

Crystal structure of a hairpin ribozyme–inhibitor complex with implications for catalysis

Peter B. Rupert & Adrian R. Ferré-D'Amaré

Division of Basic Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, Washington 98109-1024, USA

The hairpin ribozyme catalyses sequence-specific cleavage of RNA. The active site of this natural RNA results from the docking of two irregular helices: stems A and B. One strand of stem A harbours the scissile bond. The 2.4 Å resolution structure of a hairpin ribozyme–inhibitor complex reveals that the ribozyme aligns the 2'-OH nucleophile and the 5'-oxo leaving group by twisting apart the nucleotides that flank the scissile phosphate. The base of the nucleotide preceding the cleavage site is stacked within stem A; the next nucleotide, a conserved guanine, is extruded from stem A and accommodated by a highly complementary pocket in the minor groove of stem B. Metal ions are absent from the active site. The bases of four conserved purines are positioned potentially to serve as acid-base catalysts. This is the first structure determination of a fully assembled ribozyme active site that catalyses a phosphodiester cleavage without recourse to metal ions.

The hairpin ribozyme is a catalytic RNA derived from the self-cleaving and ligating domain of the negative polarity strand of the satellite RNA of tobacco ringspot virus. *In vivo*, this domain is responsible for generating unit-length circular satellite RNA during the course of its rolling-circle replication (reviewed in ref. 1). The cleavage reaction generates products with 5'-hydroxyl and 2',3'-cyclic phosphate termini (Fig. 1a), which are analogous to those produced by three other natural ribozymes that are part of circular satellite RNAs: the hammerhead, the hepatitis delta virus (HDV) and the Varkud satellite (VS) ribozymes. These four ribozymes, however, are structurally unrelated, and therefore represent independent evolutionary solutions to the same biochemical problem (reviewed in ref. 2). None of these ribozymes require proteins for activity. They are also quite small, for example the hairpin ribozyme requires about 50 nucleotides for activity. The hairpin ribozyme is an ideal experimental system to help understand how a small catalytic RNA self-assembles and forms an active site.

Biochemical experiments demonstrated that the active site of the hairpin ribozyme results from the association of two largely helical segments of the satellite RNA, stems A (one strand of which contains the scissile phosphate) and B (Fig. 1b). Both contain nucleotides necessary for catalysis¹. The stems can be synthesized as two separate RNAs and mixed *in vitro* to reconstitute an active ribozyme^{3,4}. Most biochemical experiments have been carried out using constructs where the two stems are connected by a single-stranded linker. In the satellite RNA, stems A and B are part of a four-helix junction⁵. Fluorescence resonance energy transfer (FRET) measurements demonstrated that the docking of stems A and B is greatly favoured in constructs that contain the four-helix junction, compared with systems where the two stems are connected by an extended linker^{5,6}. Docking is facilitated by the cooperative binding of two molar equivalents of alkaline earth metal ions, which have dissociation constants of 10⁻¹⁰ to 10⁻⁹ M (refs 5, 7); however, the hairpin ribozyme does not use metal ions to activate its 2'-OH nucleophile. Moreover, it has been shown to be fully active with cobalt (III) hexammine⁸⁻¹⁰, which mimics hydrated Mg²⁺, but cannot make direct coordination with water or RNA ligands¹¹. Thus, the RNA itself provides functional groups that accomplish catalysis, unlike 'classical' ribozymes such as the hammerhead or the group I intron that use metal ion cofactors¹. Evidence suggests that nucleotide bases with perturbed pK_a values

in the active sites of the HDV ribozyme¹²⁻¹⁴ and the peptidyl-transferase of the 50S ribosomal subunit¹⁵ enable those RNAs to carry out general acid-base catalysis. Such a mechanism is potentially used by the hairpin ribozyme.

As a step towards elucidating its catalytic mechanism, and to provide a structural framework for rationalizing the large body of biochemical information gathered on this catalytic RNA, we have now determined the structure of a fully assembled hairpin ribozyme (Table 1 and Methods). We crystallized a four-helix junction form of the RNA in complex with a substrate strand in which the nucleophilic 2'-OH group (Fig. 1a) was replaced with a methoxy function (Fig. 1b). Such modified substrates are competitive inhibitors of the ribozyme^{16,17}. Crystals were grown in the presence of 20 mM Ca²⁺, enough to saturate the structural metal-ion-binding sites. The cognate site for the RNA-binding protein UIA was grafted onto the structurally and functionally dispensable¹ distal end of stem B, and the engineered RNA was crystallized together with the basic protein (see Methods and Supplementary Information). The structure consists of two independent hairpin ribozyme–inhibitor complexes in different crystalline environments. Although we did not

Table 1 Crystallographic statistics

Data set	Crystal I			Crystal II
	λ1	λ2	λ3	
Diffraction data				
Resolution (Å)	30.0–2.9	30.0–2.9	30.0–2.9	100.0–2.4
Wavelength (Å)	0.9796	0.9793	0.9566	1.1000
Reflections				
Total/unique*	83,395/44,497	168,452/45,575	168,844/45,520	148,993/41,290
Completeness (%)	98.2 (89.6)	99.1 (95.0)	99.1 (94.9)	95.6 (73.9)
⟨I⟩/⟨σ(I)⟩	19.7 (1.7)	26.5 (2.3)	24.6 (2.2)	22.8 (2.1)
R _{sym} (%)†	6.4 (41.0)	8.0 (42.3)	7.7 (43.7)	6.7 (26.1)
Phasing statistics				
Phasing power‡	1.26 (0.24)	1.34 (0.36)	1.10 (0.27)	
R _{crystal} §	0.05 (0.20)	0.04 (0.15)	0.04 (0.15)	

Numbers in parentheses refer to the high-resolution shell (3.0–2.9 Å for crystal I; 2.5–2.4 Å for crystal II).

* For crystal I data, the Bijvoet pairs were not merged.

† $R_{sym} = \sum |I - \langle I \rangle| / \sum I$ where I is the observed intensity and $\langle I \rangle$ is the statistically weighted absolute intensity of multiple measurements of symmetry related reflections.

‡ Phasing power = $(F_o) / (|F_p + F_n| - |F_{pn}|)$, reported for all acentric reflections.

§ $R_{crystal} = \sum (|f_p + F_n| - |F_{pn}|) / \sum |F_p|$, reported for all reflections.

use non-crystallographic symmetry (NCS) restraints, the structures of the two catalytic cores have a root-mean-square (r.m.s.) difference of < 0.5 Å for all atoms, which is comparable to the precision of the atomic coordinates (see Methods). The significant agreement

of the refined crystallographical model with the results of biochemical experiments, and the in-line-attack conformation present at the catalytic site, suggest that this 2.4-Å resolution structure accurately depicts the hairpin ribozyme in an active conformation.

