Crystal Structures of Two Plasmid Copy Control Related RNA Duplexes: An 18 Base Pair Duplex at 1.20 Å Resolution and a 19 Base Pair Duplex at 1.55 Å Resolution†

Peter S. Klosterman,‡ Sapan A. Shah,‡,§ and Thomas A. Steitz*,‡,||,⊥

Department of Molecular Biophysics and Biochemistry, Department of Chemistry, and Howard Hughes Medical Institute, Yale University, 266 Whitney Avenue, New Haven, Connecticut 06520-8114

Received June 4, 1999; Revised Manuscript Received August 23, 1999

ABSTRACT: The structures of two RNA duplexes, whose sequences correspond to portions of the ColE1 plasmid copy control RNA I and RNA II, have been determined. Crystals containing the 18mers 5′-CA CCGUUGGUAGCGGUGC-3′ and 5′-CACCGCUACCAACGGUGC-3′ diffract to 1.20 Å resolution while those containing the 19mers 5′-GCACCGUUGGUAGCGGUGC-3′ and 5′-GCACCGCUACCAACG-GUGC-3′ diffract to 1.55 Å resolution. Both duplexes are standard A form, with Watson−Crick base pairing throughout. Use of anisotropic atomic displacement factors in refinement of the 1.20 Å structure dramatically improved refinement statistics, resulting in a final R_{free} of 15.0% and a crystallographic *R*-factor of 11.6%. Perhaps surprisingly, these crystals of the 18 base pair RNA exhibit a 36-fold static disorder, resulting in a structure with a single sugar−phosphate backbone conformation and an averaged base composition at each residue. Since the sugar−phosphate backbone structure is identical in the 36 different nucleotides that are superimposed, there can be no sequence-dependent variation in the structure. The average ribose pucker amplitude is 45.8° for the 18 base pair structure and 46.4° for the 19 base pair structure; these values are respectively 19% and 20% larger than the average pucker amplitude reported from nucleoside crystal structures. A standard RNA water structure, based on analysis of the hydration of these crystal structures and that of the TAR RNA stem [Ippolito, J. A., and Steitz, T. A. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95,* 9819−9824], has been derived, which has allowed us to predict water positions in lower resolution RNA crystal structures. We report a new RNA packing motif, in which three *pro-*S_p phosphate oxygens interact with an ammonium ion.

Recent progress in techniques of RNA synthesis and purification, together with use of synchrotron X-ray sources, has led to an increasing number of high-resolution RNA crystal structures. Examples include the r(C_4G_4) structure, with eight canonical base pairs and data to 1.46 Å (*1*), the 5S rRNA dodecamer, with 12 base pairs, 6 of which are canonical, and data to 1.5 Å (*2*), the HIV-1 TAR RNA stem structure, which contains 12 canonical base pairs plus 3 looped-out bases, with diffraction data to 1.3 Å (*3*), and the 7 base pair $tRNA^{Ala}$ acceptor stem structure, with data to 1.16 Å (*4*). These high-resolution structures provide unambiguous definition of metal ion and water molecule locations; both metal ions and water molecules play central roles in determining RNA structure and function.

The importance of water in nucleic acid structure and function is well accepted; the sequence-independent character of this water structure is evident from recent analyses of high-resolution nucleic acid crystal structures. Using crystal structure data from 14 decamer structures, Schneider and Berman (*5*) have described a local, or sequence-independent, hydration pattern for B-DNA. They found waters contacting each base atom to be well localized, with the exception of those contacting guanine N2 atoms, for which two lower occupancy clusters were reported. On the basis of analysis of the r(C_4G_4) structure, Egli (*1*) described a repeating pattern of RNA hydration in both the major and minor grooves. Auffiger and Westhof recently published a survey of RNA hydration (*6*), using crystal structures present in the Nucleic Acid Database (NDB)[1] (*7*) as of July 1, 1998. They also identified water clusters which are sequence-independent.

The structures of double-stranded RNA are less diverse than those of DNA by several measures. Double-stranded DNA has been crystallized in various forms, while double-stranded RNA crystals contain either the A or A′ forms, which are closely related (*8*). A-tract associated curvature is well established for DNA but has not been observed in RNA (*9*); a review of RNA crystal structures, published in 1981, found little correlation between RNA conformation and sequence (*10*). RNA is less flexible than DNA, since a variety of measurements indicate that double-stranded RNA has a 1.5−2.0-fold larger persistence length than DNA (*9*).

† Research funded by NIH Grant PO1-GM22778.

* To whom correspondence should be addressed. Phone: (203) 432-5619. Fax: (203) 432-3282.

‡ Department of Molecular Biophysics and Biochemistry.

§ Present address: McKinsey and Co., 600 Campus Drive, Floram Park, NJ 07932.

|| Department of Chemistry.

⊥ Howard Hughes Medical Institute.

[1] Abbreviations: CHESS, Cornell High-Energy Synchrotron Source; MAD, multiwavelength anomalous dispersion; CNS, Crystallography and NMR System; NDB, Nucleic Acid Database; ESD, estimated standard deviation.

10.1021/bi9912793 CCC: $18.00
Published on Web 10/21/1999

We report here two high-resolution RNA duplex crystal structures. Crystals of an 18 base pair duplex diffract to 1.20 Å, while those of a 19 base pair duplex diffract to 1.55 Å. The 18mer oligos produce a perfect Watson−Crick duplex with overhanging C residues at both 3′ ends; the 19mer oligos have the same sequences except for an additional 5′-terminal G, giving blunt-ended duplexes. The sequences used are portions of the ColE1 plasmid copy control RNA I and RNA II. These RNA oligomers form a pair of complementary stem−loops, which are stable in solution but have melted to duplex in the crystallization drop. In the stem−loop conformation the RNA oligos bind the plasmid-encoded four-helix bundle protein rop (*11*).

MATERIALS AND METHODS

Synthesis, Crystallization, and Data Collection. The 18mer RNA oligonucleotides rCACCGUUGGUAGCGGUGC and rCACCGCUACCAACGGUGC were synthesized by the W. M. Keck Foundation Biotechnology Resource Laboratory at Yale University, using the phosphoramidite method. The 19mer oligonucleotides rGCACCGUUGGUAGCGGUGC and rGCACCGCUACCAACGGUGC were obtained from the Keck Laboratory and were also synthesized by T7 RNA polymerase transcription (*12*). Oligos were purified using either polyacrylamide gels or ion-exchange chromatography; the latter was carried out in 6 M urea, using a Pharmacia monoQ column heated to 50 °C in order to eliminate secondary structure. Rop protein was purified as described previously (*13*).

Both 18 base pair (COPY18) and 19 base pair (COPY19) crystals were grown at 20 °C by vapor diffusion, using 5 μL drops. Drops containing 1.0 mM each RNA oligo, 1.0 mM rop protein, 1.05 M ammonium sulfate, 5 mM sodium cacodylate (pH 6.5), 1 mM magnesium chloride, and 0.02% sodium azide were equilibrated with a well solution containing 2.1 M ammonium sulfate. To form the hairpin conformation, RNA oligos were annealed separately at 90 °C, quick cooled over ice, and stored separately. Although the best crystals grew with rop protein present, crystals also grew in its absence. The COPY19 crystals are triangular prisms, the best diffracting having dimensions 0.13 × 0.13 × 0.10 mm. The COPY18 crystals are hexagonal rods; the best COPY18 crystals grew from COPY19 crystal seeds, to 0.4 mm long and 0.1 mm in diameter. Both COPY18 and COPY19 crystals were flash frozen in liquid propane using a stabilizing solution containing 1.9 M ammonium sulfate, 20% glycerol, 20% sucrose, and 15% xylitol.

