The P4−P6 Domain Directs Higher Order Folding of the *Tetrahymena* Ribozyme Core†

Elizabeth A. Doherty and Jennifer A. Doudna*

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

Received September 26, 1996; Revised Manuscript Received December 10, 1996⊗

ABSTRACT: The active site of group I self-splicing introns occurs at the interface of two proposed structural domains. In the *Tetrahymena* intron, half of the catalytic core resides within the independently-folding P4−P6 domain while the other half belongs to a putative domain that includes helices P3, P7, P8, and P9 (P3−P9). To determine whether the P3−P9 region of the intron can also fold independently, we used Fe(II)-EDTA and dimethyl sulfate to probe the solvent accessibility of separate fragments of the *Tetrahymena* intron. These RNAs self-assemble into an active complex *in trans*, enabling analysis of their structural features both alone and within the complex. Our results show that while the P3−P9 region of the intron retains its secondary structure, most of the tertiary interactions within this region do not form stably in the absence of the P4−P6 domain. This indicates that the P4−P6 domain induces folding in the P3−P9 region, organizing the catalytic cleft between them. Thus the P4−P6 domain provides a scaffold for the folding of the *Tetrahymena* intron core.

Like proteins, large catalytically-active RNA molecules may be commonly composed of multiple independent substructures (Michel & Westhof, 1990). For example, circular permutation analysis (Pan & Zhong, 1994) and chemical probing (Pan, 1995; Loria & Pan, 1996) showed that the RNA component of eubacterial ribonuclease P is composed of two separately folded domains. Similarly, the active site of group I self splicing introns resides at the junction of two proposed domains. Comparative sequence analysis led to a model in which the universally conserved helical elements P4, P5, and P6 in one domain and P3, P7, P8, and P9 in the other form the catalytic core (Michel & Westhof, 1990). In a subset of introns including the well-studied example from *Tetrahymena thermophila*, this conserved catalytic core is buttressed by less-well conserved base-paired elements. Deletion of these peripheral regions leads to intron destabilization and loss of catalytic efficiency (Beaudry & Joyce, 1990; van der Horst et al., 1991; Laggerbauer et al., 1994). A key to understanding intron evolution and catalytic mechanism lies in determining which components of the core and its periphery are stable substructures, and how these regions interact to produce an active molecule.

Analysis of the *Tetrahymena* intron RNA using the solvent-based probe Fe(II)-EDTA revealed a magnesium-dependent, compact tertiary structure with a relatively solvent inaccessible interior (Latham & Cech, 1989). Mutagenesis, cross-linking, and chemical modification demonstrated that this structure involves multiple long-range interactions (Michel et al., 1990; Downs & Cech, 1990; Wang & Cech, 1992; Pyle et al., 1992; Jaeger et al., 1993; Wang et al., 1993). Approximately half of the catalytic core belongs to the independently folded P4−P6 domain (Murphy & Cech, 1993). In addition to helices P4, P5, and P6 of the core, the P4−P6 domain contains the less-well conserved helices P6a and P6b, and helices P5a, P5b, and P5c (P5abc) found only in the IC1 and IC2 subclasses of group I introns (Figure 1). The crystal structure of the P4−P6 domain (Cate et al., 1996) reveals that the P5abc region, itself a stable subdomain (Celander & Cech, 1991; Murphy & Cech, 1993), packs alongside the core helices. The P4−P6 domain folds at lower magnesium ion concentrations (Celander & Cech, 1991) and is more thermally stable than other parts of the intron (Banerjee et al., 1993). Furthermore, kinetic folding experiments that measured RNase H sensitivity of the intron in the presence of deoxyoligonucleotide probes (Zarrinkar & Williamson, 1994) showed that P4−P6 folds before other elements of the catalytic core. These experiments, however, did not address the independent stability of regions outside the P4−P6 domain.

Components of the splice site binding pocket and an essential triple-helical scaffold of the core are contained within P4−P6. During catalysis the 5′ splice site, located within the P1 helix, is positioned for attack by the 3′ hydroxyl of a guanosine co-substrate. The guanosine binding site and residues essential for positioning the P1 duplex are located within a separate structural region called P3−P9. Many of the most highly conserved nucleotides in the intron, including the nearly invariant P7 helix, are found in P3−P9 (Figure 1). In the *Tetrahymena* intron, the P3−P9 region also contains the less-well conserved P2, P2.1, P9.1, and P9.2 duplexes. Deletion experiments (Laggerbauer et al., 1994) suggest that the core helices P3 and P7 are stabilized by interactions with the P9.1 and P9.2 stems in a manner perhaps analogous to the buttressing of P4 and P6 by the P5abc extension in the P4−P6 domain. We wondered whether the P3−P9 region is likewise an independently folding structural

† Supported by the Lucille P. Markey Charitable Trust, the Donaghue Medical Research Foundation, and NIH Grant GM 22778-21. J.A.D. is a Lucille P. Markey Scholar in Biomedical Science, a Donaghue Medical Research Foundation Young Investigator, a Beckman Young Investigator, a Searle Scholar, and a David and Lucile Packard Fellow. E.A.D. is supported by NIH Training Grant GM 07223.

* To whom correspondence should be addressed. Phone: (203) 432-3108. FAX: (203) 432-9782. E-mail: doudna@csb.yale.edu.

⊗ Abstract published in *Advance ACS Abstracts,* March 1, 1997.

S0006-2960(96)02428-2 CCC: $14.00

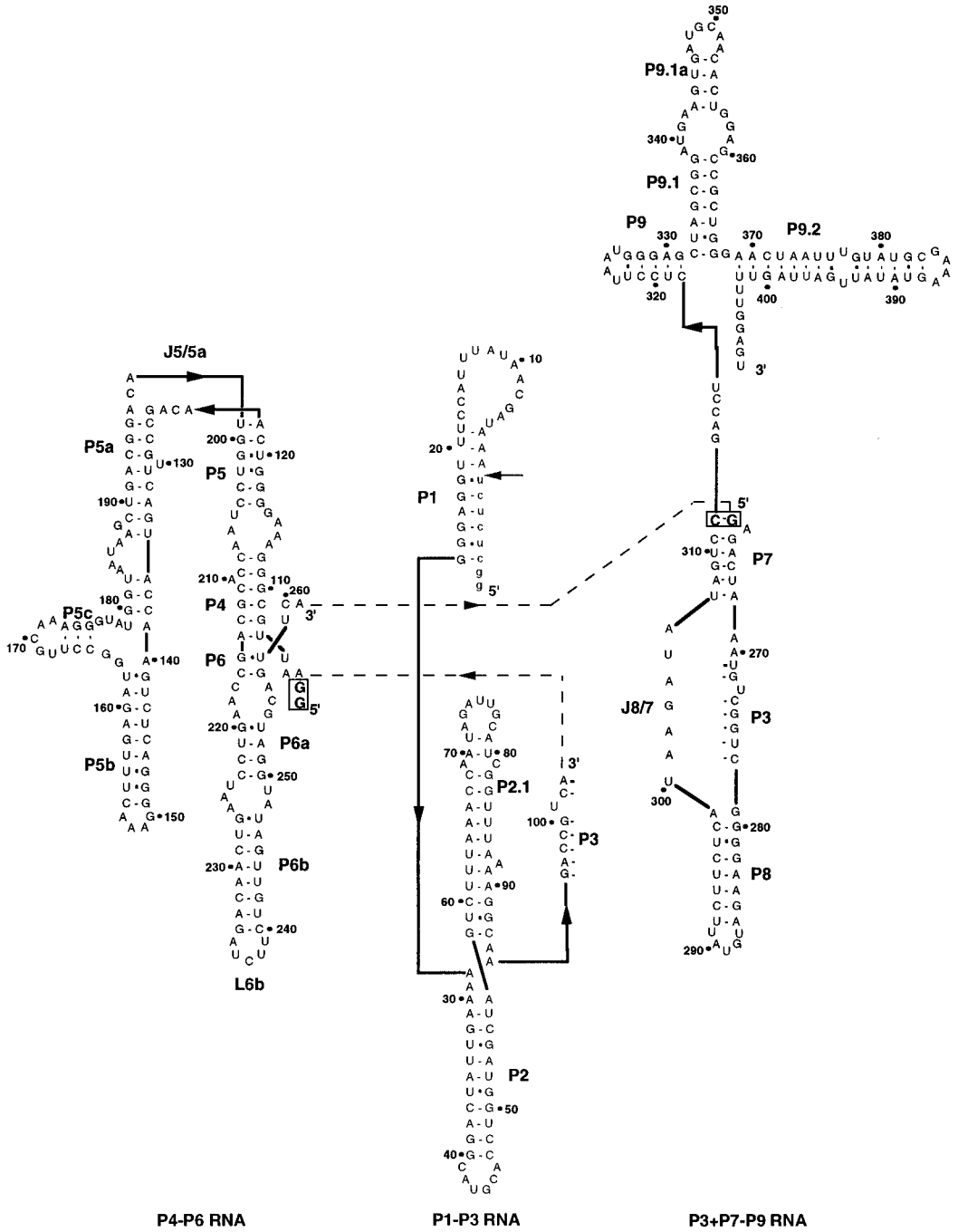

FIGURE 1: Secondary structure of the three-component *Tetrahymena* group I ribozyme (Cech et al., 1994). The numbering system is that used for the intact *Tetrahymena* intron. Solid lines show connectivities within each of the three RNAs while dashed lines show connectivities within the intact intron that were broken to produce the three-part system. Arrows superimposed on lines give 5′ and 3′ polarity. The 5′ exon sequence is shown in lower case letters and the 5′ splice site is marked with an arrow. Base-paired segments are designated "P", loops "L", and sections that join two paired regions "J". Boxed and bold nucleotides are changes from the wild-type sequence that were required to facilitate T7 transcription. (Doudna and Cech referred to P3+P7−P9 as P3−P9.)

domain within the intron or whether the P4−P6 domain influences its assembly at equilibrium.