Overview of the structure

The four helical stems that comprise the hairpin ribozyme radiate from a perfectly base-paired four-way junction. In the crystal structure, the helices form two coaxial stacks: stem D stacks on A, and stem C stacks on B. The two stacks cross at around a 60° angle, and are connected by a pair of anti-parallel crossovers (Figs 1b and 2). The junction allows the central portions of stems A and B to dock through their minor grooves and form the active site of the ribozyme. No bound metal ions are observed in the junction. In the crystal structure of an unrelated four-helix junction formed by a DNA–RNA hybrid¹⁸, a bound metal ion was found in the major groove of one of the B-form helices. Because the stems of the hairpin ribozyme are in an A-form conformation near the junction, there is no equivalent binding site in its structure. Well-ordered metal ions are observed in the non-canonical moiety of stem B (see below). These could correspond to the cooperatively bound metal ions detected by FRET^{5,7}, as tight binding of divalent ions was detected only in experiments with the intact ribozyme, but not the isolated fully-base-paired junction¹⁹.

The stacking order was correctly predicted on the basis of gel-mobility shift and FRET experiments⁵. It was proposed that the end-to-end distances of the stems would be in the order AB < CD < BD, as is observed in the crystal structure (Fig. 2). The structure is consistent with the results of hydroxyl radical probing experiments, which found that a solvent-inaccessible core is formed when the RNA adopts its active structure in the presence of polyvalent cations²⁰. The ten C4' atoms found to be solvent-inaccessible in that study (G11–A14, C25–C27, A38 and U42–43) are fully buried in the structure. Although a model of the hairpin ribozyme generated on the basis of biochemical data and long-distance crosslinking experiments²¹ is substantially different from the crystal structure (the r.m.s. difference for 66 C1' atoms is 4.3 Å), the orientation of stems A and B is the same.

Stems A and B of the hairpin ribozyme each consist of two A-form, Watson–Crick-paired segments flanking a stretch of nucleotides that adopt non-canonical conformations (Fig. 1b). The solution three-dimensional structures of isolated stems A and B have been determined^{22,23}. As these separate RNAs can be mixed *in vitro* to reconstitute a functional ribozyme^{3,4}, the nuclear magnetic resonance (NMR) structures probably represent the conformation of the two stems before docking. Comparison with the crystal structure shows that large conformational rearrangements take place in both RNA stems on assembly of the active site.

Stem A conformation

In the docked state, the non-canonical moiety of stem A consists of three single-hydrogen-bond base pairs and two unpaired nucleotides (Fig. 3a). Of the latter, the conserved, essential nucleotide following the scissile phosphate, G+1 (see Fig. 1 legend for numbering scheme), is extruded from the helical stack, and adopts a syn conformation (Fig. 3b) as proposed previously²⁴. The hairpin ribozyme tolerates any base at position –1 (ref. 24). In the crystal structure, the N3 atom of A–1 accepts a single hydrogen bond from A9. This base-pair geometry is compatible with all four bases, as both purines have an N3, and both pyrimidines have an O2 atom in the equivalent position. The geometry of the A–1·A9 pair causes A–1 to make a cross-strand stack onto G8. This and the conformation of G+1 result in a pronounced broadening of the minor groove (Fig. 3a).

The three non-canonical base pairs observed in the crystal structure are absent in the isolated stem A. In the NMR structure, an anti conformation G+1 is accommodated in the helical stack (Fig. 3c), and forms a sheared pair with A9 (ref. 22). The NMR

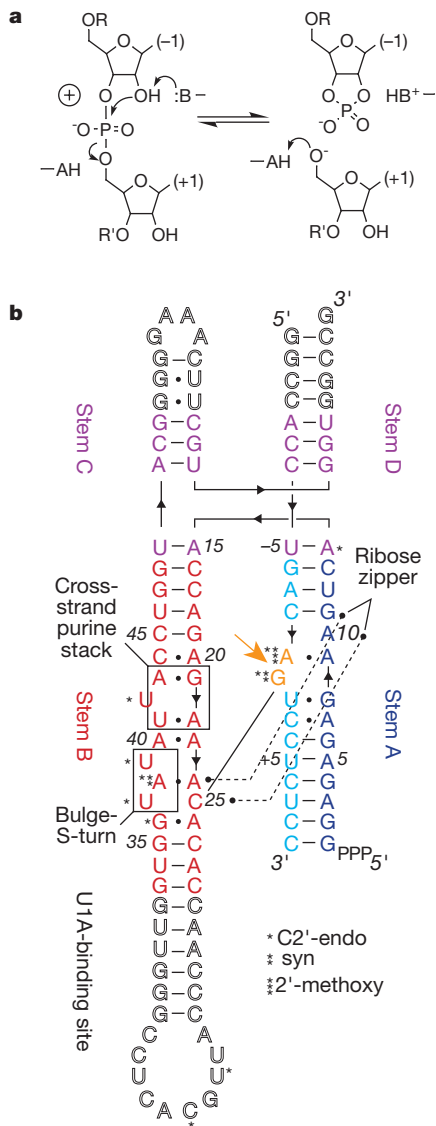


Figure 1 Sequence and reactions catalysed by the hairpin ribozyme. **a**, Reversible transesterification reaction catalysed by the ribozyme. Hypothetical acid, base and positive charge catalytic groups are indicated. **b**, Sequence and secondary structure of the crystallization construct. The RNA consists of a 92-nucleotide ribozyme and a 21-nucleotide substrate strand. In the latter, the 2'-OH adjacent to the scissile bond was substituted with a methoxy to inhibit cleavage. To facilitate crystallization²⁷, a cognate site for the RNA-binding protein U1A was grafted at the distal end of stem B. This construct was decided on after screening 21 different RNA–U1A complexes. The RNA is numbered according to convention¹. In the substrate strand, numbers of nucleotides in the 5' and 3' direction of the cleavage site (yellow arrow) are preceded by minus and plus signs, respectively. Ribozyme-strand nucleotides are numbered starting at the 5' terminus, skipping the grafted U1A-binding site. Lines with embedded arrows indicate the direction of the chains at the crossovers. Solid letters, wild-type sequence; outline letters, nucleotides added for crystallization. Thin lines indicate Watson–Crick pairs; black circles indicate non-canonical pairs involving at least one hydrogen bond between co-planar bases. Nucleotides that participate in a ribose zipper²⁵ are connected by dotted lines. Nucleotides with C2'-endo ribose pucker (asterisk), those in the syn conformation (double asterisk), and the A–1 residue that carries the 2'-methoxy modification (triple asterisk) are indicated. The same nucleotide conformations are present in both molecules in the crystallographic asymmetric unit.

construct has a C base at position -1 , and this pairs with a protonated A10. The latter pair is inconsistent with the lack of a requirement for a particular base at -1 (ref. 22). The minor groove width of the stem A conformation observed by NMR is considerably narrower than that of the docked state (Fig. 3c).

