*RNA* (2001), *7*:1425–1431. Cambridge University Press. Printed in the USA.

DOI: 10.1017.S1355838201010627

# Crystal structures of two forms of a 14-mer RNA/DNA chimer duplex with double UU bulges: A novel intramolecular U*(A•U) base triple

**JUNPENG DENG, YONG XIONG, CHELLAPPANPILLAI SUDARSANAKUMAR, KE SHI, and MUTTAIYA SUNDARALINGAM**

Biological Macromolecular Structure Center, Departments of Chemistry, Biochemistry and Biophysics Program, The Ohio State University, Columbus, Ohio 43210, USA

## ABSTRACT

**The RNA/DNA 14-mer, $(gguauuucgguaCc)_2$ with consecutive uridine bulges (underlined) on each strand has been determined in two crystal forms, spermine bound (Sp-form) and spermine free (Sp-free). The former was solved by the MAD method with three-wavelength data collected at Brookhaven National Laboratory (BNL); the later isomorphous structure was solved by the molecular replacement method using data collected on our Raxis II*c* imaging plate system. The two crystal forms belong to the space group C2 with one molecule of double-stranded 14 mer in the asymmetric unit. The Sp-form has cell constants, a = 60.06, b = 29.10, c = 52.57 Å, $\beta$ = 120.79° and was refined to 1.7 Å resolution with a final $R_{work}/R_{free}$ of 19.8%/22.7% using 8,549 independent reflections. The Sp-free structure has cell constants, a = 60.06, b = 29.58, c = 52.50 Å, $\beta$ = 120.85° and was refined to 1.8 Å with a final $R_{work}/R_{free}$ of 20.8%/23.2% using 6,285 unique reflections. The two structures are identical, except that the Sp-form has a spermine bound in the major groove, parallel to the RNA helical axis. One of the uridine bulges forms a novel intramolecular U*(A•U) base triple. The helices are in the C3′-endo conformation (A-form), but the bulges adopt the C2′-*endo* sugar pucker. Furthermore, the bulges induce a kink (30°) in the helix axis and a very large twist (55°) between the base pairs flanking the bulges. The Sp-form has one $Mg^{2+}$ ion whereas the Sp-free form has two $Mg^{2+}$ ions.**

**Keywords: base triple; kink; RNA duplex; UU-bulges; X-ray structure**

## INTRODUCTION

Bulges are among the most common nonhelical structural elements in nucleic acids and their conformations in RNA have been studied (Hermann & Patel, 2000; Sudarsanakumar & Sundaralingam, in prep.). Single and multiple bulges are one of the significant secondary structural elements observed in RNA molecules and their conformational preferences influence RNA folding and function. Bulged residues often display a strong degree of phylogenetic conservation, both in location and sequence, within functionally important larger RNA molecules. Uridine bulges are observed at key positions in many RNA molecules. A conserved uridine bulge was positioned at the center of the catalytic domain, adjacent to the site of the pre-tRNA cleavage, in the catalytically important P4 elements of all bacterial RNase-P RNAs (Frank et al., 1996; Schmitz & Tinoco, 2000). The crystal structure of the HIV-I high affinity Rev binding site shows a looped-out uridine bulge, which serves as a critical spacer between the purine–purine base pairs, providing necessary flexibility for the homopurine base pairs to orient conveniently for the binding of Rev (Iwai et al., 1992; Ippolito & Steitz, 2000).

In most cases, the bulged residues are involved in base triple interaction with symmetry related molecules either in the major or in the minor grooves (Ippolito & Steitz, 2000; Xiong & Sundaralingam, 2000). Intramolecular base triple conformations have been proposed to be involved in TAR ligand binding. Williamson and colleagues found that in the presence of argininamide, the U23 in the bulged nucleotides of TAR RNA stem loop

Reprint requests to: Muttaiya Sundaralingam, The Ohio State University, Biological Macromolecular Structure Center, Departments of Chemistry, Biochemistry and Biophysics Program, 012 Rightmire Hall, 1060 Carmack Rd., Columbus, Ohio 43210, USA; e-mail: sundaral@chemistry.ohio-state.edu.