To address this question, we took advantage of the fact that the *Tetrahymena* intron can be split into three separately transcribed RNAs which assemble primarily via tertiary interactions into an active complex (Doudna & Cech, 1995). This three-part ribozyme consists of the P4−P6 RNA and two other RNAs called P1−P3 and P3+P7−P9 (Figure 1). The P1−P3 and P3+P7−P9 segments contain all intron helices outside of the P4−P6 domain, with each contributing one strand of the P3 pseudoknot. Association of the P1−P3 and P3+P7−P9 RNAs reconstitutes P3 and thus the P3−P9 structural region (Figure 1). The three-part ribozyme provides a direct way to determine the level of independent tertiary structure outside the P4−P6 domain and how these three modules interact to form the catalytic core.

Chemical cleavage and modification with Fe(II)-EDTA[1] and dimethyl sulfate (DMS), respectively, were used to determine the level of tertiary structure in the P1−P3 and P3+P7−P9 RNAs alone, in complex with each other and

[1] Abbreviations: dNTP, deoxynucleotide triphosphate; ddNTP, dideoxynucleotide triphosphate; DTT, dithiothreitol; EDTA; ethylenediaminetetraacetic acid; PCR, polymerase chain reaction.

within the three-part ribozyme. Fe(II)-EDTA cleaves solvent accessible phosphoribose backbone groups in both single stranded and double stranded RNA, allowing detection of tertiary folding that protects specific areas of the backbone from solution (Latham & Cech, 1989). Our results indicate that a complex of P1−P3 and P3+P7−P9 has its expected secondary structure but an incomplete tertiary structure. Strikingly, most of the tertiary interactions that occur within the universally conserved structures of the P3−P9 region are induced in the presence of the P4−P6 domain. These results suggest that P4−P6 provides a scaffold to arrange the core secondary structure into its proper three-dimensional architecture.

MATERIALS AND METHODS

Materials. T7 RNA polymerase was prepared as described (Doudna & Cech, 1995).

Plasmid Template Construction. pP3P9, used to prepare the P3+P7−P9 RNA, has been described elsewhere (Doudna & Cech, 1995). Plasmid pP1P3 was constructed from pTZΔ12IVS (Woodson & Cech, 1991) by PCR amplification of the DNA sequence corresponding to positions −6 to 103. PCR primers incorporated the two G's at positions −7 and −8, a T7 RNA polymerase site, a *Bsa*I site following position 103, and *Eco*RI and *Bam*HI sites for ligation into pUC19. pP3P9.0 was constructed from pP3P9 by an analogous procedure and contains the intron sequence from 262 to 414 with a *Bsa*I site after G414. This plasmid was originally constructed to test the effect of adding nucleotides 410−414 on the structure of the three-part ribozyme. It was used here to prepare DNA for a 3′-extended P3+P7−P9 construct. pP1P3P9 was constructed from pP1P3 and pP3P9 by three PCRs (Ho et al., 1989) followed by ligation into pUC19. The sequence from position 262 to 409 was amplified with an overhanging 3′ primer containing the sequence from 410 to 414 and from 16 to 30. The sequence from 16 to 103 was amplified similarly with a 5′ overhanging primer containing the sequence from 414 to 395. These partially complementary PCR products were used as templates for a third PCR to yield a fully extended DNA sequence for ligation. Dideoxy sequencing was performed on all plasmids to confirm the correctness of the insert sequences.

PCR Template Construction. The 5′-ext P3+P7−P9 PCR template was prepared from pP2P9 (which contains the wild-type intron sequence from 28 to 414). DNA templates for 3′ and 5′ extended P1−P3 RNAs and for a P4−P6 RNA with a shortened 5′ end were amplified from pP1P3 and pP4P6 using appropriate primers.

RNA Preparation. Plasmids and PCR templates were digested with *Sca*I (pP3P9 and 5′-ext P3+P7−P9 DNA), *Hin*dIII (to create the 3′-extended P3+P7−P9 RNA), or with *Bsa*I (all others). Transcription with T7 RNA polymerase and gel purification were performed as described (Latham et al., 1990). P4−P6 and P4−P6Δ24 RNAs were obtained from Kaihong Zhou (Cate et al., 1996). Following treatment with calf alkaline phosphatase, RNAs were labeled at their 5′ or 3′ ends with ^{32}P (Latham et al., 1990; England et al., 1980).

Fe(II)-EDTA Assays. Fe(II)-EDTA assays were performed essentially as described (Latham & Cech, 1989). 25−50 nM of 5′ or 3′ end-labeled RNA was incubated at 40 °C for 15 min in an 8 μL volume containing 37.5 mM Tris-HCl, pH 7.5, 100 mM $MgCl_2$, 6.25 mM spermidine, and 1 μg of yeast tRNA^{Phe}. In order to form complexes between the P1−P3, P4−P6, and P3+P7−P9 RNAs in solution, 1−3 μM additional unlabeled RNA was added to some reactions. Generally, 1 μM of each additional RNA component was sufficient to bind more than 90% of the labeled RNA as determined by monitoring the degree of resulting Fe(II)-EDTA protections or 5′ splicing activity at increasing concentrations of the unlabeled RNA. The purpose of the tRNA^{Phe} was to standardize the level of cleavage between reactions with and without additional RNAs. Cleavage was initiated by adding 1 μL each of freshly prepared 50 mM DTT and Fe(II)-EDTA solution [10 mM $(NH_4)_2Fe(SO_4)_2$ and 20 mM Na_2EDTA, pH 8.0]. Buffer conditions during the reaction were 30 mM Tris-HCl, pH 7.5, 80 mM $MgCl_2$, 5 mM spermidine, 5 mM DTT, 2 mM Na_2EDTA, and 1 mM $(NH_4)_2Fe(SO_4)_2$. After 80 min of incubation at 40 °C, reactions were quenched with 10 mM thiourea. An equal volume of gel loading buffer was added, and denaturing polyacrylamide gels were run immediately. Control reactions were treated exactly as above except that no $(NH_4)_2Fe(SO_4)_2$ was added.

Quantitation of Fe(II)-EDTA Protection. Individual bands were quantitated using a phosphorimager (Fuji) following subtraction of background intensity and cleavage in the starting material and correction for differences in lane loading. The extent of protection at each backbone position was calculated as the intensity of the corresponding band under unfolding conditions divided by the intensity under folding conditions. Results from at least three independent experiments were averaged. The average extent of protection calculated for non-protected bands was 1.1−1.2 with a maximum standard deviation of 0.3. Backbone positions were considered to be protected if their corresponding extents of protection were at least 1.5.

DMS Assays and Primer Extension. DMS modification was performed essentially as described (Inoue & Cech, 1985; Moazed et al., 1986). 50 nM RNA was pre-incubated at 40 °C for 15 min in 200 μL of 30 mM Tris-HCl, pH 7.5, 80 mM $MgCl_2$, and 5 mM spermidine. To form complexes between RNAs, 1−2 μM concentrations of other components of the three-piece ribozyme were added to some reactions. In order to compare the amount of modification between reactions with and without additional RNAs, the total concentration of RNA in each reaction was maintained at approximately 150 ng/μL by addition of yeast tRNA^{Phe}. 4 μL of a 1:4 DMS:ethanol solution was added to each reaction after pre-incubation. Following 30 min of incubation at 40 °C, reactions were quenched with 75 μL of 1 M β-mercaptoethanol/1.5 M sodium acetate and then ethanol precipitated. Reactions were resuspended in 300 mM sodium acetate, phenol−chloroform extracted, and re-ethanol precipitated. For primer extension, 1 pmol of DMS-modified RNA was annealed to 2 pmol of a complementary DNA primer by slow-cooling from 90 °C in 50 mM Tris-HCl, pH 8.3, 60 mM NaCl, and 10 mM DTT. 70 μM of each dNTP, 3 mM $MgCl_2$, and 25 units of AMV reverse transcriptase (Boehringer Mannheim) were added and extension reactions were incubated at 42 °C for 45 min. Sequencing lanes included 14 μM of ddNTP. Reverse transcription was stopped by addition of an equal volume of gel loading buffer. Samples were heated to 90 °C for 5 min before loading onto denaturing polyacrylamide gels.

