2	d		A87	O2'	a
	O2'	d		A58	O2'	a
	O2'	a		A58	N3	a
A-2	O2'	d	J8/7	G169	O4'	a
C-3	O2'	d		A168	N1	a
C13	O2'	a		A167	O2'	d
	O2'	d		A167	N3	a
C14	O2'	d	J2/3	A39	N1	a
	O2	a		A39	N6	d
G36	N2	d		G38	O6	a
C28	O2'	a/d	P8	G164	O2'	a/d
	O2'	d		G164	O2'	a
	O4'	a		G164	O2'	d
A27	O2'	a		C147	O2'	d
	O2'	d		C147	O2	a
	N3	a		G163	N2	d
	N1	a		G163	O2'	d
A26	O2'	a/d	J8/8a	U148	O2'	a/d
A25	N1	a		A161	N6	d
	N3	a		U148	O2'	d

Note: a and d for hydrogen bond donor acceptor and donor, respectively; a/d for interactions that are ambiguous.

*Predicted interaction based upon modeling of O2' at U-1.

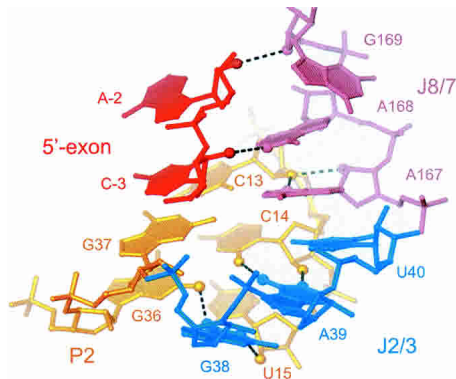


FIGURE 9. Detailed structural depiction of the J2/3 joiner segment (blue) and its tertiary interactions (black dashed lines) with the P2 helix (orange). Tertiary interaction between the first three residues in J8/7 (pink) and the P2 and P1 helices are also shown. The 5'-exon is in red. Functional groups involved in tertiary hydrogen bonds are shown as enlarged spheres.

The *Azoarcus* and *Tetrahymena* introns appear to employ the J2/3 and J8/7 regions differentially for interaction with the substrate helix (here defined as P1 in *Tetrahymena* and P1-P2 in *Azoarcus*). Comparison of biochemical and structural results suggest that there are several key differences in this critical region. The P1 helix is longer in *Tetrahymena* (6 instead of 3 bp), but it does not stack on the P2 helix as occurs in the bacterial intron. Furthermore, the *Tetrahymena* J8/7 region is seven nucleotides long, while in *Azoarcus* it is only six. This subtle, but important, feature is what distinguishes the IC1 (*Tetrahymena*) and IC3 (*Azoarcus*) intron subclasses (Michel and Westhof 1990). Biochemical data suggest that the additional *Tetrahymena* J8/7 residue (*TetU300*) makes the same tertiary contact to the substrate helix at *TetG26* as is observed between *Azoarcus* J2/3 residue A39 and C14 (Szewczak et al. 1999). Both of these residues are located 6 bp below the 5'-splice site in the substrate helix. Nucleotides in J2/3 appear to compensate for the shorter J8/7 region as well as provide contacts to the substrate helix that are not present in *Tetrahymena*. This may compensate for the shortened P1 helix as the 5'-exon must be held in place between the first and second steps of splicing. It is a striking example of how multiple architectural elements can be utilized to achieve the same functional result.

J3/4

Six strands converge at the junction of helices P4 and P6 (Fig. 2A). This includes two strands in P4, two in P6, and the two joiner strands, J3/4 and J6/7. This complex multi-strand junction was the focus of extensive phylogenetic and mutagenesis studies (Flor et al. 1989; Green et al. 1990; Michel et al. 1990). Several of the key features predicted from those investigations are present within this intron. Both stands of P4 and P6 stack to form a continuous helix.

J6/7 forms a major groove triplex with P4, while J3/4 makes a minor groove triplex with P6 (Figs. 10, 11). The tertiary pairing partners within both triplex junctions are consistent between the biochemical and structural studies (Flor et al. 1989; Green et al. 1990; Michel et al. 1990). Only the specific hydrogen bonds formed between the pairings differs in the structural model from what was predicted (Table 1).

The three nucleotides in J3/4 make A-minor interactions with the P6 helix (Fig. 10, orange; Nissen et al. 2001). A48 swings out of the P3 helical stack and enters the P6 minor groove to make a type II A minor interaction with C96. A49 makes a type I contact to the C95-G123 pair, and U50 makes an O2' mediated interaction with U124. As expected for such contacts, A48 and A49 are well conserved residues among group I introns (Cannone et al. 2002). These three residues are followed by a complete reversal of strand direction between U50 and U51, where the P4 helix begins (Fig. 10). Thus, the three residues in J3/4 are structurally disconnected from the helices that proceed and follow it;

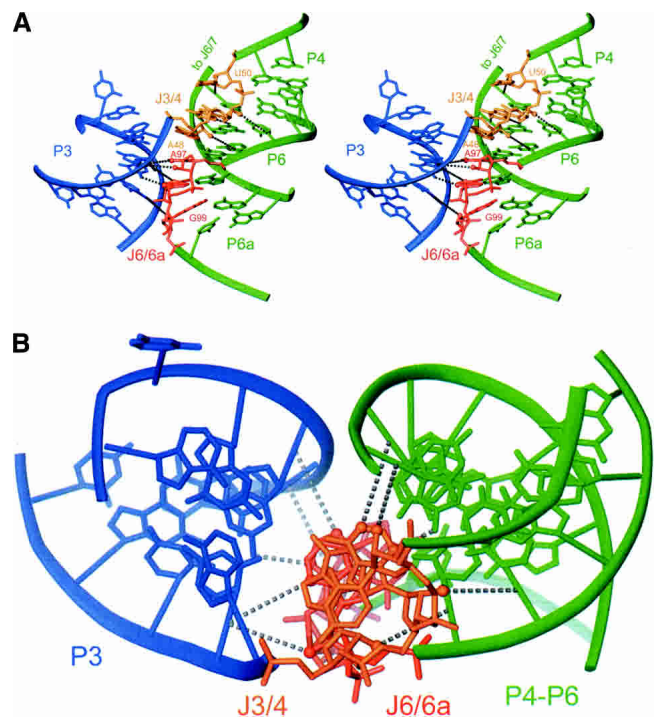


FIGURE 10. Detailed structural depiction of the J3/4 segment (orange) and its tertiary hydrogen bonding interactions (black dashed lines) with the P6 helix (green ribbon). Tertiary interactions between the J6/6a (red) and the P3 helix (blue ribbon) are also shown. Functional groups involved in tertiary hydrogen bonds are shown as enlarged spheres. Hydrogen bonds are made to the O2' of a nucleotide depicted at the midpoint of the cylinder connecting the base to the helical ribbon. The transition of the P6 strand into the J6/7 element is labeled, as are other features of the structure. Functional groups in the J3/4 or J6/6a region involved in tertiary hydrogen bonds are shown as enlarged spheres. (A) Side view shown in stereo representation. (B) Top view looking down the continuous base stack created between the J3/4 and J6/6a segments.

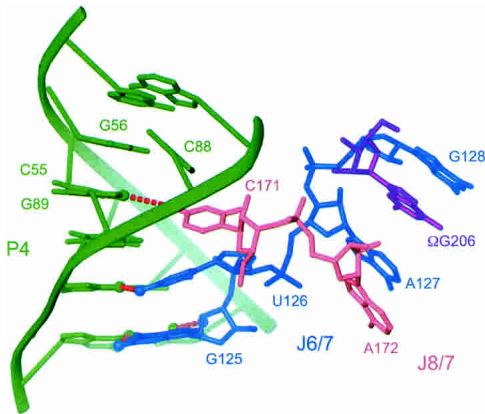


FIGURE 11. Detailed structural depiction of the J6/7 segment (blue) and its tertiary hydrogen bonding interactions (red dashed lines) with the major groove of the P4 helix (green ribbon). Tertiary interactions with the last two nucleotides of the J8/7 region (pink) and the ΩG206 (purple) are also shown.

however, they are not structurally isolated. A second set of A-minor interactions is formed immediately below J3/4 between J6/6a and helix P3 (Fig. 10, red). Here, A97 forms type II and A98 type I A minor interactions, while G99 makes an O2'-mediated contact in the P3 minor groove. A48 and A98 stack upon each other creating a pseudo-two fold symmetry between J3/4 and J6/6a. This creates a continuous single-stranded element at the interface between P3 and P6 that is most clearly visualized by looking down the helical axis (Fig. 10B). Similarly concentrated A minor patches from disparate structural elements, were reported within the ribosomal crystal structures (Gutell et al. 2000; Wimberly et al. 2000; Nissen et al. 2001).

J6/7

J6/7 contributes to two important structural elements within the intron active site (Fig. 11, blue). The first two nucleotides, G125 and U126, form a major groove triplex with P4 (Flor et al. 1989; Michel et al. 1990). J6/7 then reverses strand direction, and the second two nucleotides, A127 and G128, establish the stacking interface at the roof and the floor of the G binding site (Adams et al. 2004). The conformational distortion of J6/7 is mirrored by the fifth and sixth residues in the J8/7 joiner region (Fig. 11, pink). The major groove triplex is extended by C171 (Tanner and Cech 1997; Tanner et al. 1997), while the stacking interface in the G binding site is lengthened by A172. J6/7 and J8/7 are aligned parallel to each other and cross at a conformationally extended phosphate in each strand, which allows both strands to traverse across the active site. This creates a novel ying-yang strand juxtaposition, where C171 is stacked above U126, but A172 is stacked under A127. Of the seven residues in these two strands, only the first two nucleotides of J6/7 are stacked directly on top of each other (see stacking section below for further discussion).

