Article No. mb981789 *J. Mol. Biol.* (1998) **279**, 621–631

JMB

A General Module for RNA Crystallization

Adrian R. Ferré-D'Amaré, Kaihong Zhou and Jennifer A. Doudna*

Department of Molecular Biophysics and Biochemistry and Howard Hughes Medical Institute, Yale University P.O. Box 208114, New Haven CT 06520-8814, USA

*Corresponding author

Crystallization of RNA molecules other than simple oligonucleotide duplexes remains a challenging step in structure determination by X-ray crystallography. Subjecting biochemically, covalently and conformationally homogeneous target molecules to an exhaustive array of crystallization conditions is often insufficient to yield crystals large enough for X-ray data collection. Even when large RNA crystals are obtained, they often do not diffract X-rays to resolutions that would lead to biochemically informative structures. We reasoned that a well-folded RNA molecule would typically present a largely undifferentiated molecular surface dominated by the phosphate backbone. During crystal nucleation and growth, this might result in neighboring molecules packing subtly out of register, leading to premature crystal growth cessation and disorder. To overcome this problem, we have developed a crystallization module consisting of a normally intramolecular RNA-RNA interaction that is recruited to make an intermolecular crystal contact. The target RNA molecule is engineered to contain this module at sites that do not affect biochemical activity. The presence of the crystallization module appears to drive crystal growth, in the course of which other, non-designed contacts are made. We have employed the GAAA tetraloop/tetraloop receptor interaction successfully to crystallize numerous group II intron domain 5-domain 6, and hepatitis delta virus (HDV) ribozyme RNA constructs. The use of the module allows facile growth of large crystals, making it practical to screen a large number of crystal forms for favorable diffraction properties. The method has led to group II intron domain crystals that diffract X-radiation to 3.5 Å resolution.

Keywords: RNA; crystallization; group II intron; HDV ribozyme; tetraloop receptor

Introduction

New protein structures are now being solved daily, yet only four distinct, large RNA structures have been determined by X-ray crystallography: transfer RNA (reviewed by Holbrook & Kim, 1997), the hammerhead ribozyme (Pley *et al.*, 1994a; Scott *et al.*, 1995b), the P4-P6 domain of the *Tetrahymena thermophila* group I intron (Cate *et al.*, 1996) and the 62-nucleotide fragment I from *Escherichia coli* 5 S ribosomal RNA (Correll *et al.*, 1997). The crystallization and structure solution of short oligonucleotide duplexes has become almost routine (Berman *et al.*, 1996), as has the structure determination of small RNA fragments by nuclear magnetic resonance spectroscopy (Varani *et al.*, 1996). Despite recent refinements in methods to produce large quantities of covalently homogeneous RNA, and the development of several sparse matrix crystallization screens tailored for RNA (reviewed by Holbrook & Kim, 1997), the elucidation of the structures of large, functional RNA molecules remains a challenging endeavor because of the difficulty in obtaining large, well-ordered crystals that yield X-ray diffraction data to resolutions that allow biochemical insight.

A critical step for macromolecular crystallization is the design of the target molecule. The designer endeavors to maintain biochemical activity, while maximizing the ''crystallizability'' of the molecule of interest. Historically, protein crystallographers have used highly purified molecules from natural sources, often performing crystallization trials with homologous proteins from different species. Since homologues differ primarily in their molecular surfaces, a molecule from one source fortuitously

Abbreviations used: DLS, dynamic light-scattering; HDV, hepatitis delta virus; TR, tetraloop-tetraloop receptor.

0022–2836/98/230621–11 $25.00/0

might be easier to crystallize because of the particular juxtaposition of functional groups of neighboring molecules in a crystal. Limited proteolysis of target molecules has been employed to remove flexible, unstructured regions, increasing conformational purity, and thus, crystallizability (McPherson, 1982). The advent of molecular biological methods has made the production of molecules with unnatural ends (e.g. a single domain from a multi-domain molecule) routine. Since an unfortunate choice of molecular ends can result in conformationally heterogeneous molecules that are difficult to crystallize (Ferré-D'Amaré & Burley, 1994), domain elucidation and domain border definition have become important issues in the design of recombinant crystallization targets (Cohen, 1996).

Similar considerations apply to RNA crystallization. Since the majority of RNA targets are prepared not from natural sources, but by *in vitro* enzymatic or chemical synthesis, design decisions must be made from the outset. The double helix is the basic element of RNA structure. Because of the strength of stacking interactions in aqueous solutions, the ends of RNA duplexes are almost invariably involved in molecular contacts in crystals. Therefore, choice and systematic variation of the ends of helices is as useful in RNA crystallization as it is in DNA or DNA-protein complex crystallization (e.g. Aggarwal, 1990; Schultz *et al.*, 1990; Anderson *et al.*, 1996). Helical regions of larger RNA molecules can be predicted with some accuracy using phylogenetic covariation analysis or enzymatic probing; furthermore, the solvent accessibility of helix faces can be explored chemically. In the case of the hammerhead ribozyme, a large body of biochemical data guided the design of a number of crystallization targets varying in strand composition and connectivity, helix length and ends (Pley *et al.*, 1993; Scott *et al.*, 1995a). Knowledge of the position of helical elements within the group I intron also aided in the design of minimal functional and crystallizable domains in that system (Doudna *et al.*, 1993; Doudna & Cech, 1995). As with proteins, variation in solvent-exposed regions of RNA, such as loops, can result in improved crystal quality (Golden *et al.*, 1997).

Here, we advance a new tool for crystallizing RNA that differs fundamentally from well-established ''helix engineering''. We have placed designed intermolecular, or crystal, contacts into crystallization targets. These contacts were introduced in the form of a bipartite crystallization module consisting of an 11-nucleotide tetraloop receptor and a GAAA tetraloop, positioned to make exclusively intermolecular contacts, without adversely affecting biochemical activity of the target molecules. Introduction of the module transformed a fragment of the active site of a group II intron, and a hepatitis delta virus (HDV) ribozyme, from molecules not amenable to crystallization to RNAs that crystallized readily under a variety of conditions. Its usefulness in crystallizing two dissimilar RNAs shows that the tetraloop receptor/tetraloop interaction can be exploited in different molecular contexts. The ease of crystallization allowed us to perform a systematic search for good crystals and we report one crystal form of a 70-nucleotide RNA that diffracts X-rays to at least 3.5 Å resolution.

Results and Discussion

Group II intron domains 5 and 6 as crystallization targets

Our first crystallization target was an RNA molecule comprising domains 5 and 6 (d5 and d6) of the yeast mitochondrial group II self-splicing intron ai5γ. Group II introns are catalytic RNAs found in organellar genomes of fungi and plants, and in some bacteria (reviewed by Michel & Ferat, 1995). At elevated temperatures and ionic strength *in vitro*, they excise themselves from the transcript and ligate the flanking exons. These introns are very long, often more than 1000 nucleotides, and show minimal sequence conservation. Their sequences can, however, be arranged into a conserved pattern of six predominantly helical domains (Michel *et al.*, 1989; Figure 1(a)). Splicing proceeds through two sequential transesterifications. In the first, the 2′ hydroxyl group of a conserved ''bulged'' adenosine base in d6 attacks the 5′ exon-intron junction, forming the free 5′ exon and a lariat intron. In the second step, the terminal hydroxyl group of the 5′ intron attacks the boundary of the intron and the 3′ exon, releasing the ligated exons and a lariat intron. Water competes with the adenosine 2′ hydroxyl group as the nucleophile for the first transesterification (Daniels *et al.*, 1996). This hydrolytic pathway results in ligated exons and a linear intron.