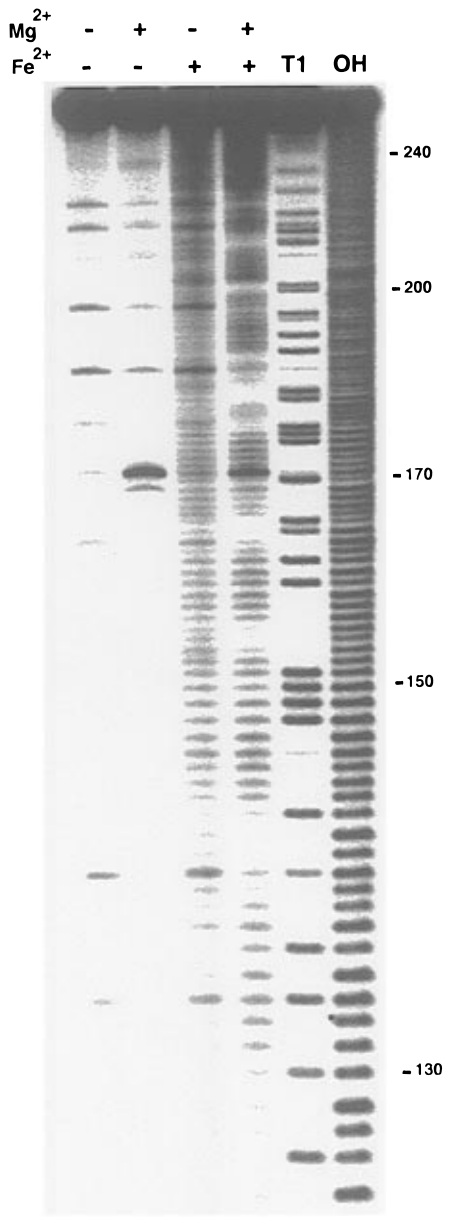

FIGURE 2: Fe(II)-EDTA cleavage of the P4−P6 RNA monitored by gel electrophoresis. Backbone positions are marked alongside the figure and are mapped relative to an RNase T1 ladder (which cleaves 3′ to G residues) and a hydrolysis ladder (OH). Fe(II)-EDTA bands are shifted up one nucleotide relative to the T1 ladder due to the different 3′ ends generated by these cleavage reactions. The error in mapping backbone positions is plus or minus one nucleotide. Lanes without Fe^{2+} are controls to show background cleavage in the starting material. 10 mM $MgCl_2$ is included where indicated.

RESULTS

Fe(II)-EDTA Analysis of the P4−P6, P1−P3, and P3+P7−P9 RNAs. To determine the level of higher order structure present in each component of the three-piece ribozyme, the individual RNAs (Figure 1) were radiolabeled at their 5′ or 3′ ends and reacted with Fe(II)-EDTA under denaturing (no Mg^{2+}) and native (10−80 mM $MgCl_2$ and 0−5 mM spermidine) conditions. The pattern of Fe(II)-EDTA cleavage at phosphoribose backbone positions at least 10 nucleotides from the 5′ and 3′ ends of the RNAs was analyzed by gel electrophoresis of the reaction products (Figures 2 and 3). Individual band intensities from at least three independent experiments were quantitated to determine the extent of protection at each backbone position (see Materials and Methods).

As shown previously by Murphy and Cech, the P4−P6 RNA is unfolded in the absence of $MgCl_2$ and a fairly uniform cleavage pattern is observed, while addition of 10 mM $MgCl_2$ leads to folding of the RNA as revealed by protection from cleavage in several areas of the backbone (Figure 2; Murphy & Cech, 1993). In contrast to P4−P6, Fe(II)-EDTA cleavage of the P1−P3 and P3+P7−P9 RNAs remains uniform, even at concentrations of magnesium ion and spermidine that support optimal catalytic activity of the three-part ribozyme (80 mM $MgCl_2$ and 5 mM spermidine, respectively; Figures 3A and 3B). However, digestion of P1−P3 and P3+P7−P9 RNAs with V1 and T1 RNases (which cleave in double- and single-stranded regions respectively) shows that these RNAs form all expected secondary structure under these conditions, though the P7 stem may be unstable as its backbone is cleaved by both enzymes (data not shown).

In order to determine the degree of independent higher order structure within the P3−P9 region, we performed Fe(II)-EDTA cleavage and RNase digestion on a complex of the P1−P3 and P3+P7−P9 RNAs (Figure 1), formed by incubating each radiolabeled RNA with an unlabeled excess of the other. Addition of a saturating concentration of unlabeled P3+P7−P9 RNA (1 μM; see Materials and Methods) to radiolabeled P1−P3 RNA leads to an average of 2-fold protection from Fe(II)-EDTA cleavage at positions 71−72 and 80−81 in the loop of the P2.1 stem (L2.1), 57−60 at the junction of P2 and P2.1 (J2/2.1), and 32−34 in P2 (Figure 3A). When excess unlabeled P1−P3 RNA is added to radiolabeled P3+P7−P9 RNA, protection is observed at positions 278−282, 298−299, and 301 near the junction of the P3 and P8 stems and at 343−351 and 360 in P9.1 and P9.1a (Figure 3B). V1 and T1 RNase digestion shows that the P3 stem and all other expected secondary structure is formed in the P1−P3/P3+P7−P9 complex, although the P7 helix may remain unstable since it is cleaved by both enzymes (not shown). The Fe(II)-EDTA protections observed in the P1−P3/P3+P7−P9 complex cluster in two areas: in or near the loops of P9.1a and P2.1 (71−72, 80−81, 343−351, and 360), which are proposed to interact within the intact *Tetrahymena* intron (Banerjee et al., 1993), and at the junction of helices P2, P2.1, P3, and P8 (32−34, 57−60, 278−282, 298−299, and 301), which may be brought close together by formation of the P3 pseudoknot. These protections are a subset of those observed in the P1−P3 and P3+P7−P9 segments of the intact intron (Latham & Cech, 1989), demonstrating that the P3−P9 region maintains some independent, native higher order structure.

To examine the influence of the P4−P6 domain on the structure of the P3−P9 region, we determined the Fe(II)-EDTA cleavage pattern of the P1−P3 and P3+P7−P9 RNAs within the complete three-part ribozyme. We assembled the three-part ribozyme by adding saturating concentrations (1 μM; see Materials and Methods) of unlabeled P4−P6 and P3+P7−P9 RNAs to radiolabeled P1−P3 RNA. This resulted in new protection in the P1−P3 RNA at positions 22−27 in P1 and at 41 and 46−48 at the base of the P2 loop (Figure 3A). Positions 41 and 46−48 in P2 are also protected from cleavage when a large excess of P4−P6 (3 μM) is added to P1−P3 alone, implying a direct interaction between P2 and P4−P6 (not shown). When excess (1 μM)

