X-ray crystallography of large RNAs: heavy-atom derivatives by RNA engineering

B. L. Golden, A. R. Gooding, E. R. Podell and T. R. Cech

RNA 1996 2: 1295-1305

RNA (1996), *2*:1295–1305. Cambridge University Press. Printed in the USA.

X-ray crystallography of large RNAs: Heavy-atom derivatives by RNA engineering

BARBARA L. GOLDEN, ANNE R. GOODING, ELAINE R. PODELL, and THOMAS R. CECH

Department of Chemistry and Biochemistry, Howard Hughes Medical Institute, University of Colorado, Boulder, Colorado 80309-0215, USA

ABSTRACT

For small RNAs, isomorphous heavy-atom derivatives can be obtained by crystallizing synthetic versions that incorporate modified nucleotides such as iodo- or bromouridine. However, such a synthetic approach is not yet feasible for RNAs greater than ~40 nt. We have been investigating P4–P6, a 160-nt domain of the self-splicing *Tetrahymena* intron whose structure was solved recently (Cate JH et al., 1996, *Science 273*:1678–1685). To incorporate iodouridine, a two-piece RNA was constructed. The 5′ segment, containing the majority of the molecule, was transcribed in vitro using a self-processing hammerhead ribozyme to cleave the nascent transcript and give a homogenous 3′ end. A synthetic 5-iodouridine-containing RNA corresponding to the remainder of the sequence was then annealed to the transcribed piece of RNA. The resulting RNA appeared structurally and functionally sound as judged by nondenaturing gel electrophoresis and RNA cleavage assays. Four versions of this two-piece system with 5-iodouridine substitutions at different positions crystallized under the same conditions as the native RNA, yielding two useful heavy-atom derivatives of P4–P6. The position of the iodine atoms for the derivatives could be determined in the absence of phase information, and an interpretable electron density map was calculated using only the data from the two iodouridine derivatives. This approach is expected to be readily adaptable to other large, structured RNA molecules.

Keywords: 5-iodouridine; group I intron; RNA structure; *Tetrahymena* intron

INTRODUCTION

Recently, crystallographic studies of RNA have enjoyed a rebirth. This is due partly to technological advances that allow milligram quantities of RNA to be transcribed in vitro or, for smaller RNAs, to be synthesized chemically. A second advance is the development of sparse matrix approaches, which have allowed RNA molecules to be screened rapidly for their ability to crystallize (Doudna et al., 1993; Baeyens et al., 1994; Scott et al., 1995a). As with proteins, there is a second major stumbling block in crystallography, the preparation of heavy-atom derivatives of the molecule. Several properties of RNA make this a particularly difficult task. RNA is a polyanion and an intrinsic chelator of metals. Therefore, it is difficult to soak crystals in heavy-atom solutions and obtain one or two specific metal binding sites per molecule.

Engineered derivatives provide distinct advantages over the traditional "soak and pray" methodologies for obtaining heavy-atom derivatives. First, the site of the modification is known with respect to the primary sequence of the macromolecule. The location of the heavy atom (such as iodine covalently bound to uracil) within the unit cell can be used either as a marker when first building the molecule into the electron density map, or as a double-check that the model is built in the correct register (Lietzke et al., 1996). Second, the modification introduced is minimal, usually substitution of a single atom. Thus, the crystals of the modified macromolecule are often isomorphous, of high quality, and behave much like native crystals in terms of their diffraction limit and mosaic spread. Finally, because the heavy atom is already incorporated to full occupancy in the macromolecule, every good crystal obtained by the procedure will be a derivative.

In protein crystallography, engineered derivatives are obtained routinely by introducing cysteine residues to create mercury binding sites or by metabolic labeling of the protein with selenomethionine (Sun et al., 1987; Hatfull et al., 1989; Hendrickson et al., 1990). Unfortunately, neither of these techniques can be applied to nucleic acids. Heavy-atom derivatives for small

Reprint requests to: Thomas R. Cech, Howard Hughes Medical Institute, Department of Chemistry and Biochemistry, Box 215, University of Colorado, Boulder, Colorado 80309-0215, USA.

RNAs, which can be synthesized chemically, are obtained readily. In these situations, 5-bromo- or 5-iodouridine can be incorporated in place of specific uracil residues (Pley et al., 1994; Scott et al., 1995b). Alternatively, it should be possible to incorporate phosphorothioates in the RNA backbone to act as mercury binding sites, in analogy to a method developed for DNA crystallography (Yang & Steitz, 1995; S. Schultz, pers. comm.).

Unfortunately, this purely synthetic approach is not currently feasible for RNA molecules larger than about 40 nt. We are investigating a domain of the self-splicing *Tetrahymena* ribosomal RNA intron that, at 160 residues, is far too large to synthesize chemically. The domain consists of residues 104–261 and is stably folded independent of the rest of the intron (Murphy & Cech, 1993). It can also be provided *in trans* to the remainder of the intron to reconstitute catalytic activity (Doudna & Cech, 1995). In the intact intron, P4–P6 is the first region to achieve tertiary structure in either a thermodynamic or a kinetic folding experiment (Laggerbauer et al., 1994; Zarrinkar & Williamson, 1994; Downs & Cech, 1996) and the rest of the intron is stabilized by its correct folding. Diffraction-quality crystals of P4–P6 were obtained several years ago (Doudna et al., 1993) and, after an extensive search, an osmium hexammine derivative was obtained that allowed the structure to be solved (Cate et al., 1996).

Here we report a strategy for engineering heavy-atom derivatives of complex RNAs such as the P4–P6 domain. The approach involves splitting the RNA into two components, one of which is short and therefore suitable for chemical synthesis (Fig. 1). The two pieces are then reannealed and crystallized without ligation, essentially introducing a nick in the RNA backbone. We have incorporated 5-iodouridine into the small oligo; crystallized the derivatized two-piece domain, and obtained diffraction data suitable for calculation of an interpretable electron density map of P4–P6. This approach is likely to be generally useful for introduction of heavy atoms into specific sites of large RNA and RNP structures.