The only stretch of highly conserved sequence in group II introns is the 34 nucleotide d5; the sequence conservation underscores its central role in forming the active site. Deletion of d5 from an intron abolishes activity; however, d5 can be added *in trans* to the rest of the intron to reconstitute splicing (Jarrell *et al.*, 1988). Deletion of d6, or of the branch-point adenosine base, results in an intron that splices using exclusively the hydrolytic pathway in the first step (Chin & Pyle, 1995). The first step of splicing can be reconstituted by mixing an RNA comprising a 5′ exon and d1 through d3 (ExD123) with d5, d5 binding to ExD123 with a dissociation constant of 800 nM (Pyle & Green, 1994). The resulting reaction is kinetically and stereochemically similar to the one that takes place in the intact intron (Pyle & Green, 1994; Podar *et al.*, 1995; Michels & Pyle, 1995). Thus, the isolated d5 appears to adopt a conformation similar to that one present in the active RNA. Examination of the role of d5, both in binding to the rest of the intron and in catalysis, has produced a picture of a helix surface densely packed with biochemically critical functional groups (Chanfreau & Jacquier, 1994;

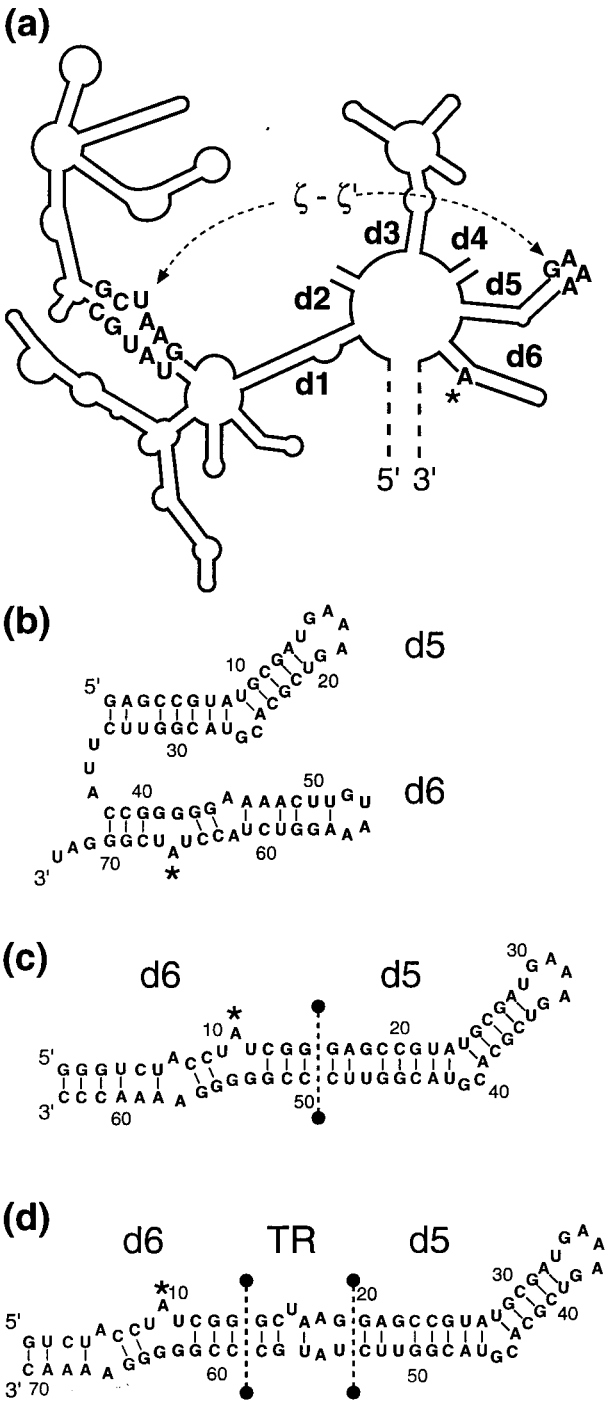

Figure 1. (a) A representation of the arrangement of helical elements of the group II intron ai5γ from yeast mitochondria, deduced from phylogenetic analysis by Michel *et al.* (1989). Six helical regions called domains 1 through 6 (d1 to d6) radiate from a central wheel. The 5′ and 3′ exons are represented as broken lines. Domains 2 and 4 are foreshortened in this Figure. The long-range interaction between the GAAA tetraloop that caps d5 and the tetraloop receptor present in d1 (ζ-ζ′, Costa & Michel, 1995) is indicated. The branch-point adenosine residue of d6 is marked with an asterisk (*). (b) Sequence and hypothetical base-pairing scheme of the wild-type d56 construct. (c) Same as (b), but for the coaxially stacked, circularly permuted d56c construct. The broken vertical line dividing d5 from d6 is conceptual; the molecule is covalently continuous. (d) Same as (c), but for the d5/TR/d6 construct engineered for crystallization by addition of the tetraloop receptor (TR). As in (c), the broken vertical lines are solely for illustration; the molecule is a covalently continuous 70 nucleotide (nt) RNA. This construct yielded the crystal shown in Figure 6(a).

Peebles *et al.*, 1995; Abramovitz *et al.*, 1996; Jestin *et al.*, 1997; Konforti *et al.*, 1998).

In preliminary experiments we found that the intact intron, the intron lariat and d1 are all poorly structured and aggregate heavily at crystallographic concentrations, making them unsuitable for crystallization (data not shown). We focused our structural studies on d5 because this short RNA forms an important part of the active site and was likely to be well structured. We decided also to pursue the crystallization of d6; this helical region provides the nucleophile for the first step of splicing and must therefore be close to the active site. In addition, biochemical studies have suggested that d6 might bind d5 (Dib-Hajj *et al.*, 1993). The bulged adenosine base that provides the nucleophile in the first transesterification catalyzed by group II introns is similar to the branch-point adenosine base of nuclear pre-mRNA splicing. Hence, the structure of d6 might also shed light on that mode of splicing. Portmann *et al.* (1996) found that a model RNA oligonucleotide with a bulged adenosine base self-cleaved in the presence of divalent cations. Since neither group II intron d6 nor spliceosomal branch-points self-cleave, this oligonucleotide structure is a poor model for the structure of the branch-point. A final reason for preferring d5-d6 (d56) constructs to d5 constructs is that short, isolated hairpins like d5 are notorious for disproportionating into duplexes as the favorable stacking interactions in the crystal drives the equilibrium in that direction (e.g. Holbrook *et al.*, 1991).

Conformational flexibility and crystallization

The d56 RNA shown in Figure 1(b) catalyzed hydrolysis of the exon-intron junction at a rate comparable to the isolated d5 (Figure 2(a), lanes 1 to 7) and formed a branched species (Figure 2(b)), as expected from previous work (Chin & Pyle, 1995). Catalysis of the second step of splicing was not examined. Inspection of Figure 1(b) raises two concerns about the crystallizability of such a molecule. First, do the three unpaired nucleotides at the 3′ end of the molecule result in aggregation? Second, are the two helices, d5 and d6, rigidly or flexibly linked together when not bound to the rest of the intron? Flexible linkage would probably preclude crystallization.

