Crystal Structure of a Brominated RNA Helix with Four Mismatched Base Pairs: An Investigation into RNA Conformational Variability‡

Amy C. Anderson,§,|| Robert H. O'Neil,|| David J. Filman,⊥ and Christin A. Frederick*,§

Committee on Biophysics, Harvard University and Dana-Farber Cancer Institute, 44 Binney Street, S1036, Boston, Massachusetts 02115

Received February 25, 1999; Revised Manuscript Received July 20, 1999

ABSTRACT: The X-ray crystal structure of a brominated RNA helix with four mismatched base pairs and sequence r(UG^{Br}C^{Br}CAGUUCGCUGGC)$_2$ was determined to 2.1 Å using the methods of multiwavelength anomalous diffraction (MAD) applied to the bromine K-absorption edge. There are three molecules in the asymmetric unit with unique crystal-packing environments, revealing true conformational variability at high resolution for this sequence. The structure shows that the sequence itself does not define a consistent pattern of solvent molecules, with the exception of the mismatched base pairs, implying that specific RNA−protein interactions would occur only with the nucleotides. There are a number of significant tertiary interactions, some of which are a result of the brominated base pairs and others that are directly mediated by the RNA 2′ hydroxyl groups. The mismatched base pairs exhibit a solvent network as well as a stacking pattern with their nearest neighbors that validate previous thermodynamic analysis.

RNA is a carrier of genetic information, a catalyst in enzymatic reactions (*1*, *2*), a regulator, either by itself (*3*−*5*) or in conjunction with proteins and, possibly, a catalyst of early enzymatic reactions in the prebiotic world (*6*, *7*). To carry out these varied interactions, an RNA molecule must be able to adopt and tolerate noncanonical conformations as well as tertiary interactions that allow the highly negatively charged backbones to approach each other. Three-dimensional structures of RNA molecules are important for elucidating these interactions and for understanding how they relate to the many roles that RNA plays in biological systems.

Obtaining structures of individual RNA motifs (secondary structures such as loops, mispaired base pairs, pseudoknots, and modified nucleotides) is also important for gaining an insight into RNA folding, RNA−protein recognition and RNA−RNA interactions. RNA motifs, in the crystallographic environment, can reveal routes of folding pathways through potential stable intermediates. Structures of RNA motifs bound to peptides provide examples of nucleotide stacking with aromatic side chains and hydrogen-bonding interactions with polar side chains (*8*). In addition, examples of RNA−RNA interactions can be extrapolated from the intermolecular interactions in the crystal lattice. The ribose 2′ hydroxyl group, in particular, is important in mediating these intermolecular contacts (*9*). Helices can achieve close packing configurations by forming hydrogen bonds between the ribose 2′ hydroxyl and other 2′ hydroxyls, bases, or phosphate oxygens.

This paper presents the X-ray crystallographic structure of a 15-nucleotide RNA sequence, r(UG^{Br}C^{Br}CAGUUCGCUGGC)$_2$. This sequence, which forms a stable hairpin loop in solution (*10*), in the crystal is primarily a self-complementary A-form helix with an internal loop of four mispaired base pairs. Three previously published crystal structures [PDB ID code 255d, r(GGACUUCGGUCC) (*11*); 165d, r(GCUUCGGC)d^{Br}U (*12*); and AR0005, r(UGAGCUUCGGCUC) (*13*)] have included this mismatched segment within different flanking sequences or under different crystallographic constraints. However, our final 2.1 Å structure shows three independently determined molecules in the asymmetric unit giving a detailed high-resolution view of the mismatched region, the brominated base pairs and the solvent structure that is unbiased by crystal packing.

EXPERIMENTAL PROCEDURES

RNA Synthesis and Deprotection. The RNA molecules were chemically synthesized using solid-phase phosphoramidite chemistry (ChemGenes, Inc.). They were deprotected and dried in vacuo. The oligonucleotides were then resuspended in 2 mL of 1 M TBAF/THF[1] (Aldrich) and incubated for 24 h at room temperature to remove the 2′ TBDMS group, and ethanol precipitated with 200 mM NaCl and 12 mL of cold ethanol to remove the silyl salt. The RNA pellet was dried and resuspended in 1 mL of water for purification.

RNA Purification. The crude RNA was purified by high-pressure liquid chromatography (Waters) using a 9 × 250

‡ Coordinates have been deposited in the PDB with accession number 1QBP.

* To whom correspondence should be addressed.

§ Harvard University and Dana-Farber Cancer Institute.

|| Current Address: Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, CA 94143-0448.

⊥ Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Avenue, Boston, MA 02115.

[1] Abbreviations: Br-UUCG, r(UG^{Br}C^{Br}CAGUUCGCUGGC)$_2$; MAD, multiwavelength anomalous diffraction; TBAF/THF, tetrabutylammonium fluoride in tetrahydrofuran; DEAE, diethylaminoethane; TBDMS, 2′-tertbutyldimethylsilyl; MPD, 2,4-methylpentanediol; EXAFS, X-ray absorption fine structure; SD, standard deviation.

10.1021/bi9904508 CCC: $18.00
Published on Web 09/09/1999

mm DEAE anion exchange column (Dionex Co.). Typically, 0.5 μmol of RNA was applied to the column and eluted with a 25 mM ammonium acetate, 1 M NaCl buffer. The main peak corresponding to the full-length species was collected and ethanol precipitated. The pure RNA was resuspended in water and stored in solution at 4 °C.

RNA Crystallization. Pure RNA was mixed with 50 mM sodium cacodylate (pH 7.0) and 5 mM $MgCl_2$ in an Eppendorf tube. This tube was placed in a beaker of 85 °C water, and the water was allowed to cool slowly to room temperature. Crystallizations were carried out by vapor diffusion using the hanging drop method. On a sterile glass coverslip, 1 μL of 1 mM RNA was mixed with 1 μL of a 2 x crystallization solution so that the diluted crystallization solution in the drop was 25 mM sodium cacodylate (pH 7), 40−50 mM $MgCl_2$, 60 mM spermine, 6.7% methylpentane diol (MPD). The coverslip was suspended over 0.75 mL reservoir of 40% MPD in a Linbro tray.

Cryoprotection and Freezing. A total of 2 μL of 30% MPD was added to the crystallization drop, and the crystals were immediately picked up in a fiber loop and flash-frozen in a liquid nitrogen bath.

MAD Data Collection. Diffraction and EXAFS data were collected at Stanford Synchrotron Radiation Laboratory (beamline 1-5). To determine the energies for data collection, fluorescence EXAFS spectra were collected from the crystal used for data collection and processed with the program DISCO (*14*). The wavelength at the minimum of the dispersive anomalous scattering factor, f', was obtained by the Kramers−Kronig transformation of the f'' data. The absolute values of f' and f'' were obtained by scaling the fluorescence spectrum to the theoretical spectrum for elemental bromine.

