Structural Basis of the Enhanced Stability of a Mutant Ribozyme Domain and a Detailed View of RNA–Solvent Interactions

Kara Juneau,* Elaine Podell,*‡ Daniel J. Harrington,*† and Thomas R. Cech*‡
*Department of Chemistry and Biochemistry
†Howard Hughes Medical Institute
University of Colorado
Boulder, Colorado 80309

Summary

Background: The structure of P4-P6, a 160 nucleotide domain of the self-splicing Tetrahymena thermophila intron, was solved previously. Mutants of the P4-P6 RNA that form a more stable tertiary structure in solution were recently isolated by successive rounds of in vitro selection and amplification.

Results: We show that a single-site mutant (ΔC209) possessing greater tertiary stability than wild-type P4-P6 also crystallizes much more rapidly and under a wider variety of conditions. The crystal structure provides a satisfying explanation for the increased stability of the mutant; the deletion of C209 allows the adjacent bulged adenine to enter the P4 helix and form an A-G base pair, presumably attenuating the conformational flexibility of the helix. The structure of another mutant (ΔA210) was also solved and supports this interpretation. The crystals of ΔC209 diffract to a higher resolution limit than those of wild-type RNA (2.25 Å versus 2.8 Å), allowing assignment of innersphere and outersphere coordination contacts for 27 magnesium ions. Structural analysis reveals an intricate solvent scaffold with a preponderance of ordered water molecules on the inside rather than the surface of the folded RNA domain.

Conclusions: In vitro evolution facilitated the identification of a highly stable, structurally homogeneous mutant RNA that was readily crystallizable. Analysis of the structure suggests that improving RNA secondary structure can stabilize tertiary structure and perhaps promote crystallization. In addition, the higher resolution model provides new details of metal ion–RNA interactions and identifies a core of ordered water molecules that may be integral to RNA tertiary structure formation.

Introduction

In the case of RNA, it has been possible to identify variant molecules with increased stability using successive rounds of in vitro selection and amplification [1, 2]. In vitro selection had been previously used to identify RNA structures that bound small molecules with high specificity, bound to the active site of a protein, or catalyzed novel biochemical reactions (see, for example, [3–5]). By creating a library of mutant P4-P6 RNA sequences, displaying them by native gel electrophoresis, isolating the minor population of extremely compact molecules, and amplifying that subpopulation by RT-PCR, mutant forms of the P4-P6 RNA with increased tertiary structure stability in solution were obtained [1]. The deletion of a single nucleotide, C209, was the mutation that arose most frequently during the in vitro selection. We now report that this mutant RNA crystallizes much more readily and provides crystals that diffract to a higher resolution than the wild-type RNA.

The structure of the wild-type P4-P6 RNA, solved in 1996, provided information about tertiary interactions that allow large ribozymes to fold into compact, globular structures [6]. Now, with higher resolution data for a P4-P6 related molecule, more information is available for the analysis of RNA–solvent interactions. It has long been known that divalent cations bind to specific sites and stabilize the folding of transfer RNA [7, 8] and ribozymes (reviewed by [9]). In addition, divalent ions are required for active site catalysis by many ribozymes, including the Tetrahymena group I intron (reviewed by [10]).

In contrast to the extensive data available regarding the magnesium dependence of RNA folding and activity, little is known about the role that water molecules play in stabilizing RNA tertiary structures. Water molecules have been located in the crystal structures of tRNA [8, 11] and a 5S rRNA fragment [12], where they appear to stabilize noncanonical base pairing interactions. Additionally, several extensive analyses of high-resolution RNA duplexes [13–15] as well as molecular modeling of solvent molecules around an RNA duplex [16] have identified general patterns for the hydration of RNA base pairs. These hydration sites may provide an extended hydrogen bonding network that could be important for the tertiary structural integrity of RNA or RNA–protein complexes. We show here that water molecules surrounding ΔC209 bind to the bases in a well ordered pattern and appear to facilitate important tertiary interactions. Thus, in addition to many examples suggesting that water molecules facilitate nucleic acid–protein interactions (for example, [17–21]), water may facilitate the RNA–RNA interactions involved in tertiary folding.

Results and Discussion

Stability of ΔC209 RNA

Nondenaturing gel electrophoresis provides a method for assessing both the stability and the conformational homogeneity of mutant P4-P6 RNAs [22–24]. The ΔC209 mutant, identified by in vitro selection and amplification [1], folded into a compact structure with high electrophoretic mobility even at 0.7 mM MgCl₂, a destabilizing condition that was insufficient to fold the wild-type RNA (Figure 1, top right). The conformational homogeneity of ΔC209 was indistinguishable from wild-type P4-P6, as both formed discrete, single bands during gel electrophoresis. We measured the energetics of the secondary-to-tertiary structure folding transition and found ΔC209...
to be 1.1 kcal/mol more stable than the wild-type P4-P6 RNA. Folding of both the mutant and the wild-type RNA showed similar cooperativity, consistent with the uptake of a minimum of four magnesium ions during the transition from secondary to tertiary structure (Figure 1) [24].