We examined the hydrodynamic behavior of d56 by dynamic light-scattering (DLS), since the usefulness of this technique in determining the crystallizability of macromolecules has been well

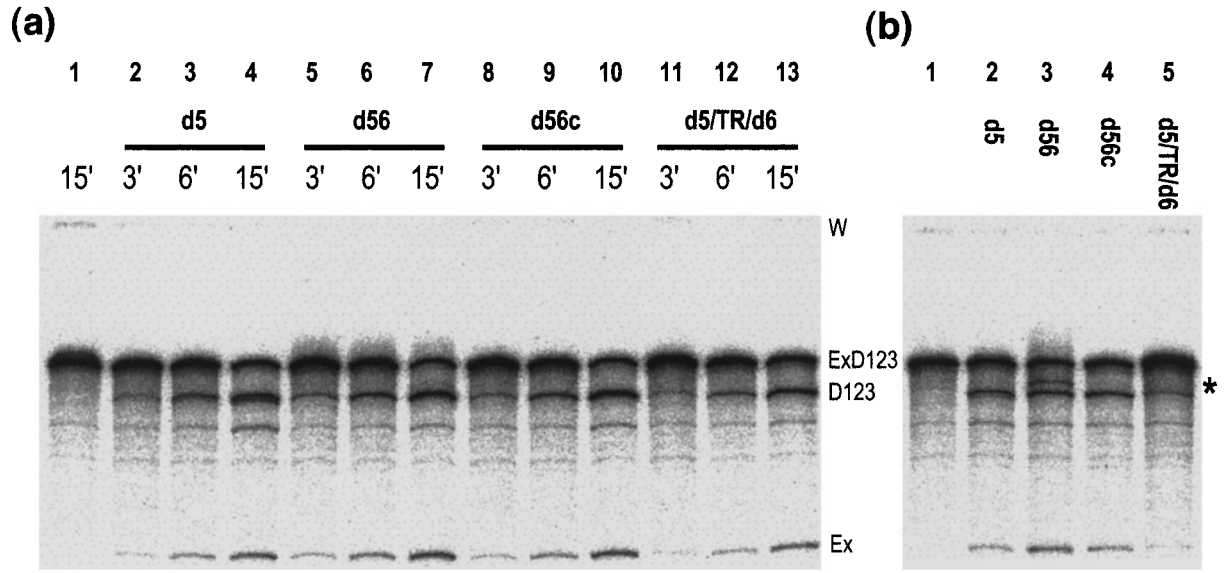

Figure 2. Comparison of biochemical activity of wild-type domain 5 (d5) with those of wild-type d56, coaxially stacked d5-d6 (d56c), and a d5-d6 construct containing the tetraloop receptor (d5/TR/d6). (a) Hydrolysis of the 5′-exon-intron boundary catalyzed by domain 5 was examined in single-turnover regime. The RNA products were resolved by electrophoresis on an 8 M urea/6% polyacrylamide gel, and visualized by autoradiography. Lane 1, body labeled substrate (ExD123) comprising a 293 nt 5′ exon (Ex) followed by the first three domains (D123, 710 nt) of the ai5γ group II intron, incubated with reaction buffer alone for 15 minutes. Lanes 2 through 4, ExD123 in reaction buffer incubated with excess d5 for the times indicated. Lanes 5 through 7, same as previous three lanes, but with wild-type d56. Lanes 8 through 10, same reactions but with the coaxially stacked d56c. Lanes 11 through 13, time-points in a reaction with d5/TR/d6. The band running below D123 is the result of a well-documented (e.g. Pyle & Green, 1994) non-catalytic degradation reaction. W denotes the position of the wells. (b) Reactions under branching conditions. ExD123 was incubated in reaction buffer alone (lane 1), or with the indicated RNAs (lanes 2 to 5) for 30 minutes, and the products of the reactions resolved and visualized as in (a). The position of the branched species is marked by an asterisk (*).

established (D'Arcy, 1994; Ferré-D'Amaré & Burley, 1994, 1997). DLS is exquisitely sensitive to higher-order aggregation, and experience suggests that if a molecule is monodisperse at modest concentrations it is likely to crystallize, while if it exists in a mixture of aggregation states or conformations, it is unlikely to do so. Figure 3(a) summarizes the results of DLS measurements on d56. The multimodal distribution of apparent molecular masses is evidence of multiple aggregation states or conformational heterogeneity. To rule out the possibility that this behavior was the result of covalent heterogeneity, a d56 molecule was produced using two ribozymes, so that both 5′ and 3′ termini would be homogeneous (Ferré-D'Amaré & Doudna, 1996). We also produced a covalently homogeneous d56 missing the last three nucleotides of the RNA shown in Figure 1(b), to examine if these were causing the polydisperse behavior. Both molecules had biochemical activity comparable to the parent d56; examination by DLS showed that they also had the same aggregation behavior (data not shown). Not unexpected, when subjected to crystallization trials, none of nine wild-type d56 molecules prepared yielded crystals.

To determine whether the polydisperse behavior of d56 results from flexible linkage of d5 and d6, we produced a series of circularly permuted RNAs that would force the two domains to stack coaxially. One such molecule is shown in Figure 1(c). This molecule catalyzes the hydrolytic scission of the ExD123 construct at a rate comparable to that of wild-type d5 or d56 (Figure 2, lanes 8 to 10); it does not, however, branch measurably. None of our coaxially stacked constructs, which differ in the position of the 5′ and 3′ ends and also in the number of ''linker'' nucleotides (one, two or three nucleotides) between d5 and d6, catalyzed branching. Thus, d5 and d6 are probably not coaxially stacked in the active site of this group II intron during the first transesterification, and no other helical element needs to stack with d5 to produce an active intron, since the d6 in d56c would have displaced it. DLS showed d56c to be monodisperse, and the apparent molecular mass was consistent with the RNA being monomeric (Figure 3(b)). We conclude that in wild-type d56, the helices d5 and d6 are flexibly linked, so that in the absence of the rest of the intron, the relative orientation of the two helices is arbitrary and variable.

Crystallization trials with ten coaxially stacked d56 constructs differing in the placement of the circularly permuted molecular ends and in the number and placement of spacer nucleotides between the two domains produced some crystalline matter. Nevertheless, despite considerable effort at optimization of growth conditions, the best that could be obtained were very long needles with a thickness

not exceeding a few micrometers. Such crystals are useless for X-ray diffraction experiments. A new design principle was needed.