The length, but not the sequence, of J6/7 is absolutely conserved (Cannone et al. 2002). It always includes three nucleotides. The short length of J6/7 combined with the large distance that it must traverse across the active site, pulls G128 out of the P7 helix and into the P7 major groove where it forms the lid to the G-binding site (Fig. 11). Among group I introns, the sequence of the first two J6/7 residues covaries with the sequence of the second and third base pair in P4 (Michel et al. 1990). Similar covariation was observed between J8/7 residue 5 (C171) and the fifth base pair in P4 (C55-G89) (Tanner and Cech 1997; Tanner et al. 1997). Mutagenesis in these regions confirmed the functional importance of these triples and led to proposals for how they may interact (Michel et al. 1990; Tanner and Cech 1997; Tanner et al. 1997). As with the J3/4 region, the triples observed within the *Azoarcus* pre-2S structure are subtly different than suggested in earlier studies, but the base-pairing partners are as predicted. G125 uses its sugar edge (O2' and N2) to straddle the Hoogsteen edge of both residues in the C52-G92 pair (Fig. 11; Table 1). The distance from C171 to the P4 helix is rather long (3.6 Å). The only possible interaction is between the C171 N4 and the Hoogsteen face (N7 or O6) of G89. The full formation of this interaction may occur differentially between the first and second step of splicing.

Unlike the other three J6/7 residues, the third nucleotide (A127) is almost universally conserved. Interference analysis suggested that its 2'-OH and N3 are important for function, a pattern expected for an A minor interaction (Soukup et al. 2002). While both of these functional groups are used in the structure, A127 does not participate in a tertiary interaction of this type. Its 2'-OH coordinates one of the two active site metal ions, metal M₂ (biochemically defined as M_C) (Adams et al. 2004). Its N3 hydrogen bonds to the 2'-OH of A172 upon which A127 stacks. While these contacts are critical, this structure does not appear to fully explain why A127 is so highly conserved. One intriguing possibility comes from the observation that A127 is coplanar with the A131-U176 pair in P7, which is also almost completely invariant (Cannone et al. 2002). However, within this structure, these residues are not sufficiently close to make a direct tertiary interaction; but only a small structural change would be needed to create such a contact, and functional groups in the A131-U176 major groove are important for intron activity. Should it form, this triple would be located immediately below the ΩG-G130-C177 triple, which might explain why all three residues are so highly conserved.

J8/7

Although J8/7 does not formally bridge between two helical domains, it contains what are arguably the six most interesting continuous nucleotides within the structure (Fig. 12). This single stranded region pirouettes through the active

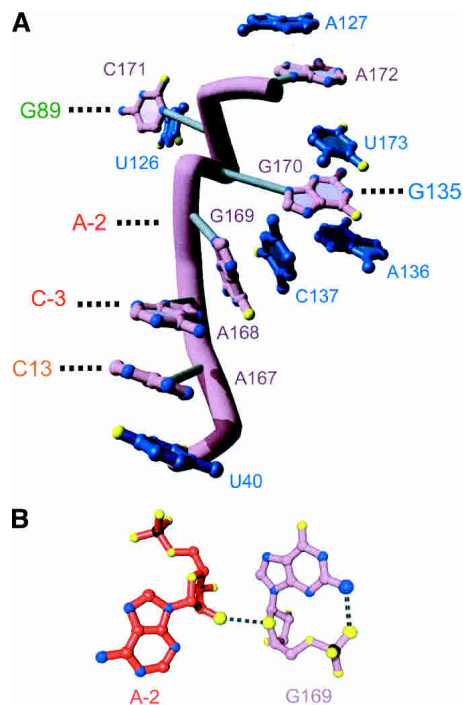


FIGURE 12. Schematic depiction of the J8/7 segment. (A) The μ shaped backbone of this single-stranded element is shown as a ribbon (pink). The six individual J8/7 bases are shown in pink with gray cylinders connecting them to the backbone. The bases of nucleotides that make stacking interactions with the J8/7 segment are shown in blue. Nucleotides that form tertiary hydrogen bonding interactions with J8/7 are not shown, but their approximate placement within the structure is indicated by a label. The color of this label follows the scheme established in Figure 1. (B) The G169 (pink) triple with A-2 (red). The intranucleotide and internucleotide hydrogen bonds are shown as black dashed cylinders to atoms depicted as enlarged spheres. Individual oxygen (yellow), nitrogen (blue), and phosphorous atoms are colored differently from the rest of the nucleotide.

site, makes critical contacts to all three helical domains, and provides ligands for active site and structural metal ions.

The first three nucleotides of J8/7 form a minor groove triplex with P1 and the adjacent base pair in P2 (Fig. 9; Szewczak et al. 1998; Soukup et al. 2002). A167 and A168 continue the single stranded base stack initiated in J2/3 (see above). These two conserved adenosines make type II and type I A-minor contacts with the internal guide sequence (C13) and the 5'-exon (C-3), respectively (Fig. 4). The third residue is stacked with the bulged C137 in the P3 helix and is in a syn conformation. This residue also contacts the P1 helix, but not in the manner predicted based upon interference suppression (NAIS) analysis (Fig. 12B). Inosine interference at *Tetrahymena* G303 (Azo G169) was suppressed by 2'-deoxy substitution at C-2 (Azo A-2) (Szewczak et al. 1998). This was interpreted to be a direct hydrogen bond between the A-2 2'-OH and the N2 amine of G303. The structure is consistent with the data, but not consistent with the interpretation. The N2 of G169 forms a hydrogen bond back to the pro- S_p oxygen of its own phosphate, and the

nucleotide is almost perpendicular to A168 (Fig. 12B). This orients the G169 O4' in the ribose ring for hydrogen bonding to the A-2 2'-OH. Thus, the interaction between the G169 amine and the A-2 2'-OH groups is indirect. Of the 15 direct and functional group specific hydrogen bonds predicted throughout the intron based upon NAIM analysis (Strobel et al. 1998; Szewczak et al. 1998; Strobel and Ortoleva-Donnelly 1999; Soukup et al. 2002; L. Ortoleva-Donnelly and S.A. Strobel, unpubl.), this is the only contact not present within this crystal structure.

Within the three residues from G169 to C171, J8/7 makes tertiary contacts to all three of the major helical elements and it does so on opposite sides of the active site (Figs. 2A, 4, 12A). The μ shaped topology mentioned above, begins at G169 and continues to U173 at the base of P7. Metal ions are coordinated at each change in strand direction, including coordination to both of the catalytic metal ions (Adams et al. 2004). G170 is fully interdigitated into the P3-P9.0 helix and forms a noncanonical pair with G135 (Figs. 4, 12A). C171 and A172 were discussed in the J6/7 section above.

J9/9.0

The asymmetric loop connecting the P9 and P9.0 helices produces a sharp right angle turn between the two helical domains that turns in the direction opposite to that expected of a standard kink turn (Klein et al. 2001; Strobel et al. 2004).

A minor interactions

The structure includes five type I and six type II A-minor motifs (Fig. 13; Doherty et al. 2001; Nissen et al. 2001). An additional type II A minor interaction (A201) mediates intermolecular contacts within the crystal. This is a remarkable abundance given that there are only 40 unpaired or mispaired A's within the RNA (exclusive of the U1A binding site). Many of the intramolecular A minor motifs occur at key places within the tertiary structure, including: (1) the P1 wobble receptor in J4/5 (A58); (2) the minor groove triplex between P1-P2 and J8/7 (residues A167 and A168); (3) the minor groove triplex between J3/4 and P6 (A48 and A49) and P6/6a and P3 (residues A97 and A98); and (4) the interactions involving both the L2 and L9 tetraloops (A25, A26, A191, and A192) (Fig. 13B). A-minor motifs are often a component within larger structural motifs, including the wobble receptor (Strobel et al. 1998), the tetraloop receptor (Cate et al. 1996a), and the K-turn (Klein et al. 2001). Several of the intramolecular A-minor interactions were predicted by Nucleotide Analog Interference Mapping and related experiments, and all seven of the A's in the central core are highly conserved, suggestive of an important functional role. Thus, the type I and type II A-minor motifs, which the ribosome uses during translation to decode the

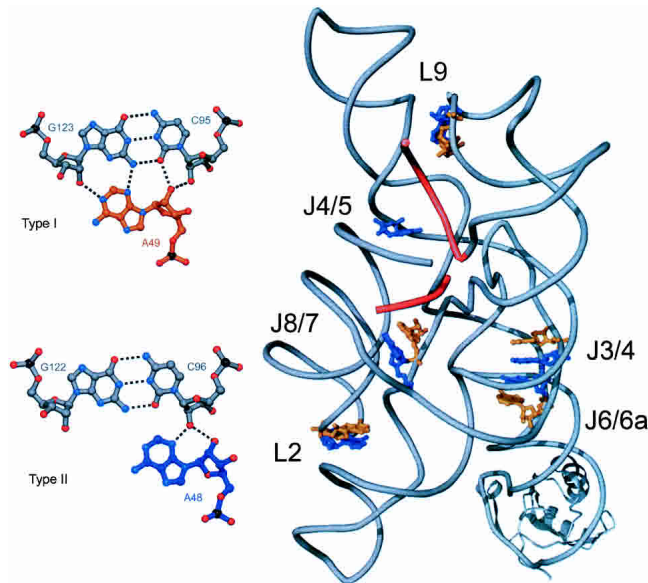


FIGURE 13. Type I and type II A minor motifs throughout the *Azarcus* group I intron. (Left) Examples of the type I (orange adenosine with green base pair) and type II (blue adenosine with green base pair). Hydrogen bonds shown as dashed black cylinders. (Right) Placement of the five type I (orange) and six type II (blue) A minor motifs within the overall structure. The entire intron and the U1A protein are shown as a gray ribbon. The 5' and 3' exons are shown as a red ribbon.

message and align the aminoacyl ends of the tRNA within the peptidyl transferase center (Carter et al. 2000; Nissen et al. 2000), are also used to select the 5'-splice site in the group I intron. This reinforces the conclusion that the A-minor motif is a versatile and ubiquitous tertiary structural element.