Stem B conformation

In the crystal structure, stem B contains six non-canonical base pairs that stack in an irregular underwound helix (Fig. 4a). A single-hydrogen-bond A20·C44 pair is followed by the sheared G21·A43 and reverse-Hoogsteen U41·A22 pairs. The latter two pairs conform closely to the cross-strand purine stack motif²⁵. In the crystal structure, U42 is extruded from the helical stack, and its base lies perpendicular to the plane of the helical base pairs. This arrangement is consistent with the observation that a photo-crosslink can be formed at high efficiency between G21 and/or A22 and U41. A crosslink to U42 was observed only at high concentrations where the ribozyme RNA formed a non-functional dimer²⁶.

The bottom of the non-canonical moiety of stem B is stretched apart by an N1–N1 pair between G36 and A26 (Fig. 4a). This is followed by an S-turn²⁷, where the direction of the ribose sugars in the backbone is inverted twice, resulting in extrusion of the U37 base into solvent. A38 lies within the stack in a syn conformation, and makes a single hydrogen bond with A24. The S-turn is followed by a second bulged pyrimidine, U39. The bulging of U39 is consistent with the observation that it can be replaced with a propyl linker without affecting ribozyme activity²⁸. The combination of an S-turn with a bulged U and a cross-strand A-stack results in a broad minor groove into which the bases of A38 and U42 protrude, to participate in the interhelical interface.

Concurrent with the wide minor groove of stem B is a narrow major groove that seems to be stabilized by the coordination of several divalent metal ions. Two of these are particularly well ordered, and engage in both inner-sphere and outer-sphere coordi-

nation (Figs 2a and 4a). In our crystals, their anomalous scattering identifies them unequivocally as Ca^{2+} (see Methods). Either calcium ion is compatible with a non-selective metal-ion-binding site identified by backbone hydrolysis with terbium (III) (ref. 29), as well as a cobalt-hexammine-binding site identified by NMR studies of the free stem B (ref. 30). Cobalt hexammine can only make outer-sphere contacts¹¹. As the calcium ions in the crystal structure make several inner-sphere contacts, this implies that the conformation of the RNA backbone in the neighbourhood of the bulged S-turn is somewhat flexible.

The conformation of the isolated stem B (ref. 23) differs greatly from that seen in the crystal structure. Only two of the seven non-canonical base pairs observed by NMR (A20·C44 and G21·A43) are retained in the docked conformation. Whereas U41 is extruded into an S-turn in the undocked stem B, it forms a base pair with A22 in the crystal structure. Superposition of the two structures (Fig. 4b, c) demonstrates the pronounced widening of the minor groove of stem B that takes place on docking.

The rearrangements of stems A and B, involving dramatic changes in base-pairing schemes and significant alteration of backbone geometry concomitant with formation of tertiary structure, are not unprecedented. For instance, NMR studies of the P5abc domain of the *Tetrahymena* group I intron have demonstrated that local changes in secondary structure take place in that RNA as it undergoes Mg^{2+} -dependent folding³¹. In contrast, when Mg^{2+} is added to the isolated stems A and B of the hairpin ribozyme in solution, neither undergoes a detectable change in conformation^{22,30}. Thus, binding energy resulting from assembly of the active site, not metal ion coordination, initiates the structural rearrangements described above.

Interhelical interface

The assembly of the hairpin ribozyme buries $1,570 \text{ \AA}^2$ of solvent-accessible surface area of the minor-groove faces of the cores of

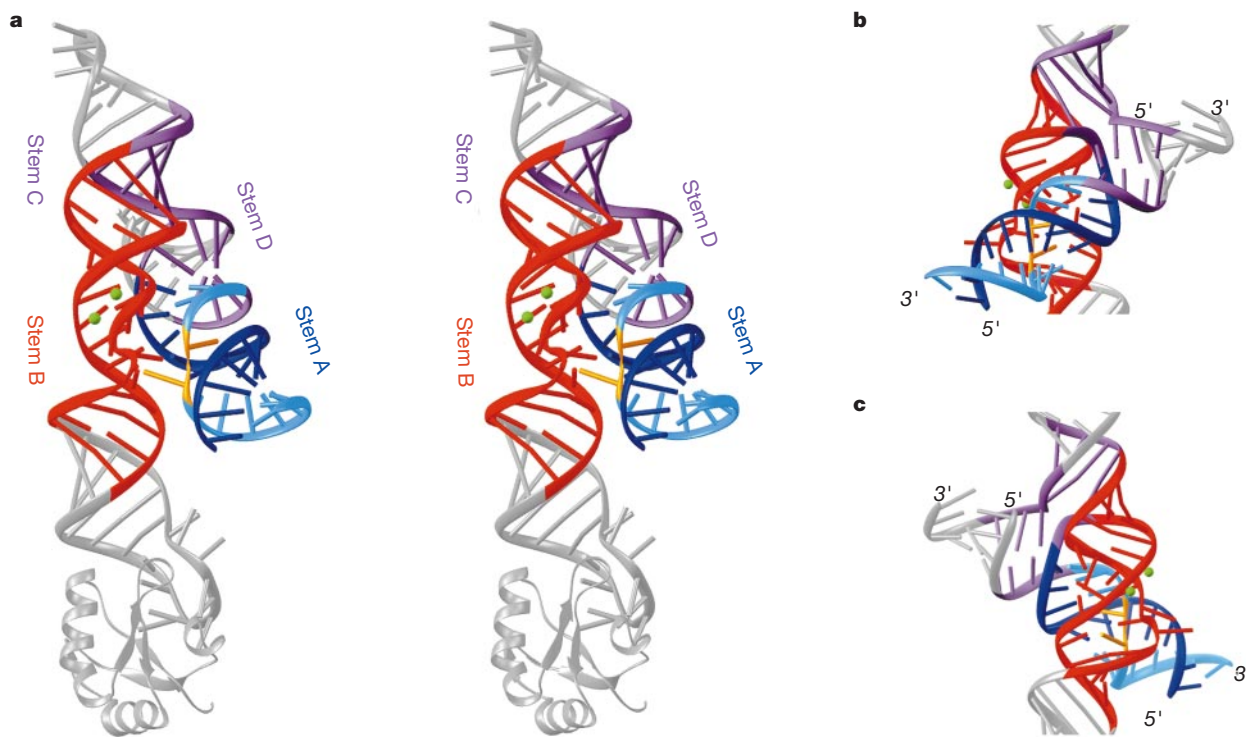


Figure 2 Architecture of the hairpin ribozyme. The backbone of the two RNA strands is depicted as ribbons; the nucleotide bases as sticks (colour code as in Fig. 1). Two tightly bound calcium ions are represented as green spheres. **a**, Stereoview of the docked conformation of the hairpin ribozyme. The scissile bond lies between the two yellow

nucleotides, which are splayed apart. The U1A protein is shown as a grey ribbon.