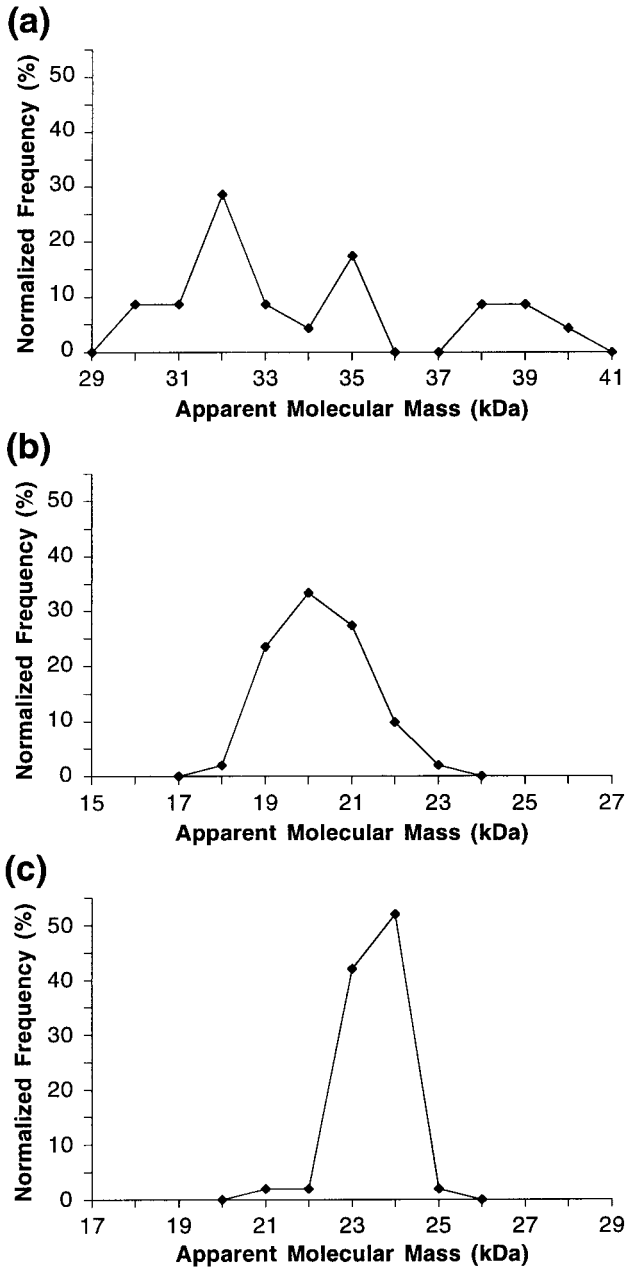

Figure 3. Histograms of the distribution of apparent molecular masses in solution computed from dynamic light-scattering measurements on three different RNA constructs comprising domains 5 and 6 of ai5γ. (a) Wild-type d56 displays a broad multimodal distribution. The covalent mass of this construct is 23 kDa. (b) Coaxially stacked d56c has a unimodal distribution centered around its covalent molecular mass (20 kDa). (c) Coaxially stacked d5-d6 construct containing the tetraloop receptor (d5/TR/d6) shows a sharp unimodal distribution with a mean close to its covalent mass (22 kDa). The means (and standard deviations) of the translational diffusion coefficients of d56c and d5/TR/d6 are 1077×10^{-13} m^2 s^{-1} (28.8) and 1011×10^{-13} m^2 s^{-1} (11.5), respectively.

Use of the GAAA tetraloop-receptor interaction as a crystallization module

We reasoned that an RNA, whether an extended helix or a globular domain, would have a molecular surface dominated by an essentially uniform array of negative charges. During crystal nucleation or growth, this might result in neighboring molecules packing subtly out of register, leading to disorder, or premature cessation of growth, or fiber-like growth in one dimension (needle-shaped crystals). If this were the case, placement of a differentiated molecular surface into the target molecule might allow neighboring molecules to bind each other stereospecifically, leading to a well-registered array of crystal contacts. RNA is well-suited to the engineering of such designed crystal contacts because the boundaries of RNA helices are relatively easy to define, and the exposure of a helix face to solvent can be established by standard biochemical methods, so that the placement of the designed contact in the target molecule can be approached rationally. Furthermore, unlike protein secondary structure elements, RNA helices are typically stable in isolation, so that placement of an extraneous element designed to make a crystal contact is unlikely to disrupt the structure of the parent molecule and, in turn, the parent molecule is unlikely to disturb the structure and function of the sequence designed to make crystal contacts.

Ideally, the engineered crystal contact would not only be stereospecific, so that neighboring molecules would be in tight register, but also directional, or asymmetric, so that growth of an infinite array or crystal, and not a closed array or cluster, would be promoted. In addition, for the introduced crystal contact to be useful in a variety of contexts in different molecules, it would be desirable that the surfaces making the contact be compact and portable (modular). Crystal growth takes place at high macromolecule concentrations, and as molecules pack together in register, driven by the designed contacts, other non-designed contacts are expected to take place and stabilize the growing crystal. Therefore, the interaction of the crystallization module does not have to be strong. Subverting a naturally occurring long-range, or tertiary, interaction (thought to be typically weak) to serve instead as a crystal contact would meet the requirements of modularity, compactness and directionality.

We decided to employ the GAAA tetraloop-tetraloop receptor (TR) interaction identified by Costa & Michel (1995). They found that a highly conserved 11 nucleotide sequence that occurs in many large catalytic RNAs specifically binds a tetraloop of sequence GAAA. In fact, d5 of ai5γ is capped by such a tetraloop, and a TR to which it appears to bind is present in d1 (Figure 1(a)). Disruption of this tetraloop or the TR sequence compromises activity, which can be restored by compensatory mutations in the two elements (Costa & Michel,

1995). This interaction meets the criteria we set forth for a crystallization module in that the two elements are small and probably functional when taken out of their original molecular contexts, and the interaction is asymmetric or directional. Furthermore, in the particular case of d5, the interaction is functionally important, so that if it were recruited to facilitate crystal growth, the crystal contact itself would be biochemically interesting. Finally, the crystal structure of the P4-P6 domain of a group I intron, which has an intramolecular tetraloop-TR interaction, has been determined (Cate *et al*., 1996), providing further support for the use of this interaction as a crystallization module. This interaction is structurally distinct from the fortuitous crystal contact made by a GAAA tetraloop in a hammerhead ribozyme crystal (Pley *et al*., 1994b).

We prepared 24 RNAs based on d56c containing a TR either between d5 and d6 or at the end of d6. The TR was placed in either orientation, and with different numbers of nucleotide base-pair spacers between the three elements d5, TR and d6. One such molecule, d5/TR/d6, is shown in Figure 1(d). This molecule catalyzes the specific hydrolysis of ExD123 with a rate similar to that of wild-type d5 and d56 (Figure 2(a), lanes 11 to 13). As expected from our findings with d56c, it does not catalyze branch formation; unexpectedly, given its activity under hydrolysis conditions, the molecule is less active than d56c under salt conditions that favor branching of the wild-type molecule (Figure 2(b)). Examination of the hydrodynamic behavior of d5/TR/d6 at approximately 0.1 mM by DLS revealed a monodisperse ϕ molecule (Figure 3(c)). The molecule appears to be monomeric, which implies that the GAAA-TR receptor interaction has a dissociation constant at least in the millimolar range. As d5 binds ExD123 with a dissociation constant of 800 nM (Pyle & Green, 1994), the tetraloop probably provides one of several contacts that stabilize the interaction.

We screened d56+TR constructs for crystallization using the sparse matrix screens of Doudna *et al*. (1993) and Scott *et al*. (1995a). The two complement each other; the former is rich in small organic precipitants, while the latter covers polyethylene glycol (PEG)/salt mixtures thoroughly. We promptly found that addition of the crystallization module had transformed d56c, a difficult crystallization target, into molecules that crystallized readily under a variety of conditions, in some cases within minutes of mixing with precipitant. In fact, virtually all d56 constructs with a TR produced crystals during the very first trials, and many of them produced several different crystal forms. Crystals grew from conditions ranging from small organic precipitants (MPD, ethanol, isopropanol), to PEG/salt mixtures, to high salt (lithium sulfate, magnesium sulfate and ammonium sulfate). A number of these crystals could be coaxed to grow to sizes suitable for X-ray diffraction.