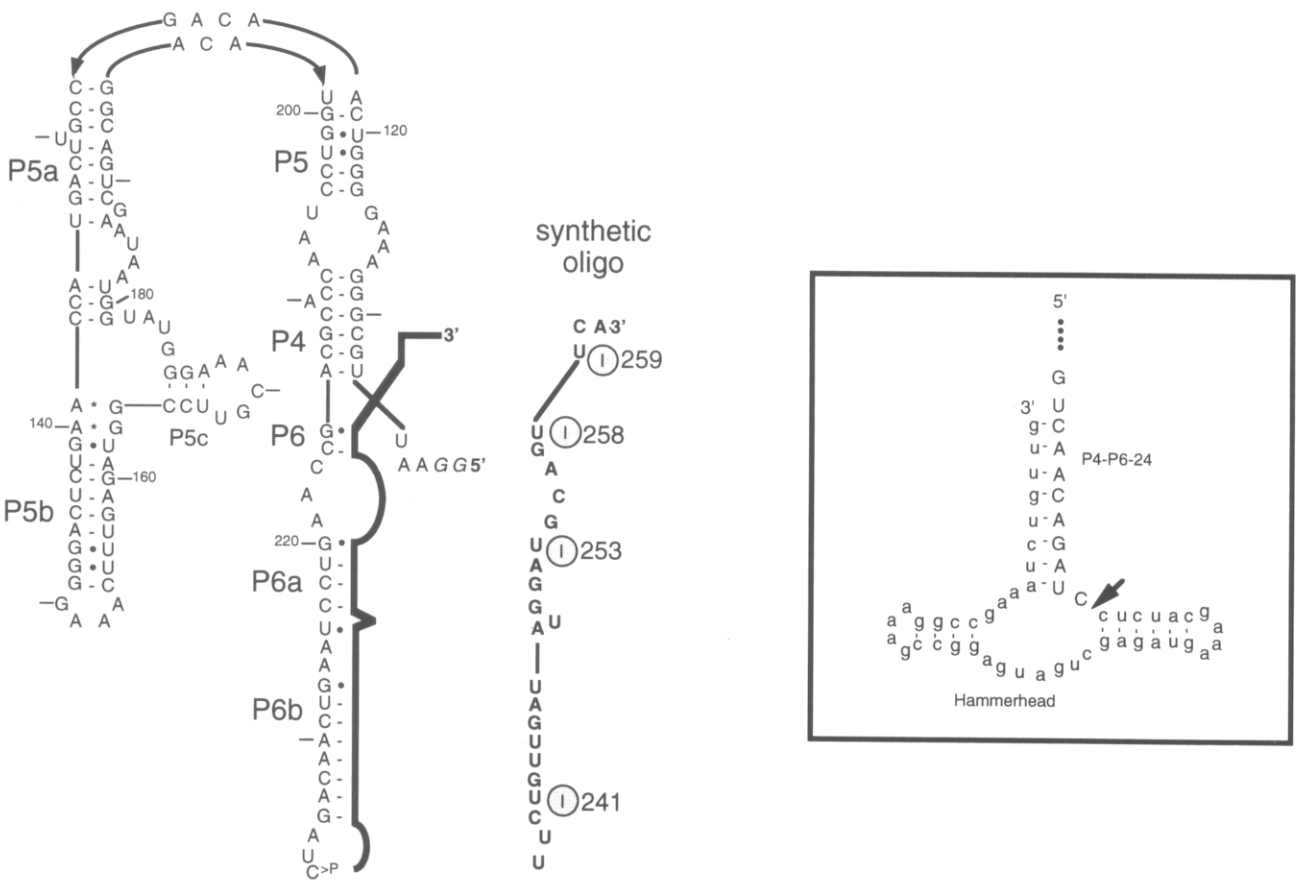


FIGURE 1. Engineered RNAs. The P4–P6 domain of the self-splicing *Tetrahymena* intron was synthesized in two pieces. Shown here is the P4P6–24 construct and the 24-residue oligo to which it anneals. The larger 5′ piece was transcribed using T7 RNA polymerase; a self-processing hammerhead sequence (boxed) produced a discrete 3′ end. The smaller RNA (bold line) was made by chemical synthesis and annealed to the larger RNA by base pairing and tertiary interactions. Crystals were obtained using oligonucleotides that contained 5-iodouracil at positions corresponding to nt 241, 253, 258, or 259 of the *Tetrahymena* intron.

RESULTS

Design of a two-piece P4–P6 system

Our semi-synthetic method for preparation of heavy-atom derivatives of P4–P6 involved transcribing a version of P4–P6 that was truncated at the 5′ or 3′ end and filling in the missing portion with a synthetic oligoribonucleotide containing the heavy atom. An initial approach, involving ligation of the two molecules with T4 DNA ligase and a DNA splint (Moore & Sharp, 1992) should, in principle, be an ideal technique for incorporating site-specific heavy atoms into RNA. However, full-length RNA prepared by this method did not crystallize well and, in some cases, was difficult to obtain in crystallographic quantities (J. Doudna, A. Gooding, & T. Cech, unpubl. results).

Therefore, a second approach was attempted, annealing a synthetic RNA to a larger transcribed piece without covalently linking the two molecules. This method produces a discontinuity in the RNA backbone. Nicks are often tolerated in RNA structures as assayed biochemically, probably because base pairing interactions are strong enough to compensate for the break. Thus, functional tRNAs and ribozymes can be generated by annealing two smaller RNA molecules or by circularly permuting the original sequences (for reviews see Long & Uhlenbeck, 1993; Pan & Uhlenbeck, 1993). The complication is that not all locations of nicks give functional molecules, and therefore, it is important to choose the site of the break carefully.

We determined sites where nicks occurred naturally in the RNA during the crystallization process, presumably by Mg^{2+}-catalyzed cleavage. RNA obtained from redissolved crystals was subjected to run-off reverse transcription. Primer extension analysis revealed stops at positions 23, 24, and 25 nt from the 3′ end (data not shown). These nucleotides reside in the terminal loop of stem P6b. Designed breaks in the chain at these positions were thought likely to have minimal effects because clearly the crystal matrix could accommodate cleavage of the full-length RNAs at these sites. The resulting 136 nt RNA, termed P4P6–24, is shown in Figure 1. A second construct, P4P6–17, terminated 17 residues from the 3′ end of the domain; it was designed to minimize the length of the synthetic RNA and yet provide a sufficient number of base pairs to anneal to the synthetic oligo.

The deletion constructs were prepared by run-off transcription using T7 RNA polymerase. When RNAs are generated by T7 RNA polymerase, there are uncoded nucleotides added onto the 3′ end of a significant percentage of the molecules (Lowary et al., 1986). Such heterogeneity may be a considerable problem in crystallography (Dock et al., 1984); furthermore, in the current situation, the uncoded nucleotides could interfere with binding of the synthetic oligos. Therefore, the truncated P4–P6 constructs were designed with a self-processing hammerhead at the 3′ terminus (Fig. 1). Processing of the hammerhead occurs during the transcription reaction and results in product RNA molecules that have a cyclic 2′-3′ phosphate at the 3′ terminus (Grosshans & Cech, 1991).

A deletion construct folds upon addition of the missing segment

The higher-order structure of the deletion constructs was evaluated initially using a mobility-shift assay. Correctly folded P4–P6 has a compact structure and therefore migrates faster than the unfolded domain on a native gel (Murphy et al., 1994). Radiolabeled P4–P6, P4P6–24, and P4P6–17 were analyzed with and without the addition of a stoichiometric amount of the appropriate compensating synthetic oligo. In the absence of the oligo, both deletion constructs migrated more slowly than full-length P4–P6 (Fig. 2). Because the deletion constructs have lower molecular weights than the intact domain, the slower mobility is presumably indicative of a less compact, unfolded structure. In the presence of the 17-mer, P4P6–17 migrated more rapidly than in its absence, but still had slower mobility than intact P4–P6. These results suggest that the P4P6–17 construct plus the 17-mer did not fold appropriately and, therefore, this molecule was not examined further. In contrast, P4P6–24 was rescued by the presence of the 24-mer. Furthermore, no additional bands that would indicate conformational heterogeneity in the two-piece system were observed, even at concentra-

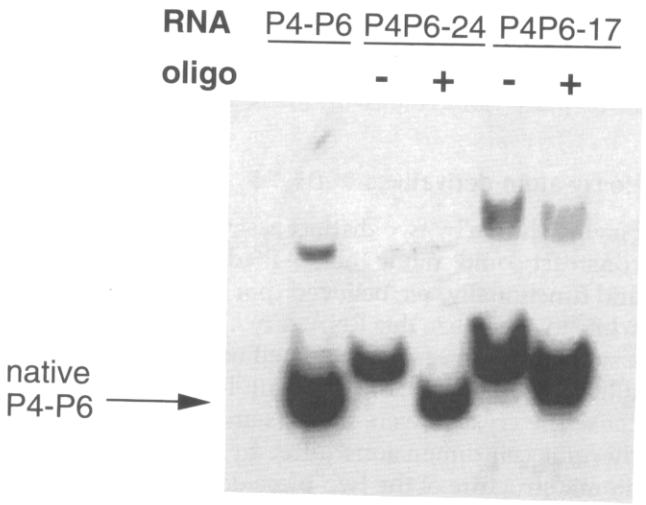


FIGURE 2. Native gel electrophoresis reveals a compact structure. Full-length P4–P6 or deletion constructs were renatured and analyzed on a 6% polyacrylamide gel containing 5 mM $MgCl_2$. Properly folded P4–P6 has a compact structure and, therefore, runs fast on this native gel. The deletion constructs are not folded correctly and have slower mobility. Structure of P4P6–24 is rescued by the addition of the appropriate 24-mer (**oligo**), but P4P6–17 runs slower than full-length P4–P6 even in the presence of a 17-mer oligo.

tions required for crystallization. Thus, by this analysis, P4P6–24 in the presence of the 24-mer appeared to have the same structure as the full-length domain.

