

# Crystal Structure of a Group I Ribozyme Domain: Principles of RNA Packing

Jamie H. Cate, Anne R. Gooding, Elaine Podell, Kaihong Zhou,  
Barbara L. Golden, Craig E. Kundrot, Thomas R. Cech,\*  
Jennifer A. Doudna\*

Group I self-splicing introns catalyze their own excision from precursor RNAs by way of a two-step transesterification reaction. The catalytic core of these ribozymes is formed by two structural domains. The 2.8-angstrom crystal structure of one of these, the P4-P6 domain of the *Tetrahymena thermophila* intron, is described. In the 160-nucleotide domain, a sharp bend allows stacked helices of the conserved core to pack alongside helices of an adjacent region. Two specific long-range interactions clamp the two halves of the domain together: a two-Mg<sup>2+</sup>-coordinated adenosine-rich corkscrew plugs into the minor groove of a helix, and a GAAA hairpin loop binds to a conserved 11-nucleotide internal loop. Metal- and ribose-mediated backbone contacts further stabilize the close side-by-side helical packing. The structure indicates the extent of RNA packing required for the function of large ribozymes, the spliceosome, and the ribosome.

RNA can both encode genetic information and catalyze biochemical reactions. The paradigm for genetic coding is the formation of a double helix involving Watson-Crick base pairs. RNA catalysts, however, form more complex three-dimensional structures whose folds remain a mystery. Even in the well-studied class of ribozymes called group I self-splicing introns, only a handful of non-Watson-Crick contacts have been identified. These introns, defined by a common secondary structure and reaction pathway (1, 2) (Fig. 1), form an active site for consecutive phosphodiester exchange reactions that produce properly spliced RNAs. Evidence for a globular conformation and a relatively solvent-inaccessible core came from free-radical cleavage experiments on the group I intron from *Tetrahymena thermophila* (3–5). Divalent metal ions play a direct role both in the formation of this structure and in catalysis (3, 5–8).

Using comparative sequence analysis, Michel and Westhof modeled the conserved core of group I introns as two sets of coaxially stacked helices juxtaposed to create the active site (9). In the *Tetrahymena*

intron, these helices reside in two domains of about equal size, P4-P6 and P3-P9 (Fig. 1). Chemical protection experiments suggested that the P4-P6 region of the *Tetrahymena* intron folds as an independent unit (3, 5) (Fig. 1). When separated from the other half of the intron, the P4-P6 RNA adopts the same secondary structure and higher order folding that it has in the intact intron (10–12). Furthermore, the P4-P6 domain assembles in trans with the remaining regions of the intron to form a catalytically active complex (13). The P4-P6 domain is also the first higher order structure to form on the kinetic folding pathway of the intron (14).

The *Tetrahymena* P4-P6 domain contains roughly half of the active site of the intron—base-paired (P) segments P4, P5, and P6 and joining (J) regions J3/4, J6/7, and J4/5. These conserved elements are found in the catalytic cores of all group I introns (15–17) (Fig. 1). The P4-P6 domain also contains the P5abc extension (P5a, P5b, and P5c) found only in the IC1 and IC2 subclasses of group I introns (9, 18). Group I introns containing the extension require it for efficient catalysis, whereas those without it typically have either additional structural elements not found in the *Tetrahymena*-like introns or essential protein cofactors that bind in the P4-P6 region (19). Deletion of P5abc from the *Tetrahymena* intron abolishes splicing activity except in the presence of high concentrations of magnesium ions (20). Splicing activity is restored at low ionic strength by supplying the P5abc region in trans (21). These re-

sults are consistent with the idea that the P5abc extension stabilizes the active conformation of the RNA through interactions near the core. Indeed, a model of the domain, built on the basis of site-specific mutagenesis and chemical probing, predicted a sharp bend in J5/5a (Fig. 1) to allow the P5abc extension to lie alongside P4, P6, and P6a (10).

The detailed structures of group I introns are unknown, although a few specific tertiary contacts have been identified on the basis of site-directed mutagenesis, chemical protection, and phylogenetic comparisons (17). Modeling efforts are limited by the small number of atomic resolution RNA structures available. The crystal structures of several transfer RNAs [~76 nucleotides (nt)] have been known for many years (22–26). More recently, two crystal structures of hammerhead ribozymes (~50 nt) have been determined (27, 28), and the solution structures of several RNA motifs have been investigated by nuclear magnetic resonance (NMR) (29). Our knowledge of the three-dimensional structure of RNA has come from these few examples of relatively small molecules.

We present here the x-ray crystal structure of the 160-nt P4-P6 domain from the *Tetrahymena* group I intron at 2.8 Å resolution. The structure reveals various long-range interactions required to stabilize large RNAs such as group I and group II introns, RNase P RNA, and ribosomal and spliceosomal RNAs. These include contacts involving noncanonically paired or “bulge” regions interspersed between helices, backbone-backbone interactions, metal binding sites, and interactions involving a GAAA tetraloop. The structure shows how RNA, though limited to four rather similar building blocks, can nevertheless assemble to produce a complex globular fold.

**Structure determination and overview.** P4-P6 RNA was synthesized in vitro with the use of T7 RNA polymerase and purified by gel electrophoresis (30). Crystals of the RNA were grown in 60 mM potassium cacodylate (pH 6), 30 mM magnesium chloride, 0.3 mM spermine, and 0.2 to 1.0 mM cobalt hexamine chloride by vapor diffusion with methylpentanediol as a precipitant. The crystals belong to space group P2<sub>1</sub>2<sub>1</sub> (30) with two P4-P6 molecules in the asymmetric unit and diffract anisotropically (2.8 to 2.5 Å resolution). The crystal structure of the P4-P6 domain RNA was solved by multiwavelength anomalous diffraction (MAD) and single isomorphous replacement (SIR) with the use of an osmium derivative (Table 1). Crystals soaked in solutions containing 0.2 to 0.4 mM osmium hexamine triflate (31) produced an osmium derivative as identified by standard dif-

J. H. Cate, K. Zhou, and J. A. Doudna are in the Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520, USA; e-mail: doudna@csb.yale.edu. A. R. Gooding, E. Podell, B. L. Golden, C. E. Kundrot, and T. R. Cech are in the Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309, USA; e-mail: cech@stripe.colorado.edu. A. R. Gooding, B. L. Golden, E. Podell, and T. Cech are also with the Howard Hughes Medical Institute at Boulder, CO 80309, USA.

\*To whom correspondence should be addressed.

ference Patterson methods. Two data sets measured from an osmium derivative crystal at wavelengths near the osmium LIII absorption edge were combined with data from a cobalt hexammine derivative crystal to calculate the initial electron density maps. The final experimental map was generated with phases from the above data sets, followed by density modification. The high quality of this map allowed correct positioning of the nucleotide sequence during model building. The register of the sequence was later confirmed by difference Fourier analysis of 5-iodouracil derivatives at nucleotides 241, 253, 258, and 259 (32). The current model consists of 154 nt in molecule A, 154 nt in molecule B, and a total of 28 metals and six waters (33). A full account of the structure determination and refinement will be published elsewhere (34).

The secondary structure of the P4-P6 domain (Fig. 1) is color coded to highlight areas of functional interest, and the crystal structure with the same color coding is shown in Fig. 2. Of the paired regions in Fig. 1, only P6 is not formed as predicted in the secondary structure. Throughout the molecule, altered major and minor groove widths characterize regions of interhelical contact as well as segments of non-Watson-Crick pairing. Only helices P6b and P5b have uninterrupted canonical A-form geometry (35).