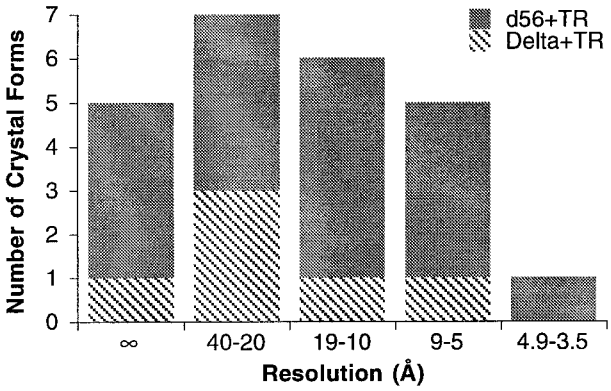

Figure 4. Histogram summarizing best diffraction observed for 24 crystal forms obtained from 16 different RNA constructs (some RNAs yielded more than one crystal form which could be grown to large dimensions) containing the tetraloop/tetraloop receptor crystallization module. Both group II intron d56 (12 RNA constructs, 18 crystal forms, gray bars) and HDV ribozyme RNA constructs (hatched bars) are represented. All data are from crystals whose smallest dimension was at least 0.1 mm. No observable diffraction is denoted "∞". Crystals were mounted in capillaries directly from their mother liquors and examined by oscillation photography at their temperature of growth, with X-rays from a rotating anode source. The value on the abscissa corresponds to the highest resolution spots (observed on an imaging plate area detector) whose positions and intensities were consistent with the diffraction pattern seen at lower resolution. Several of these crystal forms were subsequently stabilized, flash-cooled, and their diffraction examined at 110 K.

Summarized in Figure 4 are the diffraction properties of 16 RNAs that were engineered for crystallization using the GAAA/TR module. Represented are 18 different crystal forms of 12 different d56+TR RNA constructs. It is apparent that the degree of order of the crystals varies considerably, and that the majority of crystals lack enough order to make them suitable for structure determination. This variability in order presumably reflects the different number and nature of the non-designed crystal contacts that the molecules make in the different crystal types. Because these other contacts cannot be predicted and designed, use of the crystallization module still requires making a number of constructs to find one that produces well-ordered crystals. One of the d56 constructs we made, d5/TR/d6 (Figure 1(d)) produced hexagonal crystals that diffracted X-rays to at least 3.5 Å resolution (Figure 6(a) and (c)).

The GAAA/TR module can be employed for crystallizing a globular RNA

The utility of the tetraloop-TR module in producing easily crystallizable constructs of coaxially stacked d56 prompted us to investigate whether the module would also enable easy crystallization

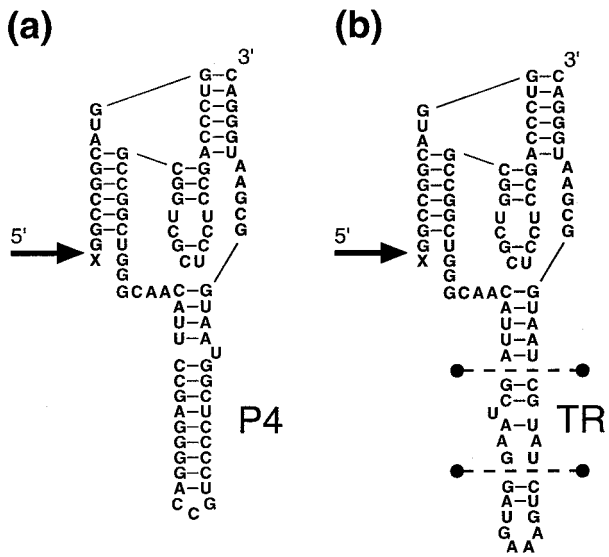

Figure 5. Engineering the hepatitis delta virus ribozyme for facile crystallization with the tetraloop receptor/tetraloop crystallization module. (a) Sequence and hypothetical secondary structure of a genomic-strand delta ribozyme (after Tanner *et al.*, 1994). Following self-cleavage, the 5′ end of the molecule is at the G residue marked with an arrow. Paired region 4 (P4) can be drastically shortened, or its sequence modified (preserving base-pairing) without affecting ribozyme activity (Been *et al.*, 1992; Tanner *et al.*, 1994). (b) Sequence and secondary structure of a delta ribozyme engineered for crystallization. P4 has been altered to consist of a tetraloop receptor (TR), and is capped with a GAAA tetraloop. The broken horizontal lines dividing the molecule into three segments are illustrative only; the molecule is a covalently continuous strand of RNA. This is the delta/TR construct that yielded the crystals shown in Figure 6(b).

of a globular RNA molecule otherwise difficult to crystallize. We chose to apply the module to the crystallization of the ribozyme from the genomic strand of the human hepatitis delta virus (HDV, Figure 5(a)).

The HDV ribozyme is responsible for processing linear multimeric viral transcripts to monomeric units during rolling circle replication. The RNA is extraordinarily stable *in vitro*, catalyzing cleavage at the phosphodiester bond preceding it even in extreme conditions such as 8 M formamide or 8 M urea (reviewed by Been, 1994; Tanner *et al.*, 1994; Lai, 1995). The minimal sequence requirements preceding the cleavage site imply that the core of the RNA derives little or no stabilization from contacts with upstream nucleotides. This suggests that the ribozyme has a similar conformation before and after cleavage, an idea supported by spectroscopic data (Duhamel *et al.*, 1996). This prompted us to attempt to crystallize RNA molecules that had cleaved themselves from the precursor. In addition to having adopted the correct fold for activity, such molecules would have a covalently homogeneous 5′ terminus. Our initial attempts at crystallization of wild-type HDV ribozymes, how-