*Abbreviations:* RNA components are represented in lower case letters (13 residues) and a DNA component in capital letters (1 residue). BNL: Brookhaven National Laboratory; DM: density modification; MAD: multiple wavelength anomalous dispersion; MPD: methyl-2,4-pentanediol; NCS: noncrystallographic symmetry averaging; Sp-form, spermine bound form; Sp-free, spermine free form.

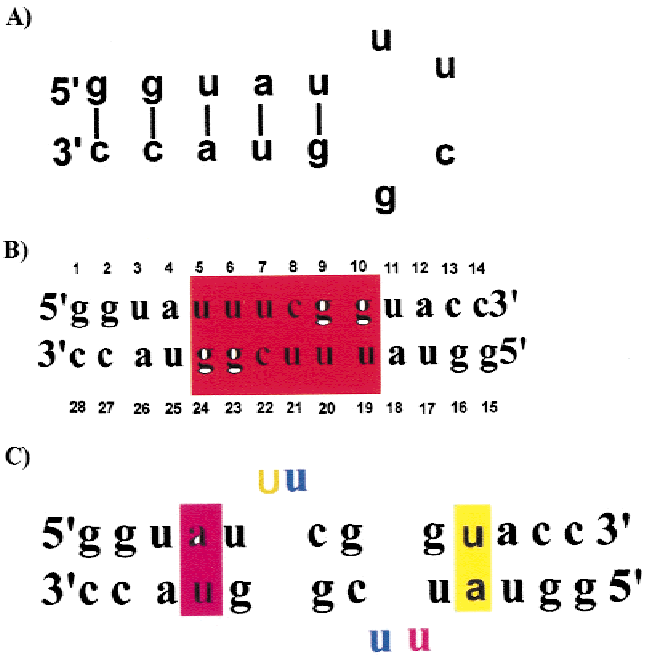


**FIGURE 1.** Schematic topology of the 14mer RNA. **A:** The proposed hairpin loop structure. **B:** Double helix with six central mispairs (red). **C:** The bulge structure. Intramolecular base triples: U6*(A18•U11) and U20*(A4•U25). U7 and U21 are both looped out.

forms a base triple interaction with A27 and U38 in solution (Puglisi et al., 1992). In the solution structure of HIV-2 TAR, which contains a UU dinucleotide bulge, the looped out U23 base was found to be involved in an intramolecular base triple interaction (Brodsky & Williamson, 1997).

The present RNA 14-mer sequence was synthesized with a hairpin with "uucg" in a tetraloop (Fig. 1A). However it is highly possible to form a continuous duplex structure with non-Watson–Crick base pairs (mispairs) in the center of the duplex (Fig. 1B), as have been reported previously (Holbrook et al., 1991). Surprisingly, the actual structure turned out to be with two uridine bases looped out in each strand (Fig. 1C). The bulged uridines in the present structures (U6 and U20) are in a unique conformation and they interact in the major groove, forming novel intramolecular U*(A•U) base triples. This is the first crystal structure of an RNA duplex with two UU-bulges, which helps in understanding the bulged conformations and their effects on the geometry of the RNA duplex and the metal ions and spermine binding.

## RESULTS AND DISCUSSION

### Conformation of the bulge: Novel intramolecular U*(A•U) base triple

The numbering scheme of the RNA duplex is shown in Figure 1B. The molecule adopts the standard A conformation with all the riboses in the C3′-*endo* conformation, except the sugar puckering in the uridine bulges, C2′-*endo*. The Sp-form and Sp-free form have an overall RMSD of 0.26 Å and are identical. All the C4′-C5′ exocyclic torsion angles and the glycosidic torsion angles are *g*+ (*gauche*) and *anti*. A total of 68 and 69 water molecules are observed in the Sp-form and Sp-free form, respectively. Most of the solvent molecules are located in the major groove. Specific interactions by bridging water molecules are found at the bulge sites. Most of the phosphate oxygen atoms are hydrated. The RNA duplexes stacks on top of each other forming a pseudo-continuous infinite column. The base pairs at the junction display ~3.4 Å rise and close to a zero twist angle.

