Article No. jmbi.1999.3398 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (2000) 295, 541–556

JMB

Crystallization and Structure Determination of a Hepatitis Delta Virus Ribozyme: Use of the RNA-Binding Protein U1A as a Crystallization Module

Adrian R. Ferré-D'Amaré[1] and Jennifer A. Doudna[2]*

[1]*Department of Molecular Biophysics and Biochemistry and*

[2]*Howard Hughes Medical Institute, Yale University P.O. Box 208114, New Haven CT 06520-8114, USA*

**Corresponding author*

Well-ordered crystals of a genomic hepatitis delta virus (HDV) ribozyme, a large, globular RNA, were obtained employing a new crystallization method. A high-affinity binding site for the spliceosomal protein U1A was engineered into a segment of the catalytic RNA that is dispensable for catalysis. Because molecular surfaces of proteins are more chemically varied than those of RNA, the presence of the protein moiety was expected to facilitate crystallization and improve crystal order. The HDV ribozyme-U1A complex crystallized readily, and its structure was solved using standard techniques for heavy-atom derivatization of protein crystals. Over 1200 Å^2 of the solvent-accessible surface area of the complex are involved in crystal contacts. As protein-protein interactions comprise 85 % of this buried area, these crystals appear to be held together predominantly by the protein component of the complex. Our crystallization method should be useful for the structure determination of other biochemically important RNAs for which protein partners do not exist or are experimentally intractable. The refined model of the complex (R-free = 27.9 % for all reflections between 20.0 and 2.3 Å) reveals an RNA with a deep active site cleft. Well-ordered metal ions are not observed crystallographically in this cavity. Biochemical results of previous workers had suggested an important role in catalysis for cytosine 75. The pyrimidine base of this residue is buried at the bottom of the active site in an environment that could raise its pK_a value. We propose that this highly conserved cytosine may be the general base that catalyzes the transesterification.

Keywords: RNA; crystallization; HDV ribozyme; U1A protein; MAD

Introduction

Despite considerable advances in techniques for RNA synthesis and characterization, the preparation of crystals of large RNAs that are suitable for crystallographic analysis remains challenging (reviewed by Holbrook & Kim, 1997). Well-ordered crystals may be difficult to obtain because the molecular surfaces of tightly-folded or ''globular'' RNAs are dominated by a regular array of negatively charged phosphate groups. During crystallization, this might result in neighboring molecules packing subtly out of register, resulting in poor long-range order. If this were the case, altering the molecular surface of globular RNAs by introducing distinctive functional groups could facilitate formation of better-ordered crystals. Indeed, Golden *et al.* (1997) showed that some changes in the sequence of a solvent-exposed loop of a group I intron fragment resulted in improved crystals. We found that introduction of a GAAA tetraloop and its 11 nucleotide (nt) receptor sequence into target molecules produced RNAs that crystallized more readily, presumably by facilitating intermolecular docking (Ferré-D'Amaré *et al.*, 1998b).

We have now expanded the systematic modification of RNA molecular surfaces to include

Present address: A. R. Ferré-D'Amaré, Division of Basic Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, WA 98109-1024, USA.

Abbreviations used: HDV, hepatitis delta virus; RBD, RNA binding domain; MAD, multiwave length anomalous dispersion; FOM, figure of merrit; S.D., standard deviation.

E-mail address of the corresponding author: doudna@csb.yale.edu

0022-2836/00/030541–16 $35.00/0

proteins, greatly facilitating crystallization of RNAs as part of the resulting complexes. Since proteins have a larger variety of surface functional groups than folded RNAs, we modified our target molecule, the genomic strand hepatitis delta virus (HDV) ribozyme, to contain a high-affinity cognate site for a sequence-specific RNA-binding protein. The engineered RNA was complexed with a strongly basic protein, and the complex subjected to crystallization trials. Well-ordered crystals were readily obtained, and the structure was solved by preparing co-crystals of the RNA with a selenomethionyl protein. We propose that this strategy, the use of an RNA binding protein and its cognate site as a crystallization module, can be employed for other large RNAs.

The HDV ribozyme is a ~85 nt long self-cleaving RNA whose activity is indispensable for the viability of the human pathogen, a satellite virus of hepatitis B virus (reviewed by Lai, 1995). The products of the self-cleavage reaction are a ''ribozyme'' moiety with a 5′-hydroxyl group, and a ''leader'' RNA segment with a 2′,3′-cyclic phosphate terminus (Figure 1(a)). These products are identical to those of the hammerhead, hairpin and VS ribozymes (McKay & Wedekind, 1999), and RNase A during its first half-reaction (Richards & Wyckoff, 1971). The reaction mechanism for the HDV ribozyme is thought to consist of an in-line nucleophilic attack on the phosphorus of the scissile phosphate by the 2′-hydroxyl on the adjacent ribose, which *via* a pentacovalent phosphorus intermediate, leads to breakage of the phosphoester bond to the 5′-hydroxyl group of the ribozyme moiety (Been & Wickham, 1997). The rate of this reaction could be enhanced in several ways, including optimal alignment of the nucleophile and the leaving group, deprotonation of the attacking 2′-hydroxyl by a base, partial neutralization of the phosphate to make the phosphorus a better electrophile, stabilization of the pentacovalent intermediate by a positively charged group, and efficient reprotonation of the leaving group by an acid (Figure 1(a)). Our structure suggests that the HDV ribozyme employs several of these means to achieve catalysis.

Our design of HDV ribozyme constructs for crystallization was guided by previous biochemical work (reviewed by Been & Wickham, 1997) that had demonstrated that one of the constituent stem-loops of the ribozyme, P4-L4 (see Figure 1 for nomenclature), could be freely altered or truncated without adversely affecting the *in vitro* activity of the RNA. We replaced P4-L4 with a stem-loop from the small nuclear RNA U1 that is a high affinity binding site for the human protein U1A. The RNA binding properties of U1A have been well characterized, and the three-dimensional structure of the U1A protein RNA-binding domain (RBD) in complex with a 21 nt RNA stem-loop has previously been determined at 1.9 Å resolution (Oubridge *et al.*, 1994). A 72 nt chimeric RNA in complex with the U1A-RBD ultimately yielded co-crystals that diffracted synchrotron X-radiation beyond 2.2 Å resolution. Selenomethionyl U1A-RBD was used to solve the structure by multi-wavelength anomalous dispersion (MAD). Some features of the structure have been presented elsewhere (Ferré-D'Amaré *et al.*, 1998a). Here we discuss crystallization strategy and crystallographic methodology, aspects of crystal packing, and some structural features that bear on the mode of action of this catalytic RNA.

Results and Discussion

Crystallization strategy

Our crystallization strategy involved generating a series of RNA-protein complexes that differed from each other systematically, and subjecting them to a conventional sparse-matrix crystallization screen of modest size. In our experience, screening numerous crystallization candidates is more expeditious in producing useful crystals than exhaustively screening crystallization conditions for one or a few candidate macromolecules.

The RNA-binding domain of U1A has a number of characteristics that make it well-suited for use as a crystallization reagent. First, its structure both free and in complex with its cognate site is known, and demonstrates that residues 1-98 comprise a compact, globular domain (Nagai *et al.*, 1990; Oubridge *et al.*, 1994). Compactness is a common characteristic of macromolecules that crystallize well (Ferré-D'Amaré & Burley, 1997; Cohen, 1996). Second, the U1A-RBD binds very tightly to its cognate site (K_d ~10^{-11} M; van Gelder *et al.*, 1993). Third, the available structures show that the protein-RNA interface comprises both polar (e.g. salt bridges between phosphate groups and basic amino acid residues) and non-polar interactions (e.g. stacking of aromatic amino acid residues between RNA bases). Therefore, the protein-RNA interaction is expected to be strong both in low and high ionic strength solutions, and the complex is unlikely to dissociate under various crystallization conditions. Fourth, U1A-RBD can bind to its cognate site in two different presentations. In addition to its helix-terminal binding site from the U1 RNA (Figure 1(a)), the protein binds to two adjacent helix-internal bulges in the polyadenylation regulatory element of the 3′ UTR of its mRNA (Allain *et al.*, 1996). These two modes can be exploited to generate RNAs engineered to contain exogenous U1A binding sites in a variety of molecular contexts. Fifth, Nagai and co-workers have prepared a number of U1A-RBD constructs with mutations in solvent-exposed residues. The mutations affect the solution and crystallization properties of the proteins (Oubridge *et al.*, 1995), providing an additional source of variation that can be exploited in the combinatorial search for good crystals.

We have shown that the genomic HDV ribozyme, which in its native form is difficult to

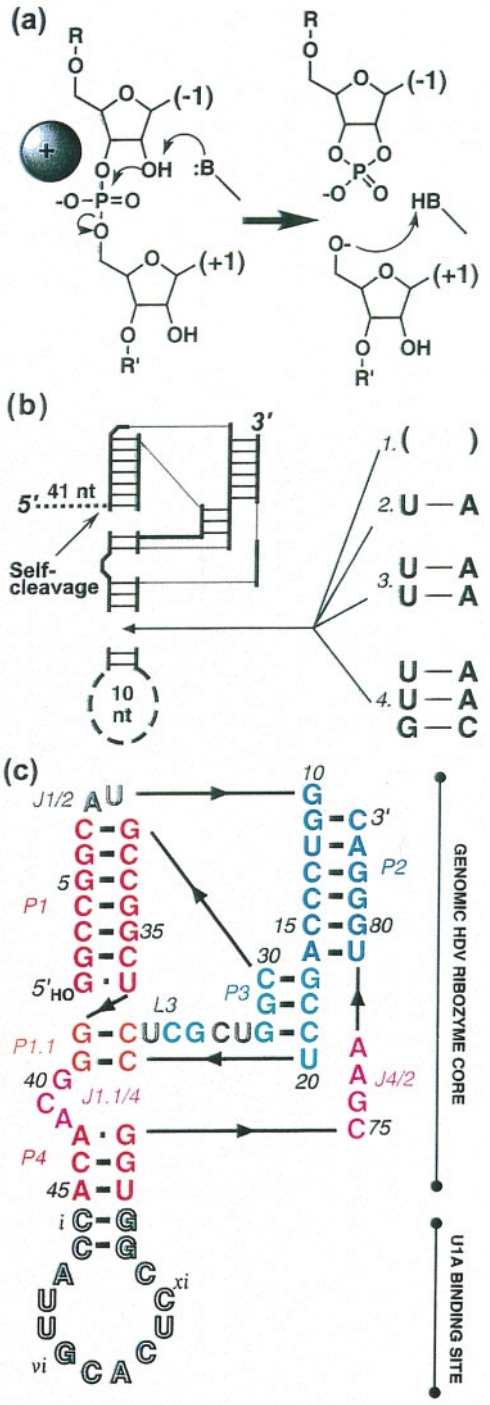

Figure 1. Engineering the U1A crystallization module into the genomic HDV ribozyme. (a) Hypothetical chemical mechanism of the reaction catalyzed by the HDV ribozyme. (b) Schematic of the length variation strategy. The secondary structure of the genomic HDV ribozyme (top left) and the U1A binding site from stem-loop II of U1 small nuclear RNA (bottom left) are shown in simplified form, separated by a horizontal gap. This gap was bridged with 0 to 3 spacer base-pairs, in order to generate multiple crystallization candidates. Constructs were transcribed *in vitro*, preceded by a 41 nt leader sequence, from which the ribozymes self-cleaved, to generate the mature RNAs used for crystallization. (c) Sequence and secondary structure of the construct (no spacer base-pairs between the ribozyme and U1A-binding site moieties) whose crystal structure was determined. Nucleotides corresponding to the genomic HDV ribozyme are numbered with Arabic numerals starting with the guanine bearing the leaving group of the self-scission reaction; nucleotides of the U1A binding site are in lower-case Roman numerals. RNA segments are named P for base-paired region, J for joining region, and L for loop, numbered from the 5′ end of the molecule, as by Ferré-D'Amaré *et al.* (1998a); the same color scheme is employed in all subsequent Figures, except for Figure 6. The riboses of residues 22, 24, 27, 41, 74-76, 77, v, vii and x have C2′-*endo* puckers.