The molecule comprises two helical regions that pack side by side with overall dimensions of about 110 by 50 by 25 Å. Helices P6b, P6a, P6, P4, and P5 form a straight column on one side of the molecule, and helices P5b and P5a are stacked on the other. A bend of ~150° between helices P5 and P5a allows the P5abc extension to interact with one helical face of the conserved core region. The three-way junction of helices P5a, P5b, and P5c is buttressed by an adenosine-rich motif referred to as the A-rich bulge. The P5c region protrudes from the plane of the helical stacks (Fig. 2, B and C). Two major sets of tertiary interactions stabilize packing of the P5abc extension against helices of the conserved core. These two contacts can be viewed as a clamp on the back of the core that presents P4 and P6 to the rest of the RNA in the intact intron. The first of these occurs between the minor groove of the P4 helix and residues in the A-rich bulge. The second links the minor groove of J6a/6b and P6a to the GAAA tetraloop at the end of P5b. The overall fold seen in the x-ray crystal structure is in agreement with previous biochemical studies, as described below.

**A-rich bulge bridges two parallel helical stacks.** Group I intron subclasses IB and IC often contain an asymmetric A-rich bulge in P5a (9) (Figs. 1 and 2). The size of the

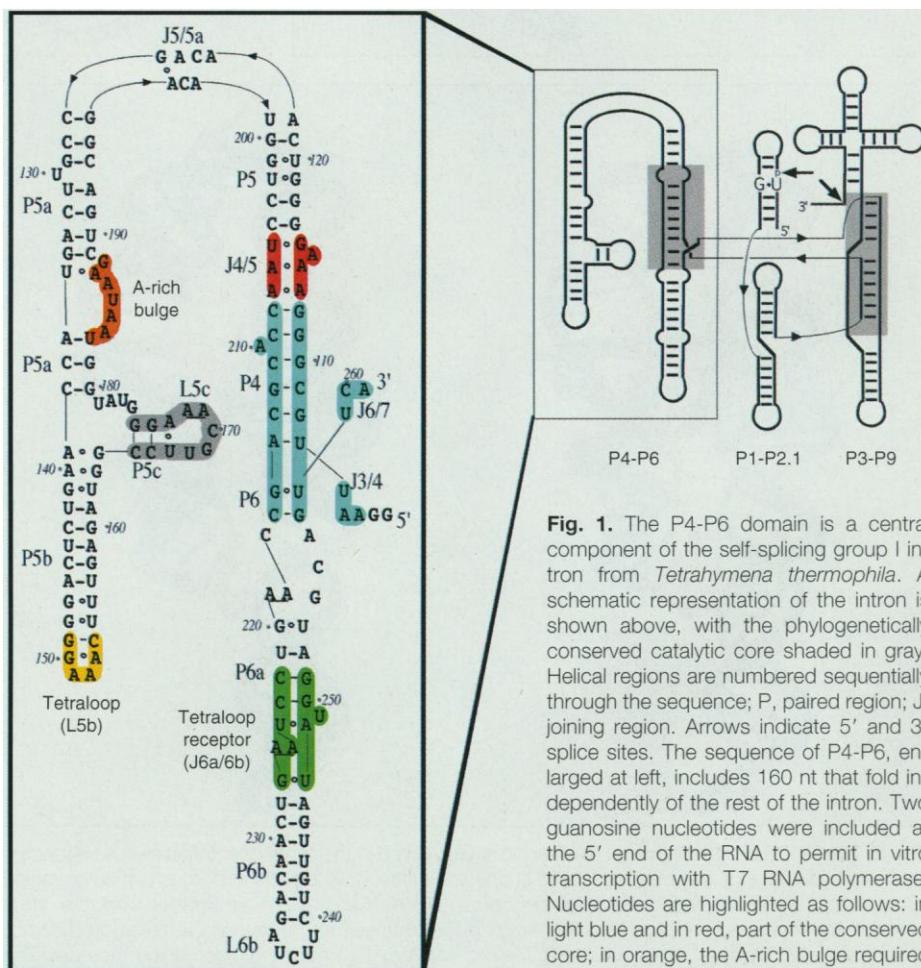
bulge and its distance from P4 are conserved (18). Within the A-rich bulge sequence, A184 and A186 are invariant, and A183 is highly conserved. Structural interactions involving the bulge are critical to the folding of the entire P4-P6 domain. For example, deletion of the A-rich bulge or a point mutation of A186 to U disrupts the global structure of the domain (10, 36). The A-rich bulge nucleates folding of a substructure within the molecule, which includes the bulge and the three-helix junction of P5a, P5b, and P5c (10, 36). Base-pair substitutions in the third base pair of P4 decrease intron splicing activity; this has been ascribed to disruption of a tertiary interaction between this base pair and A183 or a neighboring base in the A-rich bulge (37).

In the crystal structure, the backbone of the A-rich bulge makes a corkscrew turn (Fig. 3, A to C). Its bases are flipped out and interact with adjacent residues of the P4 helix on one side and the three-helix junction on the other. The four adenosines of the bulge are involved in stacking interac-

tions and an intricate network of hydrogen bonds (Table 2). Phosphates in the backbone of the bulge are packed unusually close together, with 3 Å between the closest phosphate oxygens. These phosphate oxygens directly coordinate two magnesium ions, clearly visible in the experimental electron density map (Fig. 3, A and C). The metals form an approximate axis of helical symmetry with respect to the backbone, extending from A183 to A187.

Whereas the first two adenosines bind the minor groove of P4, the last two bind in pockets at the three-helix junction. Critical to the stability of the P4-P6 domain, A186 is nestled in a pocket formed by the minor groove face of the C137-G181 base pair in P5a and residue G164 at the top of P5b (Fig. 3D). All four 2'-hydroxyl groups are involved in hydrogen bonds in this 4-nt interaction. In addition, A186 stacks on two sheared G-A base pairs at the top of P5b, and these base pairs contribute to the conformation of the three-helix junction.

The observed interactions agree well



**Fig. 1.** The P4-P6 domain is a central component of the self-splicing group I intron from *Tetrahymena thermophila*. A schematic representation of the intron is shown above, with the phylogenetically conserved catalytic core shaded in gray. Helical regions are numbered sequentially through the sequence; P, paired region; J, joining region. Arrows indicate 5' and 3' splice sites. The sequence of P4-P6, enlarged at left, includes 160 nt that fold independently of the rest of the intron. Two guanosine nucleotides were included at the 5' end of the RNA to permit *in vitro* transcription with T7 RNA polymerase. Nucleotides are highlighted as follows: in light blue and in red, part of the conserved core; in orange, the A-rich bulge required for proper P4-P6 folding; in gold, the GAAA tetraloop; in green, the conserved 11-nt tetraloop receptor; and in gray, P5c. The P5a, P5b, and P5c helices are referred to as the P5abc extension, and the junction of these three is abbreviated 3HJ.

with biochemical data for the P4-P6 domain alone and within the intact intron.