b, c, Views orthogonal to that in **a**, showing the path of the RNA chains in the four-helix junction, the crossing angle between stems A and B, and the approximately 30° bend at the site of the bulged nucleotides in stem B.

stems A and B (see Methods). A conventional ribose zipper^{12,25} stitches together the N3 atoms of A10 and A24, and the 2'-OHs of nucleotides 10, 11, 24 and 25 (Fig. 5a). Because these four 2'-OH groups were found to be important for ribozyme activity^{17,32,33}, two different proposals for ribose zippers have been made^{21,33}, neither of which predicted the correct pattern of hydrogen bonds.

U42, which is extruded from the stem B stack (Fig. 4a), packs in the interface (Fig. 5a). Its base makes five hydrogen bonds with functional groups of nucleotides in both stem A and B. Previous studies have shown that abolishing these hydrogen bonds is deleterious to ribozyme activity. For example, methylation of the N6 atom of A22 and A23, methylation of the N2 of G11 and deletion of the 2'-OH of U12 all result in reduced activity³³. Methylation of O2 of U41 reduces activity to 0.06% of the wild type³⁴. These and other modification experiments were carried out using minimal hairpin ribozymes; the effects on activity may be less pronounced for the more stable construct of the four-helix junction.

The docked conformation of stem B (Fig. 4a) creates a pocket that binds the extruded syn G+1 nucleotide of stem A (Fig. 3a). Docking of the two stems through their broad minor grooves enables G+1 to make a Watson–Crick pair with C25 (Fig. 5a, b). This was correctly

predicted on the basis of compensatory mutations³⁵. A proposed³⁵ base triple between G+1, C25 and A9 is not observed. A38, which is also in the syn conformation, and A26 form the roof and floor of the G-binding pocket. In addition to these stacking interactions, G+1 donates a hydrogen bond from its 2'-OH to G36, and receives a hydrogen bond to its N7 from the 2'-OH of A38 (Fig. 5b). Thus, the pocket is highly complementary to the chemical identity of G. Individual mutations of G+1, C25, or A38 to any other nucleotide abolish ribozyme activity^{36,37}. Substitutions that would destroy some of the hydrogen bonds in which G+1 participates, such as its replacement for inosine or methylation of O6, result in a complete loss of activity³⁸.

Active site and catalysis

Consistent with biochemical observations that the hairpin ribozyme does not require direct coordination to metal ions for catalysis^{8–10}, the crystal structure shows that the active site consists entirely of RNA. Neither MAD nor model-phased electron density maps show features that could correspond to tightly-bound metal ions in the vicinity of the scissile bond (see Supplementary Information Figs 2 and 3). The extrusion of G+1 from the helical stack of stem A, presumably stabilized by its burial into its cognate pocket in stem B, constrains the riboses of both A–1 and G+1 into C2'-endo puckers (B-form-like twists of the furanose rings). The conformation of these two nucleotides results in an in-line arrangement of the nucleophile (2'-OH of A–1) and leaving group (5'-oxo of G+1) relative to the phosphorus atom at the reaction site (Fig. 5d), as is expected for this ribozyme as it approaches the transition state of its

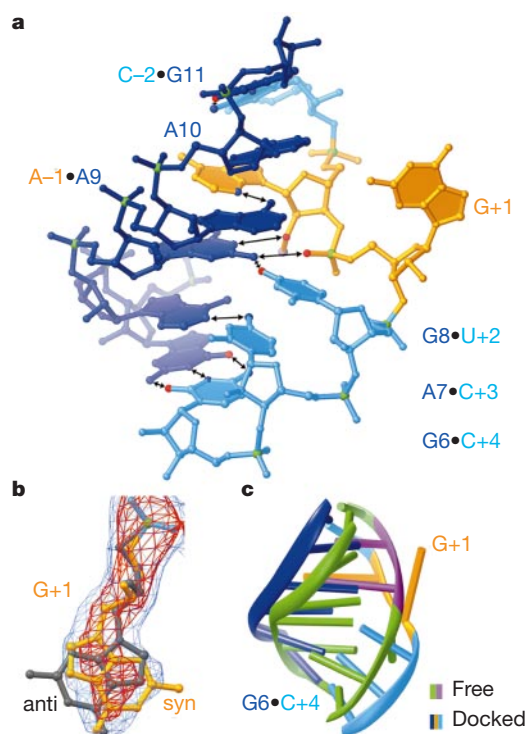


Figure 3 Structure of stem A. **a**, Conformation of nucleotides in the core of stem A in the docked ribozyme, seen from the minor groove. In this and subsequent figures phosphorus atoms (green), and oxygen and nitrogen atoms participating in noteworthy hydrogen bonding (two-headed arrows) are highlighted (red and blue, respectively). The extrusion of G+1 from the helical stack results in the A–1 torsion angles ϵ and ζ being +anticlinal (*ac*) and –antiperiplanar (*ap*), and the G+1 torsion angles α , β and γ being –synperiplanar (*sp*), –*ac* and –synclinal (*sc*). For comparison, the five angles are –*ac*, –*sc*, –*sc*, –*ap* and +*sc* in A-form RNA⁴⁵. **b**, The 2.4-Å resolution ‘solvent-flattened’ experimental electron density map corresponding to nucleotide G+1 is represented as blue and red mesh, for contour levels at one and two standard deviations above mean peak height. G+1 adopts an unusual syn conformation (yellow) in both molecules in the asymmetric unit. For comparison, the incorrect anti conformation is shown (grey). **c**, Superposition of the docked and undocked²² structures of stem A. The free state is shown in green, with nucleotides flanking the scissile bond in magenta. Note the difference in G+1 conformation. The minor groove is considerably narrower in the undocked RNA. The r.m.s. difference between the two structures (for twelve C1' atoms) is 3.6 Å.