crystallize, can be engineered to crystallize readily by introducing a crystallization module into its P4 stem. In our previous work, we employed a GAAA tetraloop and its 11 nt receptor sequence placed so as to stack coaxially and be available for intermolecular but not intramolecular association. While the engineered RNAs produced large crystals, these did not diffract X-rays beyond 7 Å resolution (Ferré-D'Amaré *et al.*, 1998b). Because our approach was partially successful, we decided to introduce the cognate site for U1A into the same stem-loop. In order to minimize any distortion of the structure of the HDV active site by introduction of the crystallization module, we placed the U1A binding site as a loop capping the distal end of P4, as far away as possible from the core of the ribozyme.

A series of RNAs were prepared in which the HDV ribozyme and helix-terminal U1A-binding moieties were separated by a variable number of spacer base-pairs (Figure 1(b)). These paired nucleotides were expected to form segments of *A*-form double helix that would coaxially stack and rigidly connect the two RNA moieties. The various constructs differed in the separation and helical twist between the two domains. This is analogous to the length-variation strategy routinely employed for DNA-protein co-crystallization (Aggarwal, 1990; Schultz *et al.*, 1990).

We employed self-cleaved ribozymes for crystallization, because HDV ribozymes are known to be prone to misfold (Been & Wickham, 1997; Perrotta & Been, 1998). Use of self-cleavage products ensured that only molecules that had successfully traversed the transition state, and thus folded correctly, were employed for crystallization (see Methods). Previous biochemical work suggested that the precursor and the product have very similar structures (Been & Wickham, 1997; Duhamel *et al.*, 1996). Nucleotides 5′ of the cleavage site are unlikely to participate in forming the architecture of the active site, since the HDV ribozyme has no sequence requirements in the 5′ leader sequence, will cleave efficiently after a single nucleotide, and the identity of this −1 position nucleotide has only modest effects on cleavage rates (Been & Wickham, 1997). Binding of U1A-RBD to the constructs did not significantly affect their self-cleavage activity

(Ferré-D'Amaré *et al.*, 1998a). The chimeric HDV ribozymes all bound U1A-RBD with 1:1 stoichiometry and migrated as homogeneous single bands in electrophoretic mobility shift assays (not shown).

Because bacteriophage polymerases add variable numbers of random nucleotides at the 3′ terminus of their transcripts, constructs incorporating either HDV or VS ribozymes (Ferré-D'Amaré & Doudna, 1996) 3′ to the ribozyme core were prepared in order to produce molecules with homogeneous 3′ termini. However, none of these successfully cleaved the RNA, possibly because of the extraordinary stability (Been & Wickham, 1997) of the HDV ribozyme moiety. Thus, the crystallization constructs were all products of HDV ribozyme self-cleavage with homogeneous 5′-hydroxyl and heterogeneous 3′ termini (Figure 1(b)).

Crystallization trials were carried out with four RNA constructs complexed to either wild-type or a double mutant U1A-RBD that had been used successfully in prior structure determinations (Oubridge *et al.*, 1994; Allain *et al.*, 1996). This set of eight complexes was subjected to the 48-condition sparse matrix by Jancarik & Kim (1991) supplemented with 1.25 mM $MgCl_2$ and 1 mM spermine in all conditions. Trials were set up in triplicate at 4 °C, 20 °C, and 30 °C, since temperature appears to have strong effects on crystallization of nucleic acids (e.g. see Schindelin *et al.*, 1995).

For all eight complexes, chunky crystals were obtained under a variety of conditions in the initial screens. Those crystals that reached sizes greater than 50 μm in each dimension were extensively washed, dissolved, and analyzed by polyacrylamide gel electrophoresis to ascertain that they contained intact RNA. Crystallization experiments under otherwise identical conditions that lacked the protein yielded, at best, very thin needles, implying participation of the protein moiety in crystal formation. Crystallization conditions were then further optimized to obtain larger crystals suitable for diffraction experiments. The first co-crystal form to reach ~100 μm in the small dimension was that shown in Figure 2(a) (crystal form I). These diffracted X-rays from a rotating anode only to ~7 Å resolution. However, crystal form II (Figure 2(b)) diffracted X-rays to 3.3 Å resolution. These tetragonal crystals, which contain the RNA shown in Figure 1(c), were then cryoprotected, flash-cooled, and taken to a synchrotron radiation source, where diffraction was observed to 2.9 Å resolution.

In order to solve the structure, we then prepared the selenomethionyl version of the double-mutant U1A-RBD construct present in the crystals. Surprisingly, the complex with the selenomethionyl protein failed to crystallize at all under conditions similar to those employed for the complex with the normal protein. It has been observed previously that due to the reduced solubility of selenomethionine relative to methionine, the concentration of precipitant used in crystallization often must be decreased (Doublié, 1997). Although this did not produce crystals in our case, further exploration of crystallization conditions eventually resulted in the new crystal form III (Figure 2(c)). These rhombohedral crystals diffracted X-rays more strongly than the tetragonal form II crystals obtained with conventional protein, and were employed for all subsequent work. Examination of several hundred crystallization experiments carried out with the sulfur protein showed that in one vapor diffusion drop, its RNA complex produced rhombohedral crystals with cell parameters similar to those of the selenomethionyl co-crystals. Thus, it appears that substitution of methionine with selenomethionine greatly favors the infrequent rhombohedral form III and suppresses formation of tetragonal crystals of this particular complex (see below). Subjecting the RNA alone to the crystallization conditions employed for obtaining either form II or III co-crystals did not yield any crystalline matter.

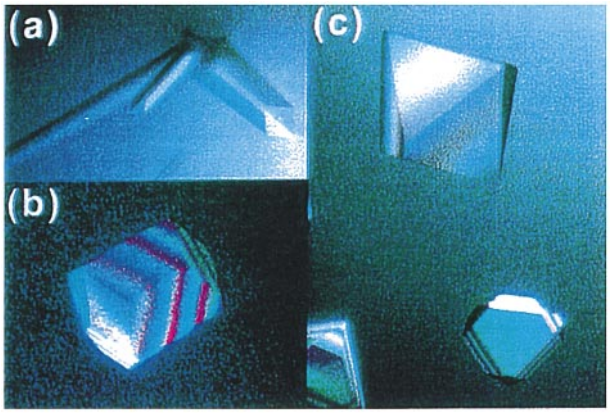

Figure 2. Examples of HDV ribozyme-U1A RBD co-crystals. (a) Tetragonal form I crystals of the HDV ribozyme self-cleavage product with a one base-pair spacer (Figure 1(a)) complexed with wild-type U1A-RBD. Crystal form I has a primitive lattice ($a = 196$ Å, $c = 143$ Å) and diffracts X-rays from a rotating anode to ~7 Å resolution. (b) Tetragonal crystal form II (space group $I422$ or $I4_122$, $a = 80.8$ Å, $c = 221.5$ Å) of the HDV ribozyme construct with no spacer base-pairs (Figure 1(a) and (b)) complexed with the double mutant (Y31H, Q36R) U1A-RBD. An 87 % complete data set to 2.9 Å (overall $R_{merge} = 6.7$ %, 4.4-fold redundancy) was collected from a similar crystal using synchrotron X-radiation. (c) Rhombohedral crystal form III (space group $R32$, $a = 109.15$ Å, $c = 190.68$ Å) of the same RNA as form II, complexed with the selenomethionyl version of the same double mutant U1A-RBD protein. Crystals such as these yielded the diffraction data used for structure determination and refinement.

Structure determination and refinement

Structure determination by MAD has been described (Ferré-D'Amaré *et al.*, 1998a). Unlike crystal form II, crystal form III grows more readily in the presence of 0.1-0.3 mM cobalt (III) hexam-

mine. Crystals of the unrelated P4-P6 domain RNA also grew better with cobalt hexammine, and the structure of that RNA was determined by growing crystals in the presence of osmium (III) hexammine. It was found there that the osmium bound tightly to a few specific sites, and a multiwavelength experiment around the osmium L_{III}-edge produced the necessary phase information (Cate *et al*., 1996). Anomalous difference Fourier syntheses calculated with amplitude data measured at the osmium L_{III} peak from a form III crystal grown in the presence of osmium hexammine, and phases from the selenomethionyl MAD experiment, failed to produce any significant peaks at locations other than at the previously located selenium atoms. Hence, despite the presence of the osmium in the crystals, which is borne out by X-ray fluorescence, the osmium hexammine is not specifically bound. Since metal hexammines are thought to replace hydrated magnesium ions (Kieft & Tinoco, 1997), this suggests that high-affinity sites for hydrated magnesium are not present in the HDV ribozyme.

Initial phasing (de la Fortelle & Bricogne, 1997) and model building (Jones *et al*., 1991) were carried out with diffraction data extending to 2.9 Å resolution measured at the bending magnet beamline X4A of the National Synchrotron Light Source. The overall figure of merit (FOM) for acentric reflections between 50 and 2.9 Å was 0.635 (0.425 between 3.2 and 2.9 Å). Data collection with 0.9209 Å radiation from the 24-pole wiggler at beamline F1 of the Cornell High-energy Synchrotron Source yielded a dataset that was 95.5 % complete to 2.3 Å resolution (see Ferré-D'Amaré *et al*., 1998a for statistics). Initially, form III crystals diffracted X-rays to beyond 2.1 Å resolution, but in order to measure the weak, high-resolution terms, long exposure times were needed, and even at a temperature of 100 K the crystals suffered substantial radiation decay by the time the 2.5 to 2.2 Å shell was 75 % complete. The 2.3 Å dataset was measured from a fresh crystal using shorter exposure times.