Crystallization of ΔC209 RNA

Wild-type P4-P6 crystals have been reported to grow in 1–2 months and, in the absence of microseeding, to grow in less than 7% of the crystallization drops [25]. In contrast, high quality, medium-sized ΔC209 crystals grew in less than 24 hr, and, in conditions optimized for single crystal nucleation, crystals grew in more than 80% of the drops. ΔC209 crystallized in a greater variety of conditions (Table 1). ΔC209 crystallized at every temperature tested (4°C, room temperature, 25°C, 30°C, and 37°C), while the wild-type RNA showed a decided preference for 30°C. Polyamines were not required for the crystallization of ΔC209, and high quality crystals were obtained from several alcohol precipitants as well as with PEG of several molecular weights. Side-by-side comparison of crystallization of ΔC209 and wild-type RNAs prepared with the same reagents and identical sparse matrices confirmed that the mutant crystallized more rapidly and under a wider variety of conditions.

The ΔC209 crystals were of the same space group and had the same unit cell dimensions (within error) as wild-type P4-P6 crystals, but they reproducibly diffracted to higher resolution. The ΔC209 structure was refined against reflections between 30.0 Å and 2.25 Å resolution, such that the R_free = 26.4% (Table 2).

Structures of the ΔC209 and ΔA210 RNAs

Although the global structure of ΔC209 remained relatively unchanged (rmsd = 1.2 Å for the backbone phosphates in the asymmetric unit of ΔC209, compared to wild-type P4-P6) (Figure 2a), analysis of the ΔC209 crystal structure revealed a local rearrangement within the P4 helix that provided a potential explanation for the mutant’s increased stability. Deletion of C209 allows 80% of the drops. Additionally, careful manipulation of magnesium where, on average, the RNA molecules are half folded at 35°C. ΔGf is the free energy of formation (kcal/mol) for folding of a mutant RNA compared to wild type (ΔGf = ΔGf_mature - ΔGf_unfolded), so the negative value indicates greater stability than wild type.

Table 1. A Comparison of Crystallization Proficiency

<table>
<thead>
<tr>
<th>Crystal Condition</th>
<th>ΔC209</th>
<th>ΔA210</th>
<th>WT</th>
</tr>
</thead>
<tbody>
<tr>
<td>MPD</td>
<td>++++</td>
<td>++++</td>
<td>++</td>
</tr>
<tr>
<td>Hexanediol</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Ethanol</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>PEG 400,600</td>
<td>+*</td>
<td>++*</td>
<td>-</td>
</tr>
<tr>
<td>PEG ≥1000</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Li2SO4</td>
<td>+</td>
<td>n.t.</td>
<td>-</td>
</tr>
<tr>
<td>Temperature (°C)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>n.t.</td>
<td>-</td>
</tr>
<tr>
<td>RT</td>
<td>++++</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>30</td>
<td>++++</td>
<td>++++</td>
<td>++</td>
</tr>
<tr>
<td>37</td>
<td>++</td>
<td>n.t.</td>
<td>-</td>
</tr>
<tr>
<td>Spermine</td>
<td>++++</td>
<td>++++</td>
<td>++</td>
</tr>
<tr>
<td>Spermidine</td>
<td>++++</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td>Putrescine</td>
<td>++++</td>
<td>n.t.</td>
<td>+</td>
</tr>
<tr>
<td>No polyamine</td>
<td>++++</td>
<td>n.t.</td>
<td>n.t</td>
</tr>
</tbody>
</table>

- The pluses marked with an asterisk denote the conditions that produced crystals that were subject to X-ray diffraction experiments.
- The highest resolution data obtained for each construct is noted under the construct headings (WT means wild-type P4-P6). MPD, methylpentanediol; PEG, polyethyleneglycol (molecular weights are specified in the table). RT, room temperature, which regularly fluctuated between 19°C and 22°C. The crystallization solutions contained sodium cacodylate at either pH 6.0 or 6.5 or Na HEPES at either pH 7.0 or 7.5, and the precipitant, temperature, and polyamine content were systematically varied.

Figure 1. Relative Stabilities of Mutant and Wild-Type P4-P6 RNAs

Top left, quantitative summary of the results; Mg2+ 1/2 is the concentration of magnesium where, on average, the RNA molecules are half folded at 35°C. ΔGf is the free energy of formation (kcal/mol) for folding of a mutant RNA compared to wild type (ΔGf = ΔGf_mature - ΔGf_unfolded), so the negative value indicates greater stability than wild type.