Base stacking interactions

A major factor in nucleic acid structural stability is base stacking. Analysis of how the intron and exon bases are stacked throughout the complex reveals information quite different from the description of tertiary hydrogen bonding interactions outlined above. A schematic representation of the base stacking interactions is shown in Figure 14. Only six residues form no stacking interaction, and five of these residues are pyrimidines. The only unstacked purine is G71 within the P5 loop, which is relatively disordered within this structure, and may make an intermolecular stacking interaction within the crystal (see below). Given that each nucleotide can make two stacking interactions—i.e., one above and one below the plane of the base—there are 424 possible interaction surfaces for stacking (excluding the residues in the U1A binding loop). Of these, 381 are stacked, which is 90% of all possible interactions.

For this discussion, the stacking interactions will be divided into three categories: (1) intrastrand stacking between

consecutive residues in a helix (Fig. 14, brown double lines); (2) interstrand or cross-strand stacking between residues on opposing strands of a duplex (Fig. 14, blue; Correll et al. 1998); and (3) tertiary strand stacking of residues distant in secondary structure (Fig. 14, red). Those residues that are left unstacked within the complex are also of interest.

P1-P10 stacking

Helix P1-P10 forms two continuous stacks that extend into the tetraloop at the bottom of P2 (Fig. 14). These helices include the functionally critical stacking of the 5' and 3'-exons between dT-1 and dA+1. The last base of the 5'-exon, C-3, stacks directly on G37 of the P2 helix. This suggests

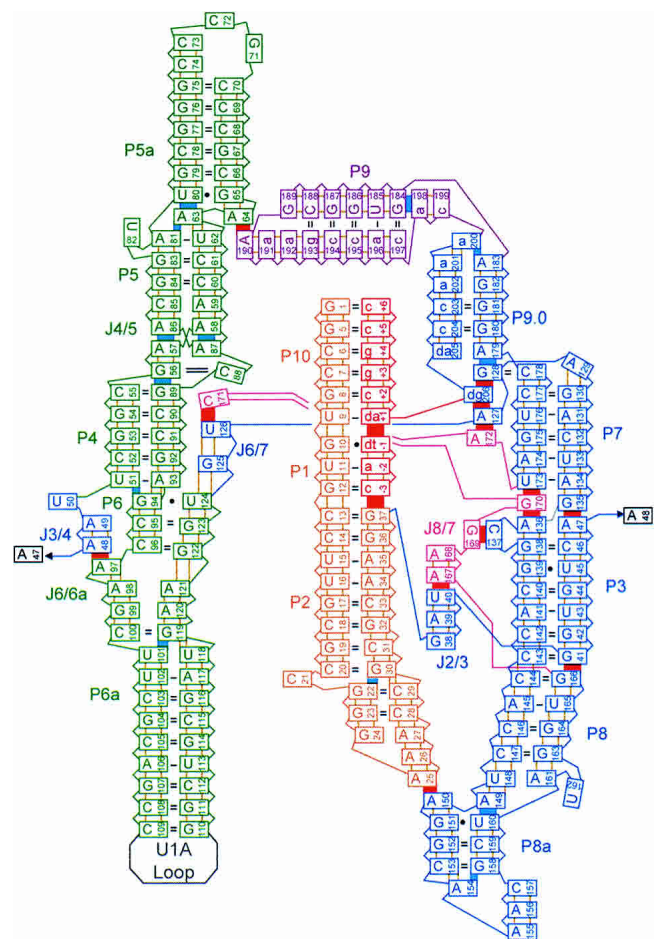


FIGURE 14. Schematic depiction of the base stacking within the *Azarcus* pre-2S complex. Each nucleotide is depicted as a rectangle labeled with the nucleotide number and sequence. The relative placement of the rectangles denotes stacking interactions. Standard stacking interactions within a helix are shown as two parallel brown lines between each nucleotide. Cross-strand stacking contacts are shown as a thick blue line between the rectangles. Tertiary strand stacking contacts are shown as a thick red line between the rectangles. No effort is made in this figure to depict tertiary contacts (Fig. 4). Individual helices and joiner elements are labeled. The connectivity of the strands is shown with colored lines. The color scheme follows that established in Figure 1.

that in the context of the full length exon, the next nucleotide (U-4) must be displaced from the P1-P10 helix. P2 stacking continues onto the bottom of the P8 helix via the tetraloop receptor where A25 stacks on A150. This is facilitated by the cross-strand stacking of A149 on U160 to form the AA platform, which is the hallmark feature of the tetraloop receptor (Cate et al. 1996b).

P4-P6 stacking

Stacking within the P4-P6 helix is unexpectedly discontinuous given what appears to be a relatively standard helical architecture (Figs. 7, 14). There are no stacking interactions that span the length of the domain (Fig. 11). An unexpected feature of the P4 helix is a cross-strand purine stack between G56 and G89 that occurs within a helix at the junction of two standard Watson-Crick base pairs (Figs. 14, 15A). It leaves C88, whose phosphate backbone coordinates a catalytic metal ion (M_1), unstacked with other residues in the P4 helix. This, in turn, leaves the underside of J4/5 residue A87 without a stacking contact. A87 makes a direct tertiary contact to the P1 helix and the 5'-splice site and its phosphate also coordinates a catalytic metal ion. Consequently, these stacking rearrangements are likely to be functionally relevant. This stacking interface is substantially different than observed within the *Tetrahymena* P4-P6 crystal structure and may reflect active site rearrangement upon P1 helix docking or metal ion binding (see below).

The second major disruption to stacking continuity within the P4-P6 domain comes in the J6/6a loop (Fig. 14). A97, the single unpaired nucleotide in the J6/6a asymmetric

loop, makes a tertiary strand stack under J3/4 residue A48. This interaction provides the foundation for the J3/4 minor groove triplex with P6, and it leaves the underside of C96 unstacked (Fig. 10).

P3-P9.0 stacking

In contrast to P4-P6, stacking within the P3-P9.0 helix is remarkably continuous, particularly given the complexity of the secondary structural interactions that make up this domain (Fig. 14). Both ends of the P3 make complete stacking contacts with P7 and P8. The junction between P3 and P7 is quite intricate and involves tertiary stacking interactions on both strands. A47 stacks under G135, which in addition to a tertiary stacking interaction, participates in a cross-strand stack, in that it is not stacked on A136. Residue G170 from the J8/7 region occupies the vacant space above A136. G170 bridges the P3 and P7 interface by stacking on A136 (P3) and under U173 (P7). This organization is quite remarkable as the next residue in J8/7, C171, also makes two tertiary stacking interactions, but on the opposite side of the molecule in P4-P6 (Fig. 11) (see above). The preceding nucleotide in J8/7, G169, makes an isolated two nucleotide tertiary stack with the P3 bulged residue C137. Within the J8/8a junction, the AA platform bifurcates the stacking within the P8a helix. The left side of the P8a stack assembles under the P2 helix via tertiary stacking with A25 (Fig. 14). The right side stack extends into P8a through a cross-strand stack with A149. Similar bifurcation occurs because of the AA platform J5/5a. In this case, the P5a stacks are divided into P9 and P5, respectively.

The stacking interactions at the apex of the P7 helix are quite unexpected and novel (Fig. 14). The stacking continuity on both P7 strands ends abruptly at the A129-G128-C178 triple. Instead of stacking on A129, G128 is coplanar with the base and positioned over the major groove of the P7 helix. There, it makes a tertiary stacking interaction with Ω G206 and a cross-strand stack with A179 of P9.0. The preceding residue, A127 stacks under the Ω G and stacks on A172, the last residue in the J8/7 segment. This complex tertiary stacking interface located in the P7 major groove, organizes the G binding site and provides ligands for catalytic metal ions. Strikingly, P9.0 does not stack directly on P7. Instead it stacks over G128 in the P7 major groove. This arrangement of bases is substantially different than any model proposed to date (Michel and Westhof 1990; Golden et al. 1998; Kitamura et al. 2002).