The highest resolution data were collected at the Cornell High-Energy Synchrotron Source (CHESS) at 100 K, using frozen crystals. COPY18 data were measured using a 2K charge-coupled device on beamline A1. These data were combined with data from a different crystal, collected using a Rigaku rotating anode source and an R-axis-IV image plate system. COPY19 data were collected at the CHESS F1 beamline, using the Area Detector Systems Corporation Quantum 4 detector. These data were also combined with data obtained from a different crystal using a rotating anode source. Data were processed using DENZO and SCALEPACK (*14*); form factors were calculated using TRUNCATE from the CCP4 program suite (*15*). Crystallographic data are summarized in Table 1.

Table 1: Native Data Set and Refinement Statistics[a]

crystal	COPY18	COPY19
space group	*P*321	*R*32
a, *b*, *c* (Å)	43.8, 43.8, 29.2	43.7, 43.7, 447.0
unique reflections	18266	22775
test set	909	1144
resolution/last shell (Å)	30−1.20/1.24−1.20	40−1.55/1.61−1.55
average I/σ	31.8/5.2	26.9/2.7
completeness (%)	93/77	91/95
R_{sym} (%)	7.3/27.4	5.2/45.2
redundancy	4.2/2.8	3.6/2.7
R (all data) (%)	11.6/20.2	15.0/26.1
R_{free} (all validation data) (%)	15.0/24.7	21.4/28.8
rms bond length (Å)	0.012	0.010
rms bond angle (deg)	1.93	1.90

[a] $R_{sym} = \sum|I - \langle I\rangle|/\sum I$. $R = \sum|F_o - F_c|/\sum F_o$. Bond and angle deviations refer to the X-PLOR DNA-RNA parameter set.

Structure Determination. The structure of the COPY18 crystals was solved by molecular replacement, using the X-PLOR program (*16*) and following a procedure used in solving the structure of crystals that contain a blunt-ended 17 base pair RNA duplex (*17*). As discussed below, both the COPY18 and 17 base pair RNA crystals exhibit static disorder.

The molecular replacement search model for the COPY18 structure contains 5.5 base pairs of A-form RNA, the contents of one asymmetric unit, with strand sequences 5′-CACCGG-3′ and 5′-CCGGU-3′. The rise per residue was adjusted to exactly fit 11 base pairs along the *c* unit cell edge. The top solution of the direct rotation search was refined as a rigid body and used in the translation search. The top translation solution was then refined as a rigid body, followed by torsion angle refinement, which yielded a crystallographic *R*-factor of 26.1% and an R_{free} of 33.9% (using data from 2.5 to 6 Å). A difference Fourier map, calculated using model phases and data from an RNA derivative containing one brominated residue, contained peaks at the predicted bromine positions in all 22 residues, due to the static disorder.

The structure of the COPY19 crystals was determined by molecular replacement, using bromine sites to position A-form helices. Crystals were grown of bromine-substituted RNA that had either 5-bromo-dC or 5-bromo-U at various positions in the sequence. Data were collected at the Cu Kα wavelength from crystals of three single bromine-substituted derivatives and one derivative containing four bromine-substituted residues. In addition, a three-wavelength multi-wavelength anomalous diffraction (MAD) data set was collected at CHESS beamline F2 from crystals of one of the singly brominated derivatives. Bromine sites were identified using SOLVE (*18*, www.solve.lanl.gov) and validated by examining difference Fourier maps using phases calculated from the bromine derivatives omitting each candidate site in turn with MLPHARE (*15*).

The dimensions of the COPY19 unit cell imply that the asymmetric unit contains one and one-half 19 base pair duplexes, oriented parallel to the *z*-axis. The position of the 9.5 base pair duplex was obtained by translating the COPY18 coordinates down one-half the *c* cell dimension of the COPY18 crystals. The coordinates of the 9.5 base pair duplex so positioned agreed well with derivative bromine positions. The 19 base pair duplex was positioned by least-squares superposition of three predicted bromine locations on deriva-

tive bromine sites, using the O program (*19*). Rigid-body refinement of this model, followed by torsion angle simulated annealing, using CNS (*20*), yielded an *R*-factor of 32.6% and an R_{free} of 35.4% (3−8 Å data). The locations of peaks in derivative difference Fourier maps calculated using model phases agreed well with the model's predicted bromine positions.

Model Refinement. Initial refinement of the COPY18 structure was performed using X-PLOR, followed by refinement with SHELXL (*21*). To accurately model the central base pair, refinement was carried out in space group *P*3. The final X-PLOR refined model contained one copy of the sugar−phosphate backbone, four fractionally occupied bases connected to each ribose C1′ carbon, and 108 waters. At this stage of refinement the *R*-factor was 23.7% and the R_{free} was 26.3% (1.20−8 Å data).

To reduce the number of parameters, SHELXL refinement was carried out using a model with one C and one G residue connected to each C1′ carbon. The refinement statistics of this model were similar to those of the model with four bases per residue. Use of anisotropic displacement factors resulted in the R_{free} decreasing by 6.5% and the *R*-factor by 6.4%. Additional water molecules were selected using the SHELXL automated water picking facility, followed by manual addition of water molecules from difference Fourier maps, using the CONTACTS program (*15*) and O to verify potential hydrogen bonds. Hydrogen atoms were added to the structure using SHELXL; this decreased the R_{free} by 0.9%. Less well ordered or lower fractional occupancy waters were retained by setting their occupancy to 0.5, according to the standard SHELXL protocol, using a threshold of $U = 0.8$ Å^2, which corresponds to a *B*-factor of 63 Å^2. The final model contains 195 waters and 4 ammonium ions in the asymmetric unit; the ammonium ions were identified by their contacts with phosphate groups, as discussed below. Of these water molecules, 126 have an occupancy of 0.5 and 23 have an occupancy between 0.9 and 1.0.

The COPY19 structure was refined in a similar fashion. As with the COPY18 structure, CNS torsion angle refinement dramatically improved refinement statistics. Water molecules were identified using difference Fourier maps, verifying potential hydrogen bonds with O; in addition, eight predicted waters were included, as discussed below. Half-occupancy waters were included; as in the COPY18 structure, the threshold for setting water occupancy to 0.5 or excluding half-occupancy waters was $U = 0.8$ Å^2. The final model contains 251 waters and 7 ammonium ions; 139 of the waters have an occupancy of 0.5. Table 1 shows final refinement statistics of both structures. The atomic coordinates and structure factors have been deposited with the Nucleic Acid Database (*7*) [NDB ID numbers AR0020 (COPY18) and AR0021 (COPY19)].