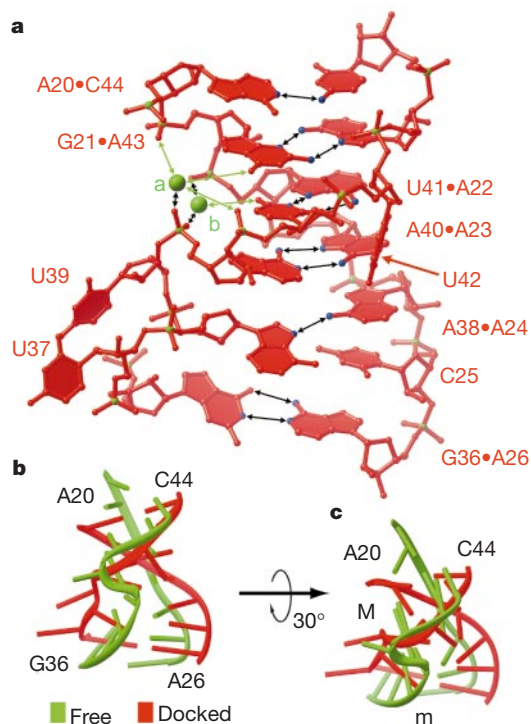


Figure 4 Structure of Stem B. **a**, Docked conformation of stem B, seen from its minor groove. For the major-groove calcium ions, inner-sphere coordination (~ 2.5 Å) is indicated by black arrows; outer-sphere coordination (~ 4.5 Å) by green arrows. The Ca^{2+} ion labelled *b* is very prominent in MAD and anomalous difference Fourier maps (4.3 and 3.8 s.d. above mean peak height). **b**, Comparison of the docked and undocked²³ conformations. The least-squares superposition of the two structures (red and green ribbons, respectively) yields a r.m.s. difference of 6.0 Å (for the 16 C1' atoms shown in **a**). **c**, View resulting from tilting the top of **b** toward the viewer. Upper and lower case ‘m’ denote the major and minor grooves, respectively. Note how the minor groove broadens and the major groove narrows on docking.

S_N2 (substitution nucleophilic, bimolecular) transesterification reaction³⁹. The in-line arrangement is observed in both molecules in the asymmetric unit. The nucleophilic 2' oxygen atom is blocked by a methyl group in our crystallization construct. Even at 2.4 Å resolution, the position of this methyl is defined in both MAD and omit maps (Fig. 5c), establishing unambiguously the RNA backbone conformation near the scissile bond.

The played substrate strand conformation that is observed in the active site of the hairpin ribozyme is markedly similar to that of the dinucleotide mimic UpcA, when bound to ribonuclease-A (ref. 40). In this structure, His 12 lies close to the nucleophile (2'-OH of U) and serves as a general base. The imidazole ring of His 119 stacks on the adenine base, and lies next to the leaving group (5'-oxo of A), serving as a general acid^{41,42}. Lysine 41 counteracts the negative charge that develops in the transition state⁴⁰. Which RNA functional groups could carry out catalysis in the hairpin ribozyme? In the crystal structure, G8 hydrogen bonds to both the pro-Rp oxygen of the scissile phosphate (through N2: 3.0 and 3.1 Å in the two independent complexes in the asymmetric unit) and the nucleophilic 2'-OH (through N1: 2.9 and 2.7 Å, Figs 3a and 5b). The purine base of A38 stacks on G+1. It does not participate in hydrogen bonding in the active site of this ribozyme-inhibitor complex, but if the 5'-oxo leaving group were to swing out, it would be within hydrogen-bonding distance of N1 of A38 (the C5' of G+1 is 3.4 Å from N1 of A38 in both molecules). Neither G8 nor A38 can be mutated to any other nucleotide without abolishing hairpin ribozyme activity^{36,37}. Replacement of G8 by inosine or 6-methyl-G impairs ribozyme activity^{33,38}. A38 is held in place by hydrogen bonding to A24 (Fig. 5b). Deletion of the N6 and N7 involved in this interaction is deleterious^{33,38}. Two additional purines present in the vicinity of the scissile bond (Fig. 5d), A9 and A10, have been implicated in ribozyme activity by mutation and functional-group

substitution^{33,37,38}. In the active conformation, neither residue uses its N1 to make hydrogen bonds. Instead, the Watson-Crick faces of these adenines are in van der Waals contacts with surrounding residues. A cavity¹³ large enough to hold a single water molecule exists here, but electron density maps show no evidence that it is occupied (not shown).

We suggest several catalytic functions for active-site residues that are consistent with the crystal structure and available biochemical data. In the ligation reaction, which the hairpin ribozyme favours over cleavage by a factor of about 30 (ref. 44), G8 and A38 could be functioning as a general acid and a general base. The positions of G8 and A38 relative to the scissile bond are very similar to those of His 12 and His 119, respectively, in ribonuclease-A. The A38-G+1 stacking interaction is markedly similar to the stacking of His 119 onto the A base of UpcA. The pK_a values of the N1 of guanosine and adenosine 5'-monophosphates are 10 and 3.9, respectively⁴⁵. Thus, a considerable change in acidity of these purines would be required for them to catalyse the reverse cleavage reaction in this manner. Alternatively, the N1 atoms of G8 and A38 could serve only to orient the nucleophilic 2'-OH and 5'-OH in the cleavage and ligation reactions, respectively. The hydrogen bond between N2 of G8 and the pro-Rp oxygen would help to relieve the developing negative charge on the scissile phosphate. If either A9 or A10 were to be protonated (as a result of a pK_a shift) they could have a function that is analogous to Lys 41 in ribonuclease-A. Such an electrostatic function would be compatible with their distance from the scissile phosphorus (the N1-P distances are 6.7 Å for A9 in both molecules in the asymmetric unit, and 8.2 and 8.4 Å for A10). Finally, A9 or A10 (with a shifted pK_a value) could activate a water molecule that would then deprotonate the nucleophilic 2'-OH in the cleavage reaction (or the 5'-OH in the ligation reaction). For this to happen, the ribozyme would have to 'breathe' to allow access to solvent.