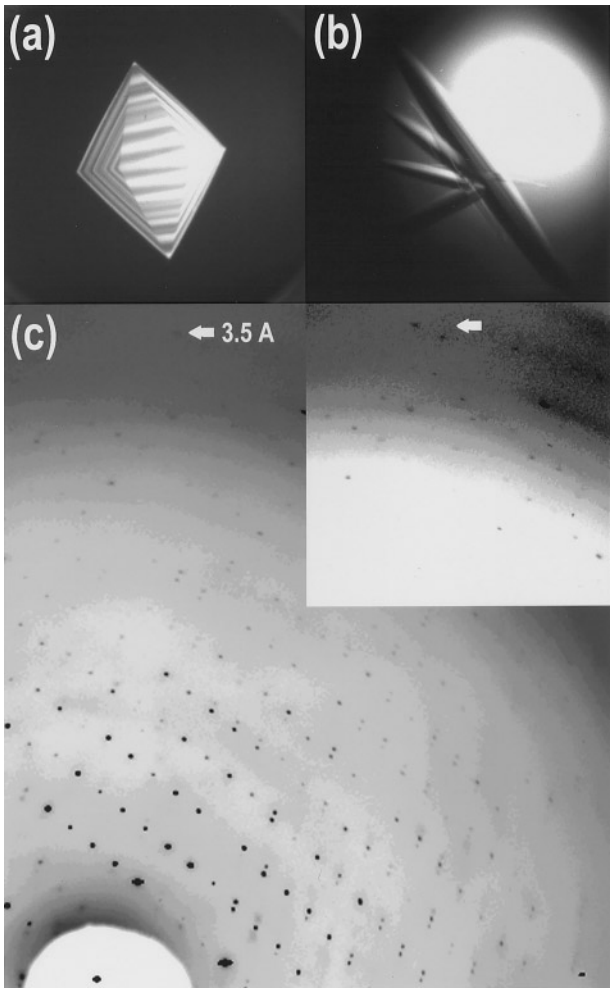

Figure 6. (a) Example of a d5/TR/d6 crystal grown from a mixture of magnesium and ammonium sulfates (see Materials and Methods). The specimen comprises a sagittal half of a hexagonal bipyramid; its longest dimension measures approximately 0.8 mm. The crystals have point-group symmetry 622 and unit cell dimensions $a = 91.4$ Å, $c = 233.9$ Å. (b) Examples of delta/TR crystals grown from polyethylene glycol/KCl (see Materials and Methods). Crystals are elongated prisms with square cross-sections, have a primitive tetragonal lattice, and unit cell dimensions $a = 98.7$ Å, $c = 75.3$ Å. The largest specimen shown has a length of over 1 mm. Samples of both crystal forms were washed extensively in stabilizing solution, dissolved with dilute EDTA, and their RNA analyzed by denaturing polyacrylamide gel electrophoresis. In both cases, the crystals contained full-length RNA. (c) Oscillation photograph taken from a flash-cooled d5/TR/d6 crystal at beamline A-1 of the Cornell High Energy Synchrotron Source (CHESS) on an Area Detector Systems (San Diego, CA) CCD detector placed at a distance of 200 mm from the crystal. The collimator diameter was 0.1 mm and the oscillation range 0.7°. Diffraction from the crystal can be seen to extend to at least 3.5 Å in the inset, which is a portion of the image adjacent to the arrow in the main panel, printed at higher contrast.

ever, failed to produce any crystalline matter whatsoever.

Previous mutational studies (Been *et al*., 1992; Tanner *et al*., 1994) demonstrated that the length and sequence of stem (or paired region) 4 of the HDV ribozyme can be varied at will, without adversely affecting the activity of the molecule. Therefore, we replaced this stem of a well-characterized genomic-strand HDV ribozyme with our crystallization module. We made constructs with the TR in either orientation, and capped the stem with the tetraloop and proximal nucleotides of ai5γ d5 (Figure 5(b)). The TR and the short stem that supports the GAAA tetraloop are coaxially stacked, so that their interaction, if it takes place, must occur *in trans*, therefore promoting formation of intermolecular contacts.

We produced delta + TR constructs by allowing them to self-cleave, confirming that introduction of the crystallization module did not affect activity. Presumably because of the compact structure of the HDV ribozyme, no other ribozyme could be used to process its 3′ terminus. Thus, these RNAs have a homogeneous 5′ end and a heterogeneous 3′ end, due to the random addition of a few 3′ nucleotides during *in vitro* transcription. Unlike the wild-type HDV ribozyme constructs that did not crystallize, delta + TR constructs crystallized readily under a variety of solution conditions. We made several constructs differing in the length of base-paired spacers separating the ribozyme core, the TR, and the GAAA tetraloop. Figure 4 summarizes the diffraction properties of six crystal forms of four RNA constructs that were grown to large dimensions. So far, our best-ordered delta + TR RNA crystals diffract X-rays from a rotating anode to 7 Å resolution (Figure 6(b), the RNA sequence is that shown in Figure 5(b)).

General crystallization modules

The striking effect on RNA crystallizability of the module in two markedly different molecules, d56 and the HDV ribozyme, supports the idea that the constituent GAAA tetraloop and the TR make intermolecular contacts in the crystals. Further supporting this notion, the identity of the TR changed the crystallization properties of RNA targets. Otherwise identical d56 + TR constructs containing either the TR sequence from the *Tetrahymena* group I intron P6 region, or the TR from ai5γ yielded distinctly different crystals under the same or related conditions. These two TR sequences, both of which are known to be active (Costa & Michel, 1995), differ in the orientation of a single base-pair (G:C *versus* C:G). This is evidence that individual TR sequences interact with GAAA tetraloops in subtly different ways, a possibility suggested by Butcher *et al*. (1997) and Costa & Michel (1997).

The degree of crystalline order of RNA crystals obtained with the use of the tetraloop/TR module varies considerably (Figure 4). The delta + TR crystals, in particular, diffract X-rays only to moderate resolution. This might be improved by making the molecules homogeneous at their 3′ termini, and by screening variant molecules with different HDV ribozyme sequences (for instance, by preparing antigenomic strand ribozyme + TR constructs). Another possibility is to place several modules in one molecule, so that crystallization selects the most favorable interactions. This could be particularly advantageous for the crystallization of very large RNAs. In addition, we are examining other classes of modules.

In preliminary experiments, we have extended the idea of RNA crystallization modules to another tertiary interaction. We have made HDV ribozymes that have their P4 stems replaced with the P4 and A-rich bulge regions of the *Tetrahymena* ribozyme. These two elements were seen to interact in the crystal structure of the P4-P6 domain (Cate *et al*., 1996). Constructs were made with P4 in either orientation, stacked coaxially with the A-rich bulge. Constructs with P4 in one orientation yielded large crystals (data not shown). This result suggests that as new RNA tertiary interactions are characterized, they might be successfully engineered to be crystallization modules.

In summary, we present a new approach to the crystallization of RNA. The introduction of a tetraloop receptor and a GAAA tetraloop into target molecules in such a way that they are available to interact intermolecularly greatly enhances the likelihood of obtaining crystals large enough for diffraction measurements of molecules that otherwise crystallize poorly or not at all. The sequences that interact constitute a self-contained module that appears to be effective in making both extended helical as well as globular RNA molecules crystallizable. Because of the stability of isolated structural elements, RNA is better suited to the development and application of crystal contact design than proteins. Use of the crystallization module is fundamentally different from the traditional approaches of sequence and helix-length variation that have been employed for nucleic acid crystallization. To our knowledge, the results presented here are the first successful demonstration that RNA can be engineered, in advance of any crystallization trials, to make crystal contacts that do not involve canonical helix stacking. The modules promise to be a valuable addition to the crystallization techniques available to RNA crystallographers.

Materials and Methods

Plasmids

Plasmid pJDI3′-673, which encodes ExD123, has been described by Jarrell *et al*. (1988), and was kindly provided by Dr Anna Marie Pyle (Columbia University). All other plasmids are pUC19 derivatives constructed to encode different RNA molecules. The coding regions, preceded by an *Eco*RI site and the phage T7 RNA polymerase promoter, and followed by a *Bam*HI site, were all constructed with the polymerase chain reaction from

overlapping synthetic DNA oligonucleotides. The identity of the inserts was confirmed by DNA sequencing. pD5F encodes a primary transcript that consists of a 34 nucleotide domain 5 corresponding to residues 1 through 34 inclusive in Figure 1(b), preceded by a hammerhead ribozyme and followed by an HDV ribozyme. pD56B encodes the wild-type d5/d6 molecule of the sequence shown in Figure 1(b). The T7 promoter is placed such that the first transcribed nucleotide is position 1, and nucleotide 73 is the last encoded by the linearized template. Transcription from pD56Q produces the molecule shown in Figure 1(c), starting on residue 1, with an HDV ribozyme following position 63. The primary transcript encoded by pCM9 consists of the d5/TR/d6 construct shown in Figure 1(d), preceded by a hammerhead ribozyme and followed by an HDV ribozyme. pD5627 encodes a transcript consisting of a d56 construct preceded by a hammerhead ribozyme and followed by the delta/TR ribozyme of the sequence shown in Figure 5(b).