Distances used in SHELXL DFIX (bond or 1,2) and DANG (angle or 1,3) constraints were obtained from published parameters for refinement of nucleic acid-containing structures (*22*), using trigonometry to convert angles to 1,3 distances. BUMP (for prevention of too close contacts), FLAT (constraints on planarity of base and associated atoms), DELU ("rigid bond" restraints on anisotropic displacement factors), and, for water and ammonium ions, SIMU (nearby atoms having similar U_{ij} components) constraints were used in refinement of both structures. In addition, EXYZ (making coordinates the same) and EADP (making atomic displacement parameters the same) constraints were used in refinement of the COPY18 structure, to model a single copy of the sugar−phosphate backbone bound to both guanine and cytosine bases at each C1′ atom. These combined to give 21 652 restraints, 17 357 unique reflection data items, and 7977 model parameters in the COPY18 structure; there were 16 965 restraints, 21 631 reflection data items, and 13 059 model parameters in the COPY19 structure.

Prediction of Water Positions. Predicted water locations were determined by successive least-squares superposition of a single residue, with associated average position waters, onto all residues of the target RNA duplex. Water molecules added to the 19 base pair portion of the COPY19 structure were refined using SHELXL, excluding those less than 1.9 Å from an existing water. Water molecules were added to the structure of the 62-nucleotide (nt) domain of *Escherichia coli* 5S ribosomal RNA (*2*), for which diffraction data extend to 3.0 Å resolution, in two refinement steps. First, CNS individual *B*-factor refinement was carried out including predicted waters, with all predicted waters having an initial *B*-factor of 70 Å^2. The average *B*-factor of RNA atoms in this structure is 108 Å^2. Individual *B*-factor refinement was then repeated, setting the *B*-factors of the RNA atoms, metals, and previously identified waters to their initial values and *B*-factors of the predicted waters to the final values of the first refinement. The *B*-factors of most of the predicted waters were 200 Å^2, the maximum allowed in the refinement.

RESULTS AND DISCUSSION

Crystal Packing. Although some aspects of the crystal packing of the COPY18 and COPY19 RNAs are very similar, they differ in significant ways. In the COPY18 crystals, the RNA helix axes are parallel to the *z*-axis and the helices are stacked end on end continuously throughout the crystal. While the helices of the COPY19 crystals are also parallel to the *z*-axis, they are not continuously stacked. In the COPY19 asymmetric unit, the 9.5 base pair helix has the same location in the unit cell as the helices in the COPY18 structure. The remaining 9.5 base pairs of this helix are generated by a crystallographic 2-fold axis. Figure 1a shows this common packing feature, in a view down the *z*-axis of the COPY18 structure. There are close interactions of symmetry-related helices around the crystallographic 3-fold axes with fractional (*x*, *y*) coordinates (2/3, 1/3) and (1/3, 2/3) and a large water channel around the 3-fold axis at (0, 0). In the remaining 19 base pair portion of the COPY19 structure (Figure 1b), the water channel is not as large and has shifted to the (2/3, 1/3) 3-fold axis; there are close contacts around the other two 3-fold axes.

Three RNA helices are in close contact around a 3-fold axis and are held together by a highly ordered ammonium ion, which contacts *pro*-S_p phosphate oxygens of the three helices and a nearby water molecule, which contacts their O2′ hydroxyls (Figure 2a). Both the ammonium ion and the water molecule mediating these contacts lie on the 3-fold axis. The molecule contacting the three phosphate oxygens is identified as an ammonium ion because a positive charge with three hydrogen bond donors is necessary to hydrogen bond oxygens of the negatively charged phosphates. The 2.81 Å distance between the phosphate oxygen and ammonium

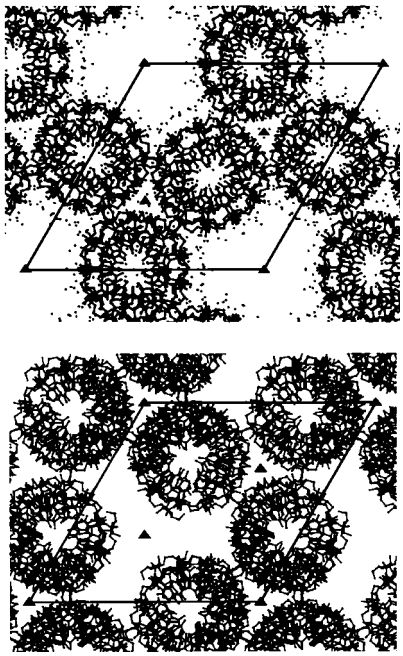

FIGURE 1: (a, top) View down the *z*-axis of the crystal packing of the COPY18 structure. The lower, 9.5 base pair portion of the COPY19 structure exhibits the same packing. (b, bottom) View down the *z*-axis of the crystal packing of the 19 base pair portion of the COPY19 structure.

ion is a good H-bond distance (*23*) but significantly longer than the 2.1 Å distance between an oxygen and a metal ion. Additional contacts (Figure 2b), away from the 3-fold axis, form a dense network holding the three helices together; these include both direct and water-mediated contacts between *pro*-R_p and *pro*-S_p phosphate oxygens and the O2′ hydroxyls of symmetry-related helices.

The stabilizing effects of these phosphate−ammonium triad interactions are reflected in the lower atomic displacement factors of the water molecules and ammonium ions having the contacts illustrated in Figure 2. Moreover, the RNA residues whose phosphates contact ammonium ions have lower atomic displacement factors than the other residues. This motif of water molecules and monovalent cations binding three RNA molecules, related by a 3-fold symmetry axis, is similar to that seen in the r(GUAUAUA)-dC purine−pyrimidine double helix (*24*) and the 16mer duplex with nonadjacent G(syn)−A^+(anti) mispairs (*25*), both of which crystallized in space group *R*3. The G−A^+ mispaired 16mer structure exhibits both a sodium ion with octahedral coordination and a water molecule lying on a 3-fold axis, both contacting three *pro*-R_p phosphate oxygens, while in the purine−pyrimidine structure the sodium ion contacts three O2 hydroxyl groups. A related motif, in which phosphate oxygens contact an ammonium ion lying on a 3-fold symmetry axis, was reported in a phosphoramidate DNA crystal structure (*26*).

The mean atomic displacement factor of the RNA atoms of the complete 19 base pair portion of the COPY19 structure is 56.4 Å^2, much higher than the mean atomic displacement factor of RNA atoms in the 9.5 base pair portion, which is 19.9 Å^2. The mean atomic displacement factor of the RNA atoms in the COPY18 structure is 20.4 Å^2, very similar to the value for the 9.5 base pair helix. The greater disorder that is observed in the 19 base pair portion is most likely