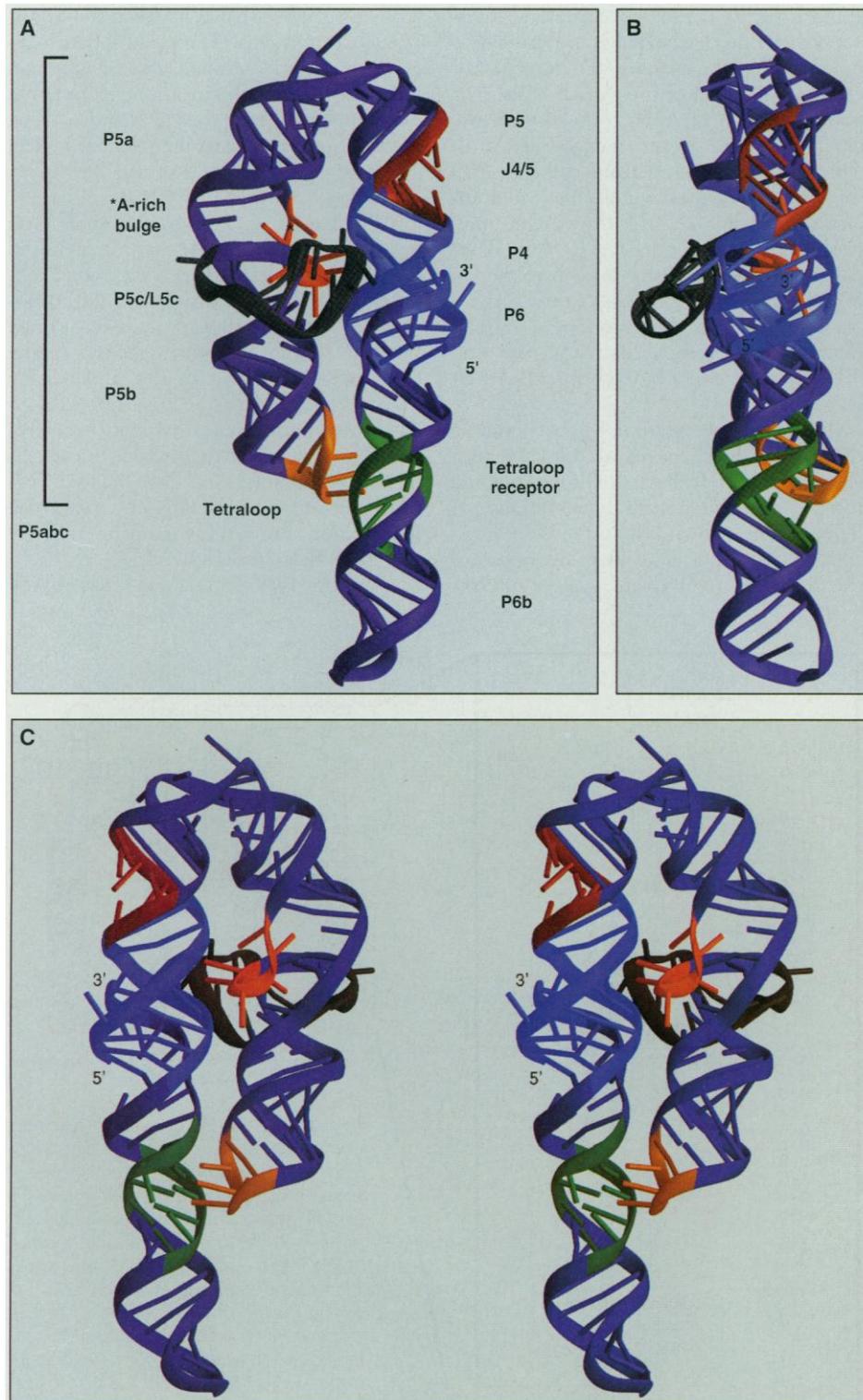
Site-directed mutagenesis and chemical protection studies (37) indicated that G212

in the P4 helix contacts either A183 or a neighbor in the A-rich bulge; the observed interaction is with A184. Methylation of the N1 of adenosine by dimethyl sulfate occurs at A183 and A187, consistent with the exposure of these bases in the structure (10). Most significantly, the extreme sensitivity of global domain structure to mutations in the A-rich bulge (10, 36) is explained by the network of hydrogen bonds that cross-stitch the two helical halves of the domain in this region.

**Tetraloop interaction with 11-nucleotide tetraloop receptor.** Large structured RNAs, including ribosomal RNA and self-splicing introns, frequently contain hairpin loops of sequence GNRA (N, any nucleotide; R, purine nucleotide) (38). Costa and Michel (39) observed that the GAAA subclass of these loops is often found together with an 11-nt motif, and they demonstrated specific interaction between these two elements in self-splicing introns. The canonical 11-nt tetraloop receptor is found in the P4-P6 domain. Centered in J6a/6b, it consists of two adjacent C-G base pairs, a 5-nt internal loop, and a G-U base pair (Fig. 1). Previous chemical protection and mutagenesis studies (10) suggested an interaction between the bottom C-G pair of P6a and the last A of the GAAA loop.

The P4-P6 crystal structure now provides an atomic-resolution view of the GAAA loop-receptor interaction (Fig. 4, A and B). The conformation of the tetraloop is virtually identical to that observed by NMR (40) and in the crystal structure of a hammerhead ribozyme (41), suggesting that it is a relatively rigid unit. All of the loop nucleotides are in the anti conformation. The loop docks in the minor groove with the P5b and P6a helices at an approximate 30° angle. The three adenine bases are stacked on the bases on the 5' side of the tetraloop receptor helix. Stacking is facilitated by adjacent adenines in the receptor internal loop that stack across the helix, forming an adenosine platform motif (42). This unusual side-by-side configuration of adenines results in a kink in the ribose backbone that opens up the minor groove of the tetraloop receptor. Although the tetraloop receptor is an asymmetric internal loop, which might have been predicted to result in a kink or bend, its structure provides nearly coaxial alignment of the flanking helices.

In addition to stacking, each adenosine in the GAAA loop makes specific hydrogen bonds to the tetraloop receptor. This reveals why the loop-receptor interaction is so sequence-specific: each A of the tetraloop is involved in hydrogen bonds specific to the adenine base. The first A of GAAA is part of an A·U·A triple (Fig. 4C). The U



**Fig. 2.** The P4-P6 crystal structure. **(A)** The molecule viewed in the same orientation as the secondary structure shown in Fig. 1. A bend of  $\sim 150^\circ$  at one end of the molecule allows helices of the conserved core (in light blue and in red) to pack against helices of the P5abc extension. Specific contacts occur between the P4 helix (light blue) and the A-rich bulge (orange), and between the tetraloop receptor (green) and the GAAA tetraloop (gold). **(B)** View of the structure from the side, facing the core. The domain is about one helix thick in this dimension ( $\sim 25 \text{ \AA}$ ), except for the P5c helix and loop (gray). **(C)** Stereo representation of the structure from the opposite side ( $180^\circ$  from Fig. 2A). This and subsequent figures, except where stated, were generated with RIBBONS (67).

dips into the plane of the two adenines and makes a reverse-Hoogsteen pair with the adenine of the receptor. The two adenines form a symmetric A·A pair as described for an internal loop of the HIV-1 Rev responsive element (43–45). The second A (GAAA) stacks on top of this triple and makes three ribose-mediated contacts (Fig. 4D). The third A (GAAA) forms a triple with the predicted C·G base pair (10) in addition to its sheared G·A pairing to the G of the tetraloop (40, 41) (Fig. 4E). Modeling of G in place of the second or third A of the loop introduces steric clashes with the tetraloop receptor, which is consistent with the observed destabilizing effects of these mutations (10). Many of the contacts between the tetraloop and the minor groove involve 2'-hydroxyl groups, as was observed in an intermolecular contact between a tetraloop and a minor groove in a hammerhead ribozyme crystal (41).

**Close packing of helices mediated by riboses and metals.** The A-rich bulge and GAAA tetraloop interactions described above bring the two helical halves of the P4-P6 domain into close proximity. The result is remarkably snug packing of the ribose-phosphate backbones. A number of riboses of the isolated P4-P6 domain are inaccessible to Fe(II)-EDTA cleavage (36, 46). The C4' atoms protected from free radical reaction lie on the inside of the surface formed by the two halves of the domain and in the region of the A-rich bulge where the backbone is buried (dark blue spheres, Fig. 5A). In contrast, the C4' atoms that are Fe(II)-EDTA-accessible in solution (red and light blue regions) cover the outside surface of the crystal structure. The light blue regions are protected in the intact intron.