The four energies (13 370, 13 472, 13 476, and 13 750 eV) were chosen to maximize both the f' and f'' differences. The monochromator used for wavelength selection has a band-pass ($\Delta E/E$) of 2×10^{-4} and was calibrated by scanning across the Au L_{III} edge (11 918 eV) from a foil that could be inserted into the beam path. The calibration was checked after each beam fill (once per day) and between changes in wavelength to check for energy drift. In no case was a significant (>1 eV) drift observed.

The crystal was mounted for data collection on a Huber four-circle diffractometer and data were collected on Fuji image plates which were scanned off line on a Fuji BAS2000 scanner. The crystal was maintained at approximately 100 K during data collection by means of a cryostream system (*15*). Due to detector and cryogenic freezing constraints, the data were collected in an almost random orientation in four sweeps, one per energy, of 171° each.

Phase Refinement. All data were indexed, and the spot intensities were integrated using the program DENZO (*16*). The data were then scaled with the program SCALEPACK (*16*) and converted to structure factor amplitudes using the program TRUNCATE from the CCP4 suite (*17*). The MAD data from all four wavelengths were put on a common scale using the program SCALEIT from the CCP4 suite, choosing the inflection data set (λ_2) as the "native".

Three types of difference Patterson maps were calculated in order to locate the bromine atoms: anomalous ($F_{\lambda 3}^{+} - F_{\lambda 3}^{-}$), dispersive ($F_{\lambda 4} - F_{\lambda 2}$), and isomorphous ($F_{\lambda 2} - F_{iso}$). The "isomorphous" data set was collected from a crystal of an RNA molecule containing only two of the four native bromine atoms. This RNA molecule included brominated cytosines at only position 4 on each strand, and the bromine positions for cytosine 3 on each strand were evident in an isomorphous difference Patterson calculation. From a combination of anomalous, dispersive, and isomorphous Patterson maps, three bromine atoms were located initially, and the remaining nine were located from anomalous or dispersive difference Fourier maps calculated using phases based on the three starting atoms. The coordinates, temperature factors, and occupancies of all 12 atoms were refined using the program MLPHARE (*18*) in the CCP4 suite.

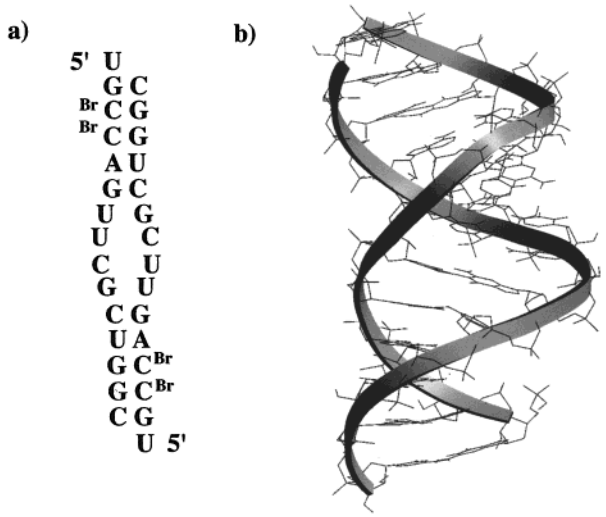

FIGURE 1: Primary and tertiary structure of Br-UUCG. (a) Sequence and predicted secondary structure of the molecule, Br-UUCG. Nucleotides are numbered as follows: molecule A 1−30, molecule B 31−60, molecule C 61−90. (b) Full model of the molecule with a ribbon emphasizing the backbone atoms.

Model Building and Refinement. Initial models of A-form helices were created in InsightII (Biosym) and preliminarily fit into the experimental electron density map at 3 Å. The model was broken into pieces and manipulated using the program O (*19*). Phases from the initial models when applied to the anomalous difference amplitudes correctly predicted the bromine positions, thus confirming the initial model. An extensive series of omit maps revealed which parts of the initial model were less certain. These parts of the structure were eliminated and gradually reintroduced later in the refinement as the electron density made their locations apparent. The model was refined with the program XPLOR (*20*) using RNA dictionaries (*21*) with the experimental phases used as refinement targets until late in the structure determination. Several cycles of conjugate gradient minimization with increasing resolution between cycles were followed by grouped temperature factor refinement (backbone and nucleotide units). A novel refinement procedure based on the "test set" reflections to distinguish authentic solvent sites from noise was used to construct the fixed solvent model in this RNA structure and in another recent RNA crystal structure (*22*).

RESULTS AND DISCUSSION

MAD Phasing. To adapt the MAD method for solving the phase problem of a brominated RNA oligonucleotide, the sequence in Figure 1 (hereafter called Br-UUCG) was

Table 1: Data Collection and Refinement Statistics[a]

	λ_1	λ_2	λ_3	λ_4	isomorphous
D_{max} (Å)	2.1	2.1	2.1	2.1	2.2
R_{symm}	5.1	5.4	6.0	5.1	6.6
completeness (%)	96.8	96.1	97.4	96.8	98.5
PhP (ac)	1.08	NA	0.39	1.65	0.91
R_{Cullis} (ac)	0.84	NA	0.97	0.71	0.88
PhP (c)	0.91	NA	0.29	1.43	0.62
R_{Cullis} (c)	0.72	NA	0.95	0.55	0.84
PhP (ano)	0.21	0.42	0.96	0.98	NA

	resolution	R_{factor}, R_{free}	reflections
model refinement	∞−2.1	0.215, 0.25	11 919

Average *B* Factors	
overall	17.45
molecule A	11.67
molecule B	10.47
molecule C	30.20

[a] R_{symm}: $\Sigma|I - \langle I\rangle|/\Sigma(I)$ where I is the observed intensity and $\langle I\rangle$ is the average intensity from multiple measurements. PhP = phasing power: rms ($|F_h|/E$), where $|F_h|$ is the heavy-atom structure factor amplitude and E is the residual lack of closure error. PhP (c) is the phasing power for the centric reflections and PhP (ac) is the phasing power for the acentric reflections. Cullis R_{factor}: ($\Sigma|E|/\Sigma||F_{PH}| - |F_P||$). NA: not applicable. The isomorphous data refers to a crystal with brominated cytosines at only C4 of each strand.