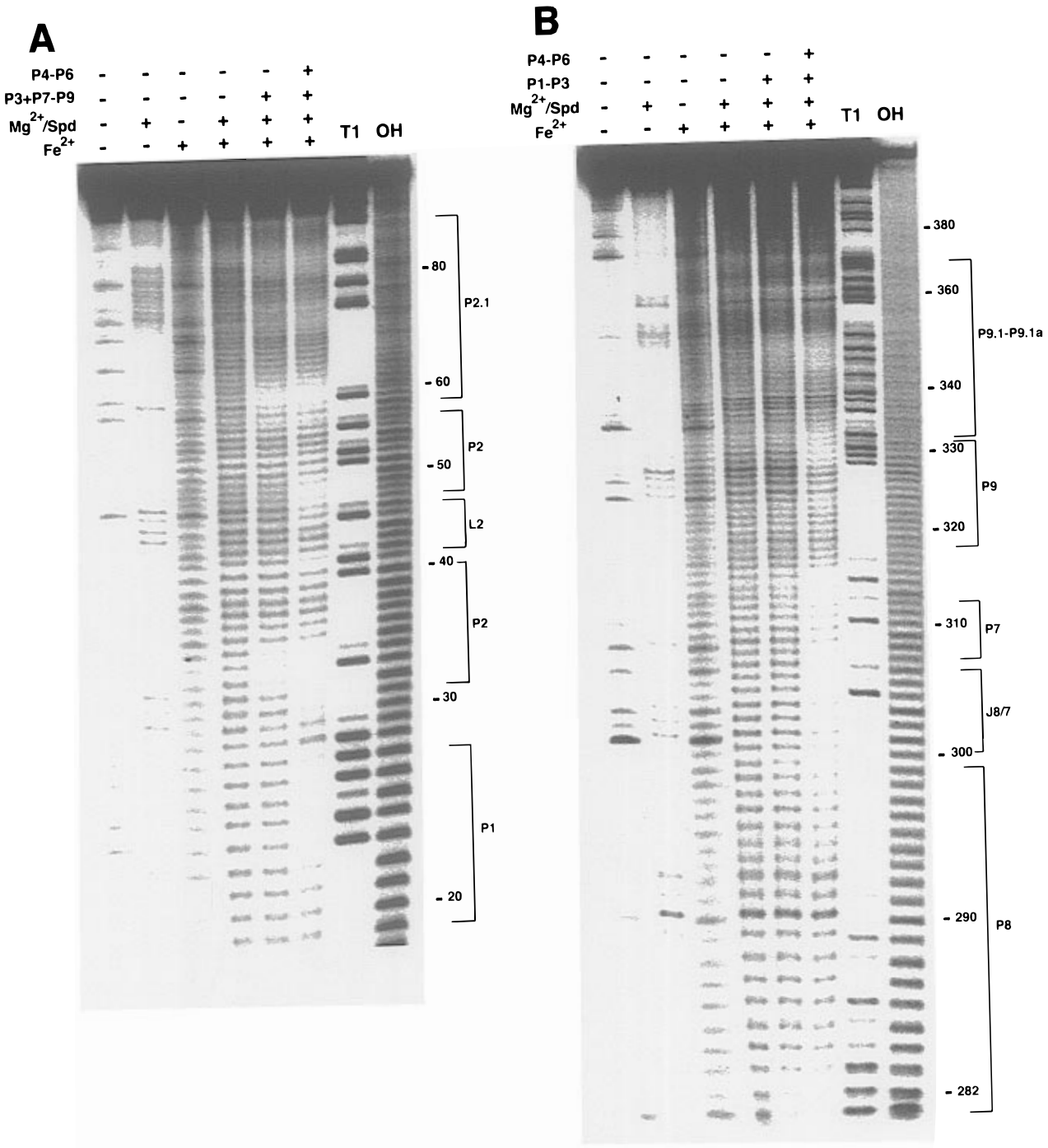

FIGURE 3: (A) Fe(II)-EDTA cleavage of the P1−P3 RNA. Lanes are similar to those in Figure 2. 80 mM $MgCl_2$ and 5 mM spermidine, 1 μM P3+P7−P9, and 1 μM P4−P6 are added where indicated. (B) Fe(II)-EDTA cleavage of the P3+P7−P9 RNA. 80 mM $MgCl_2$ and 5 mM spermidine, 1 μM P1−P3, and 1 μM P4−P6 are added where indicated.

unlabeled P4−P6 and P1−P3 RNAs are added to radiolabeled P3+P7−P9, new protection is observed at positions 300 and 302−315 in J8/7 and P7 and at 326−331 in P9 (Figure 3B). No protection is observed in the P3+P7−P9 RNA in the presence of excess (3 μM) P4−P6 but in the absence of P1−P3 (not shown). All areas of the backbone protected in the P1−P3/P3+P7−P9 complex are protected at slightly higher levels within the three-piece ribozyme (Figures 3A and 3B). In summary, the cleavage pattern of the P1−P3 and P3+P7−P9 RNAs within the three-part ribozyme closely resembles that of these segments within the intact *Tetrahymena* intron (Latham & Cech, 1989), indicating that the P4−P6 domain induces the assembly of a correctly organized catalytic core.

Fe(II)-EDTA Analysis of the 5′ and 3′ Ends of P1−P3 and P3+P7−P9. An important set of tertiary interactions within the P3−P9 region is proposed to occur between the P9.1−P9.2 helices and one face of the P7 and P3 stems (Laggerbauer et al., 1994). Specifically, it was observed that deletion of the P9.1−P9.2 extension within the intact *Tetrahymena* intron leads to a loss of Fe(II)-EDTA protection between positions 263 and 272 (Laggerbauer et al., 1994). The Fe(II)-EDTA cleavage pattern at all other positions in the deletion mutant is identical to that of the wild-type intron (Laggerbauer et al., 1994), suggesting that the P9.1−P9.2 extension alone controls the solvent accessibility of positions 263−272. However, we were unable to analyze the Fe(II)-EDTA cleavage pattern of these positions since they

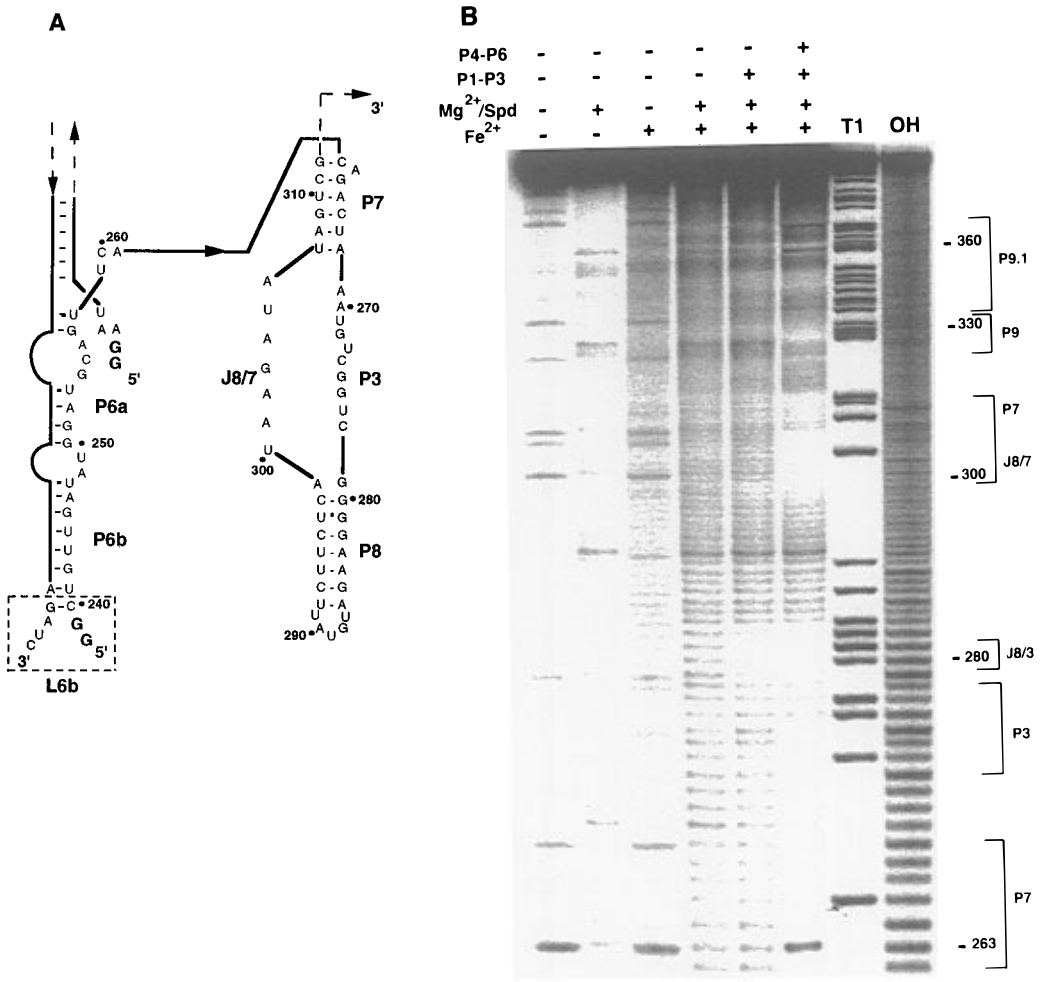

FIGURE 4: Fe(II)-EDTA analysis of the 5′-ext P3+P7−P9 RNA. (A) Secondary structure of the RNA and its interaction with the P4−P6Δ24 RNA. The L6b loop is boxed and mutated nucleotides are shown in bold. U239 and U238 were mutated to G to facilitate T7 transcription. (B) Fe(II)-EDTA analysis of the 5′-ext P3+P7−P9 RNA. Lanes without Fe^{2+} show background cleavage in the starting material. 80 mM $MgCl_2$ and 5 mM spermidine, 1 μM P1−P3, and 2 μM P4−P6Δ24 are included where indicated.

are located close to the 5′ end of the P3+P7−P9 RNA. To address this problem we designed a variant of the P3+P7−P9 RNA called 5′-ext P3+P7−P9 (Figure 4A) in which the 5′ end is lengthened by 24 nucleotides and relocated from P7 to the L6b loop in the P4−P6 domain. The single-stranded 5′ extension does not affect association with the P1−P3 RNA since the backbone positions of the 5′-ext P3+P7−P9 RNA are protected to the same degree with the same dependence on concentration of P1−P3 as those of the P3+P7−P9 RNA. To re-form a three-piece ribozyme, we complexed 5′-ext P3+P7−P9 with P1−P3 and a shortened version of P4−P6 called P4−P6Δ24 which has its 3′ end at L6b (Figure 4A). This complex was active in 5′ splicing (not shown) and was not expected to be grossly perturbed in structure since disruption of L6b does not affect activity of the *Tetrahymena* ribozyme (Price et al., 1985).