Phases were recalculated with SHARP (de la Fortelle & Bricogne, 1997) using the three energies from the original selenium MAD experiment (high-energy remote, absorption peak and rising inflection point), one energy from the ''osmium'' experiment (low energy remote), and the new 2.3 Å data, reduced without merging anomalous pairs, with the latter as the reference, high-energy remote, dataset. This resulted in a phase set with the same FOM as before between 50 and 2.9 Å but now with some phase information between 2.9 and 2.3 Å resolution (mean FOM for acentric reflections in these shells was 0.08). Density modification (Abrahams & Leslie, 1996) of these phases resulted in substantially improved electron density maps (mean overall FOM for acentrics after modification 0.81; 0.66 between 2.45 and 2.3 Å) that allowed the remaining model-building ambiguities to be resolved.

Refinement was against the 2.3 Å structure factor amplitudes, reduced merging anomalous pairs, and the unmodified phase probability distributions calculated to 2.3 Å resolution, using the maximum likelihood target with experimental phase probability distributions implemented in CNS (Adams *et al*., 1997). The standard nucleic acid parameter set of CNS with purely repulsive van der Waals terms was employed, initially keeping all ribose puckers in the canonical C3′-*endo* form. Residual difference Fourier maps were then inspected for electron density peaks near the furanose rings, and where significant density was found, the sugars were restrained to C2′-*endo* puckers for the next round of minimization. The puckers were kept if the residual density had disappeared or decreased substantially. Ultimately, 11 of the 72 nucleotides were modeled with C2′-*endo* puckers (see the legend to Figure 1). Throughout refinement, the value of *R*-free (calculated with 10 % of the data) was monitored, to the exclusion of the conventional crystallographic residual.

The final model comprises 2369 non-hydrogen atoms (protein residues 4-98, all RNA residues, 64 water molecules and 12 magnesium ions), and has a free-*R* factor of 27.9 % and a conventional residual of 24.6 %, with root-mean-square differences (r.m.s.d.) from ideal values of 0.011 Å and 1.4° for bond lengths and angles, respectively. PROCHECK (Laskowski *et al*., 1993) analysis shows that 88 amino acid residues lie in the most favored regions of the Ramachandran plot, and the remaining six lie in additionally allowed regions. Residual electron density is present near the selenium atoms in $|F_o| - |F_c|$ syntheses, and could reflect partial oxidation of the seleniums, or imperfections in the parameter set employed. RNA residues 26 and 27 are poorly ordered.

A portion of the final composite simulated-annealing omit map (Brunger *et al*., 1998) corresponding to part of the protein-RNA interface is shown in Figure 3(a). As expected at this resolution, amide carbonyls of the peptide backbone are well-resolved, as are RNA features such as 2′-hydroxyls. Twenty well-ordered water molecules in the protein-RNA interface (of which two are shown in the Figure) are conserved between the HDV-U1A complex and the U1A-RNA 21-mer complex described by Oubridge *et al*. (1994). Protein carbon atoms of the two models superimpose closely, except at the termini of the polypeptide chains (r.m.s.d. = 0.28 Å for 80 carbon pairs; Figure 3(b)). When superimposed using the protein carbon atoms, the U1A-binding domain of the HDV ribozyme and the 21-mer of the complex described by Oubridge *et al*. (1994) superimpose closely as well (except for three nucleotides, see below and Figure 3(b)), underscoring the similarity of the protein-RNA association.

The overall structure of the HDV ribozyme-U1A RBD complex is shown in simplified form in Figure 4. The RNA is comprised of five base-paired segments P1, P1.1, P2, P3, and P4, arranged in two

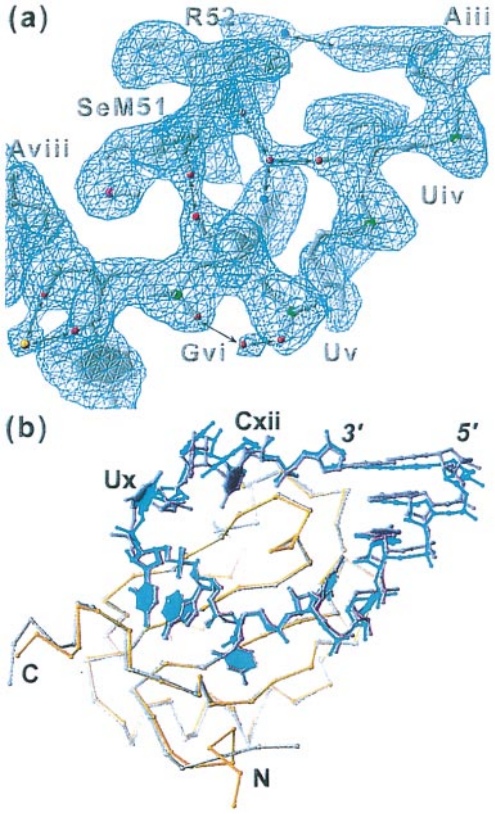

Figure 3. Protein-RNA interactions. (a) Portion of a composite simulated annealing σA-weighted $2|F_o| - |F_c|$ omit map calculated with phases from the final model, contoured at 1.5 S.D. U1A-RBD residues 48 through 52 are shown, with the exception of the side-chain of lysine 50. The selenium of selenomethionine 51 makes a van der Waals contact (3.6 Å) with the ribose O4′ of Aviii. The two water molecules shown as red spheres are conserved in the U1A-RNA 21-mer structure by Oubridge *et al*. (1994). The yellow sphere is a putative Mg^{+2} ion. (b) Superposition of the HDV ribozyme-U1A RBD complex (gray protein, carbon atoms 4 through 98, and blue RNA) on the U1A-RNA 21-mer structure by Oubridge *et al*. (1994) (yellow protein, carbon atoms 2 through 97, and magenta RNA). U1A-RBD carbon atoms of residues 10 through 90 from both structures (r.m.s.d. = 0.28 Å for 81 atom pairs) were used to superimpose the structures. Note the similarity of the RNA moieties (residues ii-xiii are shown), except for nucleotides Ux through Cxii in the upper left-hand corner (see the text and Figure 6) r.m.s.d. values for all atoms in the RNA moieties shown are 1.1 Å, and 0.7 Å for the C1′ atom pairs (when aligned using the proteins). Of the nucleotides shown in this Figure, v, vii, and x have C2′-*endo* puckers. Figures 3, 4, 6, 8-10 were generated with RIBBONS (Carson, 1991).

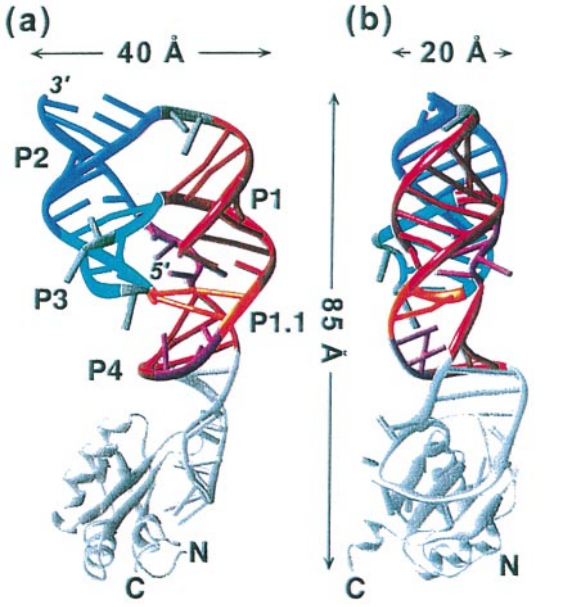

Figure 4. Ribbon representation of the three-dimensional structure of the HDV ribozyme-U1A RBD complex. The RNA backbone is displayed as a thick ribbon, with bases as short cylinders. The protein is represented by flat (β-strands) and coiled (α-helices) ribbons connected by thin tubes. (a) Front view of the complex with helical regions and chain termini marked. (b) Side view of the complex with the P1-P1.1-P4 coaxial helix stack in the foreground.

coaxial stacks. The two stacks are interconnected by five strand crossovers, resulting in a fold that can be described as a nested double pseudoknot. This fold appears to be essential for activity of both the genomic and antigenomic HDV ribozymes (Wadkins *et al*., 1999). The U1A protein and its cognate site lie at the tip of P4, away from the active site whose position, in turn, is defined by the 5′ hydroxyl leaving group of the self-scission reaction. The convoluted fold of the ribozyme produces a deep, well-defined active site cleft. This cleft may help align the scissile phosphate and the nucleophile for efficient cleavage. There is excellent agreement between the structure of the self-cleavage product and previous chemical protection and mutagenesis data. This implies that the HDV ribozyme does not undergo major conformational changes during the transesterification reaction, and that the structures of the precursor and transition-state forms of the RNA are similar to that of the reaction product (Ferré-D'Amaré *et al*., 1998a).

The poor order of nucleotides 26 and 27, which line the outside edge of the active site cavity, could reflect high mobility of this segment. Since we have crystallized the product of self-cleavage, the disorder around the active site could result from the absence of the leader sequence. We indicated above that nucleotides in the leader sequence are unlikely to participate in forming the architecture of the active site. However, the ribose of the −1 nucleotide bears the nucleophile of the transesterification reaction (Figure 1(a)), and presumably is held in place through interactions with the rest of the ribozyme, possibly with L3 (Ferré-D'Amaré

et al., 1998a). The association of the −1 ribose requires covalent attachment of the phosphoribose backbone, as the HDV ribozyme is not known to catalyze the ligation reaction (Been & Wickham, 1997). Consistent with this, when we soaked 2′:3′-cyclic CMP (the cleavage product of a minimal leader sequence) into our crystals, no binding could be detected by difference Fourier methods (not shown). It remains to be seen if L3 is better ordered in structures of HDV ribozyme precursors.

A Wilson analysis (Wilson, 1949) of the diffraction data employed for refinement (Figure 5(a)) shows that the overall *B*-factor is almost 62 Å^2. Analysis of the datasets used for phase calculation imply similarly high overall thermal parameters (not shown). This, and the fact that the value of the crystallographic residual was not allowed to diverge substantially from the *R*-free, result in a comparatively large estimate of mean coordinate error of 0.4 Å (Figure 5(b)).