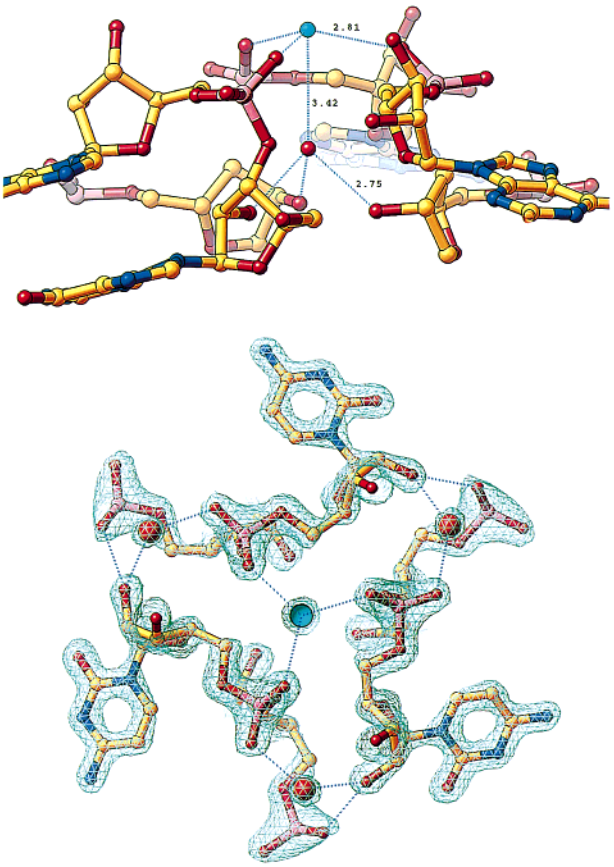

FIGURE 2: (a, top) View, perpendicular to the *z*-axis, of contacts between three symmetry-related RNA strands, an ammonium ion, and a water molecule. The ammonium ion is shown in turquoise, carbons are in yellow, oxygens are in red, and phosphorus atoms are in pink. Distances (in angstroms) are shown for potential hydrogen bonds. (b, bottom) View down the *z*-axis of the COPY19 structure electron density and contacts between symmetry-related RNA strands, water molecules, and an ammonium ion. Potential hydrogen bonds are shown as dashed lines.

due to the smaller number of phosphate−ammonium triads in this portion of the model. The 19 base pair portion of the COPY19 structure has four triads per 19 base pairs, while the 9.5 base pair portion has six per 19 base pairs. The COPY18 structure has four triads per 11 base pairs, or 6.9 per 19 base pairs.

The COPY18 crystals exhibit a static disorder that results from the structure at each base pair position being a compositional average of the entire duplex. Although the RNA duplex contains 18 base pairs, the unit cell contains only 11 base pairs. The RNA helices have identically the same backbone conformation, as required by crystal packing and resulting in diffraction beyond 1.2 Å resolution; however, they are randomly out of register with respect to base composition. The RNA forms continuous, stacked helices extending indefinitely and repeating every 18 nucleotides, whereas the unit cell periodic repeat is only 11 base pairs. Thus, there is a single copy of the sugar−phosphate backbone but the base composition is fully averaged at each residue. A similar averaging "disorder" has been reported for both RNA (*17*) and DNA (*27*) crystals; this averaging is observed when double-stranded nucleic acids form a continuous helix and an integral number of oligonucleotides do not fit in one turn of the helix. In the previously reported RNA structure (*17*), for which diffraction data extend to 1.8 Å resolution,

Table 2: Average Polar Coordinates of Waters Contacting RNA Base Atoms[a]

base	atom	r		ϕ		θ		no.
guanine	N2	3.04 (0.05)	3.21 (0.05)	−54.0 (4.7)	−48.4 (1.5)	84.1 (3.3)	81.2 (1.1)	4, 2
guanine	N3	3.05 (0.28)	2.72	1.0 (1.7)	1.0	98.1 (6.9)	95.6	3, 1
guanine	O6	2.70 (0.11)	2.87 (0.15)	40.4 (3.3)	36.6 (4.2)	81.5 (29.9)	85.6 (25.8)	6, 4
guanine	N7	2.80 (0.08)	2.85 (0.04)	4.0 (5.7)	1.2 (2.5)	81.9 (5.4)	78.5 (4.9)	6, 4
adenine	N3	2.80 (0.04)	2.76 (0.08)	10.5 (3.7)	2.1 (0.3)	96.0 (5.3)	90.5 (2.9)	3, 2
adenine	N6		3.08 (0.05)		38.6 (3.1)		82.2 (3.4)	0, 2
adenine	N7		2.82 (0.05)		0.8 (6.8)		77.0 (1.6)	0, 2
cytosine	O2	3.16 (0.33)	3.55	−29.4 (5.0)	−30.2	85.6 (5.7)	92.5	3, 1
cytosine	N4	2.98 (0.03)	3.01 (0.14)	62.8 (5.0)	60.5 (6.4)	86.6 (3.9)	82.1 (2.9)	6, 5
uracil	O2	2.77 (0.10)	2.80	7.0 (4.5)	5.7	104.6 (13.9)	107.1	2, 1
uracil	O4	2.70	3.05	62.9	69.3	68.5	82.3	1, 1

[a] Values (averages, r in angstroms, ϕ and θ in degrees, with standard deviation in parentheses) for the TAR RNA stem structure are on the left; those of the COPY19 structure are on the right.

17 base pair oligomers form continuous helices, with adjacent helices out of register with respect to the primary sequence. In the 17 base pair crystals, grown from oligomer r(CACCGGAUGGUUCGGUG), the central base pair of the resulting duplex is a G−G mismatch. Adjacent helices pack so as to avoid placing G−G bulges in contact with each other, giving a 4-fold static disorder. There is no such bulge in the COPY18 structure: the static disorder is complete (36-fold). Since the sugar−phosphate backbone structure is identical in the 36 different nucleotides that are superimposed, there can be no sequence-dependent variation in the COPY18 RNA structure.

In the COPY19 crystals the RNA molecules at both duplex positions of the asymmetric unit are present in two orientations, one with the 5′ end of strand 1 having greater z-coordinate and the other flipped 180°, with the 5′ end of strand 2 at greater z-coordinate. The 9.5 base pair duplex is explicitly averaged by the crystallographic 2-fold axis, while the averaging of the orientation of the 19 base pair duplex fully contained in the asymmetric unit was verified by calculating Fourier maps of RNA molecules that had been brominated. $F_{der} - F_{nat}$ difference maps were calculated using amplitudes from a singly substituted bromine derivative (F_{der}), native amplitudes (F_{nat}), and refined model phases. Two peaks of equal heights were seen at locations corresponding to the two RNA orientations. Although this averaging limited the number of unique bases available for analysis of hydration, the COPY19 structure was still useful because the sequences of the two COPY19 RNA strands are very similar, being identical in 14 of the 19 residues. Only those residues for which the same base is present in both orientations were included in the base hydration study. Averaging of two DNA molecules scrambled about a central dyad axis has also been previously reported (*28*).

The COPY19 crystals do not contain the continuous helices seen in the COPY18 crystals, but rather the helix axis shifts by 6.7 Å at the junction of the blunt-ended duplexes. There are, however, excellent stacking interactions between the six-membered rings of the 5′-terminal guanine bases of the adjoining helices; in addition, the O4′ oxygens of both of the 5′-terminal riboses are centered above the 3′-terminal cytosine aromatic ring of the adjacent duplex. The distance between the O4′ oxygen and the six cytosine ring atoms varies from 3.1 to 3.4 Å, close to the sum of the van der Waals radii of the interacting atoms. Similar O4′ oxygen−aromatic ring interactions are prominent in Z-DNA structures (*29*,*30*); Drew et al. describe these interactions as unfavorable. Electrostatic interactions between the partial negative charge of the O4′ oxygen and base ring π electrons would definitely be unfavorable; positioning the O4′ oxygen centered over the base ring will reduce these unfavorable interactions while allowing favorable dispersive interactions between the base and the oxygen. Egli and Gessner have pointed out that the positions of cytidine O4′ atoms relative to the guanine bases in Z-DNA are compatible with an $n \rightarrow \pi^*$ hyperconjugation between the O4′ lone pairs and the base guanidinium system (*31*).