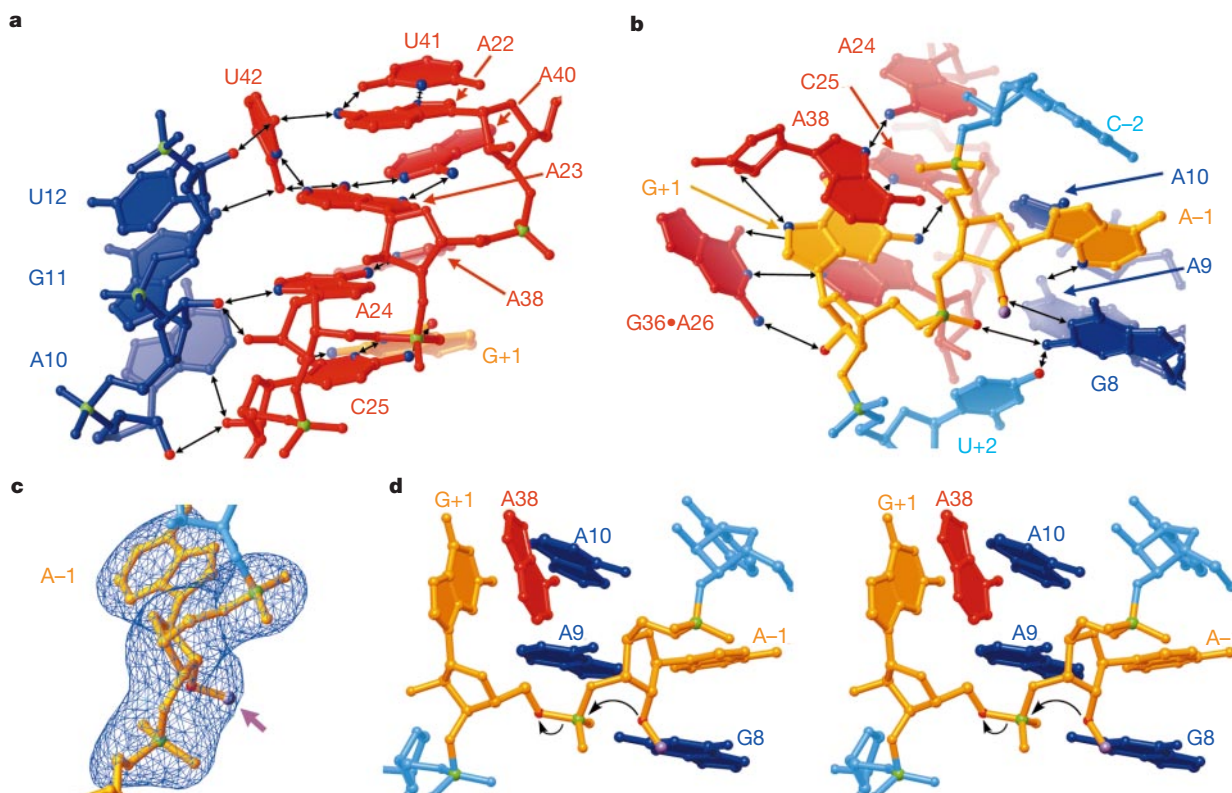


Figure 5 Active site architecture. **a**, Interface of stems A and B, seen from the side opposite to that of Fig. 2a. **b**, Detail of the G+1-binding pocket. **c**, Simulated-annealing $|F_o - F_c|$ omit density map contoured around A-1 at one s.d. The arrow points to the bulge

in the density that corresponds to the methyl group blocking the 2' oxygen. **d**, Stereoview of the active site, showing the in-line-attack conformation and several purine bases that may participate in catalysis.

The crystal structure of a hairpin ribozyme–inhibitor complex demonstrates how an RNA that catalyses a transesterification can align the 2'-OH nucleophile and the 5'-oxo leaving group of the reaction, and suggests how this nucleic acid could be carrying out acid-base catalysis. This structure represents the starting point for further crystallographic and biophysical investigation of the mechanism of action of this natural catalytic RNA. Of particular interest will be the structural determination of a ribozyme–product complex, to understand how the ligation reaction is catalysed by the same active site, the elucidation of the roles of G8, A9, A10 and A38 in catalysis, and the measurement of their pK_a values. If, indeed, the acidity of these purines is shifted, it will raise an important question: how does the organization of the active site accomplish this perturbation? □

Methods

Crystallization and data collection

The insert of plasmid pHF16, which encodes the ribozyme studied followed by a VS cleavage site⁴⁶, was prepared by PCR from overlapping synthetic DNA oligonucleotides. RNA and U1A-RNA-binding domain (RBD) preparation have been described²⁷. The substrate strand RNA was from Dharmacon Research. Grafting the U1A-RBD-binding site onto the end of stem B, and binding of U1A-RBD, did not affect catalytic activity, as judged by comparison with wild-type constructs (Supplementary Information, Fig. 1). Ribozyme and substrate strands (0.5 mM each) were mixed in 2.5 mM MgCl₂, and annealed (slow cooling from 65–22 °C). RNA was mixed 1:1 with U1A-RBD to a final complex concentration of 125 μM in 625 μM MgCl₂. Drops consisting of 5 μl each of complex and reservoir solution (20 mM CaCl₂, 220–260 mM NH₄Cl and 21% MPD) were pre-equilibrated by vapour diffusion at 22 °C. We obtained single crystals by seeding these drops with washed fragments of polycrystalline clusters grown at higher MPD concentrations. Crystals grew to 600 × 300 × 300 μm³ in several weeks. Crystals were stabilized in 5 mM spermine, 20 mM CaCl₂, 200 mM NH₄Cl and 30% MPD, and flash cooled by plunging into liquid nitrogen. The crystals (space group C2) have unit cell parameters $a = 256.3 \text{ \AA}$, $b = 44.1 \text{ \AA}$, $c = 102.4 \text{ \AA}$ and $\beta = 109^\circ$, contain two RNA–protein complexes per asymmetric unit, have a solvent content of 65%, and are stabilized by a combination of protein–protein (5% of solvent-accessible surface area buried in crystal contacts), protein–RNA (30%) and RNA–RNA (65%) interactions. All crystals contained a substrate strand with a 2'-methoxy substitution at residue A–1. Crystal I contained selenomethionine protein. Crystal II contained U1A-RBD with normal methionine, and a substrate strand with an additional 5-iodo-uracil substitution at position 5. Diffraction data (Table 1) were measured (MAD data with the inverse-beam method) at 100 K at beamline 5.0.2 of the Advanced Light Source (ALS, Lawrence Berkeley National Laboratory) and reduced with the HKL package⁴⁷. Data-sets with stronger iodine and calcium anomalous signals were collected with Cu K α radiation for crystal II, and a third crystal that contained 5-iodo-uracil at position 7 (not shown).

Phase determination and structural refinement

Six selenium sites were located in crystal I data with CNS⁴⁸, and confirmed by difference Patterson maps. Two additional sites were located in difference Fourier syntheses. Heavy atom parameters were refined and phases calculated at 2.9 Å resolution using MAD data from crystal I (Table 1). The mean overall figure of merit (FOM) was 0.42 (0.08 between 3.0 and 2.9 Å). Density modification and phase extension⁴⁸ to 2.4 Å resolution against amplitudes from crystal II produced an electron density map (Fig. 3b; see also Supplementary Information, Fig. 2), into which most of the macromolecule residues could be built unambiguously (overall FOM = 0.9; 0.99 between 2.5 and 2.4 Å) using program O (ref. 49). Correct registration of both of the protein chains was confirmed by the selenium positions. As the protein–RNA interface structure is known²⁷, this also confirmed the register of RNA residues in stem B. Anomalous difference Fourier syntheses with amplitudes from the crystals with 5-iodo-uracil at positions 5 and 7 in stem A confirmed the sequence register of this RNA moiety.