synthesized, purified, and crystallized according to previously published methods (*23*; also Experimental Procedures). The "native" RNA duplex has four brominated cytosines, does not crystallize under the same conditions as the nonbrominated RNA, and does crystallize in the space group C2 with unit cell dimensions $a = 104.45$ Å, $b = 24.27$ Å, $c = 104.45$ Å, $\beta = 118.76°$. Despite the coincidental equivalence of a and c, the data could not be successfully reduced in any higher symmetry space group. There are three molecules and 12 associated bromine atoms in the asymmetric unit. To extract phase information based on the bromine signal, a four-wavelength (λ_1= 0.920 03 Å, λ_2= 0.920 36 Å, λ_3= 0.901 68 Å, λ_4=0.927 31 Å) anomalous diffraction data set was collected at the Br K-edge. Isomorphous phasing contributions using the differences between the native (four bromines per duplex) and a duplex with only two brominated cytosines (at C4) were included. Statistics from the phase angle refinement (Table 1) show that despite concerns over the random orientation of the crystal during data collection, the anomalous and dispersive phasing power provided enough information to yield experimental maps with excellent initial connectivity for two of the three molecules. An example of the initial $2F_o - F_c$ electron density map without any density modifications such as solvent flattening or averaging is shown in Figure 2a along with an example of the final $2F_o - F_c$ map (Figure 2b). The structure was refined to 2.1 Å with the program XPLOR (*20*) to a final crystallographic R value of 21.7% and R_{free} of 25.3% with the deviations in bond distances and angles equal to 0.010 Å and 1.21 °, respectively.

Global Structure of the Three Independent Molecules. The three molecules, solved independently, are under the influence of different local packing constraints and therefore their structures provide information concerning the range of low energy conformations for this sequence in solution. The composite three structures give us an analogue of an NMR experiment but with the advantage of high resolution and multiple observations per data point. A comparison of the molecules in this structure reveals that the ends of the helices and the region of the mismatched base pairs show the largest statistical distribution and that the canonical helical regions show the least variability (see Figure 3 for a superposition of the six strands of RNA in the asymmetric unit). Molecules A and B both make extensive side-to-side contacts with neighboring molecules, are well-ordered, and with the exception of the local environments of the brominated and mismatched base pairs, have structures comparable to canonical A-form RNA. Molecule C makes fewer side-to-side packing contacts, has larger deviations from an A-form helix and larger temperature factors, thus appearing to be less well-ordered. One possible interpretation is that the conformation of A and B represents one low energy minimum for the sequence and that of C represents another energy minimum with only a slight difference in energy from the undistorted form**.** The root-mean-square (rms) differences between the backbone atoms of molecules A and B, B and C, and A and C are 0.56, 0.92, and 1.02 Å, respectively. Tables 2 and 4 list the backbone torsion angles, twist, rise and groove widths of these three helices as well as those of the three previously published UUCG duplexes: 255d (*11*), 165d (*12*), and AR0005 (*13*).

The α and γ backbone torsion angles sustain the largest degree of variability and in fact show two conformational populations. Table 2 a includes the statistics for bases that exhibit a gauche−gauche γ torsion angle (the normal low-energy form) and Table 3 includes the locations of the bases that show a gauche−trans conformation as well as a single base that exhibits a trans−gauche conformation. As the value for γ increases, α decreases in the typical "crankshaft" motion that preserves the overall duplex conformation. The bases in Table 3 can be classified into "hot spots" for altered conformation based on a pattern that exists in the majority of strands. In five of the six total molecules, one guanine base per double-stranded UUCG mismatch shows a gauche−trans conformation. The two altered guanine nucleotides of both 225d and AR0005 are equivalent due to symmetry constraints between the two strands. The 5′ cytosine flanking the mismatch region in 255d and AR0005 also shows the gauche−trans conformation. Finally, one base near the end of the Br-UUCG duplexes shows an altered conformation, perhaps due to the proximity of the brominated base pairs. These alterations in local γ angles cause an outward bulge in the helix backbone. This is particularly evident in the UUCG mismatch region where the three consecutive pyrimidines cause the helix to kink inward (evident in the ribbon diagram of Figure 1) and the altered guanine causes a compensating bulge outward.

Comparison of the helical parameters of Br-UUCG with the previously published UUCG mismatch duplexes reveals that despite the overall similarity in backbone torsion angles (Table 2), there is a significant difference in groove widths (Table 4). Global helical parameters were calculated with the programs NEWHEL93 (R. Dickerson, 1993), which uses a single helical axis, and CURVES (R. Lavery, 1996) using an optimal curvilinear helical axis. The results of both calculations display similar molecular trends. However, there are some specific differences, notably the helical twist, which can be explained by the differing algorithms. The Br-UUCG molecules have a strong similarity to 165d (*12*) and canonical

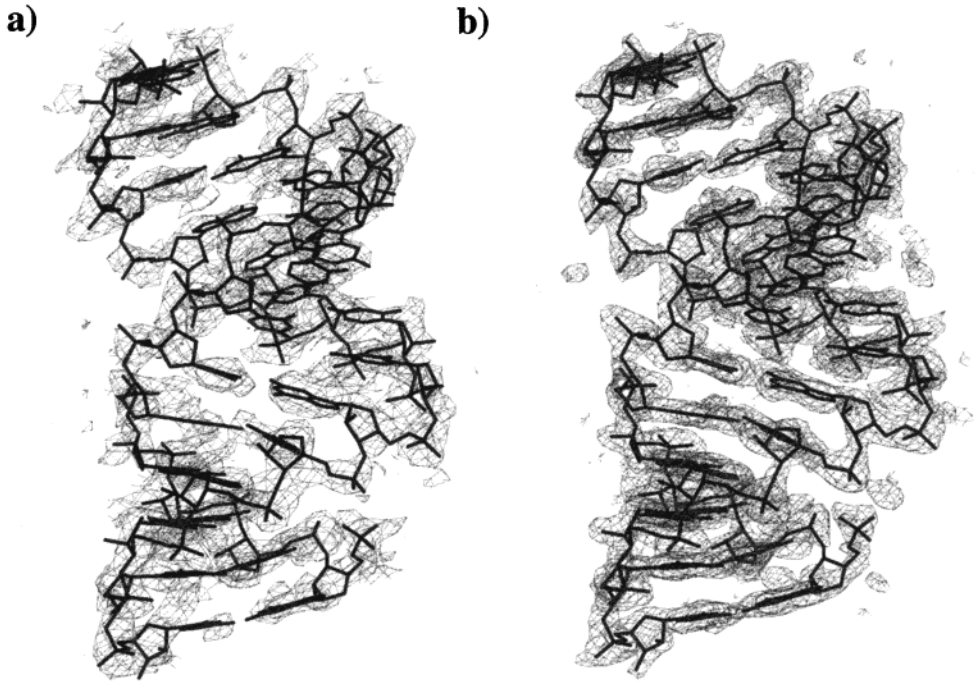

FIGURE 2: Views of the electron density maps. Sections of the (a) experimental electron density map and (b) final $2F_o - F_c$ electron density map.