All the U bulges in the two crystal forms are in the looped-out conformations (Fig. 2). One of the U bases in the Sp-form (U21) and both the bases in the Sp-free form (U7 and U21) are disordered. The bulged U bases are *anti* and the C4′-C5′ torsion angles are in the *g*+ conformations. The epsilon torsion angle of U5 is 245°, which is rather wide compared to the most commonly observed values (near *trans*). The alpha torsion angles of U7 and U8 residues are distorted and rotated to *g*+, which deviates from the usual values of *g*− conformations. Thus, the U6 base swings out from helix stacking, protrudes into the major groove, and forms Hoogsteen hydrogen bonds with A18, forming a U6*(A18•U11) base triple with considerable planarity (Fig. 3). The planar base triple is stabilized by the hydrogen bonds between base atoms O2(U6) and N6(A18), and N3(U6), and N7(A18), respectively. The bulged U20 base has a similar conformation and is Hoogsteen hydrogen bonded with the adenine A4, forming the U20*(A4•U25) intramolecular base triple in the major groove.

### Bulge induced perturbations in the backbone

The looped-out UU-residues in both the strands induce local perturbations in the sugar–phosphate backbone conformational angles. The perturbations are predominant on the 5′-flanking residues, and are limited on the 3′-flanking residues. The sugar–phosphate backbone conformation at the bulged nucleotides is stabilized by hydrogen bonds between the 2′ hydroxyl group and a phosphate oxygen atom (O1P) (O2′(U5)–O1P(C8) = 2.6 Å). Also, the sugar moiety of the bulged uracil U7 turns back with its O2′ atom close to the O1P atom of U6 forming a hydrogen bond (O2′(U7)–O1P(U6) = 2.7 Å). These two intramolecular hydrogen bonds stabilize the backbone conformations of the bulge (Fig. 4A). Only one intramolecular hydrogen bond is found in the other bulged region (O2′(U19)–O1P(C22) = 2.6 Å; Fig. 4B), which is the same as observed in both the bulged regions in the Sp-free form. The missing hydro-