RNA preparation

Plasmids were amplified in *Escherichia coli* strain JM109. Plasmid DNA was purified with Nucleobond-AX kits (Macherey-Nagel) following the manufacturer's instructions, and used without further purification. Plasmids pD5F, pD56Q, pD5627 and pCM9 were linearized with *Bsa*I; pD56B with *Eco*RV. Transcription conditions were similar to those described by Ferré-D'Amaré & Doudna (1996) with the addition of one unit/ml of inorganic pyrophosphatase from *E. coli* (Sigma, St. Louis, MO). RNA was purified on 8 M urea/polyacrylamide gels, passively eluted into water, and concentrated and desalted by successive concentrations and dilutions with water on an Amicon (Beverly, MA) ultrafiltration cell equipped with 10,000 Da nominal molecular mass cutoff membranes. RNAs were stored in aqueous solution at 4°C where they were stable for many months.

Body-labeled ExD123 was transcribed from plasmid pJDI3′-673 linearized with *Hin*dIII. Transcription reactions contained 0.1 g l^{-1} linearized template, 5 mM ATP, CTP and UTP, 25 mM GTP, 2 μM [α-^{32}P]ATP, 25 mM $MgCl_2$, 2 mM spermidine-HCl, 30 mM Tris-HCl (pH 8.1), 50 mM DTT, 0.01 % (v/v) Triton X-100, 0.1 g l^{-1} T7 RNA polymerase, and were incubated for 1.5 hours at 37°C. The RNA was purified on 8 M urea, TBE, 4% polyacrylamide gels, visualized by autoradiography, and passively eluted into a solution containing 1% (w/v) sodium dodecyl sulfate, 300 mM NaCl, 20 mM Tris-HCl (pH 7.5) and 1 mM EDTA for two hours at room temperature. It was then extracted three times with neutralized phenol, twice with chloroform, ethanol-precipitated, and taken up in 10 mM Tris-HCl (pH 7.5), 0.1 mM EDTA. The RNA was quantified by Čerenkov counting.

Domain 5 and domain 6 activity assays

All reactions contained 6 nM body-labeled ExD123 and 10 μM unlabeled d5, d56, d56c or d5/TR/d6. Reactions carried out under "hydrolysis" conditions contained, in addition, 500 mM KCl, 100 mM $MgCl_2$, 40 mM Hepes-KOH (pH 7.5) and 0.05 mM EDTA, and were incubated at 42°C. Aliquots (10 μl) were removed at the times indicated in Figure 2(a), and mixed with an equal volume of 90% (v/v) formamide, TBE. Samples were analyzed on 8 M urea, TBE, 6% polyacrylamide gels, and visualized by autoradiography on imaging plates. Reactions carried out under "branching" conditions contained, in addition to the RNAs, 500 mM NH_4Cl, 100 mM $MgCl_2$, 40 mM Hepes-KOH (pH 7.0) and 0.05 mM EDTA, and were incubated at 45°C for 30 minutes. Reactions were subsequently analyzed as for the hydrolysis reactions.

Dynamic light-scattering

DLS measurements were performed on a Protein Solutions, Inc. (Charlottesville, VA) dp-801-TC instrument at 25°C. RNA samples were annealed at a concentration of 2 g l^{-1} with 2.5 mM $MgCl_2$ and 10 mM Hepes-KOH (pH 7.0) by heating to 65°C for ten minutes and cooling slowly to room temperature. The RNA concentration was dictated by the sensitivity of the instrument. Samples were filtered through 200 Å pore-size filters (Anotop 10, Whatman, Maidstone, UK) immediately before injection into the optical cell to ensure removal of dust particles. Diffusion coefficients, hydrodynamic radii of gyration and equivalent molecular masses (based on a calibration curve constructed with measurements on globular proteins of known solution mass) were calculated from the autocorrelation function by software provided with the instrument. Fifty individual measurements were performed on each sample to obtain the data presented in Figure 3.

Crystallization

d5/TR/d6 was annealed at a concentration of 5 g l^{-1} (0.22 mM) in the presence of 2.5 mM $MgCl_2$ and 10 mM Hepes-KOH (pH 7.0) by heating to 65°C for ten minutes and cooling to room temperature over the course of two hours. For crystallization by the sitting-drop vapor diffusion technique, 2 to 6 μl of annealed RNA was mixed with half the volume (1 to 3 μl) of a reservoir solution containing 0.7 to 0.75 M $MgSO_4$, 0.25 to 0.3 M ammonium sulfate, 100 mM Hepes-KOH (pH 7.0) and 0.1 mM spermine-HCl. Crystallization setups were incubated at 30°C. Crystals appeared within a few days, and grew to maximum dimensions of 1.0 mm × 0.5 mm × 0.5 mm over the course of two to three weeks. For cryocrystallography, crystals were first equilibrated in a solution containing 0.75 M $MgSO_4$, 0.3 M ammonium sulfate, 25 mM Hepes-KOH (pH 7.0), 5 mM spermine-HCl and 20% (v/v) ethylene glycol, then mounted in nylon loops, and flash-cooled by plunging into supercooled propane maintained at liquid nitrogen temperature.

Delta/TR RNA was annealed at a concentration of 5 g l^{-1} (0.18 mM) in the presence of 20 mM Tris-HCl (pH 8.1), 5 mM $MgCl_2$, by heating to 70°C for 20 minutes and cooling to room temperature over two hours. For crystallization by the hanging-drop vapor diffusion method, 2 to 4 μl of annealed RNA was mixed with an equal volume of reservoir solution consisting of 3 to 5% (w/v) polyethylene glycol (average molecular mass 2000 Da), 0.3 to 1.0 mM spermine-HCl, 50 mM Tris-HCl (pH 8.1), 200 mM KCl, 10 mM $MgCl_2$. Crystallization was carried out at room temperature. Crystals appeared within a day, and grew over the course of a week to maximum lengths of over 1 mm. Crystals were stabilized in 10% (v/v) 2-propanol, 20% (v/v) glycerol, 15 mM $MgCl_2$, 5% (w/v) polyethylene glycol 2000, 150 mM KCl, 1.5 mM spermine-HCl, 50 mM Tris-HCl (pH 8.1) previous to flash-cooling for X-ray diffraction experiments.

Acknowledgments

We are grateful to Kaifeng Zhou for help with plasmid construction and sequencing; to T. Steitz for the use of his DLS instrument; to the staff of the Yale Center for Structural Biology for crystallographic and computational support; to the staff of CHESS for help with data collection; and to R. Batey, D. Battle, E. Doherty, J. Kieft, A. Luptak and D. Wilson for their comments on a draft of the manuscript. A.R.F. is a Fellow of the Jane Coffin Childs Memorial Fund for Medical Research. J.A.D. is a Lucille P. Markey Scholar, a Donaghue Young Investigator, a Searle Scholar, a Beckman Young Investigator, a David and Lucile Packard Foundation Fellow, and an Assistant Investigator of the Howard Hughes Medical Institute. This work was supported in part by grants from the Jane Coffin Childs Memorial Fund for Medical Research, the Lucille Markey Charitable Trust, and the Packard Foundation.