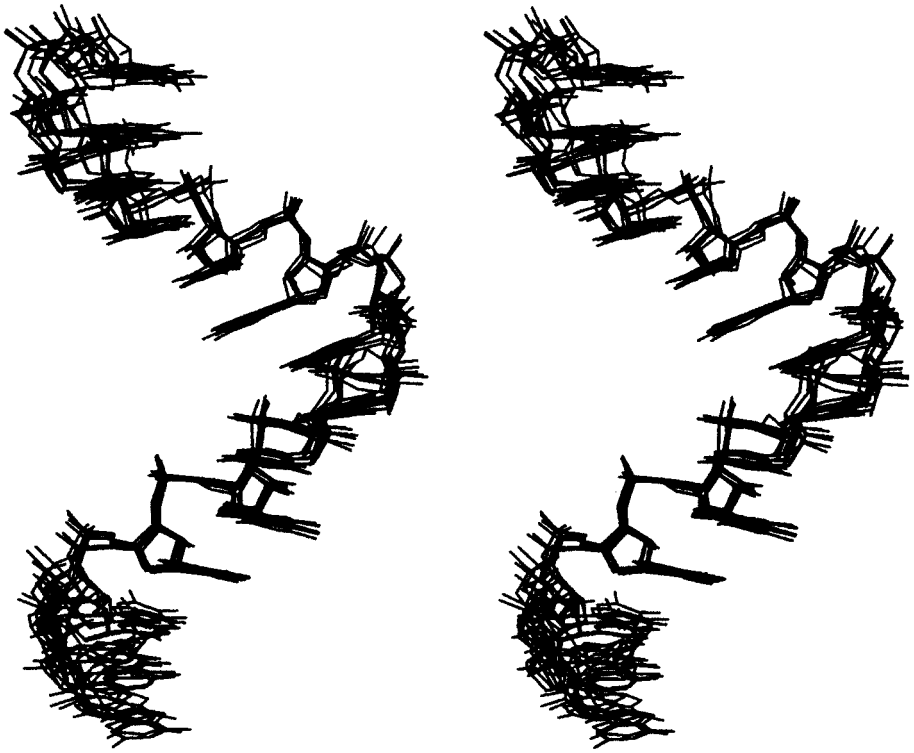

FIGURE 3: Superposition of the six crystallographically independent strands of RNA. The mismatch region and the ends of the helices have the greatest number of accessible conformations.

A-RNA, while 255d (*11*) and AR0005 (*13*), the two helices exhibiting a larger major groove width, appear more like A′-RNA. The difference in groove widths was explained in the analysis of AR0005 (*13*) by relating the change in inclination angle of the base pairs [16−20° for A-RNA and 10.5−11.2° for 255d (*11*) and AR0005 (*13*)] to the larger major groove width. However, the inclination angles of B and C in our present structure are similar to those for either 255d or AR0005, and yet A, B, and C all resemble canonical A-RNA (values for 165d could not be calculated due to the short helical length). The three molecules in this study as well as 165d exhibit full duplexes in the asymmetric unit while 255d and AR0005 have only a single strand in the asymmetric unit, the second strand of the duplex being generated by crystallographic symmetry. These crystallographic constraints may be one reason AR0005 and 255d show an altered groove width and form their own group. It must be kept in mind that calculations of global parameters

Table 2: Mean Torsion Backbone Values and Standard Deviations[a] for Critical Parameters[b]

	α		β		γ		δ		ε		ζ		χ	
RNA[c]	mean	SD	mean	SD	mean	SD	mean	SD	mean	SD	mean	SD	mean	SD
A	297	8.3	175.3	5.8	52.2	6.9	80.1	3.6	207.7	10.4	288.6	4.2	198.2	5.8
B	296.2	6.6	176.6	8.8	51.9	4.8	79.5	3.4	208.4	5.9	288.1	6.5	197.2	5.4
C	302.7	9.5	170.8	8.3	51.4	7.3	78.9	3.0	210.1	11.3	284.2	10.2	197.4	6.7
165d	295.0	5.3	170.4	11.6	58.7	3.5	74.9	3.8	207.7	12.3	290.2	10.0	198.0	4.4
255d	300.9	6.4	179.0	11.4	54.6	11.4	79.0	7.5	212.1	9.9	279.7	8.6	191.0	7.1
AR0005	293.6	9.5	178.5	9.9	55.5	10.3	80.5	3.4	208.2	7.5	286.1	7.5	193.5	6.8
Br base pairs[d]	292.8	4.7	179.0	6.4	52.6	5.3	79.0	1.7	207.8	6.1	289.6	3.5	199.3	6.9

[a] The mean and SD values exclude the outliers listed in Table 3. [b] All values taken from NEWHEL93 (Dickerson). [c] Molecules A, B, and C, 165d r(GCUUCGGC)d^{Br}U (*12*), 255d r(GGACUUCGGUCC) (*11*), and AR0005 r(UGAGCUUCGGCUC) (*13*). [d] Br base pairs refers to the brominated base pairs of Br-UUCG.

Table 3: Location of Outlying α and γ Values for Br-UUCG, 165d, 255d, and AR0005

	U	G	^{Br}C	^{Br}C	A	G	U	U	C	G	C	U	G	G	C
A										S1[a]				S2	
B		S1								S1					
C							S2	S2	S1	S1[b]		S1			S1
165d					G	C	U	U	C	G	G	C			
S2															
255d			G	G	A	C	U	U	C	G	G	U	C	C	
S1,S2	S1,S2														
AR0005		U	G	A	G	C	U	U	C	G	G	C	U	C	
S1,S2															

[a] S1 refers to strand 1, S2 refers to strand 2. [b] The γ angle for this base is trans−gauche, the other outlying γ angles are gauche−trans.

Table 4: Global Helical Parameters[a]

structure	rise (Å)	twist (deg)	inclination (deg)	groove width (Å) major	groove width (Å) minor
A	2.77 (2.71)[b]	32.6 (36.1)	11.15 (14.8)	4.2	11.3
B	2.72 (2.72)	32.6 (35.3)	11.6 (12.99)	4.1	11.7
C	2.7 (2.67)	32.4 (33.2)	11.6 (12.73)	4.2	11.5
165d	2.8	37.4	NA[c]	4.2	11.3
255d	2.92	32.1	11.2	7.8	9.4
AR0005	3.0	30.9	10.5	8.3	9.8

[a] Parameters calculated using both entire strands of the helix. [b] Values in parentheses were calculated using the program, CURVES (R. Lavery, 1996). [c] Helix too short (less than 10 bp) to calculate value of inclination.

like these suffer from the disadvantage that the helices are of different lengths and may exhibit a bias toward A-RNA statistics as the number of canonical base pairs in any given helix increases. It is not possible to calculate local helical parameters such as groove width and twist for the individual segments within each helix due to their lengths. However, visual inspection does confirm the segregation of helices into two groups, as the global parameters correctly indicate.