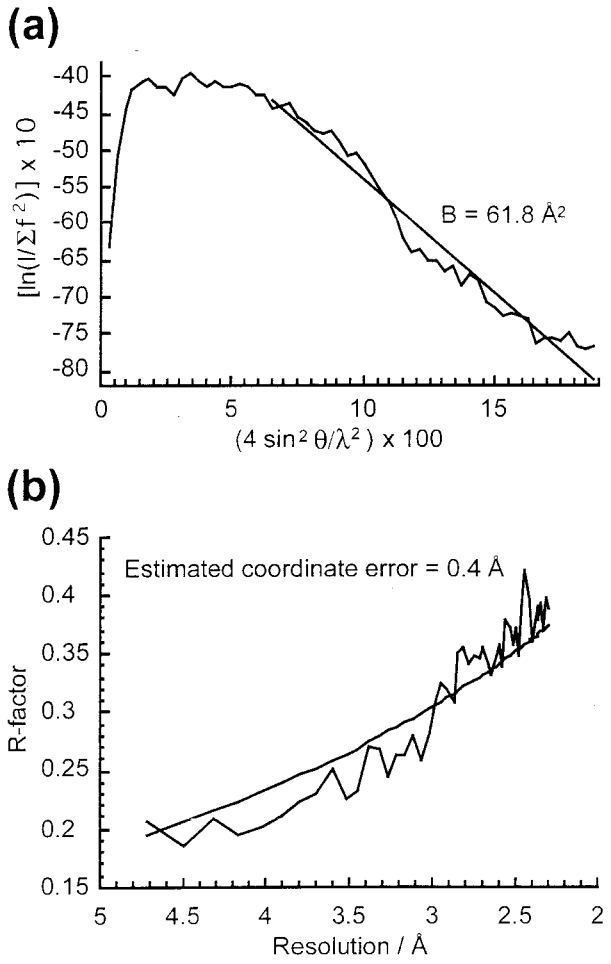

Figure 5. Wilson and Luzzati plots. (a) Wilson plot (Wilson, 1949) of the intensity data used for refinement. The overall *B*-factor estimated from a linear fit between 3.9 and 2.3 Å resolution is 61.8 Å^2. (b) Luzzati plot (Luzzati, 1952) which yields an estimate of the mean coordinate error of the final model to be 0.4 Å.

Protein-protein interactions are the dominant crystal contacts

Although the premise for using an RNA-binding protein as a crystallization module was that the basic protein would make both protein-protein and protein-RNA crystal contacts, in our HDV-U1A RBD co-crystals, the major contacts are between protein molecules in adjacent asymmetric units. The elongated (Figure 4) RNA-protein complex has a total solvent accessible surface area of 16,558 Å^2. (For comparison, the total solvent accessible surface area of the free U1A-RBD is 5858 Å^2.) Within the complex, association of U1A with the RNA buries 876 Å^2 of the solvent accessible surface of the protein. The most extensive crystal contact results from the packing across a crystallographic dyad of two U1A-RBDs from neighboring asymmetric units (Figure 6(a)). This interaction involves the entire solvent exposed surface of helix A and most of turn 4 (as defined by Nagai *et al.*, 1990), and buries 592 Å^2 of solvent accessible area on each protein molecule. An additional protein-protein crystal contact is formed around a crystallographic 3-fold where the N and C termini of three neighboring U1A-RBDs interact head-to-tail, burying 237 Å^2 in each contact (474 Å^2 per U1A-RBD, Figure 6(b)).

Neither of these crystal contacts is present in the U1A-RBD RNA 21-mer complex crystals by Oubridge *et al.* (1994). In that structure, the asymmetric unit comprises three U1A-RBDs, each bound to a 21 nt RNA stem-loop. Trimers related by a crystallographic dyad stack head-to-head. However, the dyad-related U1A proteins pack against each other employing a surface which is unrelated to that involved in the dyad interaction in the HDV-U1A RBD crystals (Figure 6(c)). The non-crystallographic 3-fold axis that relates the proteins in one asymmetric unit also results in a packing scheme which is distinctly different from that present around the crystallographic 3-fold of the HDV-U1A RBD co-crystals (Figure 6(d)). In addition to these protein-protein contacts, which together bury approximately 1000 Å^2 of the solvent accessible surface of each protein molecule, the crystals by Oubridge *et al.* (1994) are stabilized by RNA-RNA contacts between the stems of RNAs from adjacent asymmetric units, which bury 404 Å^2 of the solvent accessible surface of one nucleic acid chain, and an RNA-protein crystal contact that buries a further 432 Å^2 of the RNA (this is in addition to the cognate U1A-RNA interaction which buries ~ 850 Å^2).

The extent of RNA-protein and RNA-RNA crystal contacts is much smaller in the HDV ribozyme-U1A co-crystals. Aside from the cognate interaction, there are no significant RNA-protein contacts in these crystals. One RNA-RNA contact involves the stacking of the bases of the extruded G76 of two molecules across a dyad (Figure 7(a)). This buries 40 Å^2 of solvent accessible surface area per guanine residue. The second RNA-RNA crystal

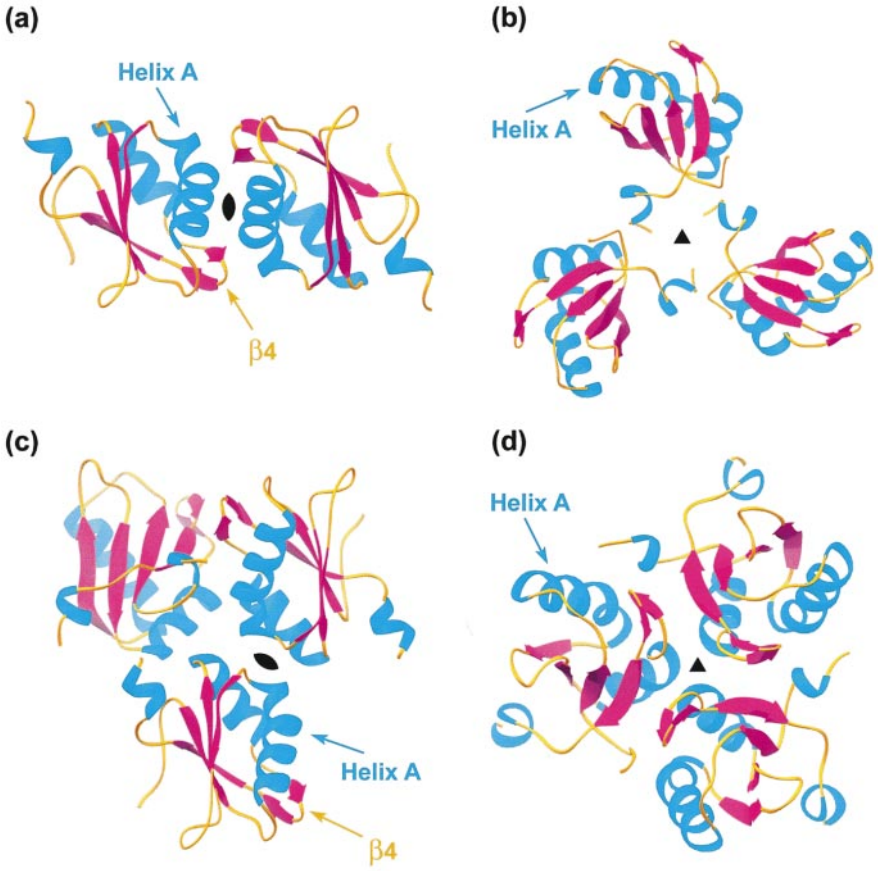

Figure 6. Comparison of protein-protein crystal contacts from the HDV-U1A RBD co-crystals (this work) and the RNA 21-mer U1A-RBD co-crystals by Oubridge *et al*. (1994). In this Figure helices, strands and coils are colored blue, pink and yellow, respectively. (a) Crystal contact involving helix A and turn 4 of two U1A RBD across a crystallographic dyad (indicated by the lune) in the HDV-U1A RBD co-crystals. (b) Crystal contacts around the crystallographic 3-fold axis (indicated by the triangle) of the HDV-U1A RBD co-crystals. (c) Crystal contacts across a crystallographic dyad in the RNA 21-mer U1A-RBD co-crystals (PDB accession number 1urn). The bottom protein is oriented similarly to the left protein in (a) to underscore the difference in packing. 155 Å^2 of the solvent accessible surface of the bottom U1A-RBD are buried in its interface with the top left protein, and 300 Å^2 in its interface with the top right U1A-RBD. (d) Three U1A-RBDs are present in the asymmetric unit of the crystals by Oubridge *et al.* (1994). The position of the approximate 3-fold axis is indicated. Approximately 500 Å^2 of the solvent accessible surface of each U1A-RBD is buried in the formation of this crystal contact. Compare with the different trimeric arrangement shown in (b).

contact also involves base stacking. Here, the three pyrimidines (Ux, Cxi, and Cxii) at the 3′ end of the U1A binding loop (Figure 1(c)), which do not interact with the protein, stack against the top of helix P2 from an adjacent HDV ribozyme RNA. This happens around a crystallographic 3-fold (Figure 7(b)), and buries 143 Å^2 on P2. The three pyrimidines are poorly ordered in the RNA 21-mer complex structure of Oubridge *et al.* (1994). In our structure, the stacking on P2 orders them. This explains the difference in the conformation of the pyrimidines in the two structures (Figure 3(b)). The selenium atom of selenomethionine 51 contacts the ribose of Aviii of its cognate RNA (Figure 3(a)). Selenium has a van der Waals radius 5% larger than sulfur. If this interaction stabilizes a conformation of the U1A-binding loop that favors the P2 crystal contact, it may have led to the selenomethionine protein crystallizing in the rhombohedral crystal form rather than the tetragonal crystal form favored by the sulfur protein (see above).

Figure 7(c) shows how Cxi of one molecule base-pairs with G10 of the RNA from the neighboring asymmetric unit, extending the P2 helix by one base-pair. This inter-molecular base-pair conforms closely to ideal *A*-form geometry (Figure 7(d)). We chose to end our HDV ribozyme constructs at position 84 of the wild-type sequence. Work of others (Been & Wickham, 1997) has shown that molecules with shorter 3′ ends have reduced self-cleaving activity, while longer molecules are not substantially more active. It has also been reported that extending the 3′ end of genomic ribozymes beyond position 86 results in slower self-cleavage (Tanner *et al.*, 1994). However, a cytidyl residue is present in position 85 of the wild-type sequence. Sequence complementarity to J1/2 ends at position 85.