References

Abramovitz, D. L., Friedman, R. A. & Pyle, A. M. (1996). Catalytic role of 2′-hydroxyl groups within a group II intron active site. *Science,* **271**, 1410–1413.

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

Anderson, A. C., Earp, B. E. & Frederick, C. A. (1996). Sequence variation as a strategy for crystallizing RNA motifs. *J. Mol. Biol.* **259**, 696–703.

Been, M. D. (1994). Cis- and trans-acting ribozymes from a human pathogen, hepatitis delta virus. *Trends Biochem. Sci.* **19**, 251–256.

Been, M. D., Perrotta, A. T. & Rosenstein, S. P. (1992). Secondary structure of the self-cleaving RNA of hepatitis delta virus: applications to catalytic RNA design. *Biochemistry,* **31**, 11843–11852.

Berman, H. M., Gelbin, A. & Westbrook, J. (1996). Nucleic acid crystallography: a view from the nucleic acid database. *Prog. Biophys. Mol. Biol.* **66**, 255–288.

Butcher, S. E., Dieckmann, T. & Feigon, J. (1997). Solution structure of a GAAA tetraloop receptor RNA. *EMBO J.* **16**, 7490–7499.

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.

Chanfreau, G. & Jacquier, A. (1994). Catalytic site components common to both splicing steps of a group II intron. *Science,* **266**, 1383–1387.

Chin, K. & Pyle, A. M. (1995). Branch-point attack in group II introns is a highly reversible transesterification, providing a potential proof-reading mechanism for 5′-splice site selection. *RNA,* **1**, 391–406.

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

Correll, C. C., Freeborn, B., Moore, P. B. & Steitz, T. A. (1997). Metals, motifs, and recognition in the crystal structure of a 5S rRNA domain. *Cell,* **91**, 705–712.

Costa, M. & Michel, F. (1995). Frequent use of the same tertiary motif by self-folding RNAs. *EMBO J.* **14**, 1276–1285.

Costa, M. & Michel, F. (1997). Rules for RNA recognition of GNRA tetraloops deduced by in vitro selection: comparison with in vivo evolution. *EMBO J.* **16**, 3289–3302.

D'Arcy, A. (1994). Crystallizing proteins–a rational approach? *Acta Crystallog. sect. D,* **50**, 469–471.

Daniels, D. L., Michels, W. J. & Pyle, A. M. (1996). Two competing pathways for self-splicing by group II introns: a quantitative analysis of *in vitro* reaction rates and products. *J. Mol. Biol.* **256**, 31–49.

Dib-Hajj, S. D., Boulanger, S. C., Hebbar, S. K., Peebles, C. L., Franzen, J. S. & Perlman, P. S. (1993). Domain 5 interacts with domain 6 and influences the second transesterification reaction of group II intron self-splicing. *Nucl. Acids Res.* **21**, 1797–1804.

Doudna, J. A. & Cech, T. R. (1995). Self-assembly of a group I intron active site from its component tertiary structural domains. *RNA,* **1**, 36–45.

Doudna, J., Grosshans, C., Gooding, A. & Kundrot, C. E. (1993). Crystallization of ribozymes and small RNA motifs by a sparse matrix approach. *Proc. Natl Acad. Sci. USA,* **90**, 7829–7833.

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. (1994). Use of dynamic light scattering to assess crystallizability of macromolecules and macromolecular assemblies. *Structure,* **2**, 357–359, 567.

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.

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.

Holbrook, S. R., Cheong, C., Tinoco, I., Jr & Kim, S.-H. (1991). Crystal structure of an RNA double helix incorporating a track of non-Watson-Crick base pairs. *Nature,* **353**, 579–581.

Jarrell, K. A., Dietrich, R. C. & Perlman, P. S. (1988). Group II intron domain 5 facilitates a trans-splicing reaction. *Mol. Cell. Biol.* **8**, 2361–2366.

Jestin, J.-L., Deme, E. & Jacquier, A. (1997). Identification of structural elements critical for inter-domain interactions in a group II self-splicing intron. *EMBO J.* **16**, 2945–2954.

Konforti, B. B., Abramovitz, D. L., Duarte, C. M., Karpeisky, A., Beigelman, L. & Pyle, A. M. (1998). Ribozyme catalysis from the major groove of group II intron domain 5. *Mol. Cell,* **1**, 433–441.

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

McPherson, A. (1982). *Preparation and Analysis of Protein Crystals,* Wiley, New York.

Michel, F. & Ferat, J. (1995). Structure and activities of group II introns. *Annu. Rev. Biochem.* **64**, 435–461.

Michel, F., Umesono, K. & Ozeki, H. (1989). Comparative and functional anatomy of group II catalytic introns–a review. *Gene,* **82**, 5–30.

Michels, W. J. & Pyle, A. M. (1995). Conversion of a group II intron into a new multiple-turnover ribozyme that selectively cleaves oligonucleotides:

elucidation of reaction mechanism and structure/function relationships. *Biochemistry,* **34**, 2965–2977.
Peebles, C. L., Zhang, M., Perlman, P. S. & Franzen, J. S. (1995). Catalytically critical nucleotides in domain 5 of a group II intron. *Proc. Natl Acad. Sci. USA,* **92**, 4422–4426.
Pley, H. W., Lindes, D. S., DeLuca-Flaherty, C. & McKay, D. B. (1993). Crystals of a hammerhead ribozyme. *J. Biol. Chem.* **268**, 19656–19658.
Pley, H. W., Flaherty, K. M. & McKay, D. B. (1994a). Three-dimensional structure of a hammerhead ribozyme. *Nature,* **372**, 68–74.
Pley, H. W., Flaherty, K. M. & McKay, D. B. (1994b). Model for an RNA tertiary interaction from the structure of an intermolecular complex between a GAAA tetraloop and an RNA helix. *Nature,* **372**, 111–113.
Podar, M., Perlman, P. S. & Padgett, R. A. (1995). Stereochemical selectivity of group II intron splicing, reverse splicing, and hydrolysis reactions. *Mol. Cell. Biol.* **15**, 4466–4478.
Portmann, S., Grimm, S., Workman, C., Usman, N. & Egli, M. (1996). Crystal structures of an *A*-form duplex with single-adenosine bulges and a conformational basis for site-specific RNA self-cleavage. *Chem. Biol.* **3**, 173–184.
Pyle, A. M. & Green, J. B. (1994). Building a kinetic framework for group II intron ribozyme activity: quantitation of interdomain binding and reaction rate. *Biochemistry,* **33**, 2716–2725.
Scott, W. G., Finch, J. T., Grenfell, R., Fogg, J., Smith, T., Gait, M. J. & Klug, A. (1995a). Rapid crystallization of chemically synthesized hammerhead RNA's using a double screening procedure. *J. Mol. Biol.* **250**, 327–332.
Scott, W. G., Finch, J. T. & Klug, A. (1995b). The crystal structure of an all-RNA hammerhead ribozyme: a proposed mechanism for RNA catalytic cleavage. *Cell,* **81**, 991–1002.
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.
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.
Varani, G., Aboul-ela, F. & Allain, F. H.-T. (1996). NMR investigation of RNA structure. *Prog. Nucl. Magn. Reson. Spec.* **29**, 51–127.

Edited by D. E. Draper

(Received 23 January 1998; received in revised form 10 March 1998; accepted 11 March 1998)