There is good correlation between the backbone accessibility determined biochemically and crystallographically (Fig. 5B). The major exception occurs in the L5c hairpin loop, where interactions between molecules in the asymmetric unit of the crystal cause internalization of the backbone. In solution, this region is not protected from Fe(II)-EDTA cleavage in the isolated domain, although it becomes protected in the intact intron.

The crystal structure reveals in detail how a solvent-inaccessible molecular interior can be constructed from ribonucleotides. Interdigitated riboses line the regions of closest contact between the helical stacks. Pairs of riboses interact by hydrogen bonding, forming "ribose zippers" in the A-rich bulge and the GAAA tetraloop long-range contacts. For example, the backbones of the A-rich bulge and the P4 helix interact through the 2'-hydroxyls of two staggered riboses (Fig. 5C). The ribose zipper is characterized by shared hydrogen bonds between the 2'-hydroxyl and

pyrimidine O2 (purine N3) of one base and the 2'-hydroxyl of its partner. At least one side of a ribose zipper has non-A-form geometry (47). An intermolecular version of a ribose zipper is present in a crystal contact between adjacent tetraloops of the hammerhead ribozyme (41, figure 1, D and E).

The close packing of phosphates from adjacent helices is mediated by hydrated magnesium ions. Experimental electron density, located about 4 Å from the nearest phosphate oxygens, is consistent with outer sphere magnesium ion complexes occurring

between P4 and J6/6a, between P5 and P5a, and between P5b and P6a (Fig. 5D). The distance between phosphate oxygens across the major groove in each helix is 8 to 9 Å, as would be expected for A-form RNA. Between helices, the phosphate oxygens are 7 to 8 Å apart and aligned to bind the hydrated metal. These coordination sites reveal one of the major roles long suspected for divalent metals in the folding of large RNA molecules (5, 6, 8, 48–50). In these cases the metal bridges the phosphate backbone, enabling close approach of adjacent

**Table 1.** Summary of crystallographic data. Crystals were grown from methylpentanediol (MPD) as described (30), with subsequent modification (34). Crystals were transferred to a solution containing 25 percent MPD, 100 mM potassium cacodylate (pH 6.0), 50 mM magnesium chloride, 0.5 mM spermine, 10 percent isopropanol, and 0.035 to 0.07 mM cobalt hexammine chloride. The osmium derivative was prepared by substituting osmium hexammine triflate (31) for the cobalt hexammine chloride. Crystals were flash-frozen in liquid propane cooled with liquid nitrogen. Initial Patterson maps were calculated from data to 5.0 Å resolution collected at  $-160^{\circ}\text{C}$  on an R-AXIS II imaging plate area detector equipped with focusing mirrors. The data set Cohex2, collected on a MacScience DIP2000 area detector equipped with focusing mirrors, was used in calculating the first traceable electron density maps. This data set was later replaced with a higher resolution cobalt data set (Cohex 1) collected at Cornell High Energy Synchrotron Source (CHESS) beamline A-1. The osmium derivative data were collected (i) at the peak ( $\lambda 1$ ) and (ii) at the first inflection point ( $\lambda 2$ ) of the osmium absorption edge at the X-4A beamline of the National Synchrotron Light Source at Brookhaven National Laboratory. All data sets were processed with DENZO and scaled with SCALEPACK. Heavy-atom sites were refined, and multiple isomorphous replacement (MIR) and anomalous phases were computed using MLPHARE (63). Initial MAD-SIR phases were improved by density modification with DM (64). A rotamer library for ribose puckering conformations (34) in the interactive model building was used for program O (65). Refinement of the model against the Cohex 1 data set was carried out with X-PLOR 3.8 (66).

Data set	Oshex ( $\lambda 1$ )	Oshex ( $\lambda 2$ )	Cohex1	Cohex2
Resolution (Å)	20.0–2.8	20.0–2.9	18.0–2.5	20.0–2.8
Reflections				
Unique	63,718	61,355	42,836	31,828
Redundancy	2.4	1.8	4.0	2.9
% Complete	95.6	91.6	87.5	90.2
$R_{\text{sym}}^*$	4.8	4.3	4.5	6.9
Phasing: ( $\lambda 1$ = native)				
$R_{\text{c}}(\text{iso})$ ac, c†		0.95, 0.86	0.89, 0.86	
$R_{\text{c}}(\text{ano})\ddagger$	0.73	0.82		
$\  \sigma(I) \ $	4.1 (3.0 Å)	27.0 (3.0 Å)		4.7 (2.8 Å)
Phasing power§				
6.5–5.1 Å				
Iso ac, c		1.2, 1.0	0.9, 0.9	
Ano	2.2	1.8		
Overall				
Iso ac, c		0.7, 0.6	0.8, 0.6	
Ano	0.9	0.6		
Mean FOM		DM-refined parameters	After final round of DM	
6.5–5.1 Å	0.68		0.90	
Overall	0.42		0.71	
Refinement	Cohex1			
Resolution (Å)	8.0–2.5			
Reflections				
Working set (N)	34,551			
Test set (N)	1,850			
Number of atoms (N)	6,824			
R-factor¶	0.242			
R-free	0.285			
rms= bond (Å)	0.010			
rms angle (°)	1.27			

\* $R_{\text{sym}} = \sum |I - \langle I \rangle| / \sum I$ . † $R_{\text{c}}(\text{iso})$ , Cullis  $R$ , ratio of lack of closure to isomorphous difference, where c is centric data and ac is acentric data. ‡ $R_{\text{c}}(\text{ano})$ , Cullis  $R$ , ratio of lack of closure to anomalous difference. §Phasing power =  $\text{rms} \sum |F_{\text{h}}| / E$ , where  $F_{\text{h}}$  is the calculated heavy-atom structure factor amplitude and  $E$  is the lack of closure error. ||FOM = figure of merit for phases from the heavy-atom derivatives. Heavy-atom parameters refined with density modification were used to generate phases with the above FOM. The MAD-SIR phases were then density-modified before generating the final electron density map. ¶R-factor and R-free =  $\sum | |F_{\text{obs}}| - |F_{\text{calc}}| | / \sum |F_{\text{obs}}|$ , where R-free includes amplitudes omitted from the refinement. =rms, root mean square.

strands, as seen in shorter range interactions in transfer RNA (51). The details of other metal binding interactions in the P4-P6 RNA are not yet completely known (52).

**The phylogenetically conserved core.** The sequence of the P4-P6 domain includes the J3/4 5' of P4 nucleotides and the J6/7 nucleotides 3' of P6. In the crystal structure, the J6/7 nucleotides interact in the major groove of P4, while the J3/4 nucleotides interact in the minor groove of P6, forming a triple helix in the core (53). Models of the intron core (9, 54) predicted the J3/4 and J6/7 strands to lie in the minor and major grooves, respectively, as seen in the structure. However, the details of the predicted interactions differ substantially from the structure. A crystal contact occurs between P4-P6 molecules just below P6, perhaps widening the helix in this region. Furthermore, an unnatural G in J3/4 (included to facilitate transcription of the RNA; Fig. 1) base pairs with C217 below P6, which may also alter native folding. Thus, it is not yet clear whether the detailed structure in this region is one that is relevant to ribozyme function.