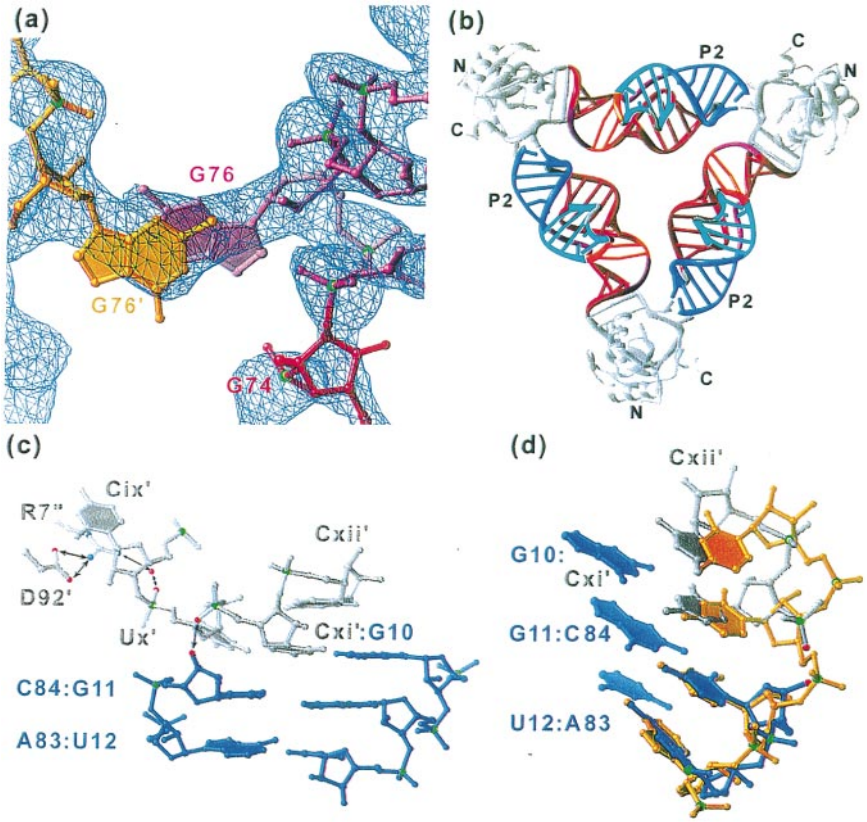

Figure 7. Crystal contacts involving base stacking. (a) Section of a 2.3 Å Fourier synthesis calculated with density modified (SOLOMON, Abrahams & Leslie, 1996) experimental phases, contoured at 1.5 S.D. and superimposed on portions of the final model from two symmetry-related molecules. The molecule on the right is colored following the convention of Figure 1(b). The residues from the neighboring asymmetric unit are shown in yellow. A crystallographic dyad runs approximately vertically between the bases of G76 and G76′. The extruded base of G76 is unlikely to interact with the rest of the ribozyme in solution, since it can be mutated to any nucleotide without affecting ribozyme activity (Tanner *et al.*, 1994), and in fact, ablation of the base results in more active RNAs (Belinsky *et al.*, 1993). (b) Complexes from three asymmetric units related by a crystallographic 3-fold axis perpendicular to the page. Note how nucleotides xi-xii from the U1A-binding loop of one molecule (gray) stack on the upper portion of P2 (dark blue) from a neighboring molecule. (c) Detail of the stacking interaction that results in the triangular arrangement in (b). Cxi from one molecule completes a base-pair with G10 of the neighboring RNA. The base of Ux lies in the major groove of P2 and does not participate in direct hydrogen bonds with either RNA. Arginine 7 from a third complex stacks under Cix of the second complex and hydrogen bonds to D92 and to a water molecule. Selected hydrogen bonds are denoted by double-headed arrows. (d) Superposition of an idealized *A*-form RNA segment (yellow) on the 3′ strand of P2. The intermolecular base-pair formed between G10 and Cxi is slightly displaced relative to where the corresponding pair of a canonical *A*-form helix would be. Cxii stacks on Cxi, but does not base-pair, and its backbone does not adopt *A*-form geometry.

Therefore, the P2 helix of the native genomic HDV ribozyme probably consists of seven base-pairs, and the 3′ termini of our crystallization constructs were fortuitously one nucleotide short, favoring formation of a crystal contact.

The foregoing discussion makes it clear that the HDV ribozyme-U1A RBD crystals are held together predominantly by protein-protein interactions. Even the RNA-RNA crystal contact between the U1A-binding loop of one complex and P2 of the neighboring complex (Figure 7(b) and (c)) would be unlikely if U1A did not bind the RNA construct and partially structure the loop. A fitting description of the crystal is that of RNA bridging protein islands (Figures 6(b) and 7(b)). Because of the paucity of RNA-RNA contacts, the RNA might be expected to have a larger degree of static or dynamic disorder than the protein. This might be exacerbated by the heterogeneity of the 3′ terminus of the RNA. While there is enough space near the 3′ terminus of the RNA-protein complex to accommodate the few random nucleotides added by the polymerase beyond C84, and no nucleotides beyond it are visible in experimental electron density maps, the close apposition of neighboring molecules (Figure 7(c)) suggests that the covalent heterogeneity could be a source of crystallographic disorder. Comparison of RNA recovered from crystals with RNA before crystallization by polyacrylamide gel electrophoresis does not reveal any enrichment in the species ending at nucleotide 84 (not shown).

The active site of the self-cleaved HDV ribozyme lacks tightly bound metal ions

The HDV ribozyme is active in even very low (<0.1 mM) concentrations of divalent cations. One molar equivalent of Mg^{2+}, Mn^{2+}, Ca^{2+}, or even Sr^{2+} is necessary and sufficient for activity (Suh *et al*., 1993). This relaxed chemical specificity led us to argue (Ferré-D'Amaré *et al*., 1998a) that the ribozyme is unlikely to have a high-affinity divalent metal ion binding site that is required for catalysis. In other RNAs that catalyze the same *trans*-esterification reaction as the HDV ribozyme, a metal ion has been proposed to activate a bound water, which in turn deprotonates the 2′-hydroxyl nucleophile on the −1 ribose (Pyle, 1993). Given the high efficiency of self-scission of the HDV catalyst (~100 times faster than a typical ribozyme, such as the hammerhead), a loosely coordinated metal ion would be unlikely to play such a central role in the reaction mechanism. Consistent with this, the structure of the HDV ribozyme does not contain well-ordered bound metal ions in the active-site cavity, and appears to be stabilized entirely by base-pairing, stacking, backbone contacts, and the connectivity of the nucleic acid chain (Ferré-D'Amaré *et al*., 1998a; Ferré-D'Amaré & Doudna, 1999). Peripheral metal ions (Lafontaine *et al*., 1999) as well as a diffuse counterion atmosphere (Misra & Draper, 1998) are, of course, expected to surround the strongly negative nucleic acid.

A phosphorothioate interference analysis to detect pro-Rp phosphate oxygen atoms that adversely affect ribozyme activity when substituted for sulfur, demonstrated only two positions with strong interference (Jeoung *et al*., 1994). Neither of these interferences could be rescued by addition of a soft metal ion, such as Mn^{2+}. If the interferences were due solely to disruption of a hard cation (e.g. Mg^{2+}) binding site, the presence of a soft cation would have been expected to restore activity to the ribozyme. One interference was detected on the phosphate of C22. In our crystal structure, this phosphate oxygen atom is seen to be participating in an intricate hydrogen-bonding network that includes C75. As this cytosine appears to play a central role in catalysis (see below) the interference at C22 can be explained by steric disruption. The other site of interference is at the scissile phosphate. If this phosphate were part of a cation binding site, the metal ion would not be expected to be ordered in our structure of the self-cleavage product, which is missing the scissile phosphate.

Figure 8(a) shows a portion of the density-modified experimental map surrounding the active site cavity. The crystals from which the reference structure factor amplitudes were measured were soaked in 10 mM $MgCl_2$ prior to flash cooling (see Methods), so that high-affinity magnesium binding sites should be well-occupied. An isolated, roughly spherical density feature is present in the active site, but its peak height is only 1.5 standard deviations (S.D.) above mean peak height. Carbon atoms of the RNA surrounding the cavity have densities as high as 3.5 S.D., implying that the isolated feature does not represent a highly occupied metal ion site. To examine this further, crystals were soaked in high concentrations of Mn^{2+} (up to 90 mM), diffraction data were measured, and an anomalous difference Fourier map calculated. A new density feature appeared in the active site cavity (Figure 8(a)) but its center is now removed

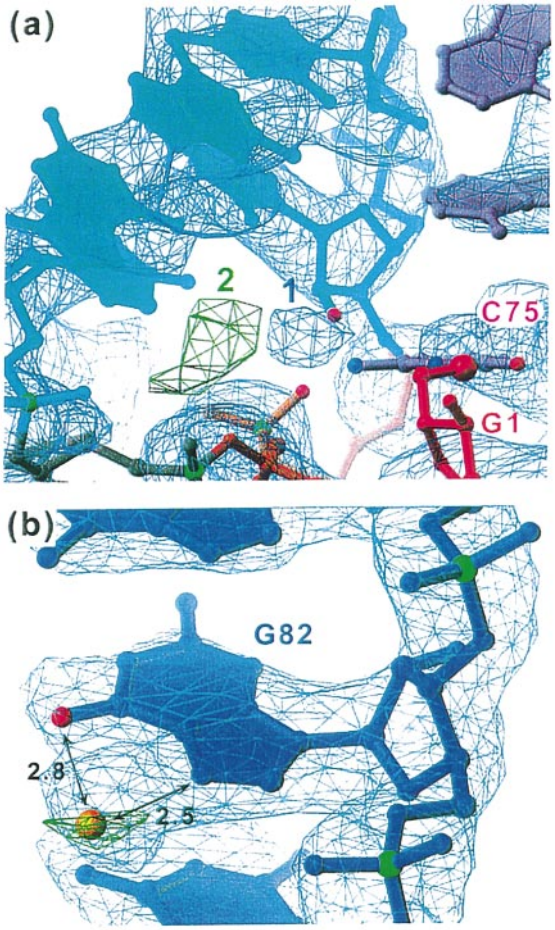

Figure 8. Two putative divalent cation binding sites. (a) View into the active-site cleft. The 5′-hydroxyl leaving group of the self-scission reaction is shown as a larger red sphere. The blue mesh is a portion of the density-modified experimental map contoured at 1 S.D. The density feature labeled 1 has a peak height of 1.5 S.D. The green mesh is a portion of the anomalous difference Fourier map calculated with data from a crystal soaked in manganese sulfate, contoured at 3 S.D. The density feature labeled 2 has a peak height of 3.6 S.D. (b) Representative major-groove divalent cation binding site. The blue mesh is a portion of a composite simulated annealing σA-weighted $2|F_o| - |F_c|$ omit map contoured at 1.5 S.D. The green mesh is from a manganese soak anomalous difference Fourier synthesis, contoured at 3.0 S.D. The yellow sphere is a putative Mg^{2+} ion which appears to make at least two inner sphere coordinations with the RNA.

by several Å from the original density feature, lies far away (>4 Å) from any potential ligands, and is not particularly strong (3.6 S.D.) Therefore, the genomic HDV ribozyme self-cleavage product does not appear to have a high-affinity magnesium binding site in which the RNA provides inner sphere coordination ligands to the metal ion. The weak density features we observe must result from the presence of a polar cleft into which cations can diffuse non-specifically. For comparison, a conventional major groove feature that has been identified as a magnesium ion, and which has at least two direct RNA ligands, is shown in Figure 8(b).

We also calculated an anomalous difference Fourier synthesis with amplitudes from the absorption peak wavelength of the osmium hexammine-containing crystals, and phases from the final model. If the metal hexammine had replaced a hydrated magnesium ion, a strong feature in the difference synthesis would be expected. The anomalous map showed positive peaks for the selenium atoms and the phosphorus atoms of the RNA, but no features consistent with a metal hexammine (not shown). It is therefore unlikely that a high-affinity binding site for a fully solvated magnesium ion is present in the active site of the self-cleaved HDV ribozyme.