The packing of both crystal forms is compact, as reflected in a calculated volume per nucleotide of 735 Å^3 for the COPY18 crystals and 720 Å^3 for the COPY19 crystals. These values are at the low end of the range of 667−937 Å^3 previously reported for RNA oligomers (*32*).

RNA Conformation. In both copy control crystal structures, the RNA forms a standard A-type helix. In the COPY18 structure and the 9.5 base pair portion of the COPY19 structure, all sugars are in the C3′-*endo* pucker. Nearly all of the residues of the 19 base pair portion also have this canonical pucker; there are two exceptions, which are in strand 2 of the 19 base pair portion of the COPY19 structure and have the nearly equivalent C2′-*exo* pucker, with the pucker phase close to the C3′-*endo* range. All α torsion angles are *gauche*$^-$, and with the exception of one residue in the 19 base pair portion of the COPY19 structure, all γ angles are *gauche*$^+$. The exception, for which γ is *trans*, is the 5′-terminal residue located furthest from the well-ordered 9.5 base pair portion of the structure; the pucker in this residue is C2′-*exo*.

Except for the pucker amplitude, which is significantly larger in the copy control structures, the conformation parameters, calculated using Curves (*33*), for these structures are very similar to those of the r(AU) fiber diffraction structure (*34*). The larger than expected pucker amplitude is not an artifact of this crystal packing, since pucker amplitudes of three additional high-resolution RNA crystal structures (Table 3) are similar to those of the copy control structures.

The previous pucker amplitude standard is based on surveys of nucleoside and mononucleotide crystal structures (*35*, *36*); the average and standard deviation of pucker amplitudes from the nucleoside survey are included in Table 3. There are sufficient high-resolution diffraction data from these RNA structures to define the pucker amplitude with precision. Estimated standard deviations (ESD's), obtained

Table 3: Sugar Conformation Parameters from Several High-Resolution RNA Crystal Structures[a]

structure	COPY18	COPY19		$tRNA^{Ala}$ stem	TAR RNA	$r(C_4G_4)$	nucleosides
high resolution limit (Å)	1.20	1.55		1.16	1.30	1.46	
space group	P321	R32		P1	P1	R32	
pucker amplitude	**45.8 (1.0)**	**46.9 (2.6)**	**46.1 (3.5)**	**45.2 (2.9)**	**45.2 (2.4)**	**45.1 (2.1)**	**38.6 (3.0)**
pucker phase	12.0 (2.8)	10.9 (4.1)	15.3 (6.8)	11.9 (4.7)	12.8 (3.3)	13.6 (5.9)	
C2′−C3′ torsion	43.8 (1.0)	44.8 (2.5)	43.1 (3.1)	43.5 (3.0)	43.1 (2.2)	43.1 (1.8)	
C2′−C3′ estimated SD	1.0	1.9	2.8		1.5		

[a] Average values (standard deviation) in degrees; sugars with puckers other than C3′-*endo* are excluded. Values for the 9.5 base pair helix of the COPY19 structure are in the left column and those for the 19 base pair helix in the right column.

Table 4: Additional Conformational Parameters for Copy Control RNA Structures

	COPY18	COPY19	
backbone torsion angles			
χ	−162.4 (2.6)	−162.0 (3.8)	−163.7 (4.6)
γ	55.1 (3.9)	56.3 (5.9)	55.8 (14.9)
δ	79.0 (2.2)	79.6 (2.3)	80.0 (5.2)
ϵ	−151.3 (3.2)	−150.9 (5.8)	−151.0 (7.1)
ζ	−70.6 (1.7)	−72.0 (2.9)	−73.0 (5.3)
α	−68.3 (3.2)	−68.0 (5.3)	−64.6 (13.8)
β	172.4 (4.5)	171.1 (5.8)	171.0 (8.0)
helicoidal parameters			
X disp (Å)	−4.77 (0.10)	−4.58 (0.15)	−4.76 (0.15)
Y disp (Å)	−0.09 (0.09)	0.12 (0.10)	−0.06 (0.18)
inclination	15.4 (1.9)	12.7 (2.2)	11.6 (2.6)
tip	1.0 (1.3)	−2.1 (1.8)	1.7 (0.7)
intra-base pair parameters			
buckle	−0.5 (1.3)	−1.6 (3.7)	−1.2 (4.2)
propeller	−15.0 (2.2)	−13.0 (3.7)	−12.6 (4.1)
inter-base pair parameters			
roll	0.0 (2.7)	1.4 (1.6)	1.2 (3.9)
rise (Å)	2.65 (0.23)	2.71 (0.25)	2.68 (0.26)
twist	33.0 (1.7)	32.1 (1.7)	31.8 (3.4)
groove parameters			
minor width (Å)	10.1 (0.3)	10.0 (0.5)	9.8 (0.2)
minor depth (Å)	0.5 (0.3)	0.5 (0.4)	0.5 (0.3)
major width (Å)	5.0 (0.8)		4.7 (1.6)
major depth (Å)	9.4 (0.1)		10.1 (0.2)
overall curvature	6.7	7.4	8.0

[a] Conformation parameters for the COPY18 structure, extended to 12 base pairs, and COPY19 structure, with parameters for the 9.5 base pair portion on the left and those for the 19 base pair portion on the right. Values are in degrees unless otherwise indicated, with standard deviations in parentheses. The 9.5 base pair portion of the COPY19 structure was not long enough to define the major groove width or depth.

from SHELXL full-matrix least-squares refinement, describe the extent to which the available crystallographic data define the structure. Although ESD's for torsion amplitude are not provided, the average ESD's for the C2′−C3′ torsion angles, which are similar in magnitude to the pucker amplitude, are given by SHELXL. Except for the less well ordered 19 base pair COPY19 helix, the ESD's are all less than 2°, much less than the difference between the oligonucleotide and nucleoside pucker amplitudes. Torsion angle restraints were not used in the SHELXL refinement.

The significance of the proposed revision of the standard pucker amplitude was evaluated by least-squares superposition of a ribose from the COPY18 structure on a ribose with a ribose from the r(AU) fiber diffraction structure. Atoms C1′, C2′, C4′, and O4′ lie nearly in a plane in both structures and superimpose with an rms fit of 0.027 Å. In the copy control ribose, the C3′ and O3′ atoms are shifted (C3′ by 0.11 Å, O3′ by 0.12 Å) above and away from that plane. The O2′ atom is shifted by 0.13 Å, below the plane and toward the C1′ atom.

As might be expected, the conformations of the three helices in the two copy control structures are very similar (Table 4). The greatest difference among the three helices is the greater variation in the conformation of residues in the 19 base pair COPY19 helix, consistent with this helix's higher atomic displacement factors.