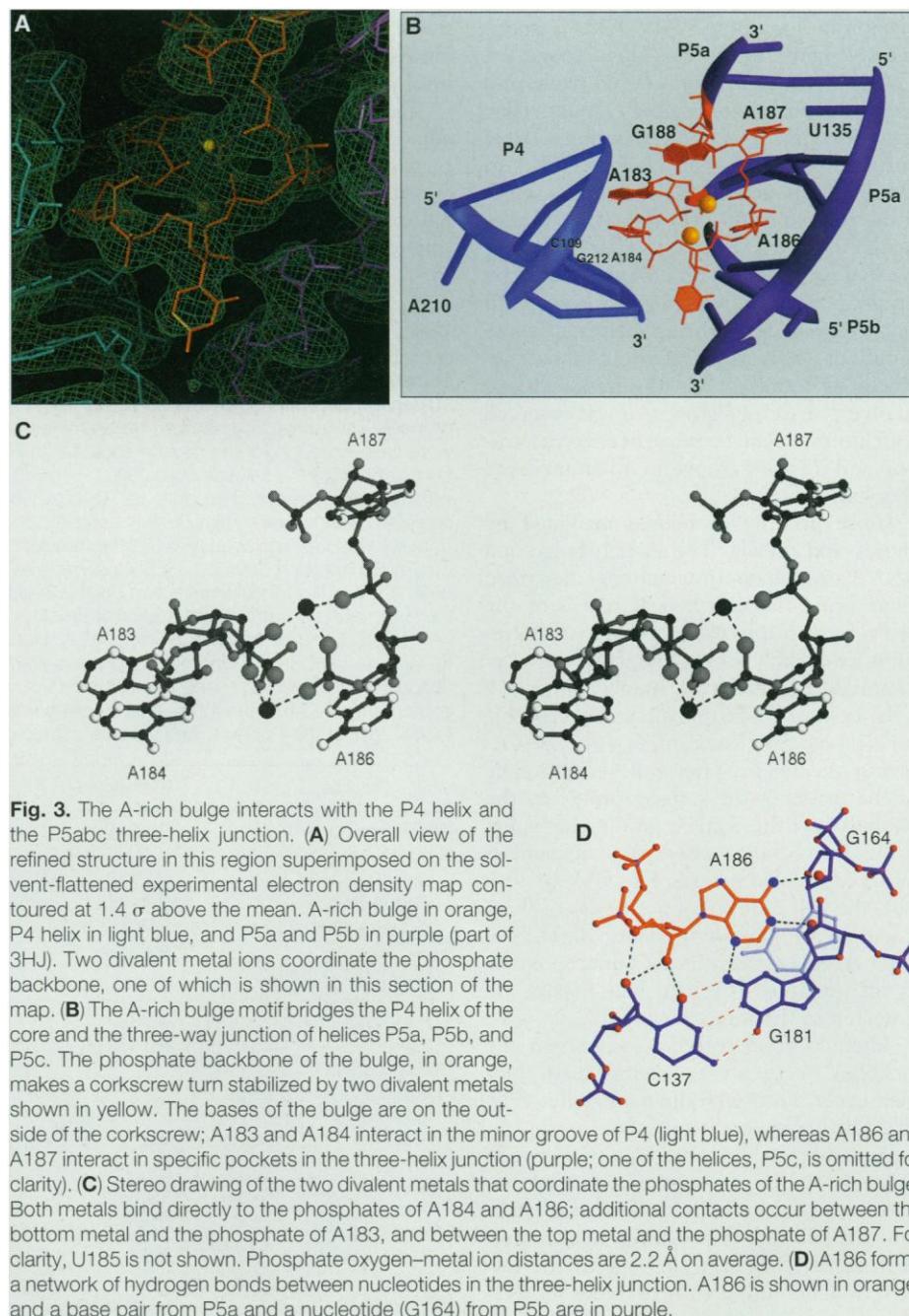
The J4/5 region, a conserved internal loop above P4, was predicted by Michel and Westhof to interact with the P1 substrate helix in the intron core (9). Experimentally, adenosines in J4/5 crosslink to the G that forms a functionally important and highly conserved G·U wobble base pair in the P1 substrate helix (55). The guanosine-nucleophile binding site in P7 (56) is in close spacial proximity to J4/5 (57), leading to the suggestion that the J4/5 internal loop may orient the nucleophile relative to the scissile phosphate in P1. Consistent with this model, mutation of A114 or A207 in J4/5 drastically decreases  $k_{cat}/K_m$  for ribozyme catalysis (58).

The crystal structure reveals J4/5 as a helical segment in which adjacent adenosines form sheared homopurine base pairs (Fig. 6A). This pairing scheme causes the N1 and N3 positions of A114 and A207 to protrude into the minor groove, distorting it away from ideal A-form geometry (Fig. 6, A and B). Molecules in the crystal lattice interact through asymmetric contacts between their respective J4/5 regions. The minor grooves are opened up and packed against each other at an  $\sim 60^\circ$  angle, with the 2'-hydroxyl of A114 in one molecule forming hydrogen bonds to the 2'-hydroxyl and N3 positions of A114 in the neighboring molecule. Packing of the backbones is mediated by intermolecular hydrogen bonds between residue A210 of one molecule and residues A117 and A118 of the other.

This intermolecular docking suggests two possible modes of P1 substrate binding

in the intact intron. In one model, we suppose A114 of one molecule to be at the

position occupied by G22 in the P1 helix of the intron. The critical 2-amino group of



**Fig. 3.** The A-rich bulge interacts with the P4 helix and the P5abc three-helix junction. **(A)** Overall view of the refined structure in this region superimposed on the solvent-flattened experimental electron density map contoured at  $1.4\sigma$  above the mean. A-rich bulge in orange, P4 helix in light blue, and P5a and P5b in purple (part of 3HJ). Two divalent metal ions coordinate the phosphate backbone, one of which is shown in this section of the map. **(B)** The A-rich bulge motif bridges the P4 helix of the core and the three-way junction of helices P5a, P5b, and P5c. The phosphate backbone of the bulge, in orange, makes a corkscrew turn stabilized by two divalent metals shown in yellow. The bases of the bulge are on the outside of the corkscrew; A183 and A184 interact in the minor groove of P4 (light blue), whereas A186 and A187 interact in specific pockets in the three-helix junction (purple; one of the helices, P5c, is omitted for clarity). **(C)** Stereo drawing of the two divalent metals that coordinate the phosphates of the A-rich bulge. Both metals bind directly to the phosphates of A184 and A186; additional contacts occur between the bottom metal and the phosphate of A183, and between the top metal and the phosphate of A187. For clarity, U185 is not shown. Phosphate oxygen-metal ion distances are  $2.2\text{ \AA}$  on average. **(D)** A186 forms a network of hydrogen bonds between nucleotides in the three-helix junction. A186 is shown in orange, and a base pair from P5a and a nucleotide (G164) from P5b are in purple.

**Table 2.** A-rich bulge hydrogen bonds.

A-rich bulge	H-bond partner	A-rich bulge	H-bond partner
A183 N3 a*	G110 O2' d	A186 O2' a	C137 O2' d
O2' d	G110 O2' a	O2' d	C137 O2 a
		N3 a	G181 N2 d
A184 N1 a	G212 O2' d	N1 a	G181 O2' d
N3 a	G212 N2 d	N6 d	G164 O2' a
O2' a	C109 O2' d		
O2' d	C109 O2 a	A187 N7 a	U135 N3 d
		N6 d	U135 O4 a

\*d and a refer to hydrogen bond donors and acceptors, respectively.

the G (59) would then be equidistant between the N3 positions of A114 and A207.

In a second model, we suppose A114 of the other molecule to be at the position of G22

in the intron. In this case, the 2'-hydroxyl of G22 would bridge the 2'-hydroxyl and N3 positions of A114 on the first molecule. Its 2-amino group would be equidistant between the 2'-hydroxyl and N3 positions of A207. These possibilities can now be tested biochemically.