At present, we cannot rule out that the active site of the HDV ribozyme precursor may have a high-affinity metal ion binding site, and that a bound cation could generate the base for activating the 2′-hydroxyl nucleophile. However, the lack of a well-ordered metal ion in the active-site cleft of our structure, together with the accumulated biochemical data indicating that a variety of metal cations are effective cofactors for the HDV ribozyme, suggests that an active site metal would not play a role in activating the 2′-hydroxyl nucleophile. In a transesterification reaction at a phosphate center, it is expected that some positively charged group will partially neutralize the negative charge of the phosphate in order to make the phosphorus a better electrophile (Westheimer, 1987). This could be the role played by the divalent cation required by the HDV ribozyme for activity. A metal ion associated with the scissile phosphate would also be consistent with the phosphorothioate interference observed at this position (Jeoung *et al*., 1994).

The trefoil turn may enable cytosine 75 to function as a general base

What activates the nucleophilic 2′-hydroxyl, if not a metal ion? We proposed that C75, which lies immediately adjacent to the 5′-hydroxyl leaving group of the self-scission reaction in our structure (Figure 8(a)) may be the general base that catalyzes the transesterification reaction (Ferré-D'Amaré *et al*., 1998a). In an extensive site-directed mutagenesis study of the genomic HDV ribozyme Tanner *et al*. (1994) found C75 to be the single most sensitive nucleotide. The equivalent C76 nucleotide in the antigenomic HDV ribozyme is equally sensitive to mutations. Photocrosslinking studies have demonstrated that C75/C76 lies very close to the scissile phosphate in analogs of both precursor ribozymes and cleavage products (Bravo *et al*., 1996; Rosenstein & Been, 1996).

Of the functional groups present in the four unmodified bases that occur naturally in RNA, N3 of cytosine has the highest pK_a value, ~4.2. While this is too low for a good general base at physiological pH, the acidity of a functional group can be strongly influenced by its environment. For instance, in a study of an unrelated aptamer RNA, Connell & Yarus (1994) found that a binding site cytosine had its pK_a value raised by nearly 4 pH units. In another example, Legault & Pardi (1994) have documented the presence of a protonated adenine in the core of a lead-dependent self-cleaving RNA by nuclear magnetic resonance spectroscopy.

Cytosine 75 is positioned at the bottom of the catalytic cleft of the HDV ribozyme by a tight turn of the RNA backbone in the J4/2 region of the molecule. This trefoil turn (Figure 9(a)) is characterized by an inversion in the orientation of two riboses (C75 and G76) relative to the direction of the chain, and an extrusion of the base of the second inverted nucleotide into solvent (the extruded base of G76 makes the crystal contact shown in Figure 6(a)). This turn can be considered a special case to the ''S-turn'' (Figure 9(b)). The trefoil turn of the HDV ribozyme is characterized by the presence of three hydrogen bonds between ribose 2′-hydroxyls and phosphate pro-Sp oxygen atoms that are made possible by a particularly tight turn of the backbone. In addition, the second (C75) and fourth (A77) bases of this turn stack on each other. In the conventional S-turn, a single backbone hydrogen bond is present, and the first two nucleotides stack (compare Figure 9(a) and (b)). The trefoil turn is held in place by the stacking of G74 under P1.1 and the ribose zipper formed by A77 and A78 triplexing in the minor groove of P3 (Figure 4; see also Ferré-D'Amaré *et al*., 1998a).

The tight trefoil turn and its position at the junction of three helical segments (P1, P1.1, and P3) bring many negatively charged phosphate groups into close proximity of C75. This high negative charge density may raise the pK_a value of the N3 of C75 by stabilizing the positive charge that would result from protonation. In addition, the N4 amino functional group of C75 is placed close to two potential hydrogen bond acceptors that would be well-positioned to stabilize a pyramidal NH_3^+ tautomer (Luisi *et al*., 1998) of the protonated cytosine, and this might further stabilize the charged state. A protonated, positively charged C75 that lies close to the scissile phosphate may also lower the activation energy of the transesterification reaction by stabilizing the increased negative charge of the pentacovalent phosphorus intermediate. Direct measurement of the pK_a value of C75, preferably in the context of an HDV ribozyme trapped close to the transition state, will be needed

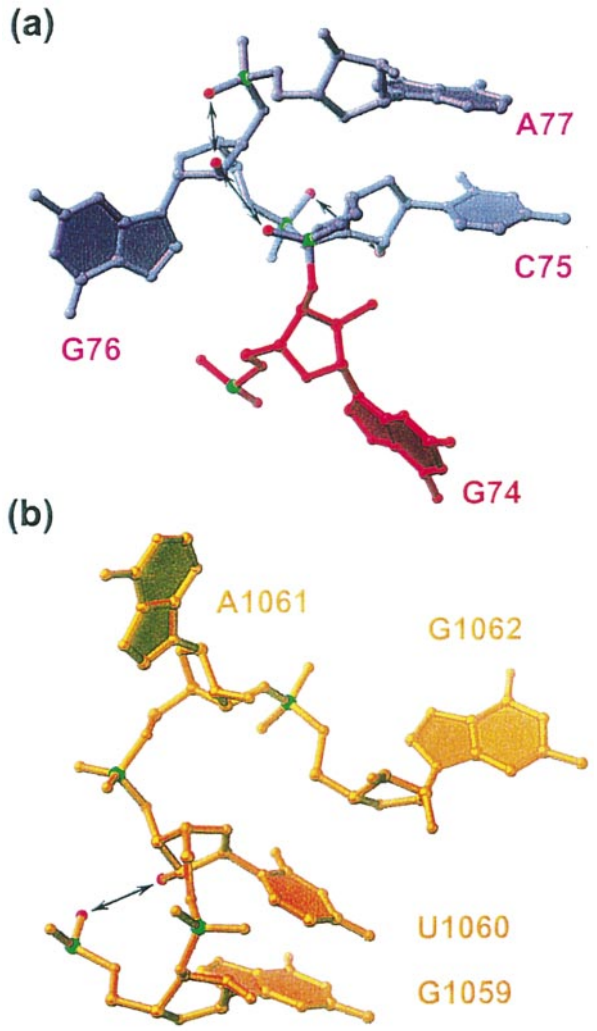

Figure 9. The trefoil turn is an especially compact form of the S-turn. (a) The four nucleotides from the HDV ribozyme that define the trefoil turn position the putative general base, C75, in the active site of the ribozyme. Note how the riboses of C75 and G76 are inverted relative to the direction of the chain. Five backbone polar groups within hydrogen bonding distance are connected by arrows. The riboses of nucleotides 74, 75 and 76 have C2′-*endo* pucker. (b) Canonical S-turn, from the L11-associated domain of the 23S rRNA of *E. coli* (Conn *et al*., 1999; see also Wimberly *et al*., 1999). As in the trefoil turn, the riboses of U1060 and A1061 are inverted. Only one hydrogen bond appears to form between groups in the sugar-phosphate backbone of the four nucleotides (arrow). The extruded bases of both turns, G76 and A1061 participate in stacking interactions. In the HDV ribozyme, this is a intermolecular crystal contact (Figure 6(a)). In the ribosomal RNA, the adenine makes an intramolecular stack with a second adenine.

to establish whether this nucleotide indeed serves as a general base. Recent biochemical work by Perrotta *et al*. (1999) provides strong support for the proposal that C75 functions as a general base.

A protonated cytosine is present as part of a base triple in the genomic HDV ribozyme

Our rhombohedral HDV-U1A co-crystals were grown at pH 7.0. The refined crystal structure contains a base-triple that demonstrates that protonation of the N3 of cytosine can be accomplished readily at this pH. The first two nucleotides of the J1.1/4 region of the genomic HDV ribozyme, G40 and C41 (Figure 1(b)) are pushed down into the narrowed major groove at the top of P4 (Figure 4). G40 makes a base-triple with the A43:G74 base-pair at the top of this helix (Ferré-D'Amaré *et al*., 1998a). C41 is pushed further down P4 so that it lies roughly coplanar with the C44:G73 base-pair (Figure 10(a)) and forms a reverse Hoogsteen pair with G73. The N3 imino nitrogen of C41 lies 2.9 Å away from the O6 carbonyl oxygen of G73, which implies that C41 is protonated. Interestingly, the O2 carbonyl of C41 does not hydrogen bond to the N4 amino group of C44, but to the N6 amino group of A43, of the base-pair above. Thus, the protonated C41 is one of a quartet of bases that from a major groove hydrogen bonding network.

Exactly the same arrangement of bases is present in the recently determined structure of the BWYV pseudoknot (Su *et al*., 1999). A superposition of the two instances of this motif is shown in Figure 10(b). We note that in both cases, a guanine precedes the protonated cytosine, and may be energetically coupled to the quadruple. In both the genomic HDV ribozyme and the BWYV pseudoknot, the bottom face of the protonated cytosine base is unstacked and exposed to solvent. This exposure would make it possible to offset the energetic cost of protonation by increased solvation of the positively charged base.

Conclusions

We have developed a new method for the crystallization of a compact, globular RNA. Since proteins have a larger variety of surface functional groups available for making crystal contacts than RNAs, we modified our target RNA to contain a high-affinity binding site for a well-characterized RNA-binding protein. The binding site was placed so as to have minimal effects on the biochemical activity of our target RNA, the genomic HDV ribozyme. By using this ''crystallization module'' in a series of constructs differing from each other in the length of an *A*-form duplex separating the module from the HDV ribozyme core moiety, we readily obtained a variety of crystal forms. Two of these diffracted synchrotron X-radiation beyond 3 Å resolution. Because these are RNA-protein co-crystals, well-established methods for derivatization of proteins could be employed to solve the structure.