While its secondary structure is intact as assayed by RNase V1 and T1 digestion, the 5′-ext P3+P7−P9 RNA is not protected from Fe(II)-EDTA cleavage at 80 mM $MgCl_2$ and 5 mM spermidine (Figure 4B). In the presence of a saturating concentration of unlabeled P1−P3 RNA, 2-fold protection from cleavage in 5′-ext P3+P7−P9 is observed at positions 269−272 between P3 and P7 (Figure 4B). When an unlabeled excess of both P1−P3 and P4−P6Δ24 is added, protection at positions 269−272 increases dramatically from 2-fold to about 8-fold while additional protection is observed at 262−268 in P7 and 273−275 in P3 (Figure 4B). These results suggest that the P4−P6 domain is required to form

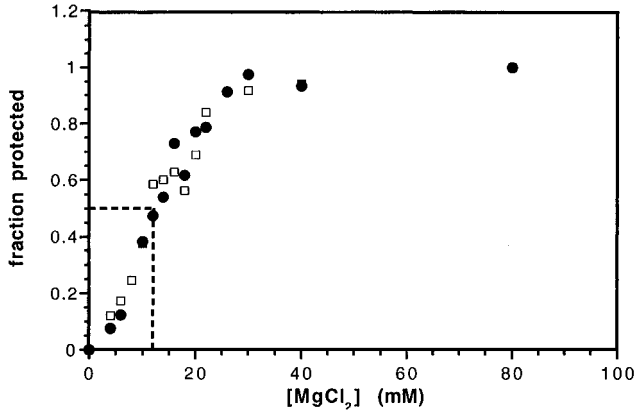

FIGURE 5: Normalized Fe(II)-EDTA protection in the P1−P3/P3+P7−P9 complex versus concentration of $MgCl_2$. Open squares, positions 346−348 in the P3+P7−P9 RNA; filled circles, positions 57−60 in the P1−P3 RNA. Four independent experiments such as that shown were run for each set of positions. The average extent of protection at 80 mM $MgCl_2$ observed for positions 346−348 is 2.0 ± 0.3 and for positions 57−60 is 2.7 ± 0.5. Hill plots of the data (not shown) were fit as described (Celander & Cech, 1991) and for positions 346−348 give an average *x*-intercept of 12 ± 1 mM $MgCl_2$ (for 50% protection) and an average slope of 2.7 ± 0.5. For positions 57−60, the average Hill plot *x*-intercept is 13 ± 1 mM $MgCl_2$ and the average slope is 2.7 ± 0.5.

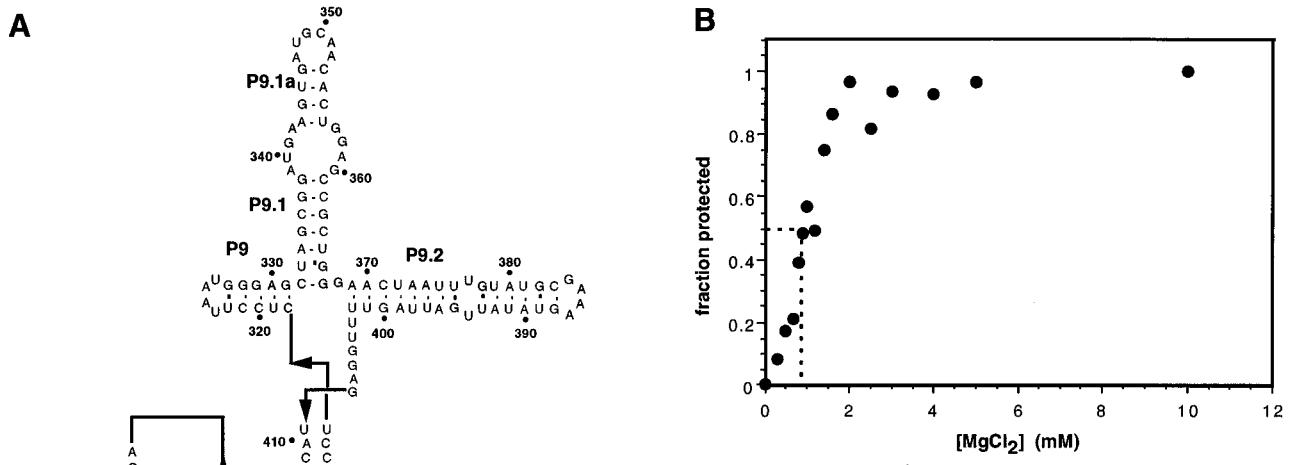

FIGURE 6: Fe(II)-EDTA protection of the P1−P3−P9 RNA versus concentration of $MgCl_2$. (A) Secondary structure of the P1−P3−P9 RNA. This RNA is a unimolecular version of the P1−P3/P3+P7−P9 complex in which the 5′ end of the P1−P3 sequence is covalently linked to the 3′ end of the P3+P7−P9 sequence between positions 16 and 414. Unlike the P1−P3/P3+P7−P9 complex, the P1 stem is not paired and the P9.0 stem is capable of forming. (B) Normalized Fe(II)-EDTA protection at positions 346−348 of P1−P3−P9 versus concentration of $MgCl_2$. The concentration of EDTA was reduced to 1.1 mM (Celander & Cech, 1991) for these experiments to more accurately determine the dependence of folding on $[Mg^{2+}]$. Four independent experiments were performed of which one is shown. The average extent of protection at these positions at 10 mM $MgCl_2$ is 2.0 ± 0.4. Hill plots of the data (Celander & Cech, 1991) give an average *x*-intercept of 1.0 mM $MgCl_2$ ± 0.1 and an average slope of 2.3 ± 0.5.

or significantly stabilize tertiary contacts involving the P3 and P7 stems. While we cannot rule out the possibility that the single-stranded extension interfered with these interactions until paired within the P4−P6 domain, this seems unlikely since the extension did not interfere with formation of the complex between 5′-ext P3+P7−P9 and P1−P3.

A similar set of RNAs was constructed to determine the protection pattern at the 3′ end of the P1−P3 RNA. This end was lengthened by 21 nucleotides, relocating it to position 124 in J5/5a (Figure 1) while the 5′ end of the P4−P6 RNA was shortened accordingly. In this case, however, the extension strongly interfered with formation of the P1−P3/P3+P7−P9 complex as no protection of the modified P1−P3 RNA was observed in the presence of as much as 10 μM unlabeled P3+P7−P9 (not shown). To determine the pattern of cleavage near the 5′ end of P1−P3 and the 3′ end of P3+P7−P9, long exon sequences were extended from these ends. The 5′ splicing activity and the association of the three-part complex were not grossly affected by these modifications, and the pattern of Fe(II)-EDTA protection of each of these RNAs was similar to that of the unmodified P1−P3 and P3+P7−P9 RNAs. However, no additional backbone protection was detected using these RNAs (not shown).

Figure 8 summarizes the Fe(II)-EDTA cleavage data of Figures 3 and 4, highlighting all of the phosphoribose backbone positions protected from Fe(II)-EDTA cleavage in P1−P3/P3+P7−P9 complexes in the presence and absence of the P4−P6 RNA.