A two-piece domain can assemble to give an active ribozyme

A functional test of proper folding of the RNA was then performed by incorporating the two-piece P4–P6 into a multipart group I intron and reconstituting activity analogous to 5′ splice-site cleavage (Fig. 3A). In the original system of Doudna and Cech (1995), the ribozyme is assembled from three RNAs: P1–P3, P3–P9, and the P4–P6 domain. We investigated whether the deletion construct in the presence of the compensating oligo would substitute for intact P4–P6. P4P6–24 and the 24-mer oligo were annealed using concentrations and protocols identical to those used for the crystallization trials. This reconstituted domain was then mixed with the P3–P9 domain and 5′ end-labeled P1–P3 RNA. Guanosine-dependent cleavage of P1–P3 by the ribozyme was monitored by appearance of the labeled 8-nt product.

P4P6–24 plus the unmodified 24-mer synthetic RNA reconstituted activity to the same extent as intact P4–P6 (Fig. 3B). Similar activity was obtained with the I258-substituted oligo, suggesting that nothing inherent to iodinated oligos interferes with ribozyme function. However, P4P6–24 annealed with the I259-substituted oligo gave an initial rate of splicing that was reduced about sevenfold relative to the other constructs. We have not investigated what step in folding, assembly, or reaction is inhibited by the presence of the iodine atom. However, it is not unexpected that an iodine at position 259 might interfere with the reaction, because a base triple interaction involving this residue (Michel et al., 1990) is absolutely required for activity of the three-piece ribozyme (Doudna & Cech, 1995).

Heavy-atom derivatives of P4–P6

Having shown by two distinct assays that the deletion construct could mimic the P4–P6 domain structurally and functionally, we believed that it would be worthwhile to crystallize this RNA. Crystals of P4P6–24 plus a synthetic 24-mer were obtained under the same conditions used to crystallize the full-length domain. Because the crystals were of the same space group and the unit cell dimensions differ by less than 1%, the atomic structure of the two-piece domain likely differs little from that of the full-length RNA. A slightly greater tendency for twinned crystals was observed with the two-piece domain. Whether this problem is inherent to this system or a result of heterogeneity in the synthetic RNA oligo was not investigated.

Crystals were obtained with iodines at four positions along the oligo, U241, U253, U258, and U259, but no crystals were obtained with an unmodified RNA oligo during these initial crystallization trials. This meant we had no "native" data for the two-piece system. Two derivatives, I253 and I258, were of sufficient quality for phase calculations (I241 and I259 did not diffract to high resolution). When the derivative data for the two-piece domain were scaled and merged with native data of the full-length domain, nonisomorphism between the crystals became apparent: the merging R-factor (R_{iso}) for native and derivative data sets was 3–4% higher than R_{iso} for the two derivative data sets. Had the native data from the full-length RNA been isomorphous, the differences between the two derivatives should be greater than the differences between one derivative and the native data. Therefore, the I253 derivative data were merged with the I258 derivative data and negative scattering factors were used to describe the iodine atoms of the I253 derivative (i.e., the iodine positions for the I253 derivative were given negative values for their occupancies). Three data sets were used to calculate phases: the merged isomorphous-difference data described above and a native anomalous difference data set for each derivative. Phasing statistics and inspection of the electron density maps confirmed that the phase information obtained in this manner was superior to the phases obtained when the derivative data were merged with native data from crystals of the one-piece RNA.

The position within the unit cell for one of the iodine atoms of the I253 derivative could be determined easily by analysis of a 3.5-Å difference Patterson map (Fig. 4). We were somewhat surprised that there were peaks in the Patterson maps corresponding to the iodine positions, because iodine is a relatively small heavy atom (53 electrons) and the asymmetric unit is very large, >100 kDa. A high degree of isomorphism between the two derivatives is indicated by the low value of R_{iso} (Table 1), and this may overcome some of the difficulties in locating an iodine atom (which is light compared to mercury or a lanthanide). Initial SIRAS phases were calculated using the position of this one iodine atom. These phases were used to calculate 3.5-Å difference Fourier maps for both iodine derivatives. The remainder of the heavy-atom sites appeared as 5–7σ peaks in these maps.

Calculation of electron density to 3.5 Å resolution using only the iodine data was straightforward (Table 1; Fig. 5). The molecular boundary was identified easily prior to any density modification. The map was solvent flattened using the automated procedure of Wang (1985) and this produced a map that was interpretable. There was clear, uninterrupted density for the backbone atoms; and side-chain density, although less strong, was also clearly present.

The current model of P4–P6 is based on data collected from a crystal grown in the presence of cobalt hexammine (Cate et al., 1996; Cate & Doudna, 1996), which is thought to be a good analogue for a hydrated magnesium ion (Tevelev & Cowan, 1995). In the elec-