Analysis of crystal packing shows that the principal interactions that stabilize the crystal are between proteins in adjacent asymmetric units. There are just two well-resolved RNA-RNA

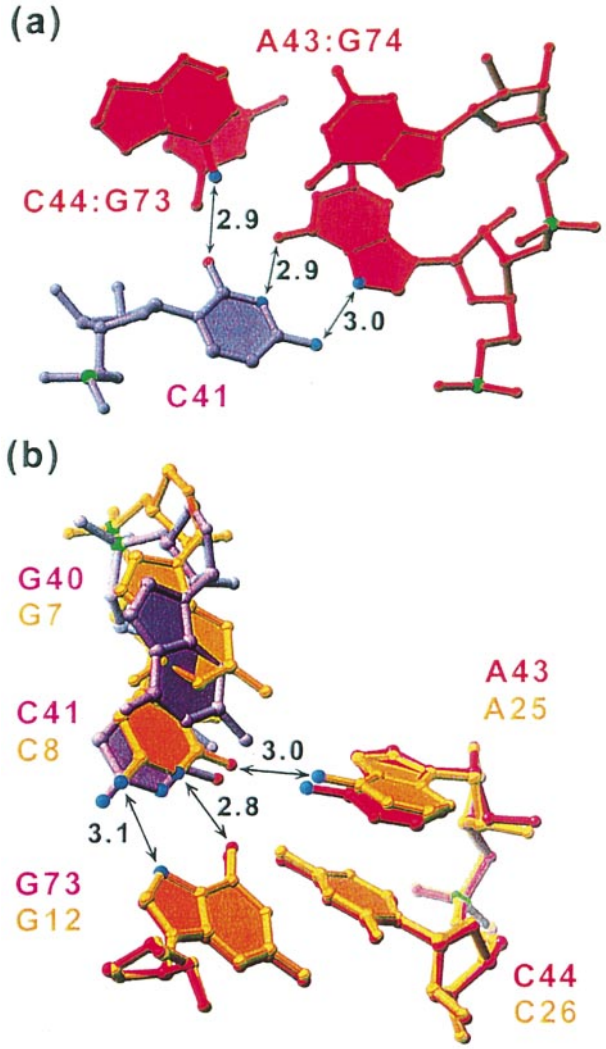

Figure 10. A base quadruple motif formed by a protonated cytosine in the major groove of a G·C pair. (a) Top view of the P4 helix of the HDV ribozyme. The base of C41 from J1.1/4 makes three major groove hydrogen bonds (black numbers are distances in Å). Two are with G73, and one with A43 which in turn base-pairs with G74. The distance between O2 of C41 and N4 of C44 is 3.4 Å. The ribose of C41 has a C2′-*endo* pucker. (b) Superposition of the reverse-Hoogsteen C^+ motif from HDV-U1A RBD on the corresponding nucleotides of the RNA pseudoknot structure by Su *et al.* (1999) in yellow. Distances shown are for the latter structure (compare with (a)). The distance between O2 of C8 and N4 of C26 is 3.2 Å. The N2 of the G preceding the protonated C (G40 in the HDV ribozyme) makes a van der Waals contact with the N6 of A (A43 and A25 in the two structures; the N-to-N distance is 3.4 Å in both).

interactions per complex, both of which bury only modest molecular surface areas. The limited nucleic acid crystal contacts are consistent with the great difficulty in obtaining crystals of wild-type HDV ribozyme (Ferré-D'Amaré *et al.*, 1998b). This demonstrates the utility of an RNA-binding protein as a crystallization reagent, and should encourage the use of this approach for the crystallization of other tightly folded RNAs. Given that the protein moiety is likely to make the majority of crystal contacts, multiple protein binding sites may be required to obtain satisfactory crystals of much larger RNA molecules.

Materials and Methods

Protein and RNA preparation

Selenomethionyl U1A-RBD was expressed in the auxotrophic strain B834 (Novagen) using the expression plasmid (A1-98, Y31H, Q36R, Oubridge *et al.*, 1994) generously provided by K. Nagai (MRC-LMB, Cambridge, UK). Freshly transformed colonies were picked from ampicillin plates and grown in M9 salts supplemented with 5% (v/v) LB and 50 mg/ml carbenicillin for six hours at 37 °C. Cells were gently pelleted, resuspended in fresh medium with the same antibiotic, and grown for a further 12 hours at 37 °C. Production cultures were grown in M9 salts supplemented with 50 mg/ml carbenicillin, 50 mg/l each of glycine and all L-amino acids except methionine, 50 mg/l of DL-selenomethionine, 1% (w/v) glucose, 2 mM magnesium sulfate, 0.1 mM $CaCl_2$, 1 mg/l thiamine, and 7.5 mg/l of $FeSO_4$ for 11 hours at 37 °C. Lysis was by freeze-thaw in the buffer described by Pognonec *et al.* (1991). Cell debris was removed by centrifugation, and the supernatant fractionated with 0.5% (w/v) neutralized polyethyleneimine (Burgess, 1991). The supernatant was further fractionated with ammonium sulfate. The protein that precipitated between 35 and 75% saturation was redissolved in, and dialyzed overnight against, buffer A (100 mM KCl, 25 mM Hepes-KOH (pH 7.5), 10 mM DTT, 0.5 mM EDTA, 0.5 mM PMSF). The dialyzed protein derived from a nine liter growth was then loaded on a 30 ml bed volume SP-Sepharose FF (Pharmacia) column equilibrated in buffer A, and eluted with a linear, 300 ml gradient to buffer B (this had the same composition as buffer A except for having 500 mM KCl). U1A-RBD elutes around 300 mM KCl. The protein was then concentrated by ultrafiltration and further purified on a 120 ml bed-volume Superdex-75 PG (Pharmacia) column equilibrated in buffer A. Phosphate buffer (pH 7.5) was added to the eluted U1A-RBD to a concentration of 10 mM. The protein was then loaded onto a 5 ml ceramic hydroxyapatite column (CHT-I, BioRad) which had been equilibrated in buffer C (10 mM potassium phosphate (pH 7.5), 50 mM KCl, 1 mM DTT). The protein was eluted by running a 20 column-volume gradient to 100% buffer D (this had the same composition as buffer C, but with the addition of 0.5 M ammonium sulfate). The protein eluted around 250 mM ammonium sulfate. The purified protein was dialyzed against a storage buffer consisting of 10 mM Hepes-KOH (pH 7.5), 1 mM DTT, 0.1 mM EDTA, concentrated to 15 g/l by ultrafiltration, and stored at −80 °C. Incubation of the protein with a stoichiometric amount of RNA for several months at room temperature did not result in detectable degradation of the nucleic acid. Matrix-assisted laser desorption-ionization mass spectrometry confirmed full substitution of methionine by selenomethionine (calculated mass difference between sulfur and selenium proteins for four methionines: 187.6 a.m.u; measured difference 185.3(±5) a.m.u.) Sulfur U1A-RBD was

purified in the same manner, except that DTT was omitted from all buffers.

RNA was prepared as described (Ferré-D'Amaré & Doudna, 1996). The sequence of the 41 nt leader preceding the ribozyme core is 5′-GGGAGGAGCCGTATGC-GATG TCGCACGTAC GGTAAATCGT C-3′. The HDV ribozyme is fully active in 8 M urea (Been & Wickham, 1997), and the stability to denaturants of ribozyme precursor and self-cleaved products appear to be similar (Duhamel *et al.*, 1996). When the HDV ribozyme self-cleavage product is purified by electrophoresis on 8 M urea gels, the RNA migrates at an anomalously fast rate (A.R.F. & J.A.D., unpublished results). The anomalous electrophoretic mobility is consistent with a compact, folded RNA. Therefore, it is likely that the HDV ribozyme RNAs we employed for crystallization were never fully unfolded after they crossed the transition state of the self-cleavage reaction.

Crystallization

Crystal form I contained wild-type 98 amino acid residue U1A-RBD and the RNA with a one base-pair spacer (Figure 1(b)). Crystals were grown by vapor diffusion at 4 °C by mixing a 0.2 mM solution of the complex (which contained 1.25 mM $MgCl_2$) with an equal volume of a reservoir solution containing 1.75 M ammonium sulfate, 2 % (w/v) polyethyleneglycol (PEG) 400, 25 mM $MgCl_2$, 0.1 mM spermine, 100 mM Hepes-KOH (pH 7.5). Crystals grew to maximum dimensions of 0.1 mm × 0.1 mm × 0.5 mm in the course of a month. For flash-cooling, crystals were soaked first in a solution of the same composition as the reservoir, except that the ammonium sulfate was increased to 1.8 M, the $MgCl_2$ to 40 mM, the spermine to 1.0 mM, and 20 % (v/v) ethylene glycol was added.

Crystal form II contained double-mutant (Y31H, Q36R) U1A-RBD and the RNA shown in Figure 1(c). Crystals were grown by vapor diffusion at 30 °C by mixing equal volumes of 0.5 mM complex solution (which contained 1.25 mM $MgCl_2$ and 0.1 mM $GdCl_3$) with a reservoir solution that contained 30 % (w/v) PEG monomethylether (MME) 2000, 150 mM Li2SO_4, 7 mM $MgSO_4$, 100 mM Tris-HCl (pH 7.75). Crystals took several weeks to appear, and grew over the course of several months to maximum dimensions of 0.8 mm^3. For flash cooling, crystals were first transferred to a stabilizer solution containing 30 % (w/v) PEG MME2000, 5 % (w/v) PEG6000, 175 mM Li_2SO_4, 10 mM $MgSO_4$, 100 mM Tris-HCl (pH 7.75), 0.5 mM $GdCl_3$, and 3 % (v/v) (−)-2,3-butanediol.

Crystal form III. Crystallization and stabilization conditions have been described (Ferré-D'Amaré *et al.*, 1998a).

Refinement

Refinement was carried out against all observed reflections between 20 and 2.3 Å resolution (17,040 and 1863 reflections in the working and test sets, respectively) as described in the text and by Ferré-D'Amaré *et al.* (1998a). Throughout the refinement, the overall anisotropic *B*-factor correction, and solvent mask were employed, and nucleotide bases were constrained to be planar. The restraints on the *B*-factors of covalently bonded atoms were adjusted to minimize the *R*-free. One protein side-chain, that of histidine 31 showed clear electron density for two alternate conformations, and has been built accordingly. This is one of the point mutations (Y31H) introduced by Oubridge *et al.* (1994) into U1A to facilitate crystallization. A number of elongated residual density features surrounding the macromolecules could represent partially ordered spermine molecules, since the crystals were stabilized in 25 mM spermine hydrochloride. However, there was no difference in *R*-free or the real-space *R*-factor when they were built as either segments of spermine or as water molecules, and therefore have been modeled as the latter. Water molecules whose *B*-factors refined to values smaller than their ligands were modeled as magnesium ions. Solvent accessible surface areas were calculated with a probe radius of 1.4 Å.

Manganese soak

A form III crystal was transferred over the course of one hour to a stabilization solution comprised of 90 mM $MnSO_4$, 20 % (w/v) PEG-MME2000, 5 % (w/v) PEG6000, 100 mM Li_2SO_4, 10 mM spermine, 100 mM Tris-HCl (pH 7.0), 3 % (−) 2,3-butanediol. After an additional hour in the stabilization buffer, the crystal was flash-cooled in super-cooled liquid propane maintained at liquid nitrogen temperature, and diffraction data were measured at 100 K on a copper rotating anode source equipped with focusing mirrors using the inverse beam method. The reduced data are 75.8 % complete between 50 and 4 Å (72.3 % between 4.14 and 4 Å) with an overall R_{merge} of 8 % (51.3 % for the last shell) and an overall redundancy of 3.9 (with anomalous pairs kept separate). The anomalous difference Fourier synthesis (Figure 8) was calculated with the experimental phases that were used for refinement and anomalous differences from this dataset.

Atomic coordinates

Atomic coordinates and structure factor amplitudes have been deposited with the Protein Data Bank with accession number 1cx0.