Mismatched Base Pairs. The structure of the mismatched base pairs, two C:U pairs (Figure 4a) and two G:U pairs (Figure 4b) per molecule, and their local, stabilizing water molecule network is remarkably similar in all three molecules of Br-UUCG as well as in 165d (*12*), 255d (*11*), and AR0005 (*13*). Specifically, the rms deviations between the mismatched regions of Br-UUCG and 165d, 255d, and AR0005 are 0.65, 1.13, and 1.2 Å, respectively. The C:U mismatched base pairs of Br-UUCG have very compromised hydrogen-bonding capacity and stacking interactions but distort the A-form helix only slightly, with values of rise, twist, and slide close to the canonical A-values. Poor stacking interactions, together with the presence of only a single hydrogen bond between nucleotides causes the C:U pair to be the least

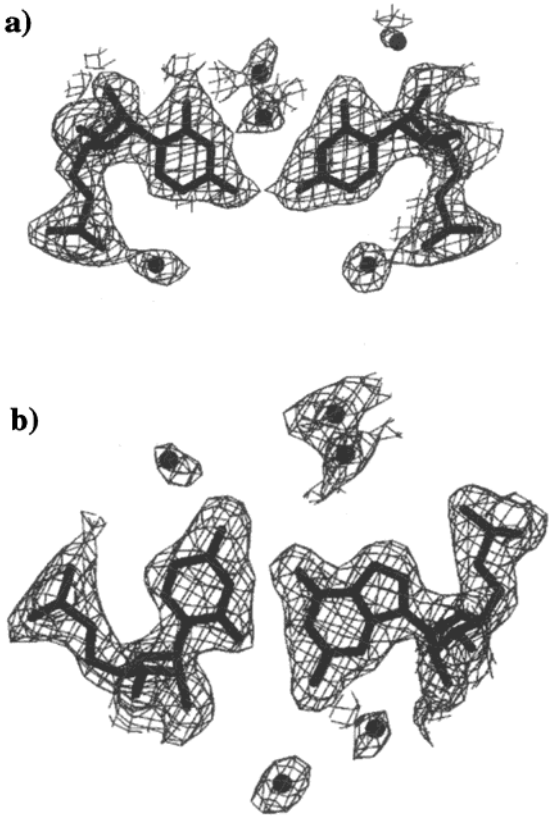

FIGURE 4: Structures of the (a) U:C base pair and (b) U:G base pair with their associated water molecules.

thermodynamically stable of the mismatched pairs (*24*, *25*). The G:U mismatch base pairs of Br-UUCG, with "wobble" type hydrogen bonding, do distort the A-form helix, as evident in the rise of the base pair and the torsion angles of the backbone. This distortion alters the geometric and charge properties of the grooves relative to the grooves composed

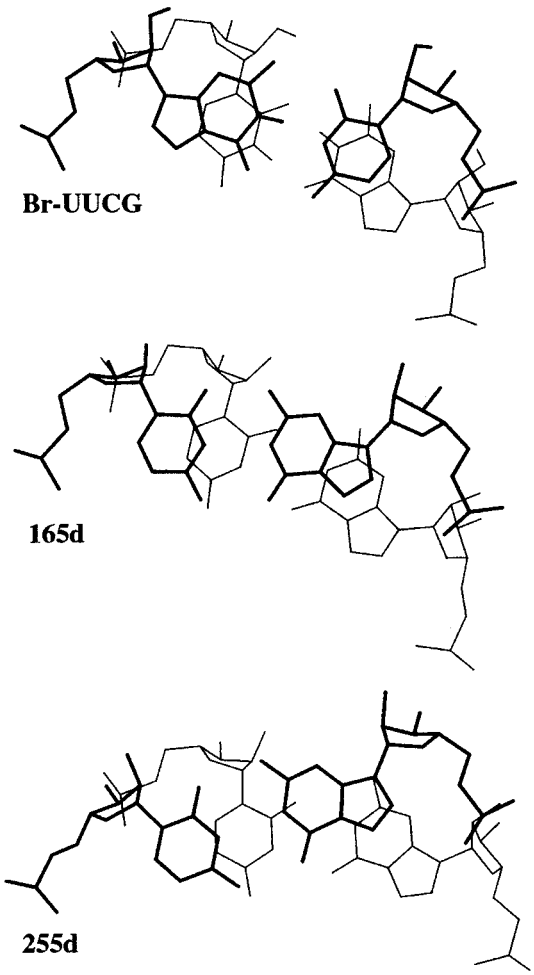

FIGURE 5: Flanking sequence comparison. The 5′ neighbor for a base pair is shown in thick lines and the 3′ base pair in thin lines. (a) 5′ flanking G:C and U:G from Br-UUCG, (b) 5′ flanking C:G and U:G from 165d, (c) 5′ flanking C:G and U:G from 255d.

of only Watson−Crick base pairs, bringing about a change in binding potential that could be a signal for protein−RNA recognition, splice site recognition, metal coordination, or RNA−RNA interaction.

Flanking Sequences. In the Br-UUCG structure, the first U of the mismatch region is flanked 5′ by the G of a G:C Watson−Crick base pair (see Table 3 for sequence alignments). This differs from all previous structures of duplexes with a mismatched UUCG internal loop that have 5′ C:G flanking base pairs. In a comparison of the stacking of the G:C/U:G base pairs versus the C:G/U:G base pairs (Figure 5), it is evident that the G:C pair facilitates a significant increase in ring overlap, leading to increased stability. In Br-UUCG, the six-member heterocycle of the flanking guanine base overlaps approximately 90% of the uracil from the neighboring U:G mismatched pair and the flanking cytosine overlaps approximately 90% of the six-member heterocycle of the guanine of the mismatch. In contrast, in both C:G flanking base pair structures (Figure 5, panels b and c) (*11*, *12*), there is no overlap between the cytosine and uracil and only partial overlap between the two adjacent guanines. The backbone torsion angles for these bases are all within the normal range, as are the λ values, a typical measure for stacking interactions, implying that the stacking of the C:G base pair is optimized without causing helix distortion. In fact, Table 3 reveals that two of the cytosines (structures 255d and AR0005) flanking the mismatched U:G exhibit the gauche−trans γ angle. Therefore, the G:C flanking base pair can be accommodated without helical distortion while the C:G pair must cause some strain in order to maximize overlap. This structural observation validates the thermodynamic analysis (*26*) that determined a 1 kcal/mol advantage for a flanking G:C base pair over a C:G base pair given a U:G mismatch.