A

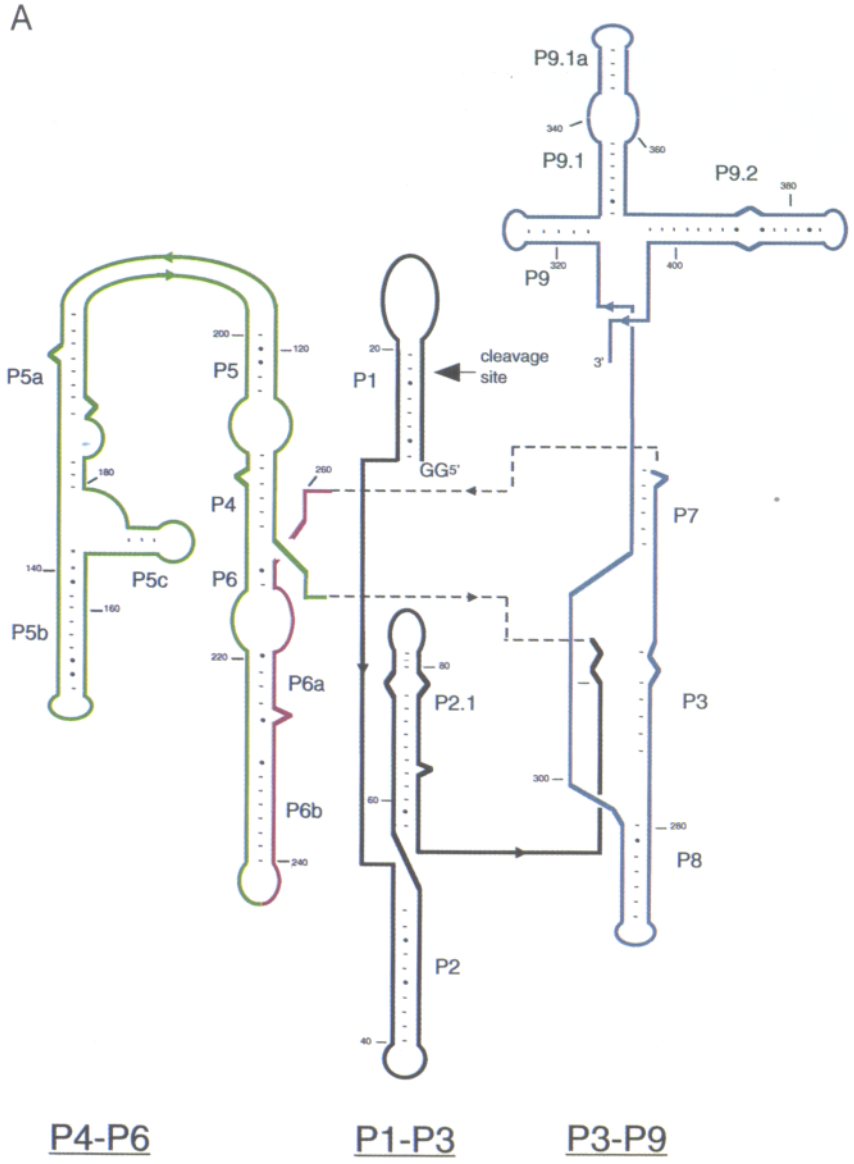
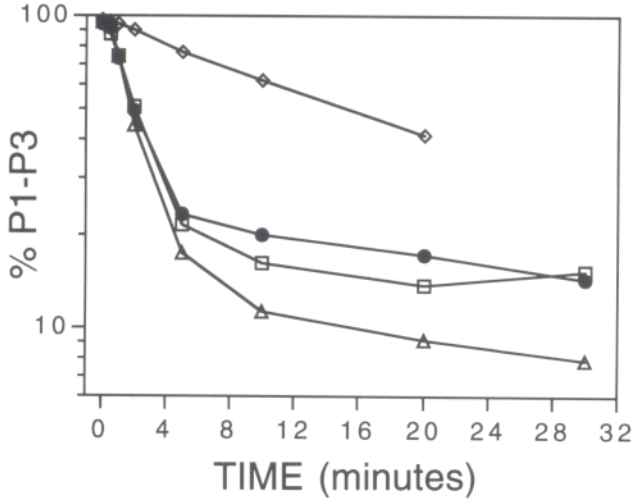


FIGURE 3. P4P6–24 can reconstitute ribozyme activity. P4P6–24 was tested for its ability to function in a three-part ribozyme system characterized previously (Doudna & Cech, 1995). **A:** The three-part ribozyme consists of three RNAs, P1–P3, P3–P9, and P4–P6. Dashed lines indicate sites of linkage in the intact ribozyme. When P4–P6 is assembled from P4P6–24 (green) and the 24-mer (violet), a four-part ribozyme is created. In the active ribozyme, P1–P3 is cleaved 8 nt from the 5′ end (indicated by a large arrow) upon the addition of guanosine. **B:** The ribozyme was reconstituted with full-length P4–P6, or with P4P6–24 plus a 24-mer synthetic RNA, and cleavage of the P1–P3 molecule was monitored. Reaction conditions were 100 nM P4–P6 (full-length or two-piece), 100 nM P3–P9, 1 nM P1–P3 substrate, 30 mM Tris, pH 7.5, 80 mM $MgCl_2$, 5 mM spermidine, and 5 mM GTP. Data are for reactions reconstituted with P4–P6 (●); P4P6–24 + an unmodified 24-mer RNA (△); P4P6–24 + an RNA oligomer iodinated at position 258 (□) or 259 (◇) of the *Tetrahymena* intron.

B



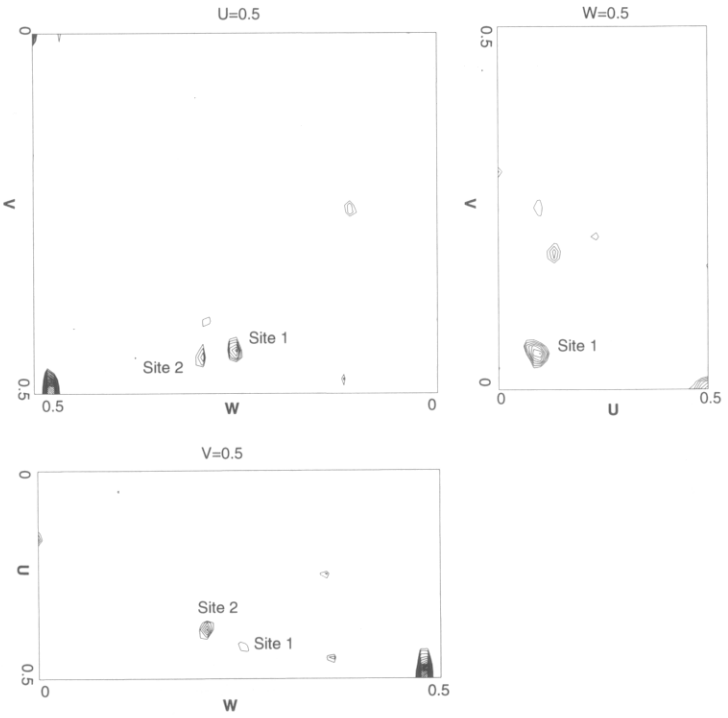


FIGURE 4. Patterson map of the I253 derivative. A difference Patterson map for the I253 derivative was calculated with data from 20 to 3.5 Å. The Harker sections at U = 0.5, V = 0.5, and W = 0.5 were contoured at 3σ and every 0.5σ. Peaks corresponding to the two iodine positions are labeled. The strong peaks near U = V = W = 0.50 are probably due to systematic build-up of signal near the symmetry axes.

TABLE 1. Crystallographic data.[a]

Space group: $P2_12_12_1$
a = 74.7 Å, b = 128.8 Å, c = 145.7 Å

	Merged iodine data sets	I253	I258
Resolution	–	3.3 Å	3.5 Å
Mosaicity	–	0.62°	0.89°
Completeness	–	100%	86%
R_{sym}	–	7.9%	7.4%
R_{iso}	8.9%	–	–
R_{Kraut}	8.1%	15.3%	11.0%
R_{Cullis}	59.0%	–	–
Figure of merit	0.441	0.361	0.308
Phasing power	2.02	1.56	1.21
Overall Figure of Merit		0.531	
After solvent flattening		0.814	

[a] Phases were calculated using a single merged isomorphous data set obtained by merging the two derivative data sets. The anomalous data were not merged and the iodine atoms were treated as native anomalous scatterers. The phases were calculated to 3.5 Å and solvent flattened using a solvent content of 60%.

$$R_{Cullis} = \frac{\Sigma\,||\,|F_{PH(obs)}| \pm |F_{P(obs)}|\,| - |F_H|\,|}{\Sigma\,||\,|F_{PH(obs)}| \pm |F_{P(obs)}|\,|};$$

for isomorphous data,

$$R_{Kraut} = \frac{\Sigma\,||\,|F_{PH(obs)}| - |F_{PH(calc)}|\,|}{\Sigma\,|F_{PH(obs)}|};$$

for anomalous data,

$$R_{Kraut} = \frac{\Sigma\,||\,|F_{PH^+(obs)}| - |F_{PH^+(calc)}|\,| + |\,|F_{PH^-(obs)}| - |F_{PH^-(calc)}|\,|}{\Sigma\,||\,|F_{PH^+(obs)}| + |F_{PH^-(obs)}|\,|};$$

phasing power = F_H/E_{RMS} (where F_H is the heavy-atom structure factor and E_{RMS} is the residual lack of closure).

tron density map calculated using only the iodine derivatives, we clearly see density in the major groove of the RNA at the two consecutive G:U wobble base pairs in P5b, one of the sites where cobalt hexammine is bound (Fig. 5). Because the data used in the present study are obtained from cobalt-free crystals, this density is interpreted as a hydrated magnesium ion visible in this solvent-flattened MIRAS map.