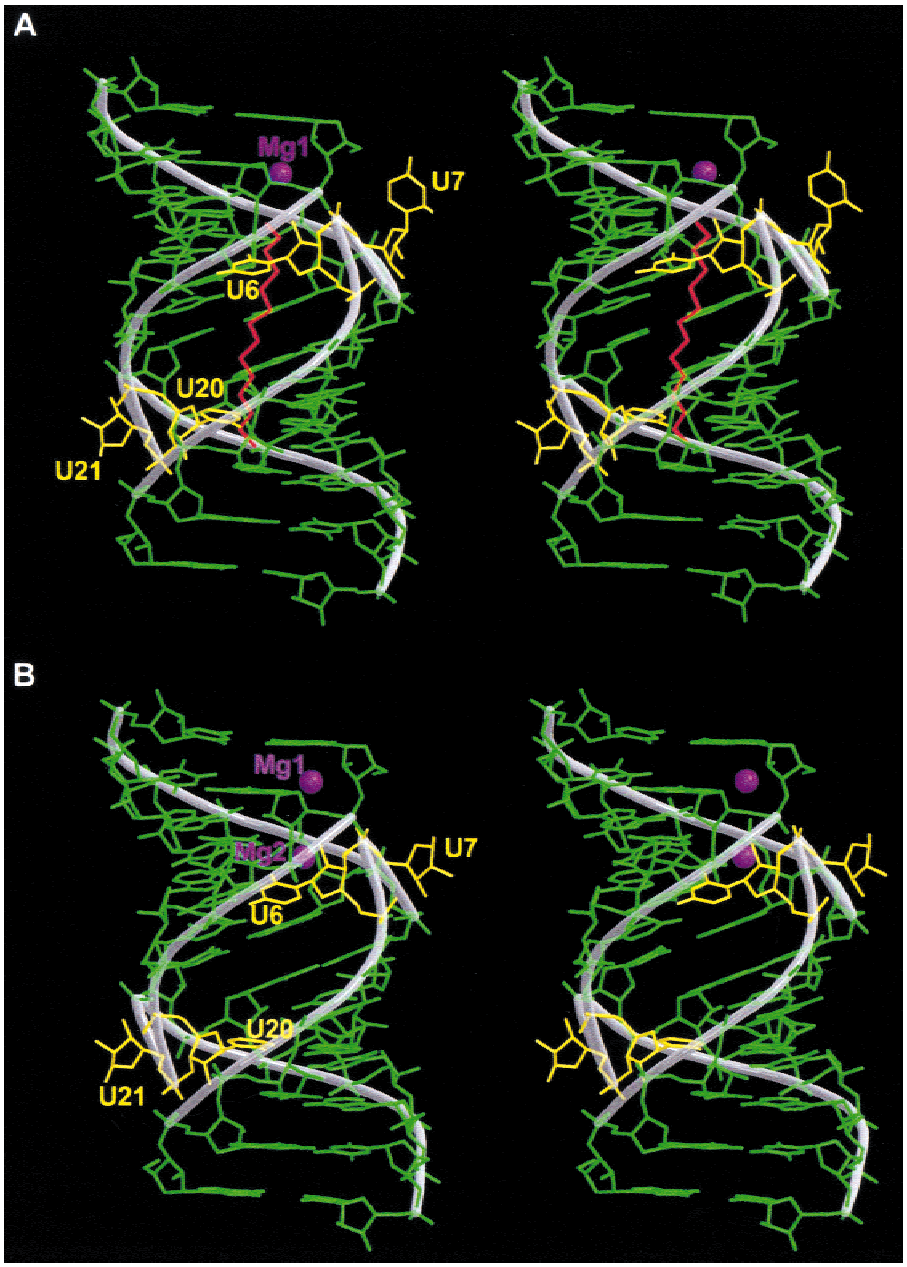


**FIGURE 2.** The overall structure. **A:** Sp-form. The bulged uridines (U6, U7, U20, and U21) are shown in yellow. One spermine molecule (shown in red) was found in the center of the RNA helix. One magnesium ion (shown in purple) was also found at one end of the spermine molecule in the bulged region. **B:** Sp-free form. All the bulges are shown in yellow. Two magnesium ions (purple) are found, one of which (Mg1) is close to the magnesium ion found in the Sp-form.

gen bonding between the O2′ atom of bulged U7(U21) residue and the O1P atom of U6(U19) may explain the disorder of these looped-out uridines (U7 in the Sp-form, both U7 and U21 in the Sp-free form). A similar backbone hydrogen bonding has been observed in the crystal structure of the single adenine bulge (Sudarsanakumar et al., 2000).

The bulged conformation is not affected by spermine and $Mg^{2+}$ binding. The molecule could be divided into three helical segments, the upper and the lower segments of 5 bp each and the middle segment of two G•C base pairs. The kink between the upper and middle segments and between the middle and lower segments are about 30° (Fig. 5). Both kinks are associated with high base pair twists (55°), and they bring the sugar phosphate atoms across the major groove closer to each other. The closest P-P separation across the major groove is 8.0 Å (Fig. 6), corresponding to a major groove width of 2.2 Å, which is profoundly narrower than the 4.7 Å commonly found in the regular A-form.