Table 1. A Comparison of Crystallization Proficiency
Table 2. Data Collection and Refinement Statistics for ΔC209

<table>
<thead>
<tr>
<th>Resolution (Å)</th>
<th>Average I/σ</th>
<th>I/σ &gt; 2 (%)</th>
<th>Completeness (%)</th>
<th>Rsym</th>
</tr>
</thead>
<tbody>
<tr>
<td>30.00-4.84</td>
<td>24.0</td>
<td>98.5</td>
<td>86.3</td>
<td>0.035</td>
</tr>
<tr>
<td>4.84-3.85</td>
<td>30.0</td>
<td>98.8</td>
<td>95.4</td>
<td>0.039</td>
</tr>
<tr>
<td>3.85-3.36</td>
<td>25.0</td>
<td>97.7</td>
<td>99.0</td>
<td>0.048</td>
</tr>
<tr>
<td>3.36-3.05</td>
<td>24.0</td>
<td>95.5</td>
<td>99.7</td>
<td>0.058</td>
</tr>
<tr>
<td>3.05-2.83</td>
<td>18.0</td>
<td>88.0</td>
<td>99.9</td>
<td>0.080</td>
</tr>
<tr>
<td>2.67-2.67</td>
<td>11.0</td>
<td>77.7</td>
<td>99.9</td>
<td>0.128</td>
</tr>
<tr>
<td>2.67-2.53</td>
<td>6.9</td>
<td>67.4</td>
<td>99.8</td>
<td>0.200</td>
</tr>
<tr>
<td>2.53-2.42</td>
<td>4.8</td>
<td>57.5</td>
<td>99.6</td>
<td>0.289</td>
</tr>
<tr>
<td>2.42-2.33</td>
<td>3.3</td>
<td>49.8</td>
<td>99.5</td>
<td>0.416</td>
</tr>
<tr>
<td>2.33-2.25</td>
<td>2.4</td>
<td>42.8</td>
<td>99.3</td>
<td>0.578</td>
</tr>
</tbody>
</table>

All reflections 14.9 77.6 98.0 0.051

Space group P2_12121
Unit cell dimensions (Å) a = 75.3, b = 125.4, c = 145.7
Multiplicity (total ref./outer shell) 4.8/4.6
Number of unique reflections 64,900
Number of atoms 7,063

Refinement
Number of reflections used in refinement 57,650
Number of reflections used for Rfree 3,017
R_{crystal} 24.4%
R_{free} 26.4%
Rmsd for bond lengths (Å) 0.0105
Rmsd for bond angles (°) 1.505

* R = \sum |F_{obs} - |F_{calc}||/\sum |F_{obs}|; R_{crystal} was calculated for all reflections in the refinement working set; R_{free} was calculated for 5% of reflections not included in the refinement. R_{sym} = \sum |\langle I(h) \rangle|/\sum |I(h)|, where \langle I(h) \rangle is the mean intensity of reflections.

We hypothesized that either the elimination of the bulge from the P4 helix or the specific formation of a G-A base pair caused the increased stability of ΔC209. To distinguish between these hypotheses, the ΔA210 mutant was cloned. By deleting A210, the bulge in P4 is removed directly, while maintaining the original G111-

Figure 2. Structural Changes Between ΔC209 and Wild-Type P4-P6
(a) Least squares superposition of the phosphate backbones of wild-type (purple) and ΔC209 (gold) RNA.
(b) Secondary structure diagram of P4-P6, with the sequences of the wild-type and the ΔC209 P4 helices shown explicitly. Nucleotides A210 and C209 are shown in orange with an arrow, labeled ΔC209, specifying the structural transformation from the wild-type P4 helix to the ΔC209 helix, as observed in the crystal structures. With C209 deleted, the normally bulged A210 folds into the P4 helix, forming a noncanonical G-A base pair.
(c) Side view of the top three base pairs of the P4 helix of ΔC209, including G110–G112 and G208–C211. The model is depicted with the carbon and phosphate atoms colored gray; oxygen atoms, red; and nitrogen atoms, blue; the 2F_o-F_i electron density map is shown as a light blue mesh contoured at 1.6σ. The side view of the P4 helix shows how the phosphate backbone is pushed out and the helix widened slightly by the purine–purine base pair, while the base pair itself is twisted and bent in a manner that minimizes helix disruption.
(d) The noncanonical G-A base pair viewed along the helical axis.
C209 base pair. Measurement of the tertiary structure stability of the \( \Delta A210 \) mutant showed that it was indeed more stable than wild-type P4-P6 although not as stable as \( \Delta C209 \) (Figure 1). It also crystallized better than wild-type RNA although, again, to a lesser degree than \( \Delta C209 \) (Table 1). The crystal structure of the \( \Delta A210 \) mutant confirmed that it had an uninterrupted P4 helix and the same overall structure as the \( \Delta C209 \) RNA (data not shown).