There are no residues stacked under nucleotide dA205 (Fig. 14), which may provide a clue into the splicing mechanism. During the first step of splicing, exogenous G rather than Ω G must occupy the G binding site (Been and Perrotta 1991). The exogenous G must be able to displace Ω G from the active site, and Ω G must be located elsewhere. It is possible that the space beneath dA205 serves as the binding site for Ω G during the first step of splicing. One desirable

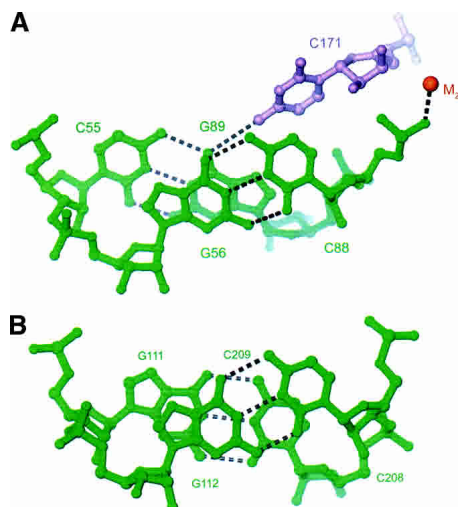


FIGURE 15. Cross strand purine stack in P4. (A) Cross strand stack in helix P4 between *Azoarcus* residues G56 and G89. Residue C171 (pink) and the catalytic metal ion (orange) coordinated to the C88 phosphate are also shown. (B) Equivalent region from the P4-P6 crystal structure showing the absence of the cross strand stack. The elements are not present in P4-P6 domain structure, which may account for the different stacking interactions in this region.

feature of this hypothesis is that it would allow exogenous G to stack under both Ω G and G128. This may provide additional stabilization, thus allowing G to compete with Ω G for active site binding. Following 5'-exon cleavage, the G would be displaced from the active site by P10 formation between the 3'-exon and the internal guide sequence, and Ω G could occupy the G binding site.

J8/7 tertiary stacking

Although it has been mentioned in the context of each helix, it is worth giving particular attention to the large number of stacking interactions made by residues in J8/7 (Figs. 12A, 14). Only two of its six residues stack upon each other. All of the other interactions are tertiary stacking contacts, including interactions with J2/3 (A167; Fig. 9), P3 in two different places (G169, G170; Fig. 12A), J6/7 in two different places (C171, A172; Fig. 11), and P7 (G170; Fig. 11). There is a total of 15 tertiary stacking interactions in the entire structure; six of these are made by residues in J8/7 (Figs. 12A, 14).

Crystal packing interactions

The interactions that make up the crystal also hold a few surprises. The U1A protein plays an entirely different role in pre-2S intron crystal formation than expected based upon previous structures (Ferre-D'Amare and Doudna 2000). There is no place where two U1A molecules make direct protein-protein interaction within the pre-2S intron crystal (Fig. 16). Even more surprisingly, there is only one intermolecular protein-RNA interaction. This occurs between the U1A of one molecule and the P5 helix of second. Although the intermolecular RNA-protein interaction involves only about 140 Å² of surface area, U1A addition was absolutely required for crystal growth.

Intermolecular RNA-RNA interactions dominate the pre-2S intron crystal packing. There are three principal points of RNA-RNA interaction between symmetry related molecules. (1) The largest intermolecular interaction (~770 Å² of surface area) involves symmetry related dimer formation between two pre-2S complexes (Fig. 16B). This is responsible for one of the two two-fold symmetry axes in the crystal. Helix P5a of one molecule sits in the minor groove of helix P2 in a second molecule, and vice versa. Some dimer formation (<5% of total RNA) was observed during native gel analysis of the pre-2S complex (Adams et al. 2004), and these interactions may be responsible for that complex. (2) The P8a loop docks against P9 and stacks on P10 of a symmetry related molecule (Fig. 16A). A155 stacks on the G1-C+6 pair, effectively extending the stacking interactions within the P10 helix. These interactions bury ~400 Å² of surface area. The same molecule that makes these interactions uses its P6a bound U1A to contact P5a (Fig. 16A). Dual intermolecular contacts involving P6a and P8a could affect the angle between the P6 and P8 helices, though hy-

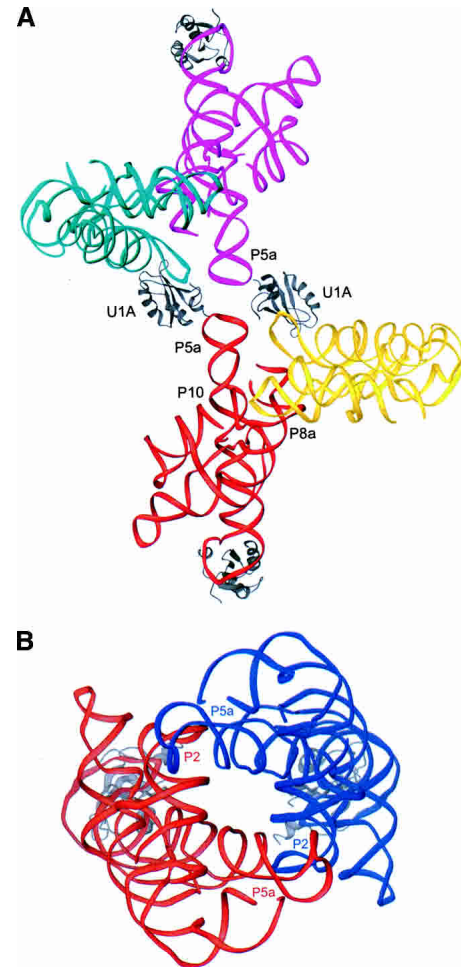


FIGURE 16. Intermolecular crystal contacts. (A) View down one of the two two-fold symmetry axes in the crystal. The U1A proteins are shown in gray. Individual pre-2S complexes are shown as ribbons of various colors. The contacts between L8a and P9 and the top of P10 are labeled. The L5a-L5a contact is also labeled. (B) View down the other two-fold axis within the crystal. This rather intimate homodimer occurs between residues in P2 and P5a.

droxyl radical footprinting of this pre-2S complex in solution argues against this possibility (see below). (3) L5a is juxtaposed opposite L5a from its symmetry mate, producing another two-fold symmetry axis with weak intermolecular contacts (Fig. 16A). G71, C72, and C73 are in close proximity (3.5–4 Å), and may interact by irregular stacking. These three bases in L5a are not well ordered, so the nature of this interaction is uncertain. The fact that none of these contacts are near active site residues argues that crystal packing is unlikely to have caused significant active site distortion.

Comparison to previous group I intron domain structures

Three previous X-ray crystal structures have revealed the architecture of *Tetrahymena* intron domains in the absence

of exons. The structure of the *Tetrahymena* P4-P6 domain, first at 2.5 Å and later at 2.25 Å resolution (Cate et al. 1996a; Juneau et al. 2001), revealed principles of RNA helix packing. This structure included helices P4, P5, and P6, which are analogous to those present in the *Azoarcus* intron, and a P5abc helical extension that does not occur within the compact bacterial intron. Subsequently, an ~5 Å resolution unrefined model of the *Tetrahymena* intron was reported for helices P3-P9 (Golden et al. 1998). This construct could be considered an apoenzyme model because it included the P4-P6, P3-P9.0, and P9 helical domains, but did not include the exons or the internal guide sequence, i.e., the P1-P10 domain.

The bacterial P4-P6 domain shows good correlation to that reported for *Tetrahymena* (Cate et al. 1996a; Juneau et al. 2001). Both domains have a similar length (~110 Å) and the core helical elements can be superimposed with a reasonably good fit (rmsd of 2.5 Å). There are some differences that result solely from the architecture of the two introns, such as the placement of the tetraloop receptors (J5/5a in *Azoarcus*, J6a/6b in *Tetrahymena*) and the orientation of helix P5a (coaxial with P5 in *Azoarcus*, antiparallel with P5 in *Tetrahymena*). The issue of greater interest involves the conformational changes that might occur upon folding of the entire intron and docking of the substrate helix (P1).

The unusual base stacking that occurs at the junction between P4 and J4/5 in the bacterial intron complex, is not observed between analogous residues (*Tet*111-113:207-209) in either the *Tetrahymena* P4-P6 domain structures or the apoenzyme structure (Fig. 15B; Cate et al. 1996a; Golden et al. 1998; Juneau et al. 2001). The key differences are the G56 G89 cross-strand stack, the complete unstacking of C88, and the partial unstacking of A87. Instead, the stacking profile is typical of that for a continuous helix. When the J4/5 regions of the structures are aligned, the absence of the cross-strand shift and C88 unstacking leaves the P4 helix, particularly the bulged residue *Tet*A210 (located between G89 and C90 in the *Tetrahymena* intron), in the space that is occupied by the IGS in the *Azoarcus* pre-2S complex (Fig. 17). This suggests that P1 helix docking, tertiary stacking of C171, and/or catalytic metal ion binding, promotes a significant conformational shift, possibly as large as 6 Å, at the interface of helix P4 and J4/5.

Comparison of the bacterial pre-2S structure with the *Tetrahymena* P3-P9 structure (Golden et al. 1998) suggests that other key regions of the intron undergo conformational changes upon substrate helix binding. Within the apoenzyme structure there is also significant steric overlap between J8/7 and the space occupied by the 5'-exon (Fig. 17). This is not likely to be due to differences between the two introns, because biochemical data suggest that the tertiary contacts to the 5'-exon are equivalent (Soukup et al. 2002). Significant changes in the J8/7 conformation from that reported in the apoenzyme structure are also consistent with biochemical studies (Szewczak et al. 1999).

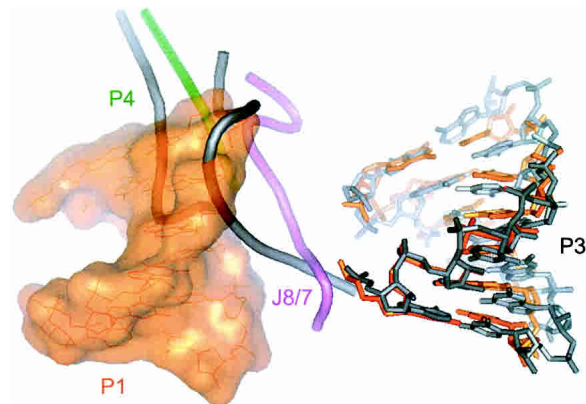


FIGURE 17. Superposition of the *Azoarcus* pre-2S and *Tetrahymena* P3-P9 apoenzyme structures. The alignment of the superposition is depicted within the P3 region, where pre-2S is shown in orange and the apoenzyme is shown in gray. The first six base pairs below the 5' splice site from the pre-2S P1-P2 helical stack is shown as a transparent orange surface. The backbone of the J8/7 (pink) and P4 (green) strands of pre-2S are shown as cylinders. The equivalent strands from the P3-P9 structure are shown in gray. Note how the J8/7 segment overlaps extensively with the 5' exon and how the P4 strand overlaps extensively with the IGS. This figure was prepared using PyMOL (DeLano 2002).