Stability of Tertiary Structure in the P1−P3/P3+P7−P9 Complex. The Fe(II)-EDTA data suggest that under conditions optimal for activity of the three-component ribozyme (80 mM $MgCl_2$ and 5 mM spermidine; Doudna & Cech, 1995), the P1−P3/P3+P7−P9 complex maintains some independent tertiary structure. We determined the dependence of folding of the P1−P3/P3+P7−P9 complex on magnesium ion concentration by analyzing protection from Fe(II)-EDTA cleavage at variable $MgCl_2$ concentration. In contrast to the higher cation concentrations required for maximum splicing activity, protection of positions in J2/2.1 and L9.1a in the P1−P3/P3+P7−P9 complex is half-complete at about 13 mM $MgCl_2$ and complete at about 30

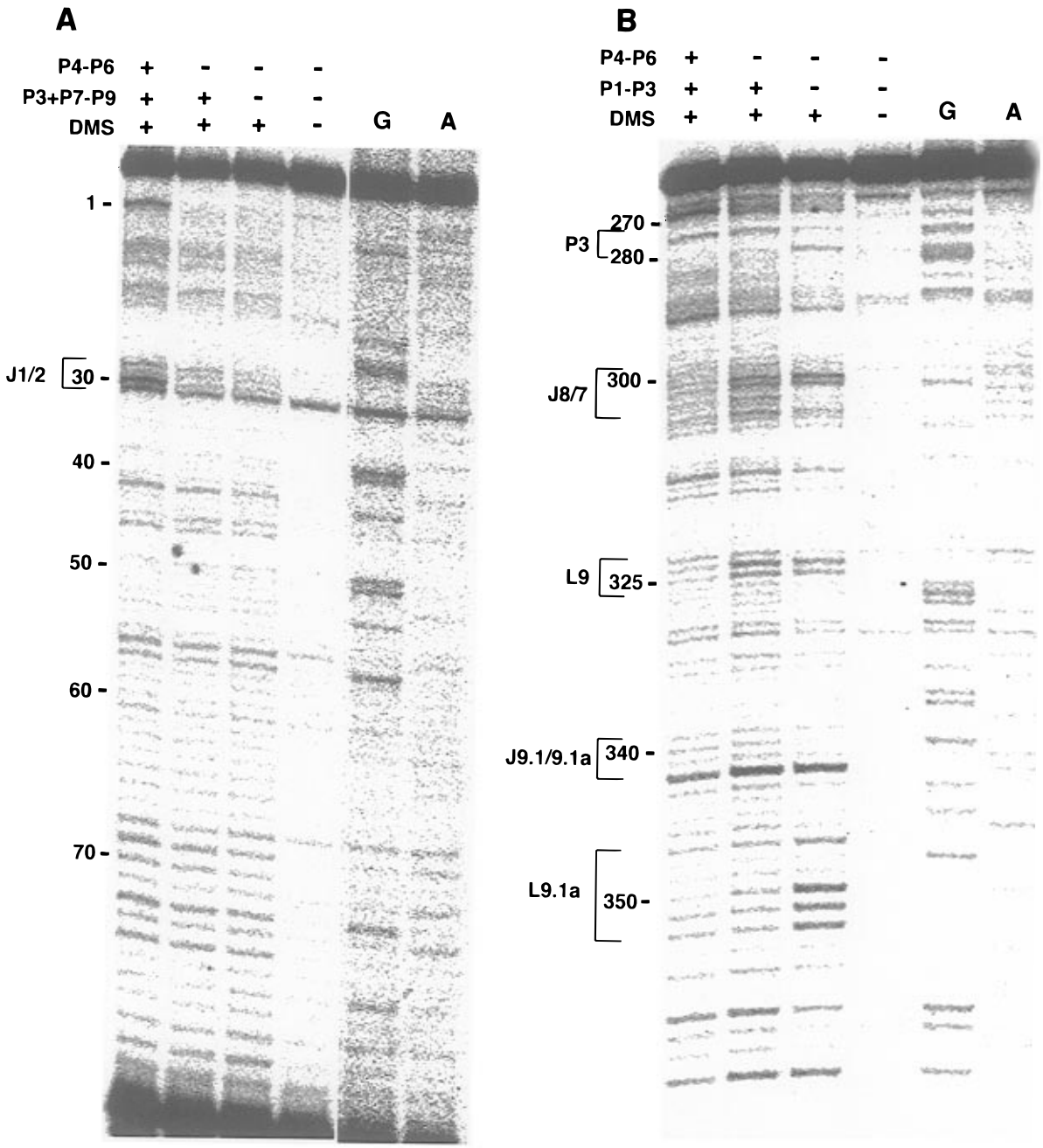

FIGURE 7: DMS modification of the P1−P3 and P3+P7−P9 RNAs. (A) DMS modification of the P1−P3 RNA detected by primer extension. All lanes contain 50 nM P1−P3 RNA. 1 μM P3+P7−P9 and 1 μM P4−P6 are included where indicated. All DMS reactions were performed at 80 mM $MgCl_2$ and 5 mM spermidine. Band positions were mapped by comparison to dideoxy sequencing ladders. Due to primer annealing, modification was detectable only from position 80 to position 1. The strong reverse transcription stop at position 1 is due to 5′ splicing activity during the DMS reaction. (B) DMS modification of the P3+P7−P9 RNA. Reaction conditions are similar to those in Figure 6A. 1 μM P1−P3 and 1 μM P4−P6 are added where indicated. Modification was detectable from nucleotides 380 to 265.

mM $MgCl_2$ (Figure 5). Furthermore, the P1−P3/P3+P7−P9 complex folds cooperatively, with a Hill coefficient of about 2.7, suggesting that binding of at least two magnesium ions accompanies the process. These results demonstrate that high magnesium ion concentrations and addition of polycations cannot substitute for RNA−RNA interactions provided by the P4−P6 domain in completing the folding of the P1−P3/P3+P7−P9 complex.

The $MgCl_2$ requirement for 50% folding of the P1−P3/P3+P7−P9 complex (Figure 5) is about 10-fold higher than that of the P4−P6 domain and the intact *Tetrahymena* ribozyme, which both require approximately 1 mM $MgCl_2$ for 50% folding (Celander & Cech, 1991; Murphy & Cech, 1993). Nevertheless, this may reflect the fact that the complex assembles *in trans*. We tested the stability of tertiary contacts in a unimolecular context by constructing an RNA called P1−P3−P9 (Figure 6A) which covalently joins the P1−P3 and P3+P7−P9 RNAs at the natural cyclization site of the *Tetrahymena* intron. This RNA is capable of reverse cyclization in the presence of P4−P6 and has a similar RNase digestion pattern and a similar Fe(II)-EDTA protection pattern as the P1−P3/P3+P7−P9 complex, but with a lower extent of Fe(II)-EDTA protection in several areas (not shown). Positions in the P9.1a loop of the P1−P3−P9 RNA acquire protection cooperatively (Hill coefficient of about 2.3) and at low concentrations of magnesium ion (50% protection at 1 mM $MgCl_2$; Figure 6B), indicating that tertiary interactions involving this loop are as stable as

those in the P4−P6 domain.

DMS Modification of the P1−P3 and P3+P7−P9 RNAs. To complement the Fe(II)-EDTA studies, DMS was used as a second probe of higher order structure of the P1−P3 and P3+P7−P9 RNAs. DMS is a small molecule which methylates non-Watson−Crick paired adenines and cytosines at the N1 and N3 positions, respectively. DMS modification depends on the solvent accessibility of these groups and is detectable by primer extension (Inoue & Cech, 1985; Moazed et al., 1986). Only A and C residues in the loop regions of the P1−P3 RNA are modified appreciably at 80 mM $MgCl_2$ and 5 mM spermidine, confirming that the secondary structure of this RNA is intact (Figure 7A). There is no change in the DMS modification pattern of P1−P3 upon addition of a saturating concentration of P3+P7−P9 (Figure 7A). However, in the presence of an excess of both P4−P6 and P3+P7−P9, A30, one of several adenines proposed to be critical for correct docking of the P1 stem into the catalytic core (Young et al., 1991), becomes heavily modified (Figure 7A). There is strong DMS modification of several bases in the single-stranded regions of the P3+P7−P9 RNA, including the P3 strand (Figure 7B). Some of these bases become protected from methylation upon addition of excess P1−P3 (Figure 7B). Protection of C278 suggests that the P3 stem is formed in the P1−P3/P3+P7−P9 complex, while dramatic reduction in methylation of four bases in L9.1a supports the involvement of this loop in tertiary interactions. Upon addition of both P4−P6 and P1−P3, bases in L9 and in J8/7 become protected from DMS modification just as the corresponding backbone positions are protected from Fe(II)-EDTA cleavage (Figure 7B). Finally, formation of the complete three-piece ribozyme leads to a large increase in the modification of A269 and a decrease in modification of A270 (Figure 7B), further suggesting that the P4−P6 domain mediates a structural change in the P3−P7 region of the core. The DMS results are summarized schematically in Figure 8.

DISCUSSION

We have examined the degree to which the P3−P9 region of the *Tetrahymena thermophila* group I intron folds independently. P3−P9 is one of two large structural elements which interact to form the active site for self splicing. The other is the independently folded P4−P6 domain, for which a crystal structure has recently been determined (Cate et al., 1996). The P3−P9 region was formed in solution as a complex of the P1−P3 and P3+P7−P9 RNAs. Our results show that each of these RNAs alone retains its expected secondary structure but has no detectable tertiary folding. A P1−P3/P3+P7−P9 complex contains some long-range tertiary interactions, but its complete folding is achieved only in the presence of the P4−P6 domain (Figures 3−8). We conclude that P4−P6 is the only independently folded domain in the *Tetrahymena* ribozyme, as discussed below.