Thus, removal of the bulged adenosine, either indirectly (\( \Delta C209 \)) or directly (\( \Delta A210 \)), stabilizes the secondary-to-tertiary transition of P4-P6. While it is expected that removal of a bulged adenosine would stabilize RNA secondary structure by 3.3 kcal/mol [29], the hierarchical model of RNA folding would not predict that this stabilization would be transmitted to tertiary stability. Our observations suggest that improving RNA secondary structure stability improves tertiary stability in this case, and there are indications that this may be a generalizable effect (see also [1] and [30]).

**Catalytic Activity**

The improvement in tertiary stability does not, in this case, translate into an improvement in catalytic activity. C209 is located in the catalytic core of the *T. thermophila* intron and is part of a base-triple interaction (C209-G111-U305) that contributes substantially to joining the two major RNA domains to form the active site [31]. When C209 is deleted from the intron, the self-splicing activity decreases 91-fold [1]. In comparison, \( \Delta A210 \) RNA retains the C209-G111-U305 interaction and still exhibits substantial activity (4- to 13-fold decrease) [32]. Since C209 and A210 are located on the periphery of P4-P6, their absence in either \( \Delta C209 \) or \( \Delta A210 \) does not grossly perturb intradomain interactions in P4-P6. In fact, the structures of wild-type P4-P6, \( \Delta C209 \), and \( \Delta A210 \) are quite similar, which suggests that structural observations made for \( \Delta C209 \) and \( \Delta A210 \) are as biologically relevant as those made for wild-type P4-P6.

**Hypotheses for the Enhanced Crystallization of \( \Delta C209 \) and \( \Delta A210 \)**

The improved crystallization of the mutant RNAs might be explained by their increased tertiary structure stability. Alternatively, in the wild-type P4-P6 crystal lattice, the A210 bases from two separate molecules stack on one another, and the elimination of this crystal contact in \( \Delta C209 \) and \( \Delta A210 \) (Figure 3) provides another potential explanation for the enhanced crystallization. However, the unit cell dimensions and the disposition of the molecules within the unit cell are virtually identical for wild-type P4-P6, \( \Delta C209 \), and \( \Delta A210 \), which makes it difficult to understand how loss of this crystal contact could be beneficial to crystallization. Also, the order of stability (\( \Delta C209 > \Delta A210 > \) wild-type) agrees with the order of crystallizability. Seven additional P4-P6 variants (six containing sequence changes near the tetraloop-receptor, and one with a covalent modification with pyrene) also show the same correlation of stability and crystallizability (E.P. et al., unpublished data). Thus, while removal of the crystal contact might (contrary to the simple expectation) contribute in some way to improved crystallization, we propose that increased tertiary structure stability makes a major contribution. This is unexpected, because the MgCl\(_2\) concentration used for crystallization (16.7 mM) is so much higher than the 2 mM required for complete folding of the wild-type RNA (Figure 1). However, even a “completely folded” RNA is a highly dynamic structure [33, 34], and rigidification of an interrupted helix like P4 may solidify the structure more globally. The finding that selection for RNA tertiary stability identified an RNA with a greatly improved ability to crystallize suggests that more stable RNAs will tend to crystallize better, much as proteins and RNA-protein complexes from extreme thermophiles are thought to be good candidates for crystallization trials.

**Analysis of Magnesium Ions Directly Coordinated to \( \Delta C209 \) RNA**

At 2.25 Å resolution, it was possible to model the hydration shells of 27 magnesium ions within the two \( \Delta C209 \) molecules (referred to as A and B) in the asymmetric unit; 13 of those were innersphere (directly) coordinated to the RNA. If one counts only the magnesium ions found in the same location in both the A and B molecules of the asymmetric unit, the \( \Delta C209 \) RNA provides specific, well occupied binding sites for six innersphere magnesium ions plus six hexahydrated magnesium ions. The criteria for identification of octahedrally coordinated magnesium ions included the intensity and shape of the \( F_\alpha-F_\beta \) electron density and the geometry of the RNA around the magnesium binding pocket (see Experimental Procedures for details). Magnesium ion complexes described here had electron densities with overall octahedral shapes and simulated annealing (SA) omit map densities greater than 2.0 \( \sigma \) (the average \( \sigma \) was 3.2 ± 0.8). The innersphere coordination distances were required to be within 1.9–2.5 Å and the outersphere coordination contacts were required to be 2.5–3.3 Å, as measured from the oxygen atoms of the coordinated water molecules. Additionally, innersphere-coordinated magnesium ions were confirmed by analyzing difference
The electron density surrounding the innersphere-coordinated magnesium ions was extremely detailed, such that the waters within the hydration shells could be confidently assigned to specific octahedral positions (Figure 5). The explicit positioning of water around the magnesium ions allowed specific assignment of outersphere coordination contacts, and, in this case, the most frequent ligands were provided by bases (58%), followed by phosphate oxygens (38%) and 2' oxygens (4%) (Table 3). These hydrogen bonding distances ranged from 2.5 to 3.3 Å, with an average of 2.9 ± 0.1 Å. (Detailed information about all metal ions is given in Table S1 of Supplementary Materials.)