Water Structure. Comparison of the hydration of the COPY18 and COPY19 structures with that of the TAR RNA stem structure (*3*) reveals a common hydration pattern. In the TAR RNA structure, the RNA binds four calcium ions; native data were collected to 1.3 Å resolution, and the *R*-factor of the refined structure, for all data, is 12.6% and the R_{free} is 17.7%. Water molecules that are involved in crystal contacts or in contacts with calcium ions were found to be in significantly variable locations, and thus these were excluded from the survey. Likewise, the water molecules bound to the full 19 base pair molecule in the COPY19 structure are not included in this summary, because the RNA atomic displacement factors are relatively high and the water molecules are poorly positioned. Otherwise, there are no significant differences between water molecule locations in the TAR stem structure and those in the COPY19 structure.

The relative positions of water molecules bound to the same base pair in different sequences were analyzed by superposition of the contacted bases (Figure 3a). The water molecules are located in largely the same place in each

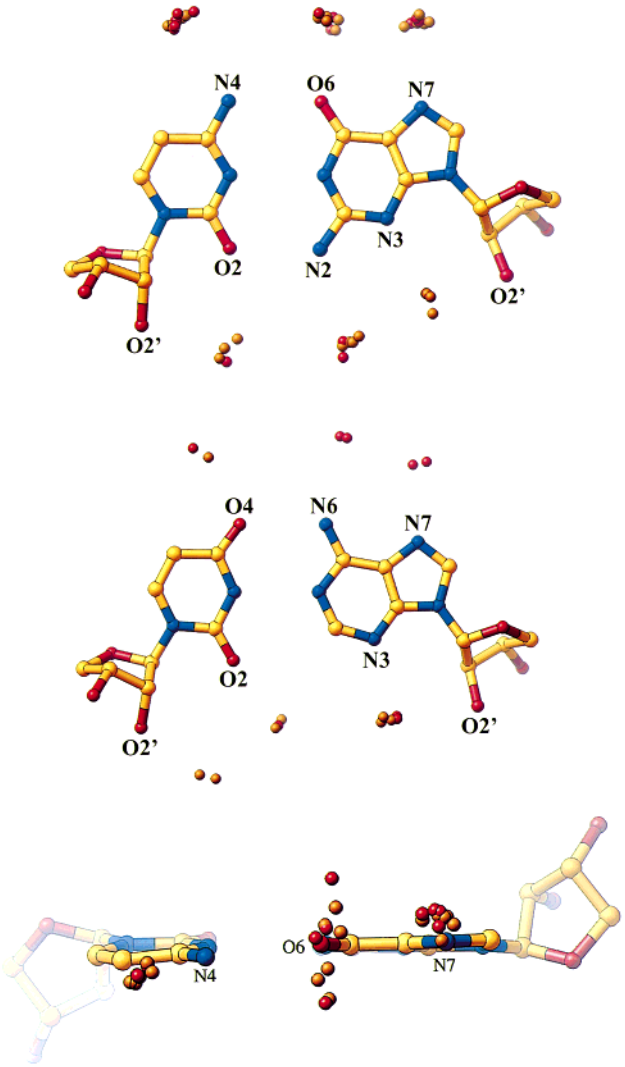

FIGURE 3: (a, top) Superposition of the canonical waters for GC base pairs. Waters from the TAR RNA stem structure are shown in orange, and those from the 19 base pair copy control structure are in red. (b, middle) Superposition of the canonical waters for AU base pairs. (c, bottom) Superposition of the major groove canonical waters for GC base pairs.

example. We observe that waters contacting exocyclic amine groups and aromatic ring nitrogens all lie near the base plane, while waters contacting ketone oxygens are found to occupy positions that vary from well above to well below the base plane (Figure 3c). When viewed perpendicular to the base plane (Figure 3a,b), waters contacting all of these base atoms appear well localized. This pattern is described quantitatively in Table 2, which summarizes, in polar coordinates, the locations of water molecules relative to the contacted base atoms. The coordinate system used places the base in the *xy* plane, with the *x*-axis along the C−N bond in the case of amino nitrogens and along the C−O bond for ketone oxygens; in the case of aromatic nitrogens, the *x*-axis bisects the C−N−C angle. Figure 4 illustrates the coordinate system used to describe contacts with guanine O6 atoms. The localization of waters not contacting ketone oxygens is reflected in the standard deviations of the polar coordinates; for these waters the standard deviations of the angular coordinates are all less than 6°.

Unlike water molecules contacting other ketone oxygens, those contacting cytosine O2 oxygens have well-defined

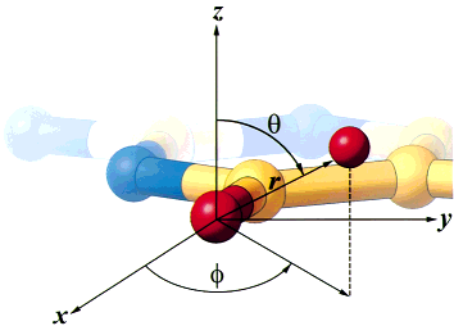

FIGURE 4: Depiction of the spherical polar coordinate system for waters contacting guanine O6 atoms. Changes in the polar angle, theta (θ), correspond to motion above and below the base plane; changes in the azimuthal angle, phi (ϕ), correspond to motion from side to side, parallel to the base plane.

positions. This is because water molecules that contact the cytosine O2 also have strong contacts with the cytosine O2′ hydroxyl group, as well as weaker contacts to the O4′ oxygen of the 3′-adjacent nucleotide.

Three well-defined clusters of water molecules are seen at both GC and AU base pairs in both the major and minor grooves. Hydration of GC base pairs in the major groove is similar to that of AU pairs. In the minor groove, the waters in the outer two clusters contact ribose O2′ hydroxyls, while those in the central cluster contact the waters in both outer clusters as well as an exocyclic base atom. The positions of minor groove waters in the outer clusters are similar in both GC and AU base pairs, while the location of the central cluster is much different (Figure 3a,b).

At GC base pairs, waters in the central cluster contact the guanine N2 nitrogen, while at AU base pairs, waters in the central cluster interact with the uracil O2 oxygen. Water molecules contacting the uracil O2 lie nearly in line with the C2−O2 bond (ϕ is less than 10°; Table 2); other water moleculess contacting ketone oxygen atoms lie at an angle of between 30° and 60° with respect to the C2−O2 bond. GC and AU base pairs also differ in the way waters contact the pyrimidine O2′ hydroxyl group. Waters contacting the cytidine O2′ hydroxyl also contact the cytosine O2 ketone oxygen, as well as the O4′ oxygen of the 3′-adjacent ribose. Waters hydrogen bonded to the uridine O2′ hydroxyl do not contact the uracil O2 oxygen but rather contact the O4′ of the 3′-adjacent ribose. These three minor groove clusters lie approximately in a plane, connecting the O2′ hydroxyls of the base-paired residues, in a contact pattern previously described as transversal (*1*). This hydration pattern agrees with that reported in Auffinger and Westhof's survey of RNA hydration (*6*).