Substantial differences are also present within the G-binding site. The P3-P9 apoenzyme structure includes an ΩG, and it was modeled into the active site within one of the two molecules in the asymmetric unit. The details of the *Azoarcus* intron G-binding site were discussed in the previous manuscript (Adams et al. 2004). In both structures the ΩG forms a Hoogsteen base triple with the universally conserved G-C pair in P7 (*Tet* G264-C311); however, all other key features of this motif are different between the two structures, including the base triple above the G-C pair, the stacking of G128 over the ΩG and the stacking of A127 under it (*Tetrahymena* residues 261–263; Figs. 1A, 4, 11). The cause of these differences is unknown, but they might result from differences between these two introns, the low resolution of the apoenzyme model, the 2'-deoxy substitution we introduced at ΩG, or significant conformational changes induced upon P1 helix docking. Higher resolution structures of folding intermediates will be needed to clarify this issue.

In addition to the two *Tetrahymena* intron crystal structures, a model of the *Azoarcus* intron was constructed based upon biochemical, phylogenetic and structural constraints (Rangan et al. 2003; Rangan et al. 2004). The overall rmsd between the *Azoarcus* structure and the model, over the 1542 sugar atoms, is 9.1 Å, but the rmsd normalized to 100 residues is 3.8 Å using LSQMAN (Carugo and Pongor 2001; Kleywegt et al. 2001). This demonstrates that it is possible to build fairly accurate RNA models using a variety of biochemical constraints and isolated structural motifs present within an increasing number of RNA crystal structures. Such models are not yet sufficiently accurate to be consid-

ered atomic resolution structures, but they certainly provide the basis for designing experiments to test atomic resolution questions.

Solvent accessibility within the pre-2S complex

One point of potentially significant divergence between the phylogenetic and crystallographic models is in the angle between the P4-P6 and the P3-P9.0 domains, which is substantially more acute in the phylogenetic model (Fig. 2B). The sharp angle was justified by observation of a solvent inaccessible Fe(II)-EDTA footprint in helix P6a (C100-C105) (Rangan et al. 2003; Rangan et al. 2004). This region appears to be fully solvent exposed within the pre-2S structure, but both the P6a and P8a helices are making intermolecular crystal contacts that might alter the angle between the helices within the crystals (see above). To explore the correlation between the crystal structure and the solution structure, we performed hydroxyl radical footprinting analysis on the pre-2S complex (Fig. 18A; Latham and Cech 1989; Cate et al. 1996a; Hampel and Burke 2003).

Within the pre-2S structure, there is excellent correlation between the degree of backbone protection and the predicted solvent accessibility (Fig. 18). The protected C4' protons are located almost exclusively along the intron core, signifying the dense packing of the ribose-phosphate backbone in these regions. The protections map precisely onto tertiary elements, including both of the TL/TLR contacts and the J8/7 region. Particularly noteworthy is J4/5 where the entire region showed strong protections. Furthermore, weak protections were observed in the U1A binding loop, but only when protein was included in the splicing complex. A total of 54 residues in the quantifiable regions were protected from reaction with hydroxyl radicals. Due to the heterogeneous ends produced from hydroxyl radical cleavage, these data are accurate to within one or two nucleotides. Based on the theoretical solvent accessibility of the C4' proton calculated from the crystal structure, 55 residues were predicted to show protection (Fig. 18). The only unpredicted protection (A134), lies on the same side of the P7 helix within a region of predicted and observed protections (C137-A141). The calculated accessibility of the C4' proton of A134 (2.1 \AA^2) is

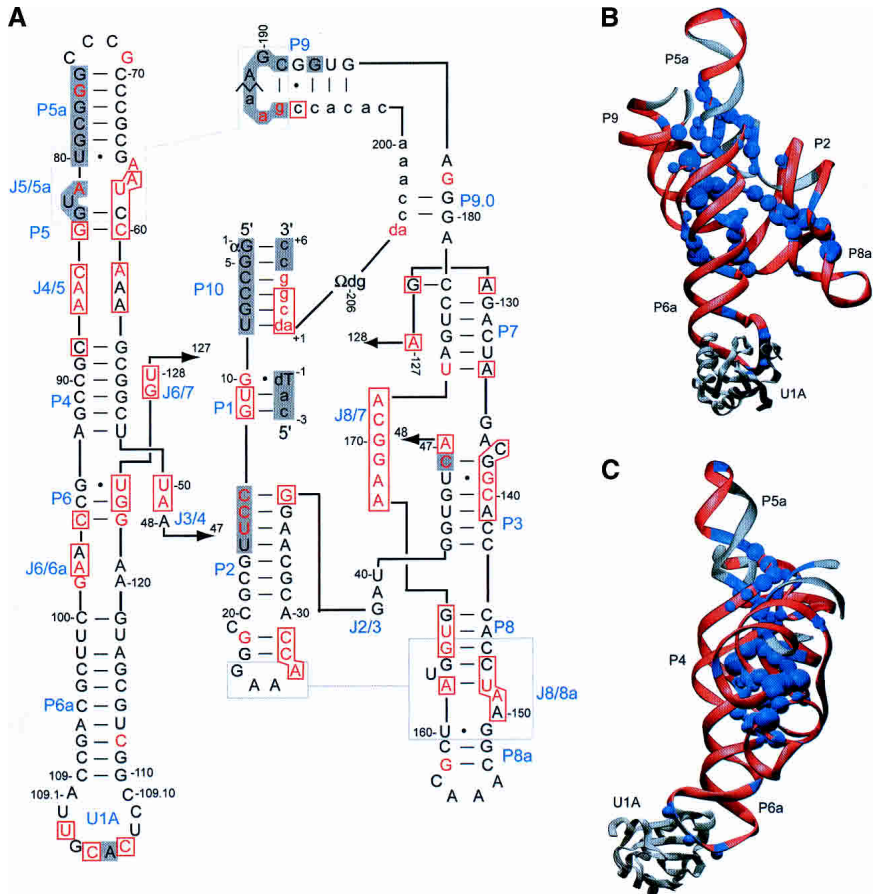


FIGURE 18. Hydroxyl radical mapping of the *Azoarcus* pre-2S complex. (A) Secondary structure map of hydroxyl radical protections. The color of the letters indicates residues whose C4' proton is predicted to be protected (red) or solvent accessible (black) within the crystal structure using a 1.4 \AA probe. Positions that were protected from hydroxyl radical reactivity in solution are enclosed in red boxes. Gray squares indicate positions that could not be quantitated due to degradation, band compression within the sequencing gel, or close proximity to the end of an RNA strand. (B) Front view of hydroxyl radical protections mapped onto the pre-2S crystal structure. The backbone color indicates areas of predicted protection (blue) and predicted solvent accessibility (red). Areas that could not be quantitated are colored gray. Blue spheres indicate C4' atoms that were resistant to hydroxyl radicals in solution. The sphere size correlates with the degree of protection (small: $1.5\text{--}2.0$, medium: $2.0\text{--}3.0$, large: >3.0). (C) As in B, but rotated 90° about the P4-P6 axis so the clustering of protected residues along the interior of the pre-2S structure can be visualized.

only slightly larger than the limit for classifying a nucleotide as protected (1.0 \AA^2). Only six residues predicted to be solvent inaccessible were reactive (Fig. 18), and each of these is isolated and located within peripheral elements. Conformational dynamics in solution could readily account for these minor exceptions.

Comparison of these data to that reported for an exonless intron lacking an internal guide sequence reveals significant differences in only two regions, J4/5 and P6a (Rangan et al. 2003). These changes may reflect differences in the intron conformation in the presence or absence of its substrates. The J4/5 region was fully solvent accessible in the earlier work, while the entire region was protected in the pre-2S complex (Fig. 18A). Helices P1 and P10 are the

principle tertiary contacts of J4/5, so it is reasonable that the absence of these helices would increase J4/5 solvent accessibility. Although protections were observed in J6/6a, none of the protections in P6a that were used to justify the acute angle between P6 and P8 were observed in the pre-2S complex (Fig. 18A; Rangan et al. 2003). The solvent accessibility of P6a within the pre-2S complex is consistent with the orientation of P6 and P8 observed in the crystal structure, which argues that the P6-P8 angle is not a crystallization artifact.

Intron modeling with full exons

Group I intron splicing proceeds by two phosphotransfer reactions separated by a conformational change (Fig. 19A; for review, see Cech and Golden 1999). In the first step, an exogenous G (α G) binds within the intron's active site and attacks the 5'-splice site, which breaks the covalent bond to

the 5'-exon. The α G covalently tethered to the 5'-end of the intron is removed from the G binding site and replaced by the Ω G. This structural rearrangement involves disruption of the P1a helix and formation of the P10 helix between the 5'-end of the intron and the 3'-exon (Golden and Cech 1996). Exon ligation is completed by nucleophilic attack of the 5'-exon on the scissile phosphate with release of the Ω G and the rest of the intron.