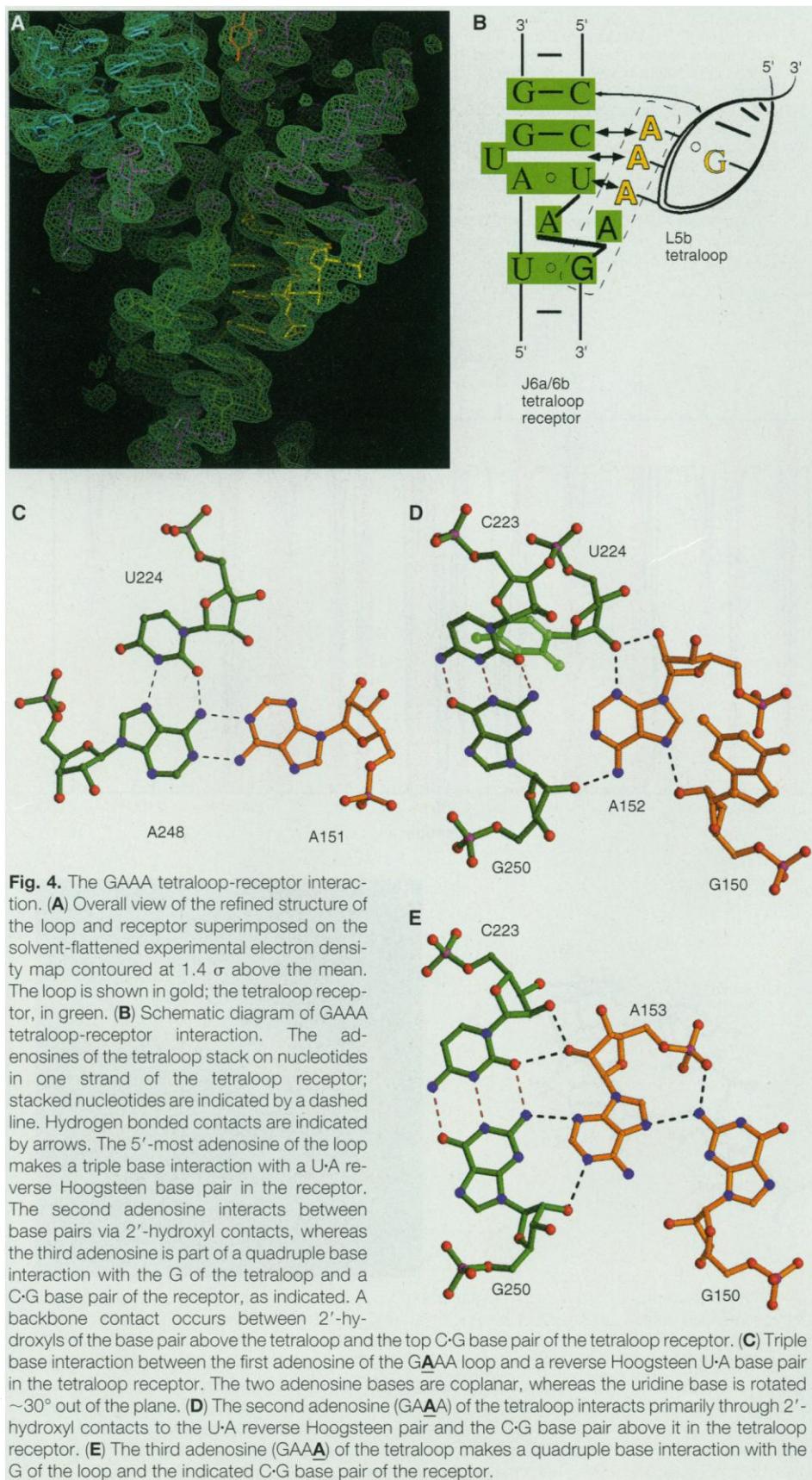
Catalysis by protein enzymes requires a globular molecule with a well-defined active site cleft, capable of precisely orienting substrates for reaction. It has long been thought that catalysis by RNA would also require a closely-packed structure, very different from the classical view of a stringlike messenger RNA (60, 61). The crystal structure of the 160-nt P4-P6 domain now provides a detailed view of the compactness of RNA folding.

As predicted, metal-phosphate coordination is abundant, bringing polyanionic helices into close proximity. Most of the bases are involved in base pairs (either Watson-Crick or noncanonical) and many participate in base triples and even quadruples. Base stacking, long known to make a major contribution to helix stability, mediates interactions between separate helical regions. The defining functional group of RNA, the 2'-hydroxyl, is a ubiquitous donor and acceptor of hydrogen bonds to phosphates and bases and to other 2'-hydroxyls, in some places forming ribose zippers. Thus, whereas single-stranded DNA may have some catalytic potential (62), it may be difficult to construct a highly efficient active site in nucleic acids without the use of 2'-hydroxyls.

Since the P4-P6 domain includes only half of the group I ribozyme active site, insights about catalysis are necessarily indirect. Intermolecular interactions in the crystal have provided some clues as to how the J4/5 internal loop, with its opened minor groove, may organize the substrate helix in the active site of the ribozyme. The next step is to test these ideas biochemically and crystallographically.

## REFERENCES AND NOTES

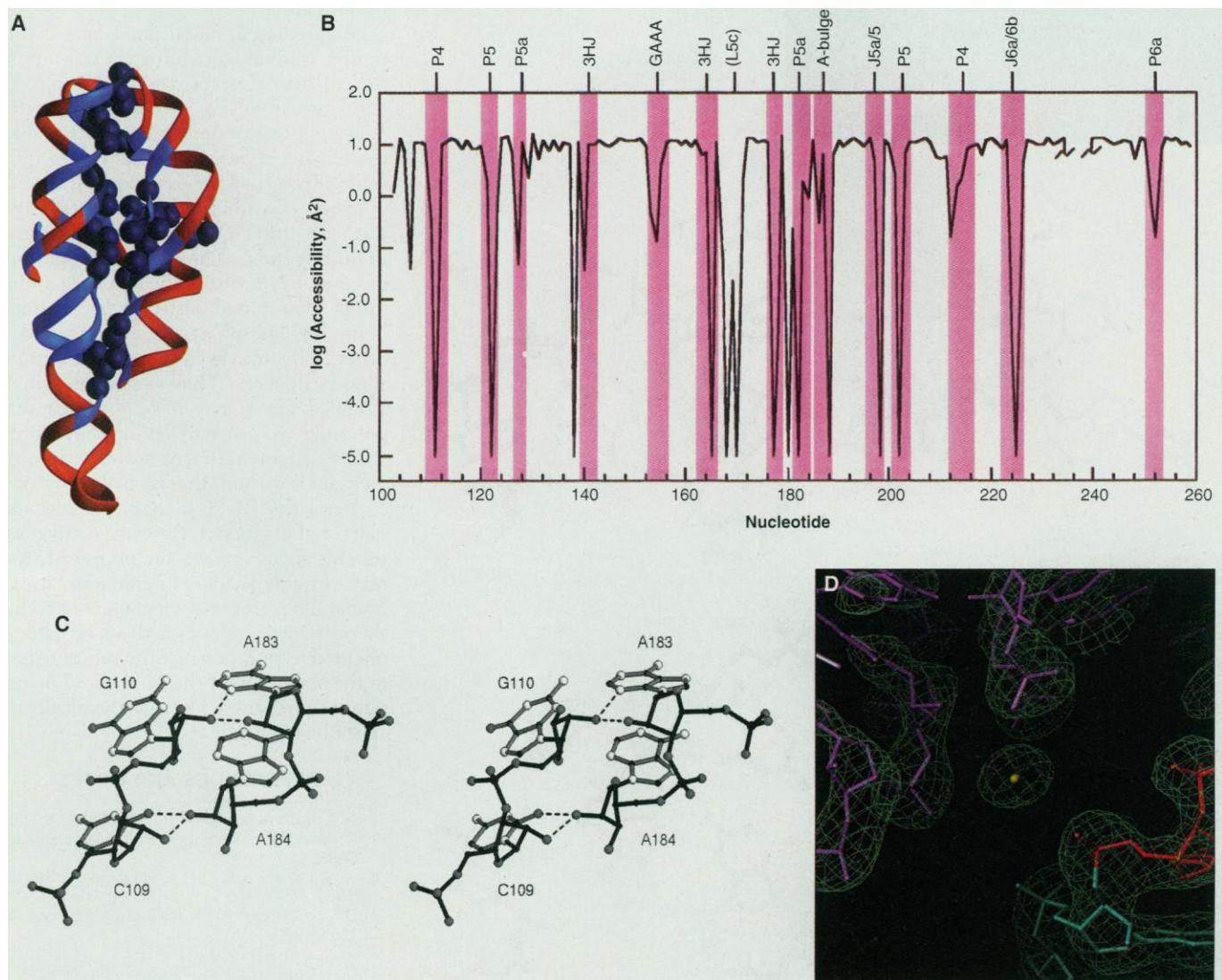
1. T. R. Cech, *Annu. Rev. Biochem.* **59**, 543 (1990).
2. R. Saldanha, G. Mohr, M. Belfort, A. M. Lambowitz, *FASEB J.* **7**, 15 (1993).
3. J. A. Latham and T. R. Cech, *Science* **245**, 276 (1989).
4. D. W. Celander and T. R. Cech, *Biochemistry* **29**, 1355 (1990).
5. ———, *Science* **251**, 401 (1991).
6. C. A. Grosshans and T. R. Cech, *Biochemistry* **28**, 6888 (1989).
7. J. A. Piccirilli, J. S. Vyle, M. H. Caruthers, T. R. Cech, *Nature* **361**, 85 (1993).
8. E. L. Christian and M. Yarus, *Biochemistry* **32**, 4475 (1993).
9. F. Michel and E. Westhof, *J. Mol. Biol.* **216**, 585 (1990).
10. F. L. Murphy and T. R. Cech, *ibid.* **236**, 49 (1994).
11. B. Laggerbauer, F. L. Murphy, T. R. Cech, *EMBO J.* **13**, 2669 (1994).
12. Y. H. Wang, F. L. Murphy, T. R. Cech, J. D. Griffith, *J.*