Acknowledgments

We are grateful to Kaihong Zhou for superb biochemical assistance during the course of this project to Craig Ogata and David Cook at beamline X4A of the National Synchrotron Light Source and the staff of the Cornell High-energy Synchrotron Source for help during data collection and to Jeff Kieft and Andrej Lupták for comments on the manuscript. A.R.F. was partly funded by a post-doctoral fellowship from the Jane Coffin Childs Memorial Fund for Medical Research. This work was supported by grants from the NIH, the David and Lucille Packard Foundation, and the Beckman Foundation to J.A.D.

References

Abrahams, J. P. & Leslie, A. G. W. (1996). Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallog. sect. D*, **52**, 30-42.

Adams, P. D., Pannu, N. S., Read, R. J. & Brunger, A. T. (1997). Cross-validated maximum likelihood enhances crystallographic simulated annealing refinement. *Proc. Natl Acad. Sci. USA*, **94**, 5018-5023.

Aggarwal, A. K. (1990). Crystallization of DNA binding proteins with oligodeoxynucleotides. *Methods,* **1**, 83-90.

Allain, F. H.-T., Gubser, C. C., Howe, P. W. A., Nagai, K., Neuhaus, D. & Varani, G. (1996). Specificity of ribonucleoprotein interaction determined by RNA folding during complex formation. *Nature,* **380**, 646-650.

Been, M. D. & Wickham, G. S. (1997). Self-cleaving ribozymes of hepatitis delta virus RNA. *Eur. J. Biochem.* **247**, 741-753.

Belinsky, M. G., Britton, E. & Dinter-Gottlieb, G. (1993). Modification interference analysis of a self-cleaving RNA from hepatitis delta virus. *FASEB J.* **7**, 130-136.

Bravo, C., Lescure, F., Laugâa, P., Fourrey, J.-L. & Favre, A. (1996). Folding of the HDV antigenomic ribozyme pseudoknot structure deduced from long-range photocrosslinks. *Nucl. Acids Res.* **24**, 1351-1359.

Brunger, A. T., Adams, P. D., Clore, G. M., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). Crystallography and NMR system (CNS): a new software system for macromolecular structure determination. *Acta Crystallog. sect. D,* **54**, 905-921.

Burgess, R. R. (1991). Use of polyethyleneimine in purification of DNA-binding proteins. *Methods Enzymol.* **208**, 3-10.

Carson, M. (1991). RIBBONS 2.0. *J. Appl. Crystallog.* **24**, 958-961.

Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Kundrot, C. E., Cech, T. R. & Doudna, J. A. (1996). Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science,* **273**, 1678-1685.

Cohen, S. L. (1996). Domain elucidation by mass spectrometry. *Structure,* **4**, 1013-1016.

Conn, G. L., Draper, D. E. E. E. L. & Gittis, A. G. (1999). Crystal structure of a conserved ribosomal protein-RNA complex. *Science,* **284**, 1171-1174.

Connell, G. J. & Yarus, M. (1994). RNAs with dual specificity and dual RNAs with similar specificity. *Science,* **264**, 1137-1141.

de la Fortelle, E. & Bricogne, G. (1997). Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472-494.

Doublié, S. (1997). Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol.* **276**, 523-530.

Duhamel, J., Liu, D. M., Evilia, C., Fleysh, N., Dinter-Gottlieb, G. & Lu, P. (1996). Secondary structure content of the HDV ribozyme in 95% formamide. *Nucl. Acids Res.* **24**, 3911-3917.

Ferré-D'Amaré, A. R. & Burley, S. K. (1997). Dynamic light scattering in evaluating crystallizability of macromolecules. *Methods Enzymol.* **276**, 157-166.

Ferré-D'Amaré, A. R. & Doudna, J. A. (1996). Use of cis- and trans-ribozymes to remove 5′ and 3′ heterogeneities from milligrams of in vitro transcribed RNA. *Nucl. Acids Res.* **24**, 977-978.

Ferré-D'Amaré, A. R. & Doudna, J. A. (1999). RNA folds: insights from recent crystal structures. *Annu. Rev. Biophys. Biomol. Struct.* **28**, 57-73.

Ferré-D'Amaré, A. R., Zhou, K. & Doudna, J. A. (1998a). Crystal structure of a hepatitis delta virus ribozyme. *Nature,* **395**, 567-574.

Ferré-D'Amaré, A. R., Zhou, K. & Doudna, J. A. (1998b). A general module for RNA crystallization. *J. Mol. Biol.* **279**, 621-631.

Golden, B. L., Podell, E. R., Gooding, A. R. & Cech, T. R. (1997). Crystals by design: a strategy for crystallization of a ribozyme derived from the *Tetrahymena* group I intron. *J. Mol. Biol.* **270**, 711-723.

Holbrook, S. R. & Kim, S.-H. (1997). RNA crystallography. *Biopolymers,* **44**, 3-21.

Jancarik, J. & Kim, S.-H. (1991). Sparse matrix sampling: a screening method for crystallization of proteins. *J. Appl. Crystallog.* **24**, 409-411.

Jeoung, Y.-H., Kumar, P. K. R., Suh, Y.-A., Taira, K. & Nishikawa, S. (1994). Identification of phosphate oxygens that are important for self-cleavage activity of the HDV ribozyme by phosphorothioate substitution interference analysis. *Nucl. Acids Res.* **22**, 3722-3727.

Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallog. sect. A,* **47**, 110-119.

Kieft, J. S. & Tinoco, I. (1997). Solution structure of a metal-binding site in the major groove of RNA complexed with cobalt (III) hexammine. *Structure,* **5**, 713-721.

Lafontaine, D. A., Ananvoranish, S. & Perreault, J.-P. (1999). Presence of a coordinated metal ion in a trans-acting antigenomic delta ribozyme. *Nucl. Acids Res.* **27**, 3236-3243.

Lai, M. M. (1995). The molecular biology of hepatitis delta virus. *Annu. Rev. Biochem.* **64**, 259-286.

Laskowski, R. J., Macarthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283-290.

Legault, P. & Pardi, A. (1994). In situ probing of adenine protonation in RNA by ^{13}C NMR. *J. Amer. Chem. Soc.* **116**, 8390-8391.

Luisi, B., Orozco, M., Sponer, J., Luque, F. J. & Shakked, Z. (1998). On the potential role of the amino nitrogen atom as a hydrogen bond acceptor in macromolecules. *J. Mol. Biol.* **279**, 1123-1136.

Luzzati, P. V. (1952). Traitement statistique des erreurs dans la determination des structures cristallines. *Acta Crystallog.* **5**, 802-810.

McKay, D. B. & Wedekind, J. E. (1999). Small ribozymes. In *The RNA World* (Gesteland, R. F., *et al.*, ed.), pp. 265-286, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Misra, V. K. & Draper, D. E. (1998). On the role of magnesium ions in RNA stability. *Biopolymers,* **48**, 113-135.

Nagai, K., Oubridge, C., Jessen, T. H., Li, J. & Evans, P. R. (1990). Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature,* **348**, 515-520.

Oubridge, C., Ito, N., Evans, P. R., Teo, C.-H. & Nagai, K. (1994). Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature,* **372**, 432-438.

Oubridge, C., Ito, N., Teo, C.-H., Fearnley, I. & Nagai, K. (1995). Crystallization of RNA-protein complexes II. The application of protein engineering for crys-

tallization of the U1A protein-RNA complex. *J. Mol. Biol.* **249**, 409-423.
Perrotta, A. T. & Been, M. D. (1998). A toggle duplex in hepatitis delta virus self-cleaving RNA that stabilizes an inactive and a salt-dependent pro-active ribozyme conformation. *J. Mol. Biol.* **279**, 361-373.
Perrotta, A. T., Shih, I. & Been, M. D. (1999). Imidazole rescue of a cytosine mutation in a self-cleaving ribozyme. *Science,* **286**, 123-126.
Pognonec, P., Kato, H., Sumimoto, H., Kretzschmar, M. & Roeder, R. G. (1991). A quick procedure for purification of functional recombinant proteins overexpressed in *E. coli*. *Nucl. Acids Res.* **19**, 6650-.
Pyle, A. M. (1993). Ribozymes: a distinct class of metalloenzymes. *Science,* **261**, 709-714.
Richards, F. M. & Wyckoff, H. W. (1971). Bovine pancreatic ribonuclease. In *The Enzymes* (Boyer, P. D., ed.), vol. 4, pp. 647-806, Academic Press, New York.
Rosenstein, S. P. & Been, M. C. (1996). Hepatitis delta virus ribozymes fold to generate a solvent-inaccessible core with essential nucleotides near the cleavage site phosphate. *Biochemistry,* **35**, 11403-11413.
Schindelin, H., Zhang, M., Bald, R., Fürste, J.-P., Erdmann, V. A. & Heinemann, U. (1995). Crystal structure of an RNA dodecamer containing the *Escherichia coli* Shine-Delgarno sequence. *J. Mol. Biol.* **249**, 595-603.
Schultz, S. C., Shields, G. C. & Steitz, T. A. (1990). Crystallization of *Escherichia coli* catabolite gene activator protein with its DNA binding site the use of modular DNA. *J. Mol. Biol.* **213**, 159-166.
Su, L., Chen, L., Egli, M., Berger, J. M. & Rich, A. (1999). Minor groove RNA triplex in the crystal structure of a ribosomal frameshifting viral pseudoknot. *Nature Struct. Biol.* **6**, 285-292.
Suh, Y.-A., Kumar, P. K. R., Taira, K. & Nishikawa, S. (1993). Self-cleavage activity of the genomic HDV ribozyme in the presence of various divalent metal ions. *Nucl. Acids Res.* **21**, 3277-3280.
Tanner, N. K., Schaff, S., Thill, G., Petit-Koskas, E. P., Crain-Denoyelle, A.-M. & Westhof, E. (1994). A three-dimensional model of hepatitis delta virus ribozyme based on biochemical and mutational analyses. *Curr. Biol.* **4**, 488-498.
van Gelder, C. W. G., Gunderson, S. I., Jansen, E. J. R., Boelens, W. C., Polycarpou-Schwarz, M., Mattaj, I. W. & van Venrooij, W. J. (1993). A complex secondary structure in U1A pre-mRNA that binds two molecules of U1A protein is required for regulation of polyadenylation. *EMBO J.* **12**, 5191-5200.
Wadkins, T. S., Perrotta, A. T., Ferré-D'Amaré, A. R., Doudna, J. A. & Been, M. D. (1999). A nested double-pseudoknot is required for self-cleavage activity of both the genomic and antigenomic hepatitis delta virus ribozymes. *RNA,* **5**, 720-727.
Westheimer, F. H. (1987). Why nature chose phosphates. *Science,* **235**, 1173-1178.
Wilson, A. J. (1949). The probability distribution of X-ray intensities. *Acta Crystallog.* **2**, 318-321,.
Wimberly, B. T., Guymon, R., McCutcheon, J. P., White, S. W. & Ramakrishnan, V. (1999). A detailed view of a ribosomal active site: the structure of the L11-RNA complex. *Cell,* **97**, 491-502.

Edited by D. C. Rees

(Received 30 August 1999; received in revised form 15 November 1999; accepted 17 November 1999)