Figure 5a shows the six canonical base-associated minor groove waters for adjacent AU and GC base pairs. The C2 carbon of adenine produces a hydrophobic gap in the center of the AU base pair, so that the central water, W2, associated with the AU base pair contacts uracil but does not form a continuous network of hydrogen bonds between waters across the minor groove. In contrast, the central water of the GC base pair, W5, contacts the N2 of guanine, resulting in a continuous series of hydrogen bonds connecting the O2′ hydroxyls of cytosine and guanine. The six water molecules associated with bases of the same two base pairs in the major groove show that those associated with nitrogen atoms (W7, W8, W10, and W12) are largely in the plane of the base

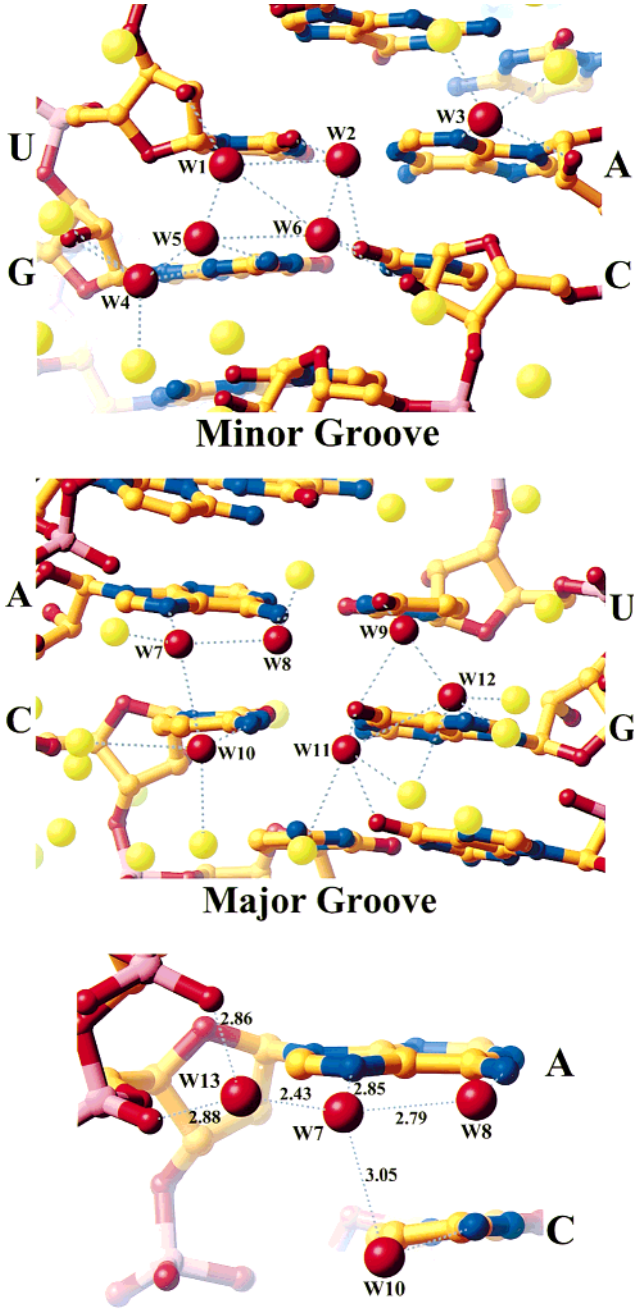

FIGURE 5: (a, top) Minor groove hydration of adjacent UA and GC base pairs. The canonical base-associated waters of these two base pairs are labeled and are in red; the remaining waters are in yellow. (b, middle) Major groove hydration of the base pairs shown in panel a, with the same labeling and color scheme. (c, bottom) Enlarged view of the major groove hydration of the adenosine residue depicted in panels a and b. Only the waters directly contacting the N7-contacting water (W7) are shown, in red, with corresponding distances (in angstroms).

pairs while those bound to keto oxygens (W9 and W11) are not (Figure 5b).

The RNA hydration pattern is very similar to that described by Schneider and Berman (*5*) for B-DNA, the most significant difference being their description of two distinct, partially occupied guanine N2 hydration sites. With the exception of waters contacting thymine O4 atoms, which are displaced by a nearby exocyclic methyl group, the positions of water molecules in the major groove of both B-DNA and RNA are nearly identical. Water molecules bound to guanine N3, adenine N3, and cytosine O2 are all shifted toward the 2′-hydroxyl group in RNA. The variation in the positions of waters bound to guanine O6 relative to the base plane was also reported in a previous study of DNA hydration (*37*); this variation in polar angle of hydrogen-bonding partners of ketone oxygens has also been seen in a survey of small-molecule crystal structures (*38*).

The seven water molecules (W3, W4, W5, W7, W8, W10, and W12) that contact nitrogen atoms of the bases in Figure 5a,b contact no other base atoms, while three of the four ketone−oxygen contacting water molecules (W2, W9, and W11) contact ketone oxygens of two adjacent base pairs. The availability of hydrogen-bonding partners from two adjacent base pairs explains the variation in the out of base plane positioning (polar angle) of water molecules contacting uracil O2 (W2) and guanine O6 (W9, W11). The localization of waters contacting exocyclic amines is explained by their forming hydrogen bonds with well-localized hydrogen atoms. We propose that the small variation in the positioning of water molecules contacting aromatic nitrogens is due to the additional contacts these waters make with the O2′ hydroxyl group, for N3-contacting waters, and to other specifically positioned waters, for water molecules contacting N7. Figure 5c illustrates the latter contact pattern; the W7 water molecule, which contacts N7, forms nearly optimal hydrogon bonds with two water molecules, W8 and W10, which contact exocyclic amines as well as the particularly well-ordered water molecule, W13, which bridges adjacent *pro-*R_p phosphate oxygens.

Other highly conserved water locations include those bridging O1P (*pro-*R_p) phosphate oxygens of adjacent residues and those bridging O2′ and O3′ oxygens of the same residue. These waters have the same locations relative to RNA atoms in the COPY18 and TAR stem structures. The variation in relative positions of O1P-bridging and O2′−O3′ bridging waters is similar to that of base waters not contacting ketone oxygens, discussed above.

Prediction of Water Molecule Locations in Lower Resolution Structures. The positions of water molecules hydrating GC base pairs and bridging O1P phosphate oxygens in the copy control and TAR stem structures were used to predict the positions of water molecules bound to structures whose crystals diffract only to resolutions at which waters cannot be directly identified. Addition of predicted waters yielded slightly improved refinement statistics in both the high atomic displacement factor 19 base pair portion of the COPY19 crystals and the structure of the 62-nucleotide domain of *E. coli* 5S ribosomal RNA.

Of the predicted waters in the 19 base pair portion of the COPY19 crystals, three O1P-bridging waters and five GC base pair waters had acceptable atomic displacement factors after refinement. Addition of both predicted O1P-bridging and predicted GC waters (in separate refinements) reduced the R_{free}; inclusion of the eight predicted waters led to a 0.15% decrease in R_{free}. Inclusion of predicted AU base pair and O2′-O3′ bridging waters (in separate refinements) yielded no waters with an acceptable atomic displacement factor and a slight increase in R_{free}.

Adding predicted waters to the model of the 62-nt 5S rRNA domain gave similar results. Including predicted O1P bridging waters led to a 0.3% improvement in R_{free} and including predicted GC base pair waters a 0.15% decrease

in R_{free}. As with the copy control structure, addition of predicted AU waters and of predicted O2′-O3′ bridging waters led to an increase in R_{free}. The difficulty in predicting AU waters is consistent with Auffinger and Westhof's conclusion (*6*) that hydration around AU pairs appears slightly less well defined than around GC pairs.