In its natural context, the exons of the *Azoarcus* intron base pair to form the anti-codon helix of Ile-tRNA (Fig. 19A; Reinhold-Hurek and Shub 1992). To facilitate structure determination, both exons were truncated, but it was relatively straightforward to model these exons into the complex using the crystal structure of Ile-tRNA from *E. coli* to estimate the exon's structure (Fig. 19C; Silvan et al. 1999). P10 formation in the natural pre-2S complex (Fig. 19A) necessitates that the last two base pairs of the anticodon helix be denatured (Reinhold-Hurek and Shub 1992).

A+1 does not base pair to the IGS (helix P10) in the natural context for this intron (though it does base pair in other examples). Thus, within this full length exon model we left A+1 in the P1-P10 stack, but shifted the base-pairing register by pairing U9 with A+2 and continuing that shift through the P10 helix (Figs. 1A, 19A). When extrapolated into the tRNA, this base-pairing shift resulted in a slight displacement of the tRNA that eliminated a modest steric clash between the aminoacyl stem and the P9 helix. In this model, the anticodon helix stacks directly on top of P10 and the aminoacyl acceptor helix straddles the P9 helix (Fig. 19C).

We next set out to visualize the RNA prior to the first step of splicing (Fig. 19B). We made the simplifying assumption that the core of the intron remains unchanged for the two steps of splicing, which is likely to be sufficiently accurate for the present purpose. All of the changes were confined to helix P1a, P10, the relative orientation of the tRNA, and which G (α or Ω) was located in the G binding site. α G and Ω G were assumed to bind equivalently. This provided a starting point for considering the conformational changes needed to distinguish the two splicing reactions. For the first set of splicing, P1a was modeled using the first two base pairs of P10 in the current structure and capping the helix with the structure of a UNCG tetraloop (Cheong et al. 1990;

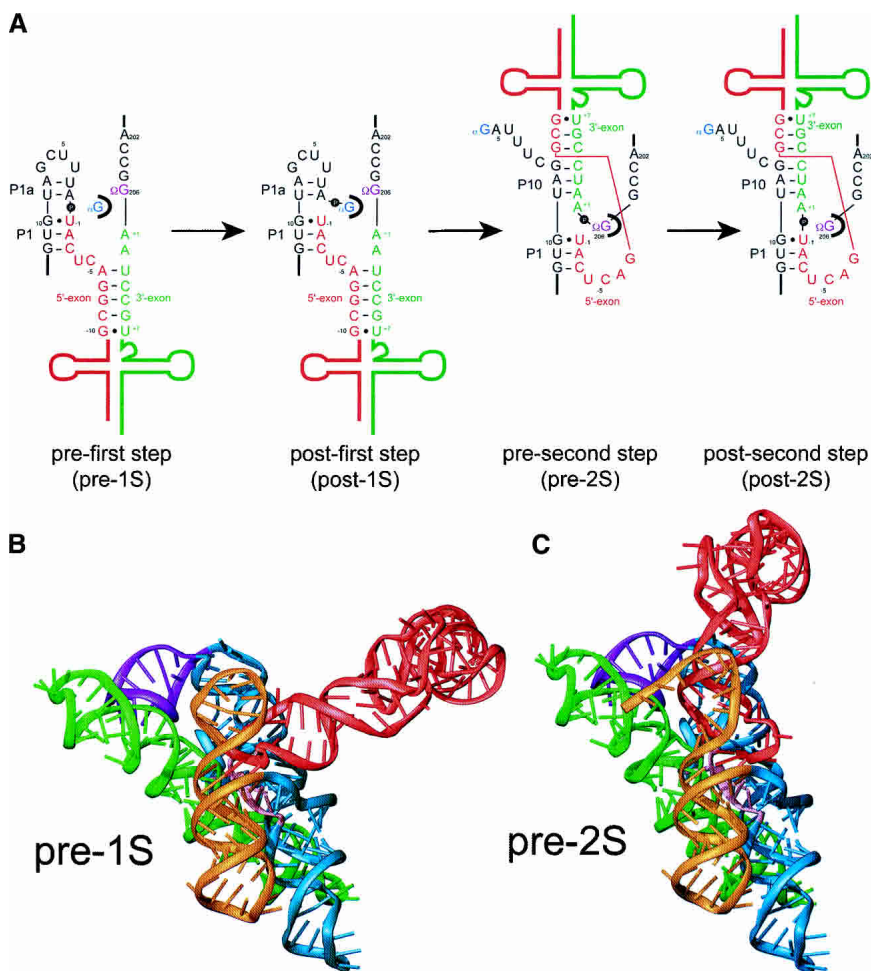


FIGURE 19. (A) Secondary structures of four intermediates in the group I intron splicing reaction: pre-1S, post-1S, pre-2S, and post-2S. These are shown with all the nucleotides in the Ile-tRNA anticodon helix. (B) Model of the pre-1S complex including a P1a helix and a full tRNA. The assumptions used to create this model are described in the text. The color scheme follows that of Figure 1 with the 5' and 3' exons extended in red ribbon. (C) Model of the pre-2S complex including the full tRNA.

Colmenarejo and Tinoco 1999). Ω G does not occupy the G binding site, so it was moved into the stacking position under A205 based upon the justifications given above. The approximate location of the 3'-exon in this pre-1S complex could be fixed by the need to form the anticodon helix between A-6 in the 5'-exon and U+3 in the 3'-exon. An angle between the anticodon helix and P1-P2 was chosen that eliminated steric clash between the tRNA and the intron. The tRNA cannot occupy the position modeled for the second step of splicing due to the presence of the P1a helix. Consequently, P10 formation during the second step of splicing may result in as much as a 60° reorientation of the tRNA relative to the intron. Although this helical rearrangement is substantially less dramatic than the conformational changes predicted to occur during pre-mRNA splicing, it provides a glimpse into the dynamic processes that must occur in that system (Staley and Guthrie 1998).

Conclusions

The group I intron has been the focus of intensive biochemical investigation since its discovery almost 25 years ago (Cech et al. 1981). Almost every one of the helical domains and joiner segments has been studied in detail, sometimes in isolation, sometimes in the context of a complete intron (Cech and Golden 1999). This structure of a group I splicing intermediate finally makes it possible to view an entire RNA splicing system in its full context. The strong correlation between the biochemical and structural results is both a credit to those who have worked on this system over the years and a demonstration that the crystallized intron is in a functionally relevant conformation. We hope in the near future to see structures of other steps in the splicing pathway, including mimics of the chemical transition state, to see if, or rather how, the RNA conformation changes during the splicing process.

MATERIALS AND METHODS

Refinement

An additional round of refinement was performed using Refmac5 to correct minor errors identified in the original structure (Murshudov et al. 1997; Adams et al. 2004). Three residues (A57, G99, and G182) were incorrectly modeled in a syn conformation. These were changed to anti, the placement of the nucleotide was adjusted, and the structure was refined. The R_{free} and R values of the resulting complex were improved significantly to 27.9% and 24.6%, respectively. The modified coordinates have been deposited in the PDB under accession number 1U6B. The metals in this deposition have been renumbered to be consistent with that reported in Adams et al. (2004).

Activity analysis

The activity of the pre-2S intron complex was tested using ribose and 2'-deoxy ribose substrates, in the absence and presence of the

U1A protein (Fig. 1B). Reactions were performed in 50 mM sodium cacodylate at pH 6.8, 10 mM MgCl_2 , 1 μM UP62 RNA, 2 μM CAU (r) or CAT (d) oligonucleotide, and 1.2 μM U1A protein. A trace amount of 5'-end ^{32}P -radiolabeled dCIRC or rCIRC was added to initiate the reaction. Aliquots of the reaction mixture were removed at various times and quenched in a formamide loading buffer. Products were separated from substrates by polyacrylamide gel electrophoresis and visualized using a Storm Phosphorimager (Molecular Dynamics).

Hydroxyl radical footprinting

Footprinting of the pre-2S crystallographic complex was performed similar to that described previously (Latham and Cech 1989). The footprinted RNA was ^{32}P 5'- or 3'-end labeled on either the UP62 transcript or the dCIRC oligonucleotide. The pre-2S crystallographic complex with U1A protein was formed at $\sim 1 \mu\text{M}$, with the labeled RNA slightly limiting, heated to 50°C and cooled slowly to room temperature. The folded and unfolded samples contained 10 mM MgCl_2 and 10 mM EDTA, respectively. Uncleaved samples were also prepared in 10 mM MgCl_2 to control for degradation and any metal induced cleavage. Footprinting conditions were performed in 5 mM DTT and 1 mM Fe(II)-EDTA for 80 min at 42°C. Reactions were quenched with 10 mM thiourea and ethanol precipitated. Samples contained about 2×10^5 c.p.m. and cleavage products were separated on sequencing gels. Transcripts containing one phosphorothioate analogue (A α S, C α S, G α S, or U α S) were cleaved in 10 mM I_2 /ethanol and used for sequence determination. Gels were imaged using a Storm Phosphorimager (Molecular Dynamics). Protections were quantitated by normalizing each lane for loading and calculating a protection factor ($\text{Pf} = \text{unfolded intensity} / \text{folded intensity}$). A position was considered protected if $\text{Pf} \geq 1.5$. Regions with significant non-Fe(II)-EDTA related degradation (C46 and A109.6), GC compressions (C13-U16 and G75-G83), or at the very end of the RNA could not be quantitated. Solvent accessibility calculations were performed on the crystal structure using the program ACCESS (Lee and Richards 1971). Solvent accessible areas $\leq 1.0 \text{ \AA}^2$ were considered protected.