The majority of the directly coordinated magnesium ions seen in both P4-P6 and ΔC209 were located in the A-rich bulge and the three-helix junction (Figure 4). The innersphere-coordinated magnesium complexes of ΔC209 were found to be deeply buried in the core of the RNA with little solvent-exposed surface (Table S1), and many were found in binding pockets having large negative electrostatic potential, as predicted by Chin et al. [36]. The magnesium ions coordinated to the RNA seem to be stabilizing the structures formed by P5a and P5c, as reflected by the lower atomic displacement factors for the residues that were in contact with magnesium ions (data not shown).

Since both innersphere and outersphere coordination sites were explicitly assigned for all identified innersphere-coordinated ions in ΔC209, it was confirmed that these ions do not bridge between the two quasihelical halves of the molecule (Figure 4, left half includes P5abc, and right half includes P4, P5, and P6). Instead, the innersphere-coordinated magnesium ions help fold important local structural motifs, which in turn form essential tertiary contacts. Interestingly, only one magnesium ion, hexahydrated m14/15, appeared to be involved in linking the two halves of P4-P6 together; m14/15 made contacts to P5a on one side and to P4 and P5 on the other (Figures 4 and 5c). (Note: in the naming convention for the ions, m denotes a magnesium, and the numbers correspond to specific ions; two numbers are given if the ion is found in both molecules of the asymmetric unit.)

The two magnesium ions located in the A-rich bulge each coordinate three nonbridging phosphate oxygens and facilitate the formation of the bulged corkscrew turn [6]. One of these (m1/2) (Figures 4 and 5a) innersphere coordinates to the nonbridging phosphate oxygens of A183, A184, and A186, as also seen in the wild-type P4-P6 structure [6]; and now, with a more detailed electron density map available, outersphere coordination contacts from m1/2 to the N7 of A184 and the N7 of A186 have been identified (Figure 5a). The other magnesium in the bulge (Figure 4, see m3/4) coordinates the nonbridging phosphate oxygens of A184, A186, and A187 and is now seen to outersphere coordinate O2P of A183 and O1P of A188. These two magnesium ions work together to bend the sugar-phosphate backbone around themselves into a corkscrew turn (Figure 5a), which pushes the adenine bases out so they are available for making tertiary interactions with P4 [6]. A third magnesium, m11/12 (Figures 4 and 5b), uses both innersphere
Figure 5. Three Magnesium Ions Seen in ΔC209

The magnesium ions are shown as yellow spheres, the coordinating waters as smaller red spheres. The innersphere coordination contacts are depicted as thick black lines, the outersphere coordination contacts as thin orange lines. The electron density maps are shown in blue and are contoured at either 1.6 σ (a and b) or 1.4 σ (c). Atoms are colored as in Figure 2, except A183 is completely gray for simplicity.

(a) m2 in the A-rich bulge, stereo representation. The sugar-phosphate backbone is bent around the magnesium so that it can make three innersphere coordination contacts to the nonbridging phosphate oxygens of A183, A184, and A186, while making two outersphere contacts to the N7 of bases A184 and A186.

(b) m12, directly coordinated to the O6 group of G188, occupies a binding pocket between the A-rich bulge and the three-helix junction. This magnesium makes five additional outersphere contacts to O1P of U182, O1P of C165, O2P of U168, O1P of U168, and O1P of A183 (2.7 Å, not explicitly shown).

(c) m15, a hexahydrated magnesium, bridges the two sides of ΔC209. Ion m15 makes water-mediated contacts to the nonbridging phosphate oxygens of G112, A113, C128, U202, and U203 as well as to the O2 of C127. This magnesium coordinates P4, P5, and P5a and may help stabilize the bend at J5/5α.

Table 3. Summary of Coordination Contacts

<table>
<thead>
<tr>
<th></th>
<th>P–O</th>
<th>O6-G</th>
<th>N7-G</th>
<th>N7-A</th>
<th>N6-A</th>
<th>O4-U</th>
<th>O2-U</th>
<th>2′OH</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inner</td>
<td>9</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>Outer</td>
<td>27</td>
<td>14</td>
<td>7</td>
<td>7</td>
<td>4</td>
<td>7</td>
<td>2</td>
<td>3</td>
<td>71</td>
</tr>
</tbody>
</table>

Inner, innersphere ligands; Outer, outersphere (water-mediated) contacts. P–O denotes coordination contacts from a magnesium ion or a magnesium-coordinated water molecule to a phosphate oxygen; O6-G and N7-G to the O6 or N7 of guanine, N7-A and N6-A to the N7 or N6 of adenine, O4-U and O2-U to the O4 or O2 of uracil, 2′OH to the 2′ hydroxyl of any nucleotide. Most of the metal ions occur in the same position in both molecules of the asymmetric unit; in such cases their ligands are listed only once in the table. The outersphere contacts listed had hydrogen bonding distances ≤3.3 Å and were not required to meet the criterion of being within 25° of the plane of a base, as was the case in Figure 6a.
ersphere coordination contacts could be assigned for the hexahydrated magnesium ions placed in ΔC209, new information regarding magnesium–RNA interactions is now available even for previously seen magnesium ions such as m14/15.