Difference Fourier maps were calculated using phases back-calculated from the P4–P6 model (Cate et al., 1996). This produces maps that are independent of the phases calculated using the iodine derivatives, and this prevents the occurrence of "ghost" peaks at the iodine positions due to residual phase error. These maps revealed density at the 5-position of the uracil base as expected (Fig. 6) for all four iodine derivatives.

In summary, production of a series of semisynthetic RNAs with single 5-iodouridine substitutions led to the identification of two usable iodine derivatives of the P4–P6 domain. Each of these derivatives introduced only one heavy atom per molecule of P4–P6 (two heavy atoms per asymmetric unit). Very few metal compounds can be soaked into a crystal and bind specifically to one site in an RNA molecule of this size. Even osmium hexammine, which provided the derivative that was ultimately used to solve the crystal structure, bound to eight sites within the asymmetric unit (Cate et al., 1996; Cate & Doudna, 1996).

DISCUSSION

We have developed a method for site-specific introduction of heavy atoms into large RNA molecules suitable

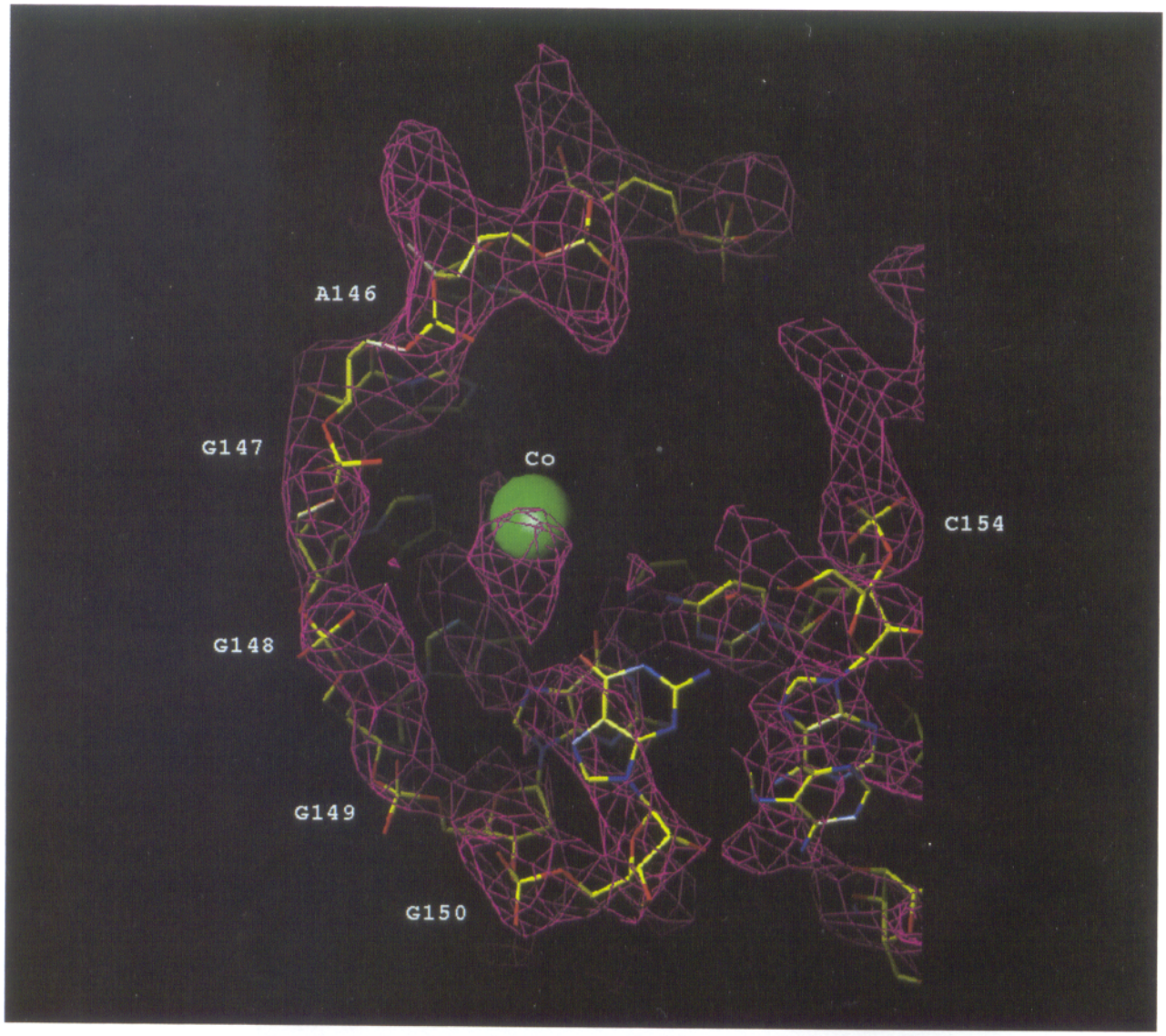


FIGURE 5. Electron density map. The solvent-flattened map of P4–P6 (magenta net) was calculated using only the isomorphous and anomalous differences from the iodine derivatives and is contoured at 1.1σ. It is superposed on the previously determined model (Cate et al., 1996). There is clear, unbroken density for the backbone in this region (near G147). Also, there is density at the site where cobalt hexammine is present in the model. The cobalt position is indicated by a green ball. Because there was no cobalt in the crystals examined in this study, the density probably corresponds to a hydrated magnesium atom. Figure displayed using the program O (Jones et al., 1991).

for use in X-ray crystallographic studies. The resulting derivatives meet the requirements of minimal perturbation of the structure, introduction of a sufficiently heavy atom, and ability to be synthesized on a milligram scale. Using this methodology, we were able to generate a 3.5-Å electron density map using only two iodine derivatives.

Adaptation of this approach to other large, structured RNAs may be accomplished through the following steps. (1) Identify a site within 40 nt of one end of the RNA where introduction of a nick gives no observable perturbation of the folded structure. (2) Synthesize the large component of the RNA by in vitro transcription of an appropriately engineered DNA template. (3) Synthesize the smaller component chemically, incorporating 5-iodouridine at successive uridine positions. (4) Renature stoichiometric mixtures of the large and small RNA components, testing for structural and functional integrity. (5) Crystallize the two-piece RNA using conditions successful for the intact parental molecule. Using this technique, we have recently obtained crystals of a 221-nt catalytic RNA that incorporated a synthetic oligomer (unpubl. results), suggesting that this method may be applicable to many RNA crystals.