Rounds of manual rebuilding, interspersed with positional, torsion-angle-simulated annealing, and restrained individual B-factor refinement⁴⁸ (restraints adjusted to minimize R_{free}), produced the current model ($R_{\text{free}} = 28.9\%$, $R_{\text{work}} = 24.7\%$; $R_{\text{free}} = 41.6\%$ and $R_{\text{work}} = 40.6\%$ between 2.5 and 2.4 Å). Refinement was against all observed crystal II amplitudes (37,148 and 4,142 in the work and test sets) and crystal I MAD phase-probability distributions, using a maximum-likelihood target⁴⁸. A solvent mask and an overall B-factor correction were used throughout; NCS restraints were not. The model contains all RNA atoms for both copies of the ribozyme–inhibitor complex (except for both 5'- γ -phosphates, which are disordered), U1A-RBD residues 6–97 for both chains, 48 water molecules, and 17 calcium and 2 chloride ions (6,405 non-hydrogen atoms). The average B-factor for all RNA atoms in the catalytic core (defined as residues –5 to +3, 7–14, 21–27 and 35–43) is 80.1 Å². Despite this, electron density is well defined for the ribozyme core (Figs 3b and 5c; see also Supplementary Information, Figs 2 and 3). The average real-space R-factor (in a composite simulated-annealing omit map) for all RNA residues in the catalytic core is 6.7%. For comparison, the protein atoms have an average B-factor of 49.9 Å², and have an average real-space R-factor of 4.4%. Ribose puckers were restrained to either C2'-endo or C3'-endo to account for residual features in $|F_o| - |F_c|$ syntheses. There are no breaks in the backbone electron density of the ribozyme core at 0.8 s.d. in a

composite σ_A -weighted simulated-annealing omit $2|F_o| - |F_c|$ map. The cross-validated σ_A mean coordinate error is 0.51 Å. The current model has r.m.s deviations of 0.0114 Å and 1.41° from ideal bond lengths and angles. Of the protein residues, 89.9% lie in the most favoured regions of the Ramachandran plot, with the remainder in additional allowed regions. Figures 2–5 were prepared (using RIBBONS⁵⁰) with molecule B, which is the best-ordered in the ribozyme core. Stems C and D are better ordered in molecule A. The r.m.s. difference between the two complexes in the asymmetric unit is 1.62 Å for all macromolecule atoms, and 0.47 Å for all RNA atoms in the core. A probe radius of 1.4 Å was used to measure solvent-accessible surfaces.

Received 21 December 2000; accepted 12 February 2001.

- Fedor, M. J. Structure and function of the hairpin ribozyme. *J. Mol. Biol.* **297**, 269–291 (2000).
- McKay, D. B. & Wedekind, J. E. in *The RNA World* (eds Gesteland, R. F., Cech, T. R. & Atkins, J. F.) 265–286 (Cold Spring Harbor Press, Cold Spring Harbor, 1999).
- Butcher, S. E., Heckman, J. E. & Burke, J. M. Reconstitution of hairpin ribozyme activity following separation of functional domains. *J. Biol. Chem.* **270**, 29648–29651 (1995).
- Shin, C. *et al.* The loop B domain is physically separable from the loop A domain in the hairpin ribozyme. *Nucleic Acids Res.* **24**, 2685–2689 (1996).
- Murchie, A. I. H., Thomson, J. B., Walter, F. & Lilley, D. M. J. Folding of the hairpin ribozyme in its natural conformation achieves close physical proximity of the loops. *Mol. Cell* **1**, 873–881 (1998).
- Walter, N. G., Burke, J. M. & Millar, D. P. Stability of hairpin ribozyme tertiary structure is governed by the interdomain junction. *Nature Struct. Biol.* **6**, 544–549 (1999).
- Walter, F., Murchie, A. I. H., Thomson, J. B. & Lilley, D. M. J. Structure and activity of the hairpin ribozyme in its natural junction conformation: effect of metal ions. *Biochemistry* **37**, 14195–14203 (1998).
- Hampel, A. & Cowan, J. A. A unique mechanism for RNA catalysis: the role of metal cofactors in hairpin ribozyme cleavage. *Chem. Biol.* **4**, 513–517 (1997).
- Nesbitt, S., Hegg, L. A. & Fedor, M. J. An unusual pH-independent and metal-ion independent mechanism for hairpin ribozyme catalysis. *Chem. Biol.* **4**, 619–630 (1997).
- Young, K. J., Gill, F. & Grasby, J. A. Metal ions play a passive role in the hairpin ribozyme catalyzed reaction. *Nucleic Acids Res.* **25**, 3760–3766 (1997).
- Cowan, J. A. Metallobiochemistry of RNA. Co(NH₃)₆³⁺ as a probe for Mg²⁺ (aq) binding sites. *J. Inorg. Biochem.* **49**, 171–175 (1993).
- Ferré-D'Amaré, A. R., Zhou, K. & Doudna, J. A. Crystal structure of a hepatitis delta virus ribozyme. *Nature* **395**, 567–574 (1998).
- Perrotta, A. T., Shih, I. & Been, M. D. Imidazole rescue of a cytosine mutation in a self-cleaving ribozyme. *Science* **286**, 123–126 (1999).
- Nakano, S. -I., Chadalavada, D. M. & Bevilacqua, P. C. General acid-base catalysis in the mechanism of a hepatitis delta virus ribozyme. *Science* **287**, 1493–1497 (2000).
- Muth, G. W., Ortoleva-Donnelly, L. & Strobel, S. A. A single adenosine with a neutral pKa in the ribosomal peptidyl transferase center. *Science* **289**, 947–950 (2000).
- Chowrira, B. M. & Burke, J. M. Binding and cleavage of nucleic acids by the “hairpin” ribozyme. *Biochemistry* **30**, 8515–8522 (1991).
- Walter, N. G., Hampel, K. J., Brown, K. M. & Burke, J. M. Tertiary structure formation in the hairpin ribozyme monitored by fluorescence resonance energy transfer. *EMBO J.* **17**, 2378–2391 (1998).
- Nowakowski, J., Shim, P. J., Prasad, G. S., Stout, C. D. & Joyce, G. F. Crystal structure of an 82-nucleotide RNA–DNA complex formed by the 10–23 DNA enzyme. *Nature Struct. Biol.* **6**, 151–156 (1999).
- Walter, F., Murchie, A. I. H. & Lilley, D. M. J. Folding of the four-way RNA junction of the hairpin ribozyme. *Biochemistry* **37**, 17629–17636 (1998).
- Hampel, K. J., Walter, N. G. & Burke, J. M. The solvent-protected core of the hairpin ribozyme–substrate complex. *Biochemistry* **37**, 14672–14682 (1998).
- Earnshaw, D. J. *et al.* Inter-domain cross-linking and molecular modeling of the hairpin ribozyme. *J. Mol. Biol.* **274**, 197–212 (1997).
- Cai, Z. & Tinoco, I. J. Solution structure of loop A from the hairpin ribozyme from tobacco ringspot virus satellite. *Biochemistry* **35**, 6026–6036 (1996).
- Butcher, S. E., Allain, F. H. -T. & Feigon, J. Solution structure of the loop B domain from the hairpin ribozyme. *Nature Struct. Biol.* **6**, 212–216 (1999).
- Chowrira, B. M., Berzal-Herranz, A. & Burke, J. M. Novel guanosine requirement for catalysis by the hairpin ribozyme. *Nature* **354**, 320–322 (1991).
- Moore, P. B. Structural motifs in RNA. *Annu. Rev. Biochem.* **68**, 287–300 (1999).
- Butcher, S. E. & Burke, J. M. A photo-cross-linkable tertiary structure motif found in functionally distinct RNA molecules is essential for catalytic function of the hairpin ribozyme. *Biochemistry* **33**, 992–999 (1994).
- Ferré-D'Amaré, A. R. & Doudna, J. A. Crystallization and structure determination of a hepatitis delta virus ribozyme: use of the RNA-binding protein U1A as a crystallization module. *J. Mol. Biol.* **295**, 541–556 (2000).
- Schmidt, S. *et al.* Base and sugar requirements for RNA cleavage of essential nucleoside residues in internal loop B of the hairpin ribozyme: implications for secondary structure. *Nucleic Acids Res.* **24**, 573–581 (1996).
- Walter, N. G., Yang, N. & Burke, J. M. Probing non-selective cation binding in the hairpin ribozyme with Tb(III). *J. Mol. Biol.* **298**, 539–555 (2000).
- Butcher, S. E., Allain, F. H. -T. & Feigon, J. Determination of metal ion binding sites within the hairpin ribozyme domains by NMR. *Biochemistry* **39**, 2174–2184 (2000).
- Wu, M. & Tinoco, I. RNA folding causes secondary structure rearrangement. *Proc. Natl Acad. Sci. USA* **95**, 11555–11560 (1998).
- Chowrira, B., Berzal-Herranz, A., Keller, C. F. & Burke, J. M. Four ribose 2'-hydroxyl groups essential for catalytic function of the hairpin ribozyme. *J. Biol. Chem.* **268**, 19458–19462 (1993).
- Ryder, S. P. & Strobel, S. A. Nucleotide analog interference mapping of the hairpin ribozyme: implications for secondary and tertiary structure formation. *J. Mol. Biol.* **291**, 295–311 (1999).
- Young, K. J. *et al.* The role of essential pyrimidines in the hairpin ribozyme-catalysed reaction. *J. Mol. Biol.* **288**, 853–866 (1999).
- Pinard, R. *et al.* Structural basis for the guanosine requirement of the hairpin ribozyme. *Biochemistry* **38**, 16035–16039 (1999).