Of the predicted O1P-bridging waters in the 5S rRNA structure, the two with the lowest *B*-factors both contact metal ions. The predicted metal ion hydration is confirmed in the structure of the 5S loop E duplex dodecamer, for which diffraction data extend to 1.5 Å (*2*). Among the predicted GC base pair waters with a reasonable *B*-factor is one that contacts both a guanine N7 atom and a metal ion; this metal ion hydration is also confirmed by the 5S dodecamer structure. The agreement between predicted water positions and positions observed in a higher resolution structure implies that waters bridging O1P atoms and those contacting guanine N7 atoms are in fixed locations and can be considered an integral part of RNA structure. This conclusion is supported by results of molecular dynamics simulations; in a report of 3 ns of simulations of an RNA hairpin, Auffinger and Westhof (*39*) reported remarkably long residence times for waters bridging O1P atoms.

Conclusion. These high-resolution COPY18 and COPY19 crystal structures exhibit close interactions between phosphate oxygens and ammonium ions around 3-fold axes. The static disorder seen in these structures is evidence for the lack of correlation between primary sequence and double-stranded RNA structure. Hydration analysis has identified well-localized water positions; significantly, these can be used to predict water locations in lower resolution crystal structures. Analysis of these structures furthermore suggests that oligoribonucleotide pucker amplitude is significantly larger than that of nucleosides and mononucleotides.

ACKNOWLEDGMENT

We are grateful to Drs. Nenad Ban, Temple Burling, Graham Cheetham, Don Crothers, Joe Ippolito, Satwik Kamtekar, Anna Lee, John Marino, Paul Predki, Tobias Restle, Charles Scafe, William G. Scott, Alex Szewczak, Jose Tormo, and Jimin Wang for helpful discussions and to Axel Brünger for reviewing and commenting on the manuscript. Drs. Chad Brautigam, Carl Correll, Soo-Hyun Eom, Joe Ippolito, and Laura Silvian provided assistance in collecting synchrotron data. Special thanks are due to Paul Predki for several gifts of rop protein early in the project and also to Carl Correll and Joe Ippolito for providing coordinates and structure factors for the 5S RNA and TAR RNA stem structures, respectively.

REFERENCES

1. Egli, M., Portmann, S., and Usman, N. (1996) *Biochemistry 35*, 8489−8494.
2. Correll, C. C., Freeborn, B., Moore, P. B., and Steitz, T. A. (1997) *Cell 91*, 705−712.
3. Ippolito, J. A., and Steitz, T. A. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 9819−9824.
4. Mueller, U., Schübel, H., Sprinzl, M., and Heinemann, U. (1999) *RNA 5*, 670−677.
5. Schneider, B., and Berman, H. M. (1995) *Biophys. J. 69*, 2661−2669.
6. Auffinger, P., and Westhof, E. (1998) *J. Biomol. Struct. Dyn. 16*, 693−707.
7. Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S. H., Srinivasan, A. R., and Schneider, B. (1992) *Biophys. J. 63*, 751−759.
8. Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.
9. Hagerman, P. J. (1997) *Annu. Rev. Biophys. Biomol. Struct. 1997*, 139−156.
10. Holbrook, S. R., Sussman, J. L., and Kim, S.-H. (1981) *Science 212*, 1275−1277.
11. Eguchi, Y., and Tomizawa, J. (1990) *Cell 60*, 199−209.
12. Milligan, J. F., Groebe, D. R., Witherell, G. W., and Uhlenbeck, O. C. (1987) *Nucleic Acids Res. 15*, 8783−8798.
13. Predki, P. F., Nayak, L. M., Gottlieb, M. B., and Regan, L. (1995) *Cell 80*, 41−50.
14. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol. 276*, 307−326.
15. CCP4 (1994) *Acta Crystallogr. D50*, 760−763.
16. Brünger, A. T. (1992) *X-PLOR (Version 3.1): A System for X-ray Crystallography and NMR*, Yale University, New Haven, CT.
17. Shah, S. A., and Brünger, A. T. (1999) *J. Mol. Biol. 285*, 1577−1588.
18. Terwilliger, T. C., and Berendzen, J. (1997) *Acta Crystallogr. D53*, 571−579.
19. Jones, T. A., Zou, J.-Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. A47*, 110−119.
20. Brünger, A. T., Adams, P. D., DeLano, W., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Pannu, N. S., Read, R. J., Rice, L. M., and Simonson, T. (1998) *Acta Crystallogr. D54*, 905−921.
21. Sheldrick, G. M. (1997) *SHELX-97, a Set of Programs for Crystal Structure Determination*, University of Göttingen, Germany.
22. Parkinson, G., Vojtechovsky, J., Clowney, L., Brünger, A. T., and Berman, H. M. (1996) *Acta Crystallogr. D52*, 57−64.
23. Ondik, H., and Smith, D. (1962) in *International Tables for X-ray Crystallography* (Macgillavrly, C. H., and Rieck, G. D., Eds.) Vol. III, pp 257−274, Birmingham.
24. Wahl, M. C., Ban, C., Sekharudu, C., Ramakrishnan, B., and Sundralingam, M. (1996) *Acta Crystallogr. D52*, 655−667.
25. Pan, B., Mitra, S. N., and Sundralingam, M. (1999) *Biochemistry 38*, 2826−2381.
26. Tereshko, V., Gryaznov, S., and Egli, M. (1998) *J. Am. Chem. Soc. 120*, 269−283.
27. Zhang, H., van der Marel, G. A., van Boom, J. H., and Wang, A. H. (1992) *Biopolymers 32*, 1559−1569.
28. DiGabriele, A. D., and Steitz, T. A. (1993) *J. Mol. Biol. 231*, 1024−1039.
29. Wang, A. H.-J., Quigley, G. J., Kolpak, F. J., Crawford, J. L., van Boom, J. H., van der Marel, G., and Rich, A. (1979) *Nature 282*, 680−686.
30. Drew, H., Takano, T., Tanaka, S., Itakura, K., and Dickerson, R. E. (1980) *Nature 286*, 567−573.
31. Egli, M., and Gessner, R. V. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 180−184.
32. Holbrook, S. R., and Kim, S. H. (1997) *Biopolymers 44*, 3−21.
33. Lavery, R., and Sklenar, H. (1988) *J. Biomol. Struct. Dyn. 6*, 63−91.
34. Chandrasekaran, R., and Arnott, S. (1989) in *Landolt-Börnstein Numerical Data and Functional Relationships in Science and Technology*, (Saenger, W., Ed.) Vol. VII/1b, pp 31−170, Springer-Verlag, Berlin.
35. de Leeuw, H. P. M., Haasnoot, C. A. G., and Altona, C. (1980) *Isr. J. Chem. 20*, 108−126.
36. Gelbin, A., Schneider, B., Clowney, L., Hsieh, S.-H., Olson, W. K., and Berman, H. M. (1996) *J. Am. Chem. Soc. 118*, 519−529.
37. Schneider, B., Cohen, D., and Berman, H. M. (1992) *Biopolymers 32*, 725−750.
38. Lommerse, J. P. M., Price, S. L., and Taylor, R. (1997) *J. Comput. Chem. 18*, 757−774.
39. Auffinger, P., and Westhof, E. (1997) *J. Mol. Biol. 269*, 326−341.

BI9912793