Although we have obtained an interpretable map using this technique, we expect that there is still room

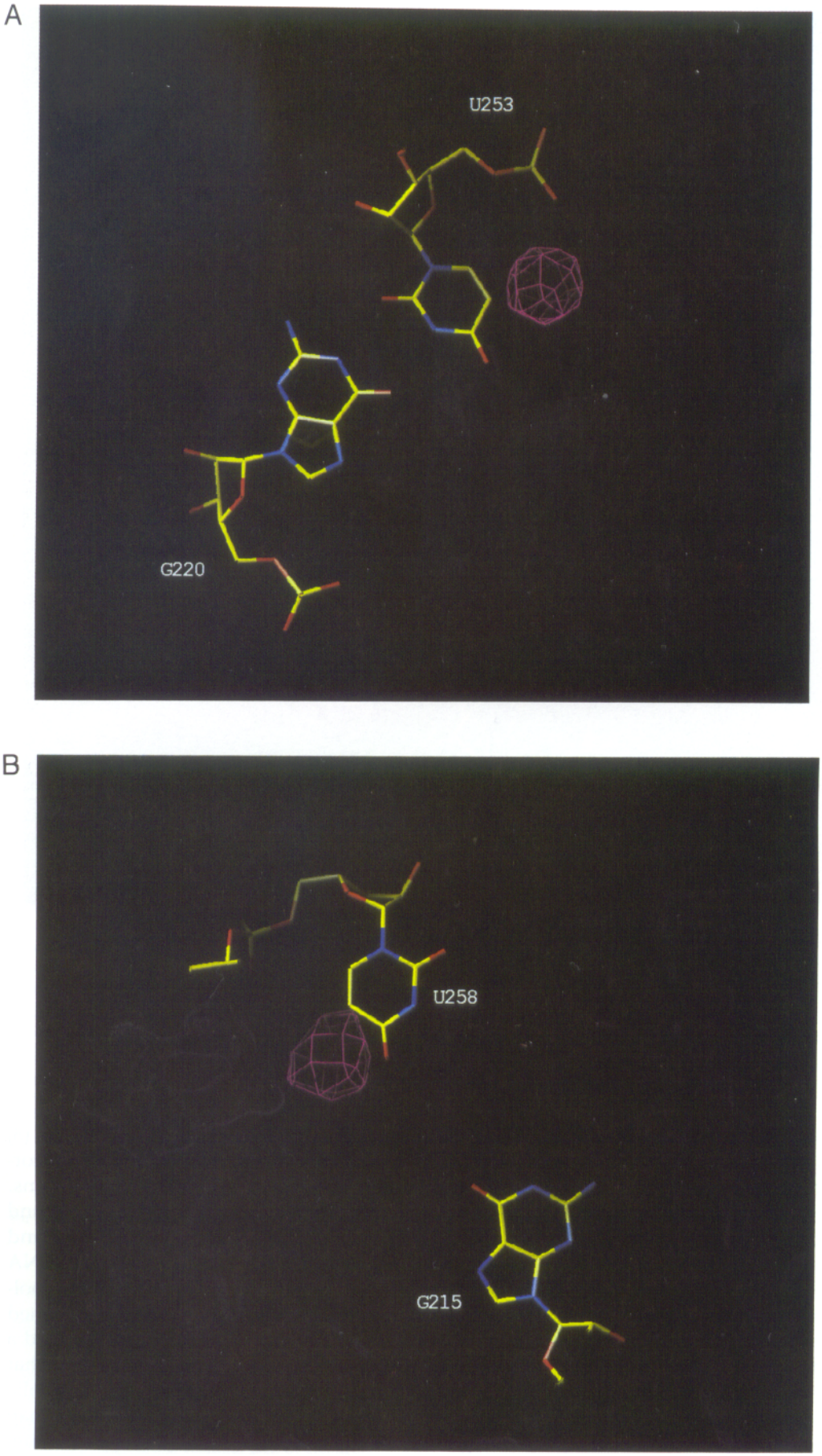


FIGURE 6.

for improvement. All of the crystallographic data used in this study were collected on a home source, CuK_α radiation from a rotating anode X-ray generator. Use of synchrotron radiation is expected to give even better results by extending the resolution, and perhaps the accuracy of the derivative data. The asymmetric unit of these crystals contains a dimer of P4–P6 domains. There are two iodine positions in the asymmetric unit and these positions could be used to determine the noncrystallographic symmetry operator that relates the two molecules within the dimer. This is valuable because it is allows averaging of the electron density for the noncrystallographically related molecules in the asymmetric unit. This process can improve the quality of the electron density map dramatically in some cases.

The two useful derivatives were obtained by substitution of uridines predicted to be involved in G:U wobble pairs. (The I253 site forms a genuine wobble pair, whereas the I258 G:U wobble is splayed apart in the crystal structure.) It seems possible that the wobble geometry better accommodates iodine substitution. If this does prove to be the case, G-C base pairs can be redesigned as wobble base pairs by substituting appropriate cytidine residues in the synthetic oligo with 5-iodouridine. Alternative strategies involve incorporation of iodine as 5-iodocytidine, which is also commercially available in the deoxy-form, and introduction of phosphorothioates in the RNA backbone to bind mercury.

MATERIALS AND METHODS

Plasmid construction

PCR was used to construct plasmids encoding 3′-terminal deletions of the P4–P6 domain. The 5′ PCR primer included the T7 promoter sequence and the 3′ PCR primer contained a sequence encoding a hammerhead ribozyme (Grosshans & Cech, 1991). PCR products were purified on 4% agarose gels and then cleaved with *Eco*R I and *Xba* I restriction endonucleases. These fragments were ligated into *Eco*R I/*Xba* I digested pUC19 using T4 DNA ligase. The ligation reactions were used to transform competent *Escherichia coli*, strain XL-1 Blue, using a Bio-Rad gene pulser set to 1.8 kV and 400 Ω. DNA was isolated from white colonies growing on LB/ampicillin/IPTG/X-gal plates. Clones containing the correct sequence were identified by restriction endonuclease digestion and sequencing.

Two constructs were made: pP4P6–17 generated an RNA that lacked the last 17 residues of P4–P6 after hammerhead processing, and pP4P6–24 generated an RNA that had 24 residues deleted from the 3′ end.

RNA synthesis and purification

P4–P6, P4P6–17, and P4P6–24 RNAs were made by runoff transcription using T7 RNA polymerase; plasmid template was digested with the appropriate restriction endonuclease (*Ear* I for the full-length construct and *Sma* I for the deletion constructs) and phenol/chloroform extracted prior to use. Typical reactions contained 4 mM each ATP, CTP, GTP, UTP, 30 mM $MgCl_2$, 2 mM spermidine, 40 mM DTT, 40 mM Tris, pH 8.0, 20 μg/mL linearized plasmid, and T7 RNA polymerase purified from *E. coli* strain BL21 pAR1219 (Davanloo et al., 1984). Reaction proceeded at 25 °C for 16 h. The RNA was recovered by ethanol precipitation and purified on 4–6% polyacrylamide gels. Unless otherwise indicated, polyacrylamide gels contained 7 M urea, 1.0× TBE (0.1 M Tris base, 0.83 M boric acid, and 1 mM EDTA), 29:1 acrylamide:(bis)acrylamide. The band containing RNA was identified by UV shadowing and excised from the gel. Gel slices were crushed and the RNA was eluted overnight into TEN (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 250 mM NaCl). Following removal of the polyacrylamide by filtration, the RNA was recovered by ultrafiltration using an Amicon concentrator. The concentration of unlabeled RNA was determined by UV absorbance.