*Mol. Biol.* **236**, 64 (1994).  
 13. J. A. Doudna and T. R. Cech, *RNA* **1**, 36 (1995).  
 14. P. P. Zarrinkar and J. R. Williamson, *Science* **265**, 918 (1994); W. Downs and T. R. Cech, *RNA* **2**, 718 (1996).  
 15. F. Michel and B. Dujon, *EMBO J.* **2**, 33 (1983).  
 16. J. M. Burke *et al.*, *Nucleic Acids Res.* **15**, 7217 (1987).  
 17. T. Cech, S. Damberger, R. Gutell, *Nature Struct. Biol.* **1**, 273 (1994).  
 18. R. A. Collins, *Nucleic Acids Res.* **16**, 2705 (1988).  
 19. G. Mohr, M. G. Caprara, Q. Guo, A. M. Lambowitz, *Nature* **370**, 147 (1994).  
 20. G. F. Joyce, G. van der Horst, T. Inoue, *Nucleic Acids Res.* **17**, 7879 (1989).  
 21. G. van der Horst, A. Christian, T. Inoue, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 184 (1991).  
 22. S. H. Kim *et al.*, *Science* **179**, 285 (1973).  
 23. J. D. Robertus *et al.*, *Nature* **250**, 546 (1974).  
 24. R. W. Schevitz, A. D. Podjarny, N. Krishnamachari, J. J. Hughes, P. B. Sigler, *ibid.* **278**, 188 (1979).

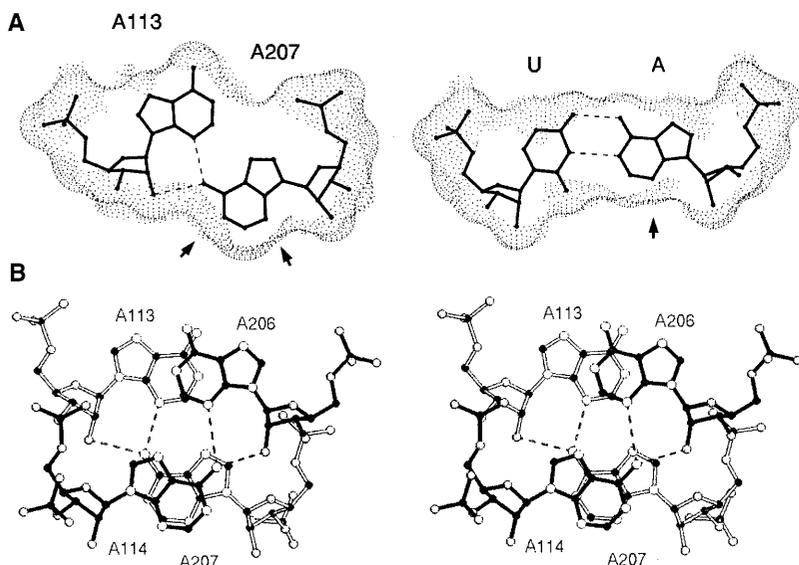
25. D. Moras *et al.*, *ibid.* **288**, 669 (1980).  
 26. N. H. Woo, B. A. Roe, A. Rich, *ibid.* **286**, 346 (1980).  
 27. H. W. Pley, K. M. Flaherty, D. B. McKay, *ibid.* **372**, 68 (1994).  
 28. W. G. Scott, J. T. Finch, A. Klug, *Cell* **81**, 995 (1995).  
 29. Reviewed by J. R. Wyatt and I. Tinoco Jr., in *The RNA World*, R. F. Gesteland and J. F. Atkins, Eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1993), p. 465; L. X. Shen, Z. Cai, I. Tinoco Jr., *FASEB J.* **9**, 1023 (1995).  
 30. J. A. Doudna, C. Grosshans, A. Gooding, C. E. Kundrot, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 7829 (1993).  
 31. P. A. Lay, R. H. Magnuson, H. Taube, *Inorg. Chem.* **28**, 3001 (1989).  
 32. B. L. Golden, A. R. Gooding, E. Podell, T. R. Cech, in preparation.  
 33. Both molecules A and B contain residues 103 through 260. In the crystals, ~50 percent of the RNA is nicked in loop L6b between residues C237 and U238 [deter-

mined by direct RNA sequencing and primer extension analysis of RNA in crystals (unpublished data)]. For both molecules L6b has fairly continuous density in the experimental map for the backbone; the bases for nucleotides 236 through 239 have not been modeled. In each molecule, there are three to four residues where the phosphate density is visible in the map but the base or ribose and base are not. In these cases, the full nucleotide has been included for refinement, but the occupancy for the missing component has been set to zero.  
 34. J. H. Cate and J. A. Doudna, *Structure*, in press.  
 35. N. E. Kallenbach and H. M. Berman, *Q. Rev. Biophys.* **10**, 138 (1977).  
 36. F. L. Murphy and T. R. Cech, *Biochemistry* **32**, 5291 (1993).  
 37. P. J. Flor, J. B. Flanagan, T. R. Cech, *EMBO J.* **8**, 3391 (1989).  
 38. C. R. Woese, S. Winker, R. R. Gutell, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 8467 (1990).



**Fig. 5.** Close packing of helices in the P4-P6 domain. **(A)** Ribbon representation of the P4-P6 crystal structure, color coded to reflect experimentally determined solvent accessibility of C4' atoms in the backbone. In dark blue, C4' positions protected from cleavage in the domain alone and in the intact intron (C4' atoms are shown as spheres); in light blue, C4' positions protected from cleavage in the intact intron but not in the isolated domain; in red, C4' positions always unprotected (36, 68). **(B)** Graph comparing experimental data and solvent accessible surface in the structure. Protection of the isolated P4-P6 domain from Fe(II)-EDTA cleavage (36, 68) indicated by pink bars. Accessibility

of C4' atoms to a probe of 1.6 Å radius was computed from molecule B of the x-ray structure. Region around nucleotide 236 was not calculated because of a nick in the P6b hairpin loop. **(C)** Stereo diagram of interdigitated riboses between the backbones of the P5b and P6a helices. These ribose zippers are characterized by directed hydrogen bonds between 2'-hydroxyls and bases, as shown. **(D)** Metal ions coordinate phosphates in the backbone between P5 and P5a. Solvent-flattened experimental electron density, contoured at 2.2  $\sigma$  above the mean, shows this magnesium ringed by phosphate oxygens. The density for the presumed magnesium is more than 5  $\sigma$  in the map.