ACKNOWLEDGMENTS

We thank C. Hobartner and R. Micura for the gift of a 2'-selenomethyl substituted dCIRC oligonucleotide, M. Becker and the staff of the X-25 beamline at Brookhaven NSLS for assistance with data collection, J. Pflugrath for helpful discussions, C. Berman for assistance with pre-1S modeling, and the staff in the Yale Center for Structural Biology for extensive technical assistance. We also thank L. Szweczek, J. Cochrane, R. Anderson, E. Westhof, and A. Lescaute-Philipps for comments on this manuscript. This work was supported by NSF grant MCB-0315329 and NIH grant GM54839 to S.A.S.

Received July 30, 2004; accepted October 4, 2004.

REFERENCES

- Adams, P.L., Stahley, M.R., Kosek, A.B., Wang, J., and Strobel, S.A. 2004. Crystal structure of a self-splicing group I intron with both exons. *Nature* **430**: 45–50.
- Bass, B.L. and Cech, T.R. 1984. Specific interaction between the self-

- splicing RNA of *Tetrahymena* and its guanosine substrate: Implications for biological catalysis by RNA. *Nature* **308**: 820–826.
- . 1986. Ribozyme inhibitors: Deoxyguanosine and dideoxyguanosine are competitive inhibitors of self-splicing of the *Tetrahymena* ribosomal ribonucleic acid precursor. *Biochemistry* **25**: 4473–4477.
- Basu, S., Rambo, R.P., Strauss-Soukup, J., Cate, J.H., Ferre-D'Amare, A.R., Strobel, S.A., and Doudna, J.A. 1998. A specific monovalent metal ion integral to the A-A platform of the RNA tetraloop receptor. *Nat. Struct. Biol.* **5**: 986–992.
- Been, M.D. 1994. *Cis-* and *trans-*acting ribozymes from a human pathogen, hepatitis delta virus. *Trends Biochem. Sci.* **19**: 251–256.
- Been, M.D. and Perrotta, A.T. 1991. Group I intron self-splicing with adenosine: Evidence for a single nucleoside-binding site. *Science* **252**: 434–438.
- Bevilacqua, P.C. and Turner, D.H. 1991. Comparison of binding of mixed ribose-deoxyribose analogues of CUCU to a ribozyme and to GGAGAA by equilibrium dialysis: Evidence for ribozyme specific interactions with 2' OH groups. *Biochemistry* **30**: 10632–10640.
- Burke, J.M., Belfort, M., Cech, T.R., Davies, R.W., Schweyen, R.J., Shub, D.A., Szostak, J.W., and Tabak, H.F. 1987. Structural conventions for group I introns. *Nucleic Acids Res.* **15**: 7217–7221.
- Cannone, J.J., Subramanian, S., Schnare, M.N., Collett, J.R., D'Souza, L.M., Du, Y., Feng, B., Lin, N., Madabusi, L.V., Muller, K.M., et al. 2002. The Comparative RNA Web (CRW) Site: An Online Database of Comparative Sequence and Structure Information for Ribosomal, Intron, and other RNAs. *BioMed Central Bioinformatics* **3**: <http://www.rna.icmb.utexas.edu/>
- Carson, M. 1991. Ribbons 2.0. *J. Appl. Crystallogr.* **24**: 958–961.
- Carter, A.P., Clemons, W.M., Brodersen, D.E., Morgan-Warren, R.J., Wimberly, B.T., Ramakrishnan, V. 2000. Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. *Nature* **407**: 340–348.
- Carugo, O. and Pongor, S. 2001. A normalized root-mean-square distance for comparing protein three-dimensional structures. *Protein Sci.* **10**: 1470–1473.
- Cate, J.H., Gooding, A.R., Podell, E., Zhou, K., Golden, B.L., Kundrot, C.E., Cech, T.R., and Doudna, J.A. 1996a. Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science* **273**: 1678–1685.
- Cate, J.H., Gooding, A.R., Podell, E., Zhou, K., Golden, B.L., Szweczkak, A.A., Kundrot, C.E., Cech, T.R., and Doudna, J.A. 1996b. RNA tertiary structure mediation by adenosine platforms. *Science* **273**: 1696–1699.
- Cech, T.R. 1986. The generality of self-splicing RNA: Relationship to nuclear mRNA splicing. *Cell* **44**: 207–210.
- . 1990. Self-splicing of group I introns. *Ann. Rev. Biochem.* **59**: 543–568.
- Cech, T.R. and Golden, B.L. 1999. Building a catalytic active site using only RNA. In *The RNA World*, 2d ed. (eds. R.F. Gesteland, T.R. Cech, and J.F. Atkins), pp. 321–349. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Cech, T.R., Zaug, A.J., and Grabowski, P.J. 1981. In vitro splicing of the ribosomal RNA precursor of *Tetrahymena*: Involvement of a guanosine nucleotide in the excision of the intervening sequence. *Cell* **27**: 487–496.
- Cheong, C., Varani, G., and Tinoco, I. 1990. Solution structure of an unusually stable RNA hairpin, 5'GGAC(UUCG)GUCC. *Nature* **346**: 680–682.
- Colmenarejo, G. and Tinoco Jr., I. 1999. Structure and thermodynamics of metal binding in the P5 helix of a group I intron ribozyme. *J. Mol. Biol.* **290**: 119–135.
- Correll, C.C., Munishkin, A., Chan, Y.L., Ren, Z., Wool, I.G., and Steitz, T.A. 1998. Crystal structure of the ribosomal RNA domain essential for binding elongation factors. *Proc. Natl. Acad. Sci.* **95**: 13436–13441.
- Costa, M. and Michel, F. 1995. Frequent use of the same tertiary motif by self-folding RNAs. *EMBO J.* **14**: 1276–1285.
- Damberger, S.H. and Gutell, R.R. 1994. A comparative database of group I intron structures. *Nucleic Acids Res.* **22**: 3508–3510.
- DeLano, W.L. 2002. The PyMOL user's manual. DeLano Scientific, San Carlos, CA.
- Doherty, E.A., Batey, R.T., Masquida, B., and Doudna, J.A. 2001. A universal mode of helix packing in RNA. *Nat. Struct. Biol.* **8**: 339–343.
- Doudna, J.A., Cormack, B.P., and Szostak, J.W. 1989. RNA structure, not sequence, determines the 5' splice-site specificity of a group I intron. *Proc. Natl. Acad. Sci.* **86**: 7402–7406.
- Ferre-D'Amare, A.R. and Doudna, J.A. 1996. Use of *cis-* and *trans-*ribozymes to remove 5' and 3' heterogeneities from milligrams of in vitro transcribed RNA. *Nucleic Acids Res.* **24**: 977–978.
- . 1999. RNA folds: Insights from recent crystal structures. *Annu. Rev. Biophys. Biomol. Struct.* **28**: 57–73.
- . 2000. Crystallization and structure determination of a hepatitis delta virus ribozyme: Use of the RNA-binding protein U1A as a crystallization module. *J. Mol. Biol.* **295**: 541–556.
- Flor, P.J., Flanagan, J.B., and 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.
- Golden, B.L. and Cech, T.R. 1996. Conformational switches involved in orchestrating the successive steps of group I RNA splicing. *Biochemistry* **35**: 3754–3763.
- Golden, B.L., Gooding, A.R., Podell, E.R., and Cech, T.R. 1998. A preorganized active site in the crystal structure of the *Tetrahymena* ribozyme. *Science* **282**: 259–264.
- Green, R., Ellington, A.D., and Szostak, J.W. 1990. In vitro genetic analysis of the *Tetrahymena* self-splicing intron. *Nature* **347**: 406–408.
- Grosshans, C.A. and Cech, T.R. 1991. A hammerhead ribozyme allows synthesis of a new form of the *Tetrahymena* ribozyme homogeneous in length with a 3' end blocked for transesterification. *Nucleic Acids Res.* **19**: 3875–3880.
- Gutell, R.R., Cannone, J.J., Shang, Z., Du, Y., and Serra, M.J. 2000. A story: Unpaired adenosine bases in ribosomal RNAs. *J. Mol. Biol.* **304**: 335–354.
- Hampel, K.J. and Burke, J.M. 2003. Solvent protection of the hammerhead ribozyme in the ground state: Evidence for a cation-assisted conformational change leading to catalysis. *Biochemistry* **42**: 4421–4429.
- 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.
- Herschlag, D. and Cech, T.R. 1990. DNA cleavage catalysed by the ribozyme from *Tetrahymena*. *Nature* **344**: 405–409.
- Herschlag, D., Eckstein, F., and Cech, T.R. 1993. The importance of being ribose at the cleavage site in the *Tetrahymena* ribozyme reaction. *Biochemistry* **32**: 8312–8321.
- Höbartner, C. and Micura, R. 2004. Chemical synthesis of selenium-modified oligonucleotides and their enzymatic ligation leading to an U6 snRNA stem-loop segment. *J. Am. Chem. Soc.* **126**: 1141–1149.
- Juneau, K., Podell, E., Harrington, D.J., and Cech, T.R. 2001. Structural basis of the enhanced stability of a mutant ribozyme domain and a detailed view of RNA-solvent interactions. *Structure (Camb)* **9**: 221–231.
- Kitamura, A., Muto, Y., Watanabe, S., Kim, I., Ito, T., Nishiya, Y., Sakamoto, K., Ohtsuki, T., Kawai, G., Watanabe, K., et al. 2002. Solution structure of an RNA fragment with the P7/P9.0 region and the 3'-terminal guanosine of the tetrahymena group I intron. *RNA* **8**: 440–451.
- Klein, D.J., Schmeing, T.M., Moore, P.B., and Steitz, T.A. 2001. The kink-turn: A new RNA secondary structure motif. *EMBO J.* **20**: 4214–4221.
- Kleywegt, G.J., Zou, J.Y., Kjeldgaard, M., and Jones, T.A. 2001. Around O. In *International Tables for Crystallography* (eds. M.G. Rossman and E. Arnold), pp 353–356, 366–367. Kluwer Academic Publishers, The Netherlands.