The extensive outersphere coordination contacts visible for m14/15 in the ΔC209 structure reveal its role as a structurally important hexahydrated magnesium that may stabilize the tight bend made by J5/5a. Ion m14/15 makes clear outersphere contacts to the phosphoryl oxygen atoms of G112, A113, C128, U202, and C203 (Figures 4 and 5c), presumably neutralizing the negative charge of the phosphate backbone. Magnesium ions m16/17 and m18/19 may also stabilize the J5/5a turn; they are bound in the major grooves of P5 and P5a, respectively, and could provide additional charge neutralization. In addition, m16/17 may aid the compaction of J4/5, a functionally important region with a large number of noncanonical base pairs [6], into a helical fold. Another magnesium with the potential for providing structural stability is m24/25; it is located near the A-rich bulge centered on the major groove face of two base pairs contorted by large propeller twists (A136–U182 and C137–G181). Ion m24/25 may bind and hold the contorted base pairs such that bases A136 and C137 can more readily stack on base pair U135–A187 and bases G181 and U182 can stack on base pair C138–G180. In contrast, ions m20/21 and m22/23 appear to be less involved with tertiary structure formation. They are both located in the major grooves of A-form helices; m20/21 is centered between base pairs U144–A159 and C145–G158 in P5b, and m22/23 is centered between base pairs A230–U244 and A231–U243 in P6b.

In summary, fully hydrated magnesium ions mainly bind ΔC209 in the major groove of an A-form helix or in positions that resemble a major groove, but they are also found in areas with complex RNA geometry.

Magnesium Substitution by Manganese or Cobalt Hexamine

Cobalt (III) hexammine shares many RNA binding characteristics with hexahydrated magnesium, including size, a preference for major grooves, and octahedral coordination chemistry. Cobalt hexammine neither helped nor hindered the growth of high quality crystals of ΔC209, unlike the wild-type RNA where it aided crystallization [25]. Thus, the interchangeability of hydrated magnesium and cobalt hexammine could be investigated with ΔC209. As described in more detail in Supplementary Materials, our data indicate that magnesium and cobalt hexammine do not bind RNA identically and suggest that there are specific cobalt hexammine binding sites in the RNA that are not usually occupied by magnesium.

In crystals where minimal magnesium was present, manganese substituted for hexahydrated magnesium ions at only 4 sites (m24/25 and m26/27) out of 14. In contrast, under the same conditions, manganese readily substituted for all 13 magnesium ions that were innersphere coordinated to the RNA and gave strong Fourier difference map intensities (average: 5.3 ± 0.9 σ). The observation that manganese ions substituted well only for innersphere-coordinated magnesium ions could be due to a preference of Mn2+ for innersphere coordination [9, 38–40].

Water Molecules in the Interior of the RNA

The water molecules placed in the ΔC209 structure had positive Fo-Fc densities greater than 3.5 σ, refined 2Fo-Fc densities greater than 1.2 σ, and hydrogen bonding distances less than 3.3 Å. Although the buffer used to crystallize ΔC209 RNA contained sodium, no sodium ions were modeled, because sodium ions and water molecules are isoelectronic, both have pliable and unpredictable coordination geometries, and at 2.25 Å resolution, one cannot unambiguously differentiate between the two. In theory, water binding sites in crystal structures of nucleic acids could be occupied by waters, sodium ions, or, in a hybrid-solvent model [41, 42], fractionally occupied by both; thus, electron density that could correspond to either water or sodium was conservatively modeled as water.

The majority of ordered water molecules bound to the RNA were found in the interior of the molecule (compare the solvent-exposed surface in Figure 6a to the interior of the RNA in Figure 6b). Over a quarter of the water molecules placed (28%) were found in both molecules in the asymmetric unit. Many water molecules hydrogen bonded to the bases at preferred positions. The pattern
Figure 7. Hydration Pattern Around the Major and Minor Groove Sides of RNA Bases in P4-P6

Using a least squares fit, water molecules and their contacted bases are superimposed onto either a standard U-A base pair or a standard C-G base pair. Single water molecules hydrogen bonded to base pairs are colored orange, single water molecules hydrogen bonded to single bases are yellow, water molecules coordinated to magnesium ions and hydrogen bonded to base pairs are dark purple, and magnesium-coordinated water molecules hydrogen bonded to single bases are lavender. To the right, side views of the same data with the pyrimidines facing toward the viewer so that the vertical distribution of water molecules can also be visualized.