The P4−P6 domain was shown to be an independently folded region of tertiary structure primarily by Fe(II)-EDTA analysis (Murphy & Cech, 1993). There is excellent correlation between the backbone solvent accessibility determined by examination of the crystal structure and by Fe(II)-EDTA (Cate et al., 1996). The P4−P6 domain is characterized by a sharp bend that allows the P5abc helices

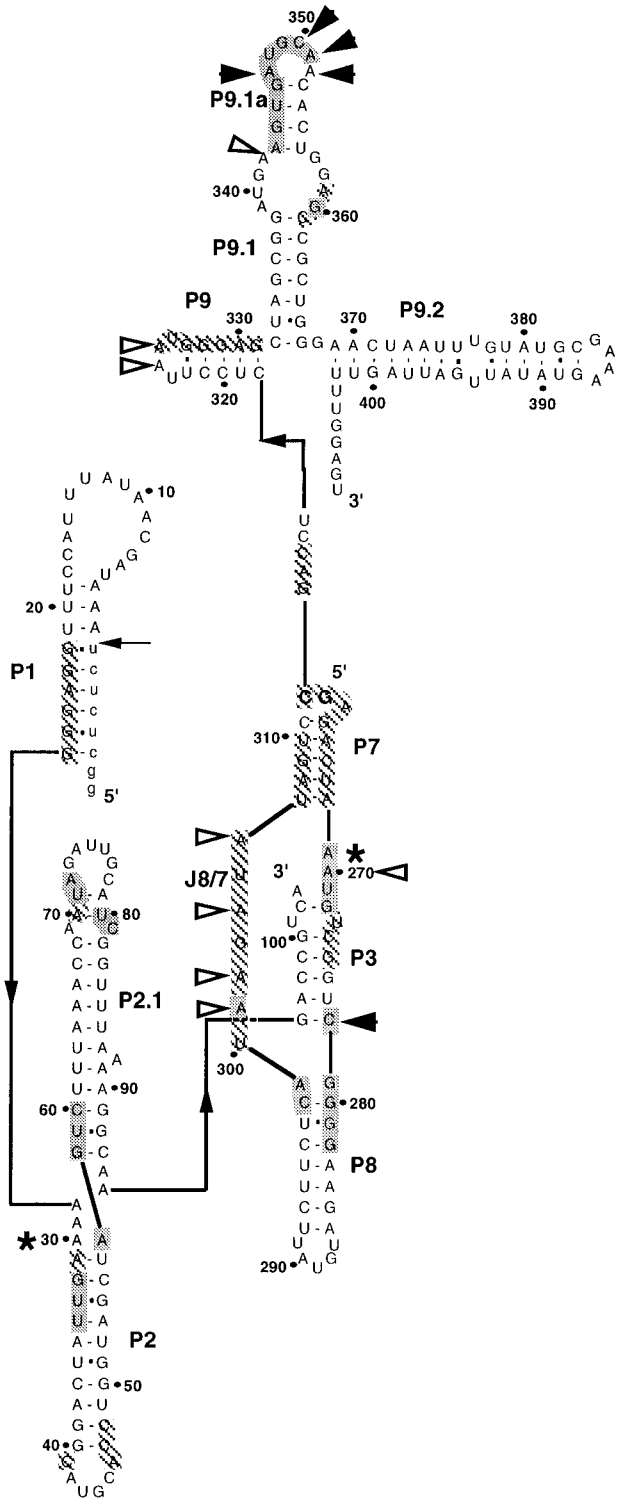

FIGURE 8: Summary of Fe(II)-EDTA cleavage and DMS modification of complexes of P1−P3 and P3+P7−P9 RNAs in the absence and presence of P4−P6. Shaded positions are protected from Fe(II)-EDTA cleavage by at least 1.5-fold in P1−P3/P3+P7−P9 complexes. Hatched positions are protected from Fe(II)-EDTA cleavage at least 1.5-fold in P1−P3/P3+P7−P9 complexes only in the presence of excess P4−P6. The Fe(II)-EDTA cleavage pattern of positions 89−103 could not be determined. Filled arrowheads show protection from DMS modification in the P1−P3/P3+P7−P9 complex. Open arrowheads show DMS protection induced by addition of P4−P6 to the P1−P3/P3+P7−P9 complex. Asterisks show DMS modification in the P1−P3/P3+P7−P9 complex which is increased in the presence of P4−P6.

to pack tightly against the P4, P6, and P6a stems, stabilizing their coaxially stacked orientation. This packing produces

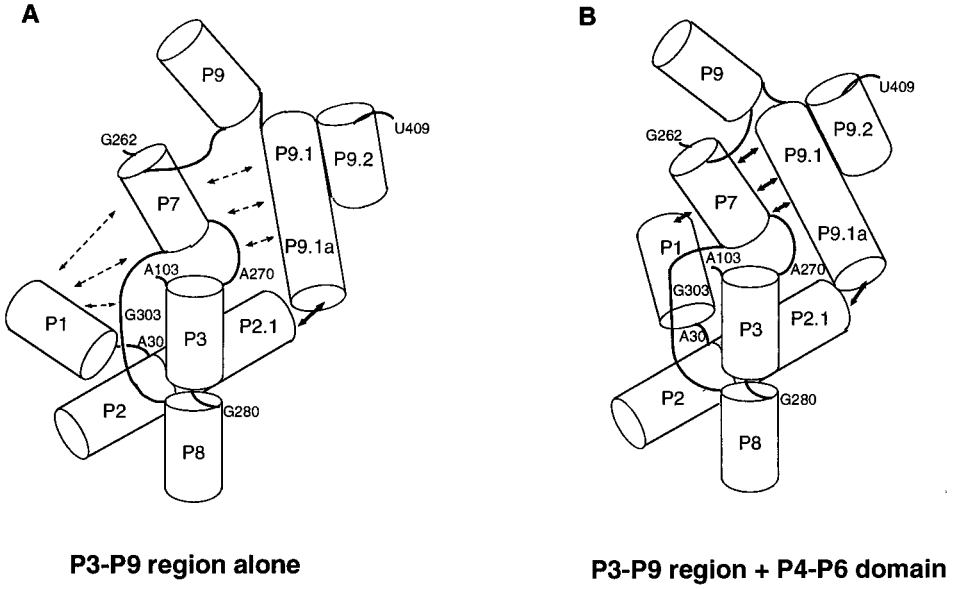

FIGURE 9: Model of the organization of the P3−P9 region in the absence (panel A) and presence (panel B) of the P4−P6 domain. Dashed arrows show weak or non-existent tertiary contacts while solid arrows show strong tertiary contacts. This figure is not meant to represent the overall structure of the P3−P9 region since the orientation of the helical elements in most cases is unknown. However, the P2 and P2.1 stems are proposed to stack coaxially based on cross-linking data (Downs & Cech, 1990). In the phylogenetic model of the catalytic core (Michel & Westhof, 1990), the P3 and P8 stems are coaxially stacked while the P3 and P7 stems are bent relative to one another.

a relatively flat molecule approximately two helices thick in one dimension and one helix thick in the other (Cate et al., 1996).

Less-well conserved segments that may perform a function analogous to that of P5abc also occur within the P3−P9 region of the *Tetrahymena* ribozyme and in other classes of group I introns. Long-range interactions in the *Tetrahymena* P3−P9 region have been proposed between the P9.1−P9.2 extension and the P3−P7 region (Laggerbauer et al., 1994) and between the P9.1a and P2.1 loops (Banerjee et al., 1993) based on phylogenetic covariation, mutagenesis, and chemical probing. Protection from Fe(II)-EDTA cleavage in L9.1a and L2.1 in the P1−P3/P3+P7−P9 complex suggests that in the absence of P4−P6 (Figures 8 and 9) interactions between these loops are intact. However, lack of strong Fe(II)-EDTA protection along one face of the P3 and P7 stems in the P1−P3/5′-ext P3+P7−P9 complex suggests that contacts between the P3−P7 and P9.1−P9.2 segments are weak or not formed in the absence of P4−P6 (Figures 8 and 9). Additionally, strong Fe(II)-EDTA protection was observed at several locations near the junction of the P2, P2.1, P3, and P8 stems in the P1−P3/P3+P7−P9 complex alone (Figure 8). Whether the lack of solvent accessibility at these positions results from short-range interactions caused by formation of the P3 pseudoknot or from longer-range contacts to less-well conserved areas within the P3−P9 region is unknown.