36. Siwkowski, A., Shippy, R. & Hampel, A. Analysis of hairpin ribozyme base mutations in loops 2 and 4 and their effects on cis-cleavage *in vitro*. *Biochemistry* **36**, 3930–3940 (1997).
37. Shippy, R., Siwkowski, A. & Hampel, A. Mutational analysis of loops 1 and 5 of the hairpin ribozyme. *Biochemistry* **37**, 564–570 (1998).
38. Grasby, J. A., Mersmann, K., Singh, M. & Gait, M. J. Purine functional groups in essential residues of the hairpin ribozyme required for catalytic cleavage of RNA. *Biochemistry* **34**, 4068–4076 (1995).
39. van Tol, H., Buzayan, J. M., Feldstein, P. A., Eckstein, F. & Bruening, G. Two autolytic processing reactions of a satellite RNA proceed with inversion of configuration. *Nucleic Acids Res.* **18**, 1971–1975 (1990).
40. Richards, F. M. *et al.* Protein structure, ribonuclease-S, and nucleotide interactions. *Cold Spring Harbor Symp. Quant. Biol.* **36**, 35–43 (1971).
41. Findlay, D., Herries, D. G., Mathias, A. P., Rabin, B. R. & Ross, C. A. The active site and mechanism of action of bovine pancreatic ribonuclease 7. The catalytic mechanism. *Biochem. J.* **85**, 152–153 (1962).
42. Roberts, G. C. K., Dennis, E. A., Meadows, D. H., Cohen, J. S. & Jardetzky, O. The mechanism of action of ribonuclease. *Proc. Natl Acad. Sci. USA* **62**, 1151 (1969).
43. Kleywegt, G. J. & Jones, T. A. Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Crystallogr. D* **50**, 175–177 (1994).
44. Fedor, M. J. Tertiary structure stabilization promotes hairpin ribozyme ligation. *Biochemistry* **38**, 11040–11050 (1999).
45. Saenger, W. *Principles of Nucleic Acid Structure* (Springer, New York, 1984).
46. Ferré-D'Amaré, A. R. & Doudna, J. A. Use of *cis*- and *trans*-ribozymes to remove 5' and 3' heterogeneities from milligrams of *in vitro* transcribed RNA. *Nucleic Acids Res.* **24**, 977–978 (1996).
47. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
48. Brünger, A. T. *et al.* Crystallography and NMR system: a new software system for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
49. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
50. Carson, M. Ribbons. *Methods Enzymol.* **277**, 493–505 (1997).

Supplementary information is available on *Nature's* World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of *Nature*.

Acknowledgements

We thank T. Earnest, L. Hung and G. McDermott for help at ALS beamline 5.0.2; C. Hoang for biochemical support; J. Bolduc, P. Heath and B. Shen for computational and crystallographic support; K. Nagai for U1A plasmids; and S. Biggins, M. Holmes, P. Li, M. Rosenberg, S. Ryder, S. Sigurdsson, B. Stoddard, S. Strobel, R. Strong, G. Varani, D. Wilson and K. Zhang for discussions. This work was supported by institutional funds from the Fred Hutchinson Cancer Research Center (FHCRC). Access to ALS beamline 5.0.2 as part of the principal research consortium was made possible by general support from the FHCRC. P.B.R. is a post-doctoral trainee of the Chromosome Metabolism and Cancer training grant from the National Cancer Institute to the FHCRC.

Correspondence and requests for materials should be addressed to A.R.F. (e-mail: aferre@fhcrc.org). Atomic coordinates have been deposited in the Protein DataBank under accession number 1HP6.