- Kuo, L.Y. and Piccirilli, J.A. 2001. Leaving group stabilization by metal ion coordination and hydrogen bond donation is an evolutionarily conserved feature of group I introns. *Biochim. Biophys. Acta* **1522**: 158–166.
- Kuo, L.Y., Davidson, L.A., and Pico, S. 1999. Characterization of the *Azoarcus* ribozyme: Tight binding to guanosine and substrate by an unusually small group I ribozyme. *Biochim. Biophys. Acta* **1489**: 281–292.
- Latham, J.A. and Cech, T.R. 1989. Defining the inside and outside of a catalytic RNA molecule. *Science* **245**: 276–282.
- Lee, B. and Richards, F.M. 1971. The interpretation of protein structures: Estimation of static accessibility. *J. Mol. Biol.* **55**: 379–400.
- Leontis, N.B. and Westhof, E. 2001. Geometric nomenclature and classification of RNA base pairs. *RNA* **7**: 499–512.
- Mei, R. and Herschlag, D. 1996. Mechanistic investigations of a ribozyme derived from the *Tetrahymena* group I intron. Insights into catalysis and the second step of self-splicing. *Biochemistry* **35**: 5796–5809.
- Michel, F. and 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.
- Michel, F., Ellington, A.D., Couture, S., and Szostak, J.W. 1990. Phylogenetic and genetic evidence for base-triples in the catalytic domain of group I introns. *Nature* **347**: 578–580.
- Moran, S., Kierzek, R., and Turner, D.H. 1993. Binding of guanosine and 3' splice site analogues to a group I ribozyme: Interactions with functional groups of guanosine and with additional nucleotides. *Biochemistry* **32**: 5247–5256.
- Murphy, F.L. and Cech, T.R. 1993. An independently folding domain of RNA tertiary structure within the *Tetrahymena* ribozyme. *Biochemistry* **32**: 5291–5300.
- Murphy, F.L., Wang, Y.-H., Griffith, J.D., and Cech, T.R. 1994. Coaxially stacked RNA helices in the catalytic center of the *Tetrahymena* ribozyme. *Science* **265**: 1709–1712.
- Murshudov, G.N., Vagin, A.A., and Dose, E.J. 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr.* **D53**: 240–255.
- Nissen, P., Hansen, J., Ban, N., Moore, P., and Steitz, T. 2000. The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**: 920–930.
- Nissen, P., Ippolito, J.A., Ban, N., Moore, P.B., and Steitz, T.A. 2001. RNA tertiary interactions in the large ribosomal subunit: The A-minor motif. *Proc. Natl. Acad. Sci.* **98**: 4899–4903.
- Oubridge, C., Ito, N., Evans, P.R., Teo, C.-H., and Nagai, K. 1994. Crystal structure of 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature* **372**: 432–438.
- Pan, J. and Woodson, S.A. 1998. Folding intermediates of a self-splicing RNA: Mispairing of the catalytic core. *J. Mol. Biol.* **280**: 597–609.
- Perez-Salas, U.A., Rangan, P., Krueger, S., Briber, R.M., Thirumalai, D., and Woodson, S.A. 2004. Compaction of a Bacterial Group I Ribozyme Coincides with the Assembly of Core Helices. *Biochemistry* **43**: 1746–1753.
- Price, J.V. and Cech, T.R. 1988. Determinants of the 3' splice site for self-splicing of the *Tetrahymena* pre-rRNA. *Genes & Dev.* **2**: 1439–1447.
- Pyle, A.M. and Cech, T.R. 1991. Ribozyme recognition of RNA by tertiary interactions with specific ribose 2'-OH groups. *Nature* **350**: 628–631.
- Rangan, P. and Woodson, S.A. 2003. Structural requirement for Mg²⁺ binding in the group I intron core. *J. Mol. Biol.* **329**: 229–238.
- Rangan, P., Masquida, B., Westhof, E., and Woodson, S.A. 2003. Assembly of core helices and rapid tertiary folding of a small bacterial group I ribozyme. *Proc. Natl. Acad. Sci.* **100**: 1574–1579.
- . 2004. Architecture and folding mechanism of the *Azoarcus* group I pre-tRNA. *J. Mol. Biol.* **339**: 41–51.
- Reinhold-Hurek, B. and Shub, D.A. 1992. Self-splicing introns in tRNA genes of widely divergent bacteria. *Nature* **357**: 173–176.
- Rupert, P.B. and Ferre-D'Amare, A.R. 2001. Crystal structure of a hairpin ribozyme-inhibitor complex with implications for catalysis. *Nature* **410**: 780–786.
- Sclavi, B., Sullivan, M., Chance, M.R., Brenowitz, M., and Woodson, S.A. 1998. RNA folding at millisecond intervals by synchrotron hydroxyl radical footprinting. *Science* **279**: 1940–1943.
- Shan, S., Yoshida, A., Sun, S., Piccirilli, J.A., and Herschlag, D. 1999. Three metal ions at the active site of the *Tetrahymena* group I ribozyme. *Proc. Natl. Acad. Sci.* **96**: 12299–12304.
- Shan, S., Kravchuk, A.V., Piccirilli, J.A., and Herschlag, D. 2001. Defining the catalytic metal ion interactions in the *Tetrahymena* ribozyme reaction. *Biochemistry* **40**: 5161–5171.
- Silvian, L.F., Wang, J., Steitz, T.A. 1999. Insights into editing from an ile-tRNA synthetase structure with tRNA^{ile} and mupirocin. *Science* **285**: 1074–1077.
- Sjogren, A.S., Pettersson, E., Sjoberg, B.M., and Stromberg, R. 1997. Metal ion interaction with cosubstrate in self-splicing of group I introns. *Nucleic Acids Res.* **25**: 648–653.
- Soukup, J.K., Minakawa, N., Matsuda, A., and Strobel, S.A. 2002. Identification of A-minor tertiary interactions within a bacterial group I intron active site by 3-deazaadenosine interference mapping. *Biochemistry* **41**: 10426–10438.
- Staley, J.P. and Guthrie, C. 1998. Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell* **92**: 315–326.
- Strauss-Soukup, J. and Strobel, S. A. 2000. A chemical phylogeny of group I introns based upon interference mapping of a bacterial ribozyme. *J. Mol. Biol.* **302**: 339–358.
- Strobel, S.A. and 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.
- . 1995. Minor groove recognition of the conserved G-U pair at the *Tetrahymena* ribozyme reaction site. *Science* **267**: 675–679.
- Strobel, S.A. and Ortoleva-Donnelly, L. 1999. A hydrogen bonding triad stabilizes the chemical transition state of a group I ribozyme. *Chem. & Biol.* **6**: 153–156.
- Strobel, S.A., Ortoleva-Donnelly, L., Ryder, S.P., Cate, J.H., and Moncoeur, E. 1998. Complementary sets of noncanonical base pairs mediate RNA helix packing in the group I intron active site. *Nat. Struct. Biol.* **5**: 60–66.
- Strobel, S.A., Adams, P.L., Stahley, M.R., and Wang, J. 2004. RNA kink turns to the left and to the right. *RNA* (this issue).
- Szewczak, A.A., Ortoleva-Donnelly, L., Ryder, S.P., Moncoeur, E., and Strobel, S.A. 1998. A minor groove triple helix in the active site of the *Tetrahymena* group I intron. *Nat. Struct. Biol.* **5**: 1037–1042.
- Szewczak, A.A., Ortoleva-Donnelly, L., Zivarts, M.V., Oyelere, A.K., Kazantsev, A.V., and Strobel, S.A. 1999. An important base triple anchors the substrate helix recognition surface within the *Tetrahymena* ribozyme active site. *Proc. Natl. Acad. Sci.* **96**: 11183–11188.
- Szewczak, A.A., Kosek, A.B., Piccirilli, J.A., and Strobel, S.A. 2002. Identification of an active site ligand for a group I ribozyme catalytic metal ion. *Biochemistry* **41**: 2516–2525.
- Tanner, M.A. and Cech, T.R. 1996. Activity and thermostability of the small self-splicing group I intron in the pre-tRNA^{leu} of the purple bacterium *Azoarcus*. *RNA* **2**: 74–83.
- . 1997. Joining the two domains of a group I ribozyme to form the catalytic core. *Science* **275**: 847–849.
- Tanner, M.A., Anderson, E.M., Gutell, R.R., and Cech, T.R. 1997. Mutagenesis and comparative sequence analysis of a base triple joining the two domains of group I ribozymes. *RNA* **3**: 1037–1051.
- Waring, R.B., Towner, P., Minter, S.J., and Davies, R.W. 1986. Splice-site selection by a self-splicing RNA of *Tetrahymena*. *Nature* **321**: 133–139.
- Wimberly, B.T., Brodersen, D.E., Clemons Jr., W.M., Morgan-Warren, R.J., Carter, A.P., Vonnrhein, C., Hartsch, T., and Ramakrishnan, V. 2000. Structure of the 30S ribosomal subunit. *Nature* **407**: 327–339.
- Zarrinkar, P.P. and Williamson, J.R. 1994. Kinetic intermediates in RNA folding. *Science* **265**: 918–924.