Structure of the Brominated Base Pairs. The geometry of each of the brominated C:G base pairs is very similar to canonical Watson−Crick C:G base pairs, and the distances of the hydrogen bonds are well within the range of regular Watson−Crick bonds. λ values for the bases, which represent the angle between the glycosidic bond relative to the C1′−C1′ vector, are also within the normal range 51−60°. Average values and standard deviations (averaged over all 12 brominated bases) for the backbone torsion angles (Table 2) as well as the standard parameters describing the base pair are all within the normal values despite high standard deviation values in some cases.

The values for the helical twist of the brominated base pairs, however, fall outside the normal ranges and reflect a high degree of underwinding in this part of the helix. These base pairs are underwound by an average of 6.95°. In a superposition of the Br-UUCG helix with a canonical A-RNA model generated with matching global rise and twist, the 3′ end of the sequence (the nonbrominated end) matches the canonical A-RNA strand very closely. However, the 5′ end of the sequence with the brominated cytosines diverges from the model with the canonical parameters.

There are two possible explanations for the divergence in helical twist. First, the distortion might result from an attempt to avoid steric hindrance between the bromine substituent and the amino group of the neighboring cytosine. The van der Waals' radius of a bromine atom (1.95 Å) is equivalent to the radius of an amino group. In a normal stacking pattern for this base pair step (Figure 6a), the bromine atom would fall directly under the amino group (*27*). Since the rise between base pairs at this step is approximately 3.3 Å, the two groups would clash if the helix retained its usual geometry. By underwinding the base pair, it is possible to avoid this steric hindrance and to still maintain good helical stacking. The second reason that the brominated cytosine twists the helix is to optimize the positive dispersion forces between the induced dipole of the aromatic ring and the bromine atom (Figure 6b). The halogen can induce a dipole in the ring (and vice versa), similar to the effects of stacking two aromatic rings (*28*, *29*). The additional stacking energy provided by the halogen and the aromatic ring stabilizes the arrangement.

This underwinding of the helix in the region of the brominated bases causes the minor groove to expand. This expansion is visually apparent, though due to its location near the end of the helix, it was not possible to measure the groove width directly using phosphate-phosphate distances in the region of the brominated base pairs. The increased width is also stabilized by intermolecular contacts at the helix ends. The expanded minor groove exposes the ribose 2′ hydroxyls and allows them to hydrogen bond with other 2′ hydroxyls or nucleotide bases from adjacent helices and thereby form close crystal contacts. The additional hydrogen

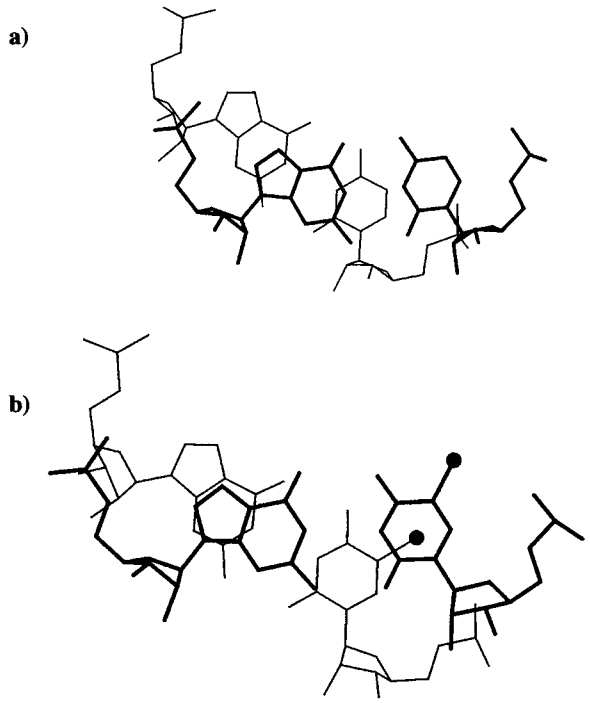

FIGURE 6: Arrangement of the brominated base pairs. **(**a) Normal stacking pattern for an unmodified C:G base pair (b) Stacking of the brominated C:G base pair. The bromine atoms are shown in cpk representation and the 5′ base pairs are shown in thick lines. Base pairs have normal values of tip (3.4°, SD 4.3), roll (10.1°, SD 10.2), inclination (3.59°, SD 3.3), propeller twist (−9.19°, SD 3.2), tilt (0.92°, SD 3), rise (3.34 Å, SD 0.22), slide (−1.74 Å, SD 0.7), and displacement ($x = -4.5$ Å, SD 0.7, $y = -1.76$ Å, SD 0.32).

bonds strengthened the lattice interactions thus yielding improved crystals for the brominated version of the sequence.

Solvent Structure. In the context of DNA−protein recognition, it has been suggested that there may be a specific water molecule network associated with nucleic acids which mediates the specificity of such interactions (*30*). To test this theory for these helices at 2.1 Å resolution in the absence of protein, the positions of all water molecules associated with molecules A and B (which share a similar packing environment) were compared. Molecule C, while sharing a similar structure to molecules A and B, appears to show greater static disorder and does not bind as many ordered solvent molecules as either A or B.

Molecule A has 49 ordered first shell water molecules associated with it, of which 20% is bound to 2′ OH groups, 24.5% is bound to the phosphate oxygens, 2% is bound to phosphate ester oxygens, and 40.8% is bound to nucleotides. Molecule B has 55 ordered first shell water molecules associated with it, of which 20% is bound to 2′ OH groups, 32.7% is bound to phosphate oxygens, 2% is bound to phosphate ester oxygens, and 36.4% is bound to nucleotides. Despite the overall similarity in the statistics between the two molecules, in specific, very few of the fixed water positions superimpose when the RNA backbones themselves are superimposed. This result implies that the overall sequence, with the exception of the mismatch region, fails to specify the water network uniquely. However, there is a solvent network which stabilizes the U:C and U:G base pairs and whose water molecules occupy essentially the same positions independent of the rest of the crystallographic structure.