Synthetic RNA oligonucleotides were made on an ABI 394 Synthesizer using phosphoramidites purchased from Glenn Research. The column matrix was then incubated overnight in 3:1 (v:v) NH_4OH:ethanol at 50 °C (iodinated oligos were incubated for 24 h at room temperature) to cleave the RNAs from the solid support and to deprotect the bases. The RNA was decanted from the solid matrix, dried under vacuum, and resuspended in 1 M tetrabutylammonium fluoride in tetrahydrofuran (Aldrich) for 24 h at room temperature to deprotect the 2′-hydroxyl. The RNAs were recovered by ethanol precipitation and purified on 12% polyacrylamide gels as described above. Iodinated RNAs were protected from ambient light during synthesis and purification.

RNA to be 5′ end-labeled was treated with calf intestinal phosphatase (New England Biolabs) to remove the triphosphate group at the 5′-terminus. The RNA was then phosphorylated with γ-^{32}P-ATP and T4 polynucleotide kinase (New England Biolabs) and separated from unincorporated label by gel purification as described above.

Reverse transcription mapping

Crystals of P4–P6 were grown as described below and washed multiple times with stabilization buffer (25% 2-methyl-2,4-pentanediol, 25 mM $MgCl_2$, 50 mM potassium cacodylate,

FIGURE 6. Difference Fourier maps. $F_o - F_c$ maps for the iodine derivatives were obtained using XPLOR (Brünger, 1993) and superposed on the model of P4–P6 (Cate et al., 1996). The phases for this map were back calculated from the model of full-length P4–P6 so that none of the density would correspond to "ghost" peaks at the iodine sites. Difference maps (magenta net) for I253 (**A**) and I258 (**B**) were displayed and contoured at 5σ and 3σ, respectively. Although U253 is involved in a G:U wobble base pair as predicted, U258 is quite distant from its proposed base pairing partner (Michel & Dujon, 1983; Waring et al., 1983). In all four cases where derivative data were obtained, density was clearly seen at the 5-position of the appropriate uracil base. Figure displayed using the program O (Jones et al., 1991).

pH 6.0, 0.5 mM spermine, and 10% isopropanol) before redissolving in 10 μL of a buffer containing 0.5 M β-mercaptoethanol and 0.75 M sodium acetate, pH 5.5. The RNA was ethanol precipitated and resuspended in 100 μL water.

Reverse transcriptase mapping was performed using a 5′ end-labeled DNA primer complementary to residues 240–261 of P4–P6. Annealing reactions contained 1.25 pmol of primer and 1 μL of redissolved RNA in 50 mM Tris, pH 8.3, 60 mM NaCl, and 10 mM DTT. After heating to 95 °C for 1 min, the reactions were cooled to room temperature and 0.14 U AMV reverse transcriptase, deoxynucleoside triphosphates (0.4 mM each dATP, dCTP, dGTP, and dUTP) and 6 mM $MgCl_2$ were added to synthesize cDNAs. Reverse transcription reactions containing 1.2 μg uncrystallized RNA and 80 μM ddATP, ddCTP, ddGTP, ddTTP, or no dideoxynucleoside triphosphate were performed to generate sequencing and background lanes. Products were analyzed on 8% polyacrylamide sequencing gels. Gels were dried and exposed to phosphor storage screens, which were imaged and quantitated using a Molecular Dynamics PhosphorImager.

Gel shift analysis

RNAs (typically 100 nM each molecule) were preincubated in 5 mM $MgCl_2$, 5% glycerol, and 50 mM potassium cacodylate, pH 6.0, at 60 °C for 10 min and equilibrated at room temperature for 1 h. Samples were then loaded onto native gels containing 6% acrylamide, 1.0× TBE, and 5 mM $MgCl_2$. Gels were run at 10 W at 20 °C, dried, and exposed to film.

Functional analysis

The tripartite ribozyme system described previously (Doudna & Cech, 1995) is comprised of 3 RNAs: P1–P3 (containing two nonnatural guanosines at the 5′ terminus, the last six residues of the 5′ exon, and the first 103 residues of the intron), P4–P6 (residues 104–261 of the intron), and P3–P9 (residues 262–409 of the intron). (The plasmid encoding P1–P3 was kindly provided by E. Doherty and J. Doudna.) Full-length P4–P6 (or P4P6–24 plus a 24-mer synthetic RNA) was first renatured in 62.5 mM potassium cacodylate, pH 6.0, and 6.25 mM $MgCl_2$ by heating to 60 °C for 10 min and slow cooling to room temperature in 1.5 h. The ribozyme was reconstituted by annealing 100 nM P4–P6 (or P4P6–24 + 24-mer), 100 nM P3–P9, and ~1 nM 5′ end-labeled P1–P3 in 30 mM Tris, pH 7.5, 5 mM $MgCl_2$, 5 mM spermidine at 50 °C for 10 min. The reactions were equilibrated at the reaction temperature, 40 °C, for 5 min. The reaction was initiated by simultaneous addition of 5 mM GTP and $MgCl_2$ to a final concentration of 80 mM. Aliquots of the reaction were removed and quenched by the addition of two volumes stop buffer (95% formamide, 50 mM EDTA, 0.1× TBE, 0.025% each xylene cyanol and bromophenol blue). Products were separated on 20% denaturing acrylamide gels, which were dried and exposed to phosphor storage screens. Release of the eight-residue product generated by ribozyme cleavage of the P1–P3 RNA was then quantitated as described above.

Crystallization and data collection

Crystals of full-length P4–P6 or the two-piece domain were grown essentially as described previously (Doudna et al., 1993). RNA was renatured in 50–62.5 mM potassium cacodylate and 5–6.25 mM $MgCl_2$ by heating for 10 min at 60 °C and cooled to room temperature over the course of an hour. The solution was adjusted to 1.25–2.5 mg/mL RNA, 25 mM potassium cacodylate, pH 6.0, 10 mM $MgCl_2$, 0.15–0.25 mM spermine, and 9.5–11.5% 2-methyl-2,4-pentanediol. Sitting drops of this solution were equilibrated at 30 °C against reservoirs containing 19–23% 2-methyl-2,4-pentanediol. Crystals typically grew to maximum dimensions of 500 × 300 × 250 μm over the course of 4–6 weeks.

For data collection, crystals were transferred to stabilization buffer for 24 h. The crystals were then mounted on a nylon loop, flash frozen in liquid propane, and stored in liquid nitrogen. Intensity data were collected at −170 °C with a R-AXIS II image plate system using CuK_α radiation from a Rigaku RU300 rotating anode X-ray generator (Molecular Structures Corporation, Houston, Texas). The crystal-to-screen distance was 150 mm and data were collected as 1.5° oscillation images. The data were reduced and scaled using the packages DENZO and SCALEPACK (Otwinowski, 1993).

Data processing

The software package PHASES (Furey & Swaminathan, 1996) was used for manipulation of crystallographic data, phase calculation, and density modification. Phase calculations were complicated by slight nonisomorphism between native data obtained from crystals of full-length P4–P6 and derivative data obtained from crystals of the two-piece P4–P6 molecule. No crystals were obtained for the two-piece domain when an unmodified oligonucleotide was incorporated; therefore, there was no native data for the two-piece domain. One of the derivatives, I253, was used as the native data set, and negative scattering factors were explicitly input for the iodine atoms of the I253 derivative data set.