The simplest explanation for these results is that the P3−P9 region is not an independently folded domain but instead requires assistance to assume its native conformation in the ribozyme complex and, by analogy, in the intact intron. Some of the changes in Fe(II)-EDTA cleavage and DMS modification that we observe upon addition of P4−P6 to the P1−P3/P3+P7−P9 complex represent a direct footprint of the P4−P6 domain on the P3−P9 region (e.g., positions 302−315; Figure 8). Other such changes, however (at positions 262−275; Figure 8), are almost certainly due to contacts within the P3−P9 region which are strengthened by the P4−P6 domain. Many independent lines of evidence show that the single-helix thick face of the stacked P5, P4, P6, and P6a stems of P4−P6 interacts with the core elements of the P3−P9 region (Michel & Westhof, 1990; Pyle et al., 1992; Wang & Cech, 1992; Murphy & Cech, 1993; Wang et al., 1993; Cate et al., 1996; Tanner & Cech, 1996). It is difficult to imagine how this rigid helical structure could directly protect both faces of the P3−P9 core helices from solvent simultaneously.

A series of studies to determine the folding pathway of the *Tetrahymena* ribozyme core both kinetically and at equilibrium show that the P4−P6 domain folds prior to the P3 and P7 helices (Celander & Cech, 1991; Banerjee et al., 1993; Laggerbauer et al., 1994; Zarrinkar & Williamson, 1994, 1996b; Banerjee & Turner, 1995; Downs & Cech, 1996) and that the P9.1−P9.2 extension assists assembly of the active site (Laggerbauer et al., 1994; Zarrinkar & Williamson, 1996a). Our results begin to elucidate the role of each of these substructures in the formation of the catalytic core. The P4−P6 domain induces two structural changes within the P3−P9 region (Figure 9). First, contacts between the P3−P7 region of the core and the P9.1−P9.2 extension form or are significantly strengthened. Second, the P3−P9 region interacts with P4−P6 to assemble the catalytic center and the docking site for the P1 substrate duplex. Each of these structural changes results largely from the formation of tertiary interactions as all of the secondary structure within the P3−P9 region is capable of forming independently. RNase cleavage data suggest that the P7 helix may be weakly paired in both the P3+P7−P9 RNA and in the P1−P3/P3+P7−P9 complex. The P7 helix may be unstable due to the difficulty of maintaining a base-paired stem formed by two strands distant in the sequence or due to fraying caused by locating the 5′ end of the P3+P7−P9 RNA at P7 in our constructs. While the P4−P6 domain is not required for formation of the P3 stem at equilibrium, stable tertiary interactions within the P3−P9 region may be necessary to form helix P3. The P3+P7−P9 RNA does not associate

strongly with an oligonucleotide complementary to its P3 strand (Doudna & Cech, 1995). Likewise, a P3+P7−P9 RNA with a deletion of the P9.1 and P9.2 helices is incapable of forming a complex with the P1−P3 RNA (Doudna & Cech, 1995; E. A. Doherty and J. A. Doudna, unpublished). Thus, in addition to stabilizing the P3−P7 region of the core within the complete ribozyme, the P9.1−P9.2 extension may assist formation of the P3 duplex.

The incomplete folding of the P3−P9 region in the absence of the P4−P6 domain may be rationalized in part by examination of its secondary structure. The P3 and P7 helices are joined together on only one strand by a three-base bulge, while the other strand of P7 is linked to P8 by the long J8/7 segment (Figures 8 and 9). This linkage may lead to considerable flexibility in the relative orientations of P3 and P7, explaining the need for tertiary interactions along each helical face for optimal catalytic efficiency. In contrast, the P2 and P2.1 helices and the P3 and P8 helices are proposed to be coaxially stacked on the basis of cross-linking (Downs & Cech, 1990) and phylogenetic data (Michel & Westhof, 1990), respectively. The stability derived from coaxial stacking as well as the proximity of the P2.1 and P3 helices may explain the ability of this portion of P3−P9 to assemble without significant assistance from the P4−P6 domain.

Does the apparent flexibility within the P3−P9 region of the core play a role in self splicing? Several studies have suggested that large motions occur within the *Tetrahymena* ribozyme active site during the splicing reaction (Herschlag & Cech, 1990; Herschlag, 1992; Bevilacqua et al., 1992; Wang et al., 1993) to allow the docking and release of duplexes carrying the 5′ and 3′ cleavage sites. Perhaps alterations in the relative orientations of the P3 and P7 helices aid in opening and closing the active site so that splicing may proceed smoothly.

Independently folded domains also occur within other large catalytic RNAs. Fe(II)-EDTA analysis of two eubacterial RNase P RNAs demonstrates that they assemble by the association of two domains (Pan, 1995; Loria & Pan, 1996) which, intriguingly, are organized differently from those of the *Tetrahymena* group I intron. One domain comprises the entire active site while the other includes residues important for contacting pre-tRNA substrates. Each of these domains may be broken down into smaller structural modules which might fold independently (Loria & Pan, 1996). It will be interesting to compare the roles of such subdomains in formation of the RNase P active site with the role of P4−P6 in assembly of the *Tetrahymena* intron catalytic core.

ACKNOWLEDGMENT

We thank Luke Rice for preparing plasmid pP1P3 and for help in initiating this work, Scott Strobel, Pat Zarrinkar, and members of the Doudna lab for helpful discussions, and Jamie Cate, Adrian Ferré-D'Amaré, Scott Strobel, and Dan Herschlag for critical review of the manuscript.

REFERENCES

Banerjee, A. R., & Turner, D. H. (1995) *Biochemistry 34*, 6504−6512.
Banerjee, A. R., Jaeger, J. A., & Turner, D. H. (1993) *Biochemistry 32*, 153−63.
Beaudry, A. A., & Joyce G. F. (1990) *Biochemistry 29*, 6534−9.
Bevilacqua, P. C., Kierzek, R., Johnson, K. A., & Turner, D. H. (1992) *Science 258*, 1355−8.
Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Kundrot, C. E., Cech, T. R., & Doudna, J. A. (1996) *Science 273*, 1678−1685.
Cech, T. R., Damberger, S. H., & Gutell, R. R. (1994) *Nat. Struct. Biol. 1*, 273−80.
Celander, D. W., & Cech, T. R. (1991) *Science 251*, 401−7.
Doudna, J. A., & Cech, T. R. (1995) *RNA 1*, 36−45.
Downs, W. D., & Cech, T. R. (1990) *Biochemistry 29*, 5605−13.
Downs, W. D., & Cech, T. R. (1996) *RNA 2*, 718−732.
England, T. E., Bruce, A. G., & Uhlenbeck, O. C. (1980) *Methods Enzymol. 65*, 65−74.
Herschlag, D. (1992) *Biochemistry 31*, 1386−99.
Herschlag, D., & Cech, T. R. (1990) *Nature 344*, 405−9.
Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., & Pease, L. R. (1989) *Gene 77*, 51−59.
Inoue, T., & Cech, T. R. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 648−52.
Jaeger, L., Westhof, E., & Michel, F. (1993) *J. Mol. Biol. 234*, 331−46.
Laggerbauer, B., Murphy, F. L., & Cech, T. R. (1994) *EMBO J. 13*, 2669−76.
Latham, J. A., & Cech, T. R. (1989) *Science 245*, 276−82.
Latham, J. A., Zaug, A. J., & Cech, T. R. (1990) *Methods Enzymol. 181*, 558−69.
Loria, A., & Pan, T. (1996) *RNA 2*, 551−563.
Michel, F., & Westhof, E. (1990) *J. Mol. Biol. 216*, 585−610.
Michel, F., Ellington, A. D., Couture, S., & Szostak, J. W. (1990) *Nature 347*, 578−80.
Moazed, D., Stern, S., & Noller, H. F. (1986) *J. Mol. Biol. 187*, 399−416.
Murphy, F. L., & Cech, T. R. (1993) *Biochemistry 32*, 5291−300.
Pan, T. (1995) *Biochemistry 34*, 902−909.
Pan, T., & Zhong, K. (1994) *Biochemistry 33*, 14207−14212.
Price, J. V., Kieft, G. L., Kent, J. R., Sievers, E. L., & Cech, T. R. (1985) *Nucleic Acids Res. 13*, 1871−89.
Pyle, A. M., Murphy, F. L., & Cech, T. R. (1992) *Nature 358*, 123−8.
Tanner, M. A., & Cech, T. R. (1997) *Science* (in press).
van der Horst, G., Christian, A., & Inoue, T. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 184−188.
Wang, J. F., & Cech, T. R. (1992) *Science 256*, 526−9.
Wang, J. F., Downs, W. D., & Cech, T. R. (1993) *Science 260*, 504−8.
Woodson, S. A., & Cech, T. R. (1991) *Biochemistry 30*, 2042−50.
Young, B., Herschlag, D., & Cech, T. R. (1991) *Cell 67*, 1007−19.
Zarrinkar, P. P., & Williamson, J. R. (1994) *Science 265*, 918−24.
Zarrinkar, P. P., & Williamson, J. R. (1996a) *Nucleic Acids Res. 24*, 854−858.
Zarrinkar, P. P., & Williamson, J. R. (1996b) *Nat. Struct. Biol. 3*, 432−438.

BI962428+