In addition, there is no fixed arrangement of waters in the grooves that is consistent from base pair to base pair. This lack of a fixed arrangement potentially allows solvent molecules to take advantage of the differences in local helical structure and form specific interactions in the presence of other RNA or protein molecules. The lack of consistent groove hydration is in contrast to the $U(UA)_6A$ structure (*31*) in which the solvent molecules of the major groove form a continuous column. It is also in contrast to the $r(CCCCGGGG)_2$ structure (*32, 33*) in which two water molecules are bound to each of the 2′ hydroxyls, forming an extensive regular hydration spine throughout the entire minor groove. These authors suggest that this spine of hydration is an important contribution to the selectivity of RNA−protein interactions. However, this degree of hydration is not universal in RNA structures. In fact, independent of our Br-UUCG structure, there are several other reports of RNA molecules without a noticeable spine of hydration in either groove (*12, 34, 35*). Using a particular hydration pattern for sequence-specific intermolecular recognition would obviously require the hydration pattern to be specified uniquely by the structure of the sequence. The Br-UUCG structure does not rule out this possibility for some types of RNA recognition, however, it also demonstrates another important alternative. Although the primary sequence of the molecule positions the polar atoms of the RNA in a specific way, independent of packing environment, it does not position their associated water molecules in a single sequence-specific way. Presumably potential water-binding sites are not isolated from one another, and thus in the course of interactions with other macromolecules, the solvent network is free to rearrange.

Packing. The Br-UUCG crystals contain approximately 53.8% solvent, a value that corresponds to a volume per base pair of 1382 Å^3, a reasonable value for typical A-RNA crystal packing (*32*,*36*). The three molecules form a pseudo-continuous column mediated by nucleotide stacking at the ends of each molecule (Figure 7). Two of the three molecules (A and B) interact tightly with their symmetry-related copies along the length of their backbones and the third molecule (C) is held in place by the stacking of the column and a few side-to-side contacts (residues 62−63 and 71−84). The column formed by the three molecules, however, is not straight and does not have a single helical axis common to all three molecules, a feature typical of other RNA duplex structures (*11, 12, 34, 36*−*38*). Each column bends molecule A out of the cell and creates an almost 90° helical crossing angle between each copy of molecule A and its symmetry mate. This type of packing is unusual for a helix with a close approximation to A-form. It may be explained by the close contacts created between molecules A and B and the predicted location of the overhanging base. Although the electron density for the overhanging base is disordered, residual density in the difference map (coefficients $F_o - F_c$) predicts that the overhanging bases on the 5′ end of each strand are packed against the minor groove. Therefore, the packing may partially be explained by the necessity of the overhanging bases to assume the same position with respect to each terminal base pair. Many of the crystal contacts between molecules A and B and their symmetry mates are

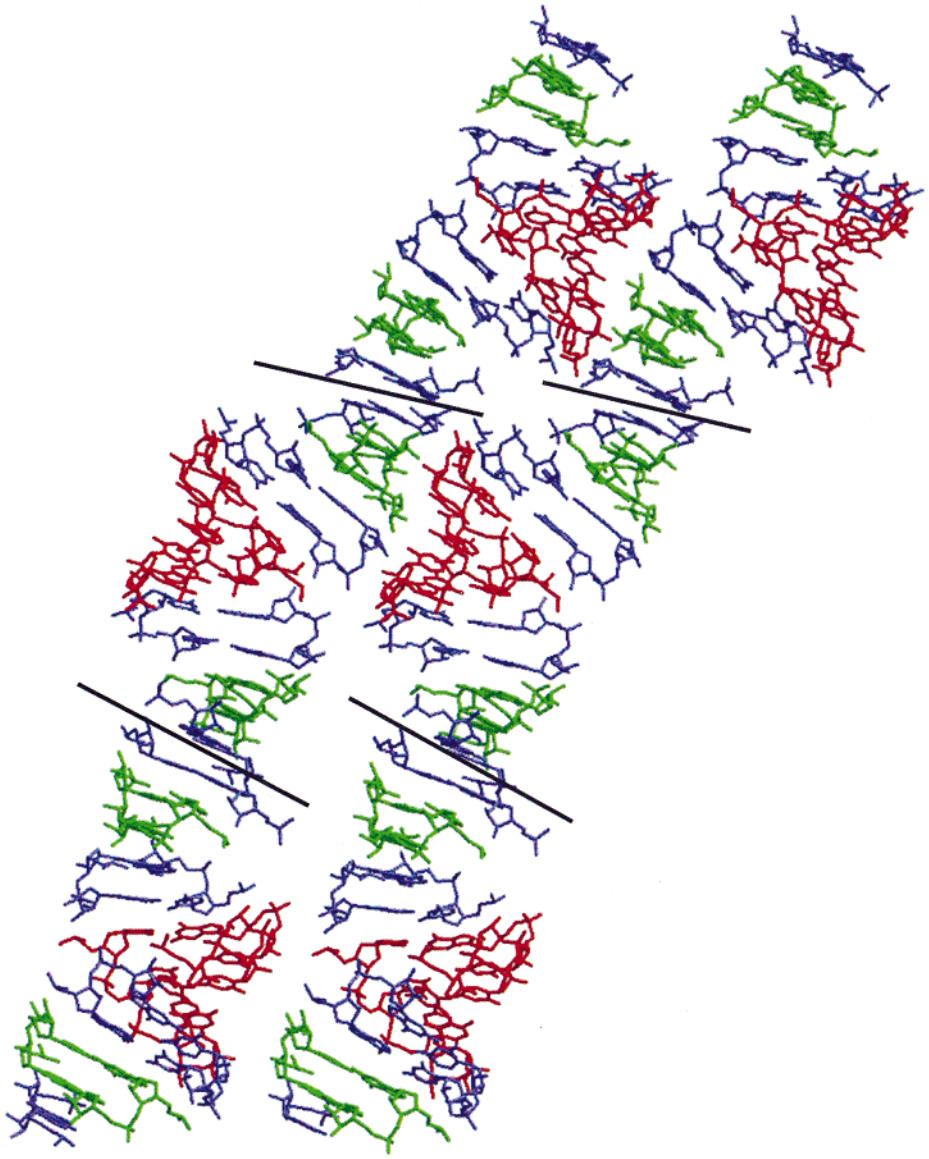

FIGURE 7: Packing of the RNA molecules in the cell. The three molecules are colored blue, the brominated cytosines are highlighted in red and the mismatched region is highlighted in green.

mediated by the ribose 2′ hydroxyl groups hydrogen bonding to other 2′ OH groups, bases, or phosphate oxygens. The contacts occur primarily between the backbone of the mismatched base pairs and the backbone of the brominated base pairs.

With the knowledge of the packing scheme, it is possible to understand why replacing one of the brominated cytosines with a standard cytosine (at C3) allows an isomorphous crystal form and replacing the other (C4) creates a higher degree of disorder. The distortion that the bromine on C4 causes for the helix is essential to the packing scheme since the backbone in the area of the C4 base abuts directly against, and forms hydrogen bonds to, the backbone of the mismatch region of a symmetry-related helix. The C3 base is further from this essential region and changes in the backbone at C3 do not affect the packing as much as changes at C4.