**Fig. 6.** The core region. **(A)** At left, sheared A·A base pair in J4/5 with solvent-accessible surface overlaid; at right, A·U base pair shown for comparison. Arrows indicate N1 and N3 positions exposed in the minor groove. This A·A differs from a sheared G·A pair (40, 41) in that the bottom A is slightly closer to the purine above it, bringing the N6 of the bottom A within hydrogen-bonding distance of N3 of the top purine. This figure was produced with the use of MidasPlus (69) and RIBBONS (67). **(B)** Stereo drawing of sheared homopurine base pairs in J4/5, viewed down the helical axis. Stacking and hydrogen bonding interactions stabilize this unusual conformation.

39. M. Costa and F. Michel, *EMBO J.* **14**, 1276 (1995).  
 40. H. A. Heus and A. Pardi, *Science* **253**, 191 (1991).  
 41. H. Pley, K. Flaherty, D. McKay, *Nature* **372**, 111 (1994).  
 42. J. H. Cate *et al.*, *Science* **273**, 1696 (1996).  
 43. D. P. Bartel, M. L. Zapp, M. R. Green, J. W. Szostak, *Cell* **67**, 529 (1991).  
 44. R. D. Peterson, D. P. Bartel, J. W. Szostak, S. J. Horvath, J. Feigon, *Biochemistry* **33**, 5357 (1994).  
 45. J. L. Battiste, R. Tan, A. D. Frankel, J. R. Williamson, *ibid.*, p. 2741.  
 46. Free radicals produced by Fe(II)-EDTA in solution (presumably ·OH) cleave RNA without substantial specificity for base sequence or single- or double-strandedness and therefore are useful for probing the solvent accessibility of elements within complex folded RNAs (3–5).  
 47. Intermolecular hydrogen bonds between O2' atoms of duplex RNAs in a crystal have also been reported by H. Schindelin *et al.*, *J. Mol. Biol.* **249**, 595 (1995).  
 48. C. Guerrier-Takada, K. Haydock, L. Allen, S. Altman, *Biochemistry* **25**, 1509 (1986).  
 49. D. Smith, A. B. Burgin, E. S. Haas, N. R. Pace, *J. Biol. Chem.* **267**, 2429 (1992); D. Smith, and N. R. Pace, *Biochemistry* **32**, 5273 (1993).  
 50. A. M. Pyle, *Science* **261**, 709 (1993).  
 51. A. Jack, J. E. Ladner, D. Rhodes, R. S. Brown, A. Klug, *J. Mol. Biol.* **111**, 315 (1977); S. R. Holbrook, J. L. Sussman, R. W. Warrant, G. M. Church, S. H. Kim, *Nucleic Acids Res.* **4**, 2811 (1977).  
 52. J. H. Cate and J. A. Doudna, in preparation.  
 53. The two molecules (A and B) in the asymmetric unit of the crystal show different degrees of order at the 5' end. In molecule B, clear density is seen in the solvent-flattened experimental map for all but the first G of the sequence; in molecule A, the density is weak for the first 4 nt of the sequence.  
 54. F. Michel, A. D. Ellington, S. Couture, J. W. Szostak, *Nature* **347**, 578 (1990).  
 55. J. F. Wang, W. D. Downs, T. R. Cech, *Science* **260**, 504 (1993).  
 56. F. Michel, M. Hanna, R. Green, D. P. Bartel, J. W. Szostak, *Nature* **342**, 391 (1989).  
 57. J. F. Wang and T. R. Cech, *Science* **256**, 526 (1992).  
 58. A. M. Pyle, F. L. Murphy, T. R. Cech, *Nature* **358**, 123 (1992).  
 59. S. A. Strobel and T. R. Cech, *Biochemistry* **35**, 1201 (1996).  
 60. F. H. Crick, *J. Mol. Biol.* **38**, 367 (1968).  
 61. T. R. Cech, *Gene* **135**, 33 (1993).  
 62. B. Cuenoud and J. W. Szostak, *Nature* **375**, 611 (1995); R. R. Breaker and G. F. Joyce, *Chem. Biol.* **1**, 223 (1994); *ibid.* **2**, 655 (1995).  
 63. Z. Otwinowski, in *Isomorphous Replacement and Anomalous Scattering*, W. Wolf, P. R. Evans, A. G. W. Leslie, Eds. (SERC Daresbury Laboratory, Warrington, UK, 1991), pp. 80–86.  
 64. Collaborative Computing Project Number 4, *Acta Crystallogr. D* **50**, 760 (1994).  
 65. T. Jones *et al.*, *Acta Crystallogr. A* **47**, 110 (1991).  
 66. A. Brünger, *X-PLOR Manual, Version 3.1: A System for X-ray Crystallography and NMR* (Yale Univ. Press, New Haven, CT, 1993).  
 67. M. Carson, *J. Appl. Cryst.* **24**, 958 (1991).  
 68. F. L. Murphy, Y. H. Wang, J. D. Griffith, T. R. Cech, *Science* **265**, 1709 (1994).  
 69. The MidasPlus program was developed at the Computer Graphics Laboratory, University of California, San Francisco (supported by NIH RR-01081).  
 70. We thank H. Taube of Stanford University for the osmium hexammine triflate; T. Steitz, S. Schultz, A. Brünger, A. Ferré-d'Amaré, C. Correll, D. Engelman, A. Friedman, T. Griffin, J. Jaeger, A. M. Pyle, V. Rath, L. Rice, L. Silvin, S. Strobel, J. Wang, and members of the Center for Structural Biology at Yale for many helpful discussions; V. Carperos, L. Doherty, A. Ferré-d'Amaré, C. Correll, and J. Jaeger for help with data collection; V. Rath and Pfizer Inc. of Groton, CT, for use of data collection equipment; C. Ogata and W. Hendrickson for useful discussions and data collection time at beamline X-4A at the National Synchrotron Light Source, Brookhaven National Laboratory, Upton, NY; the staff at Cornell High Energy Synchrotron Source (CHESS) beamline A-1; A. Thompson and the staff at beamline BL-19 at the European Synchrotron Radiation Facility (ESRF), Grenoble, France; and A. Ferré-d'Amaré, V. Rath, and S. Strobel for review of the manuscript. Supported by the Lucille P. Markey Charitable Trust Scholars Program, the Donaghue Medical Research Foundation, NIH grant GM22778-21, the Searle Scholars Program, and the Beckman Young Investigators Program (J.A.D.), NIH training grant 5T32GM08283-07 (J.H.C.), American Cancer Society postdoctoral fellowship (B.L.G.), NSF grant MCB-9221307 (C.E.K.), Howard Hughes Medical Institute (T.R.C.), and the Keck Foundation (C.E.K. and T.R.C.).

15 July 1996; accepted 16 August 1996

## Location. Location. New Location...

Discover SCIENCE On-line at our new location and take advantage of these features...

- Fully searchable database of abstracts and news summaries
- Interactive projects and additional data found in the Beyond the Printed Page section
- Classified Advertising & Electronic Marketplace

Tap into the sequence below and see SCIENCE On-line for yourself.

NEW URL

<http://www.sciencemag.org>

SCIENCE