Figure 8. Solvent Molecules Hydrating and Contributing to the Structure of \( \Delta C209 \) RNA

(a) A structural water molecule hydrogen bonded to both the tetraloop and the tetraloop receptor. The water molecule is shown as a small red sphere, hydrogen bonds are colored orange, and the \( 2F_o - F_c \) electron density map is shown in light blue, contoured at 1.4 \( \sigma \).

(b) The solvent network bridging P4, P5c, and the A-rich bulge. The magnesium ions are shown as large yellow spheres, inspheres on the base and within 25\( \text{Å} \) from the plane of the base coordination contacts as thick black lines, and water-mediated hydrogen bonding interactions within 3.5\( \text{Å} \) as thin orange lines. The \( 2F_o - F_c \) electron density map shown in light blue is contoured at 1.4 \( \sigma \).

of water binding sites around the RNA bases may allow bases to extend their hydrogen bonding to stabilize RNA base pair formation as well as facilitate protein recognition and RNA tertiary structure formation.

In addition to forming a regular hydration shell around the RNA bases, water molecules also formed internucleotide hydrogen bonds that may stabilize \( \Delta C209 \) folding. For example, a solvent molecule located between the tetraloop receptor (nucleotides C223 and G251) and the GAAA tetraloop (A153) (Figure 8a) may contribute to their interaction, which is essential for the stable folding of the entire P4-P6 RNA [22]. This solvent binding site was observed in the same position in both molecules of the asymmetric unit. Notably, this is a site with strong electronegative potential [36] and thus is a good
candidate for a hybrid-solvent [42] or sodium ion site. Another solvent molecule seen in both molecules in the asymmetric unit is within hydrogen bonding distance to C170, A184, and G212, linking P5c, the A-rich bulge, and P4, respectively (Figure 8b, top left red sphere). A structural water molecule located at this site could be acting along with the surrounding solvent network to facilitate the formation of important tertiary interactions within the core of C209. Approximately 20% (26 out of 126) of the ordered water molecules placed in the structure participate in internucleotide hydrogen bonding that is more complex than simple base hydration or secondary structure stabilization. Structural waters were defined as waters that were within hydrogen bonding distance to two or more nucleotides simultaneously, so they could potentially facilitate the folding of RNA. Such bridging water molecules have also been seen in tRNA [8]. Another 17% of the water molecules form bridges between the RNA and additional solvent molecules, forming a solvent network that may also help to stabilize RNA folds.

In summary, waters may work independently (Figure 8a) or in concert with magnesium ions (Figure 8b, upper right-hand water) or even as a part of a large network of solvent molecules (Figure 8b, center) to facilitate RNA structure formation. The ordered solvent is found within the interior of the RNA (Figures 6a, 6b, and 8b) and can be considered to be an integral component of the RNA tertiary structure.

Biological Implications

Determining the structural interactions necessary to fold large RNAs at the atomic level is essential for improving the functional understanding of ribozymes and protein–RNA complexes. Crystallographic analysis of large RNAs is dependent on obtaining high quality crystals, which is hampered by the homogeneous polyanionic surface of these biomolecules [43] and their conformational dynamics [33, 34]. Additionally, crystal growth is dependent on the existence of a structurally homogeneous population of molecules, which is particularly challenging for RNA [44], and on their ability to form suitable crystal contacts. As shown here, in vitro selection for stable RNAs can facilitate the rational identification of candidates for crystallization trials. In general, RNA molecules can be randomly mutated and then selected based upon their ability to fold into a uniform, stable structure. Some of the selected mutant RNAs may prove to be better at crystallization due to their adapted characteristics of being more rigid or folding more homogeneously or because they fortuitously possess better crystal contacts.

Structural analysis of in vitro–selected mutant RNAs also contributes to a more complete understanding of RNA structure and folding. The ways in which the structure of a macromolecule responds to a mutation have been well studied in the case of proteins. For example, proteins can adjust to point mutations with backbone shifts or side chain rotation to compensate for structurally destabilizing volume changes caused by specific mutations [45]. In other situations, large conformational adjustments within the protein are not utilized and the mutant protein will assume a structure that closely resembles the wild-type protein [46]. In the case of RNA, much is known on a gross scale about the sorts of mutations that tend to be disruptive to structure, but it is rare that the detailed structures of wild-type and mutant forms of the same RNA can be compared on an atomic level [47, 48]. The ΔC209 RNA structure determined here shows how the potentially destabilizing deletion of C209 is compensated for by a local rearrangement within the RNA; this rearrangement reestablishes some of the lost hydrogen bonds, while decreasing the overall flexibility of the P4 region, stabilizing the RNA tertiary fold. In comparison, the ΔA210 mutation is purely subtractive, resulting in no major rearrangement of the remainder of the RNA. These are two examples in a long list of ways that mutations could affect RNA structure. It is encouraging that such mutational consequences can now be analyzed at atomic resolution.