The heavy-atom position for one of the iodine atoms was determined by analysis of difference Patterson maps calculated using the I258 data set as "native" structure factors. Using this initial iodine site, rough initial phases were determined and these were used to calculate 3.5 Å difference Fourier maps for both I253 and I258 derivatives. All of the heavy-atom positions could be determined from these difference Fourier maps. Both isomorphous and anomalous differences were used in phase refinement and calculating the electron density map at 3.5 Å resolution. The initial electron density map was then solvent flattened using a solvent content of 60% by the automated method of Wang (1985).

ACKNOWLEDGMENTS

We thank Steve Schultz, Craig Kundrot, Evelyn Jabri, and the University of Colorado Structural Biology Group for invaluable advice and discussion; Vasili Carperos for help with data collection; Jennifer Doudna and Jamie Cate for coordinates of the P4–P6 model. We acknowledge Art Zaug for help with reverse transcription and Alex Szewczak for providing several RNAs used in the functional analysis. This work was supported by the Howard Hughes Medical Institute and the W.M. Keck Foundation. B.L.G. is American Cancer Society postdoctoral fellow and T.R.C. is an Investigator of the Howard Hughes Medical Institute and an American Cancer Society Professor.

Manuscript accepted without revision October 21, 1996

REFERENCES

Baeyens KJ, Jancarik J, Holbrook SR. 1994. Use of low-molecular-weight polyethylene glycol in the crystallization of RNA oligomers. *Acta Crystallogr D 50*:764–767.

Brünger A. 1993. *X-PLOR manual, version 3.1: A system for X-ray crystallography and NMR*. New Haven, Connecticut: Yale University Press.

Cate JH, Doudna JA. 1996. Metal binding sites in the major groove of a large ribozyme domain. *Structure*. Forthcoming.

Cate JH, Gooding AR, Podell E, Zhou K, Golden BL, Kundrot CE, Cech TR, Doudna JA. 1996. Crystal structure of a group I ribozyme domain: Principles of RNA packing. *Science 273*:1678–1685.

Davanloo P, Rosenberg AH, Dunn JJ, Studier FW. 1984. Cloning and expression of the gene for bacteriophage T7 RNA polymerase. *Proc Natl Acad Sci USA 81*:2035–2039.

Dock A, Lorber B, Moras D, Pixa G, Thierry J, Giege R. 1984. Crystallization of transfer ribonucleic acids. *Biochimie 66*:179–201.

Doudna JA, Cech TR. 1995. Self-assembly of a Group I intron active site from its component tertiary structural domains. *RNA 1*:36–45.

Doudna JA, Grosshans CA, Gooding AR, Kundrot CE. 1993. Crystallization of ribozymes and small RNA motifs by a sparse matrix approach. *Proc Natl Acad Sci USA 90*:7829–7833.

Downs WD, Cech TR. 1996. Kinetic pathway for folding of the *Tetrahymena* ribozyme revealed by three UV-inducible crosslinks. *RNA 2*:718–732.

Furey W, Swaminathan S. 1996. PHASES-95: A program package for the processing and analysis of diffraction data from macromolecules. *Methods Enzymol*. Forthcoming.

Grosshans CA, Cech TR. 1991. A hammerhead ribozyme allows synthesis of a new form of the *Tetrahymena* ribozyme homogenous in length with a 3′ end blocked for transesterification. *Nucleic Acids Res 19*:3875–3880.

Hatfull GF, Sanderson MR, Freemont PS, Raccuia PR, Grindley NDF, Steitz TA. 1989. Preparation of heavy-atom derivatives using site-directed mutagenesis: Introduction of cysteine residues into γδ Resolvase. *J Mol Biol 208*:661–667.

Hendrickson WA, Horton JR, LeMaster DM. 1990. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): A vehicle for direct determination of three-dimensional structure. *EMBO J 9*:1665–1672.

Jones TA, Zou JY, Cowan SW, Kjeldgaard M. 1991. Improved method for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A 47*:110–119.

Laggerbauer B, Murphy FL, Cech TR. 1994. Two major tertiary folding transitions of the *Tetrahymena* catalytic RNA. *EMBO J 13*:2669–2676.

Lietzke SE, Barnes CL, Berglund JA, Kundrot CE. 1996. The structure of an RNA dodecamer shows how tandem U-U base pairs increase the range of stable RNA structures and the diversity of recognition sites. *Structure 4*:917–930.

Long DM, Uhlenbeck OC. 1993. Self-cleaving RNAs. *FASEB J 7*:25–30.

Lowary P, Sampson J, Milligan J, Groebe D, Uhlenbeck OC. 1986. A better way to make RNA for physical studies. In: van Knippenberg PH, Hilbers CW, eds. *Structure and dynamics of RNA. NATO ASI Ser., Ser. A. 110*:69–76.

Michel F, Dujon B. 1983. Conservation of RNA secondary structures in two intron families including mitochondrial-, chloroplast- and nuclear-encoded members. *EMBO J 2*:33–38.

Michel F, Ellington AD, Couture S, Szostak JW. 1990. Phylogenetic and genetic evidence for base-triples in the catalytic domain of group I introns. *Nature 347*:578–580.

Moore MJ, Sharp PA. 1992. Site-specific modification of pre-mRNA: The 2′-hydroxyl groups at the splice sites. *Science 256*:992–997.

Murphy FL, Cech TR. 1993. An independently folding domain of RNA tertiary structure within the *Tetrahymena* ribozyme. *Biochemistry 32*:5291–5300.

Murphy FL, Wang YH, Griffith JD, Cech TR. 1994. Coaxially stacked rna helices in the catalytic center of the *Tetrahymena* ribozyme. *Science 265*:1709–1712.

Otwinowski Z. 1993. DENZO. In: Sawyer L, Isaacs N, Bailey S, eds. *Data collection and processing*. Warrington, England: SERC Daresbury Laboratory. pp 56–62.

Pan T, Uhlenbeck OC. 1993. Circularly permuted DNA, RNA and proteins—A review. *Gene 125*:111–114.

Pley HW, Flaherty KM, McKay DB. 1994. Three-dimensional structure of a hammerhead ribozyme. *Nature 372*:68–74.

Scott WG, Finch JT, Grenfell R, Fogg J, Smith T, Gait MJ, Klug A. 1995a. Rapid crystallization of chemically synthesized hammerhead RNAs using a double screen procedure. *J Mol Biol 290*:327–332.

Scott WG, Finch JT, Klug A. 1995b. The crystal structure of an all-RNA hammerhead ribozyme: A proposed mechanism for RNA catalytic activity. *Cell 81*:991–1002.

Sun DP, Alber T, Bell JA, Weaver LH, Matthews BW. 1987. Use of site-directed mutagenesis to obtain isomorphous heavy-atom derivatives for protein crystallography: Cysteine-containing mutants of phage T4 lysozyme. *Protein Eng 1*:115–123.

Tevelev A, Cowan JA. 1995. Metal substitution as a probe of the biological chemistry of magnesium. In: Cowan JA, ed. *The biological chemistry of magnesium*. New York: VCH Publishers. pp 53–83.

Wang BC. 1985. Resolution of phase ambiguity in macromolecular crystallography. *Methods Enzymol 115*:90–112.

Waring RB, Scazzocchio C, Brown TA, Davies RW. 1983. Close relationship between certain nuclear and mitochondrial introns. Implications for the mechanism of RNA splicing. *J Mol Biol 167*:595–605.

Yang W, Steitz TA. 1995. Crystal structure of the site-specific recombinase γδ resolvase complexed with a 34 bp cleavage site. *Cell 82*:193–207.

Zarrinkar PP, Williamson JR. 1994. Kinetic intermediates in RNA folding. *Science 265*:918–924.