CONCLUSIONS

The crystal structure of this RNA molecule provides details concerning the range of accessible conformations for a given sequence of RNA. The canonical regions of the helix are relatively similar to each other and to A-RNA, with low standard deviations for positional differences. The mismatch region and the ends of the helices, however, experience several different conformations. The structure provides an unbiased view of the possible energy minima of the mismatched G:U and C:U base pairs and the bound, stabilizing water molecules since there are three independent copies of the molecule constrained by different crystal-packing environments. The 5′ flanking G:C base pair provides added stability due to increased overlap between its nucleotides and those in the U:G mismatch. The modified nucleotides exhibit further helical stabilization and a local underwinding due to the halogen-induced dipole in an adjacent aromatic ring. Finally, the structure reveals the strong dependence on the RNA-unique 2′ hydroxyl group for mediating both RNA−RNA and RNA−solvent interactions. Many of the contacts between the tightly packed molecules are between 2′ OH groups, further emphasizing the importance of this feature in carrying out the varied functions of an RNA molecule.

ACKNOWLEDGMENT

The authors wish to thank Tricia Takahara, Rene Lachicotte, and Marie Lavere-Wright for their assistance in data collection. Also, Henry Bellamy at beamline 1-5AD at Stanford Synchrotron Radiation Laboratory was extremely helpful in both data collection and processing and Jaishree Trikha assisted in writing the X-PLOR scripts necessary for refinement.

REFERENCES

1. Cech, T. (1987) *Science 236*, 1532−1539.
2. Cech, T., and Uhlenbeck, O. (1994) *Nature 372*, 39−40.
3. ten Dam, E., Pleij, K., and Draper, D. (1992) *Biochemistry 31*, 11665−11676.
4. Asano, K., Kato, A., Moriwake, H., Hama, C., Shiba, K., and Mizobuchi, K. (1991) *J. Biol. Chem. 266*, 24549−24556.
5. Shen, L., and Tinoco, I. (1995) *J. Mol. Biol. 247*, 963−978.
6. Green, R., and Szostack, J. (1992) *Science 258*, 1910−1915.
7. Eckland, E., and Bartel, D. (1995) *Science 269*, 364−370.
8. Oubridge, C., Ito, N., Evans, P., Teo, C., and Nagai, K. (1994) *Nature 372*, 432−438.
9. Pyle, A., and Cech, T. (1991) *Nature 350*, 628−631.
10. Cheong, C., Varani, G., and Tinoco, I. (1992) *Nature 346*, 680−682.
11. Holbrook, S., Cheong, C., Tinoco, I., and Kim, S.-H. (1990) *Nature 353*, 579−581.
12. Cruse, W., Saludjian, P., Biala, E., Strazewski, P., Prange, T., and Kennard, O. (1994) *Proc. Natl. Acad. Sci. 91*, 4160−4164.
13. Tanaka, Y., Fujii, S., Hiroaki, H., Sakata, T., Tanaka, T., Uesugi, S., Tomita, K., and Kyogoku, Y. (1999) *Nucleic Acids Res. 27*, 949−955.
14. Eichhorn, K. D., DISCO. (1985) Stanford Synchrotron Radiation Laboratory.
15. Bellamy, H. D., Soltis, S. M., Phizackerly, R. P., and Hope, H. (1994) *J. Appl. Cryst. 27*, 967−970.
16. Otwinowski, Z. (1993) *DENZO 1-56-62*, SERC Daresbury Laboratory, Warrington, U.K.
17. Collaborative Computational Project Number 4 (1994) *Acta Crystallogr., Sect. D 50*, 760−763.
18. Otwinowski, Z. (1991) in *Isomorphous Replacement and Anomalous Scattering, Proc. CCP4 St. W.* (Wolf, W., Evans, P., and Leslie, A., Eds.) pp 80−86, Daresbury Laboratory, Warrington, U.K.
19. Jones, T., Zou J.-Y., Cowan, S., and Kjeldgaard, M. (1991) *Acta Crystallogr., Sect. A 47*, 110−119.
20. Brünger, A. (1993) *X-PLOR. Version 3.1 A System for X-ray Crystallography and NMR*, Yale University Press, New Haven, CT.
21. Parkinson, G., Vojtechovsky, J., Clowney, L., and Brunger, A. (1996) *Acta Crystallogr., Sect. D 52*, 57−64.
22. Trikha, J., Filman, D., and Hogle, J. (1999) *Nucleic Acids Res. 27*, 1728−1739.
23. Anderson, A., Scaringe, S., Earp, B., and Frederick, C. (1996) *RNA 2*, 110−117.
24. Xia, T., McDowell, J., and Turner, D. (1997) *Biochemistry 36*, 12486−12497.
25. Limmer, S. (1997) *Prog. Nucleic Acids Res. 57*, 1−39.
26. He, L., Kierzek, R., SantaLucia, J., Walter, A., and Turner, D. (1990) *Biochemistry 30*, 11124−11132.
27. Saenger, W. (1984) in *Principles of Nucleic Acid Structure*, (Cantor, C., Ed.) Springer-Verlag, New York.
28. Iball, J., Morgan, C., and Wilson, H. (1968) *Proc. R. Soc. London. Ser A 302*, 225−236.
29. Harris, D., and MacIntyre, W. (1964) *Biophys. J. 4*, 203−225.
30. Shakked, Z., Guzikevich-Guerstein, G., Frolow, F., Rabinovich, D., Joachimiak, A., and Sigler, P. (1994) *Nature 368*, 469−73.
31. Dock-Bregeon, A. C., Chevrier, B., Podjarny, A., Johnson, J., de Bear, J. S., Gough, G., Gilham, P., and Moras, D. (1989) *J. Mol. Biol. 209*, 459−474.
32. Portmann, S., Usman, N., and Egli, M. (1995) *Biochemistry 34*, 7569−7575.
33. Egli, M., Portmann, S., and Usman, N. (1996) *Biochemistry 35*, 8489−8494.
34. Baeyens, K., De Bondt, H., and Holbrook, S. (1995) *Nat. Struct. Biol. 2*, 56−62.
35. Lietzke, S., Barnes, C., Berglund, J., and Kundrot, C. (1996) *Structure 4*, 917−930.
36. Portmann, S., Grimm, S., Workman, C., Usman, N, and Egli, M. (1995) *Chem. Biol. 3*, 173−84.
37. Schindelin, H., Zhang, M., Bald, R., Furste, J., Erdmann, V., and Heinemann, U. (1995) *J. Mol. Biol. 249*, 595−603.
38. Leonard, G., McAuley-Hecht, K., Ebel, S., Lough, D., Brown, T., and Hunter, W. (1994) *Structure 2*, 483−94.

BI9904508