The relationship between RNA and the solvent molecules that surround it needs to be thoroughly characterized as well if one is to fully understand RNA tertiary structure formation. The ΔC209 RNA is a large, 159 nucleotide RNA with 27 octahedrally coordinated magnesium ions, 15 additional magnesium cations, and 129 water molecules placed within the asymmetric unit of the crystal. The innerness-coordinated magnesium ions help fold local RNA motifs, which then form important RNA–RNA tertiary interactions; the hexahydrated magnesium ions also help fold complex RNA motifs and can be found binding the major grooves of A-form helices. Water molecules form a complex ice-like network on the inside of the folded RNA, similar to water molecules seen previously at the interfaces between nucleic acids and bound proteins. Collectively, the RNA–solvent interactions seen in the ΔC209 structure represent a sample of those that may be prevalent in many folded RNAs.

Experimental Procedures

Measurement of RNA Stability

Native gel electrophoresis was carried out as described [1], and ΔG° was calculated as described [24].

Crystalization and Data Collection

Crystalization conditions were identified by sparse matrix screening [49]. Diffraction data were collected from a crystal grown at room temperature by vapor diffusion; the annealed RNA solution (3.5 mg/mL RNA, 10 mM NaCl, 25 mM MgCl₂, and 5 mM HEPES [pH 7.5]) was mixed in a 2:1 ratio with a reservoir solution containing 21% MPD (methylpentanediol), 50 mM sodium cacodylic acid (pH 6.5), and 0.37 mM spermine. The crystal was stabilized in a solution containing 25% MPD, 10% isopropanol, 50 mM sodium cacodylic acid (pH 6.5), 25 mM MgCl₂, and 0.37 mM spermine before flash freezing in liquid propane. Data were collected at the Advanced Light Source beamline 5.0.1 (Lawrence Berkeley National Laboratory) using 1.1 Å synchrotron-generated radiation. Data were reduced using DENZO and SCALEPACK [50].

We tested the integrity of the 2.25 Å cutoff by conducting parallel refinement strategies on three potential resolution cutoffs: 30–2.3, 30–2.25, and 30–2.2 Å. These preliminary refinements included positional refinement, simulated annealing, group B factor refinement, and individual B factor refinement. The R_cystal and R_free values were compared, and electron density maps were analyzed. The ΔC209 RNA and ΔA210 values were quite similar for all three cutoffs: (R_cystal and R_free, respectively) 27% and 33% for 30–2.2 Å, 33% and 32% for 30–2.25 Å, and 32% and 32% for 30–2.3 Å.
for 30–2.3 Å. The maps showed the best quality and least noise when calculated using data from the resolution range 30–2.25 Å.

**Structure Refinement**

The model was refined using CNS 0.5–0.9 [51]. ∆C209 was initially placed in the electron density by rigid body refinement using observed reflections between 8–4 Å and preliminary coordinates derived from the wild-type P4-P6 model (Protein Data Bank entry 1GID) stripped of solvent. The Rcrystal and Rfree values decreased from 53% to 39% and 53% to 38%, respectively, during rigid body refinement. Rounds of positional refinement, simulated annealing, conservative B factor refinement, and manual rebuilding using the program O [52] brought the Rcrystal and Rfree values for the RNA model to 27% and 30% before solvent molecules were placed (see Supplementary Materials). The final model has Rcrystal = 24.4% and Rfree = 26.4%.

**Cobalt Hexammine, Manganese, and Water Binding**

∆C209 RNA was cocrystallized with 0.1–1 mM cobalt hexammine (present in the well, the drop, and the annealing buffer). Alternatively, ∆C209 crystals grown to their maximum size were stabilized and then soaked in a manganese stabilizing buffer. Details are given in Supplementary Materials. Superposition analysis of water molecules around RNA bases is also described in Supplementary Materials.

**Supplementary Material**

Supplementary material including Table S1 is available at http://current-biology.com/suppmat/supmatin.htm.

**Acknowledgments**

We thank Thomas Earnest and Keith Henderson at the Advanced Light Source (Lawrence Berkeley National Laboratory) and Hal Lewis at the National Synchrotron Light Source (Brookhaven National Laboratory) for help with data collection. The Howard Hughes Medical Institute and National Institutes of Health grant GM28039 to T.R.C. supported this work. We thank the W. M. Keck Foundation for generous support of the program in molecular and cellular structure on the Boulder campus.

Received: July 10, 2000
Revised: December 21, 2000
Accepted: January 8, 2001

**References**


Accession Numbers

The atomic coordinates and structure factors for ΔC209 have been deposited in the RCSB Protein Data Bank with accession number 1HR2.