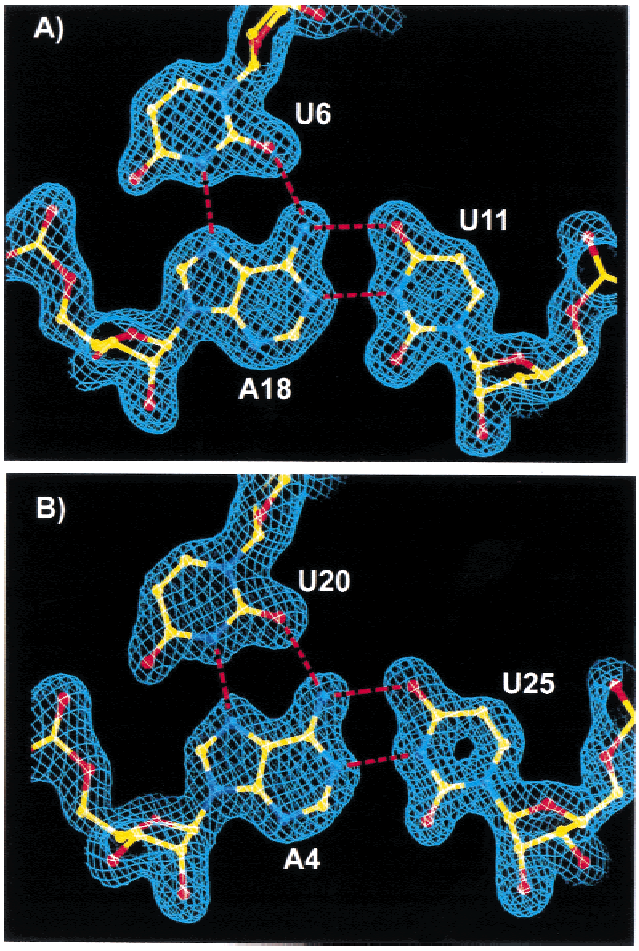


**FIGURE 3.** The two intramolecular U*(A•U) [U6*(A18•U11) and U20*(A4•U25)] base triples in the major groove. The 2Fo-Fc maps are shown in blue at 1.0 $\sigma$ level.

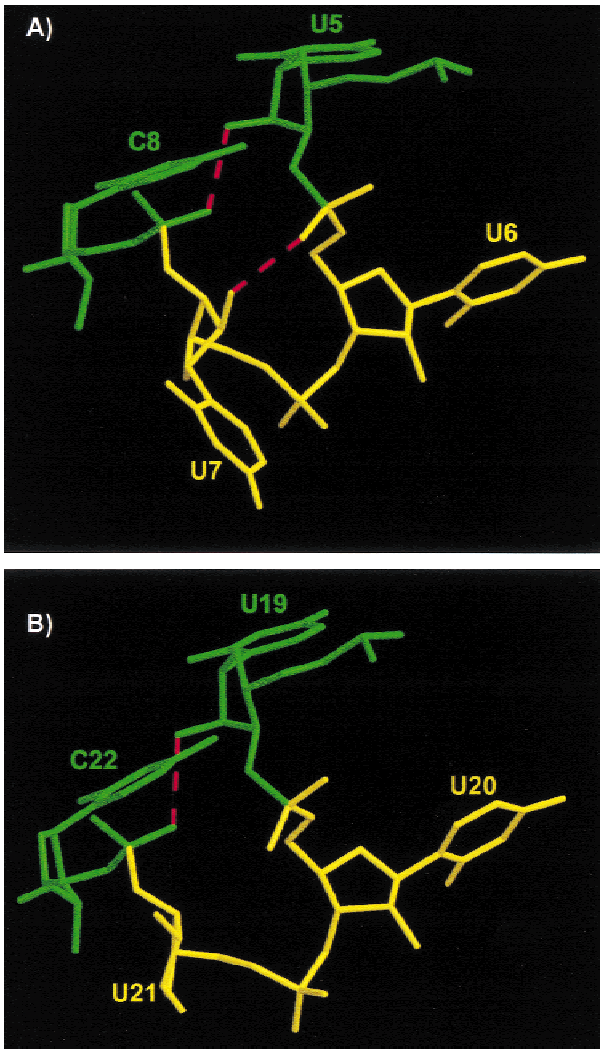


**FIGURE 4.** The bulged regions in the Sp-form. **A:** Looped out U6 and U7 residues (yellow). The backbone in the bulged region was stabilized by two intramolecular hydrogen bondings: O2′(U5)–O1P(C8) = 2.6 Å; O2′(U7)–O1P(U6) = 2.7 Å. **B:** Looped out U20 and U21 residues (yellow). Only one intramolecular hydrogen bonding was found: O2′(U19)–O1P(C22) = 2.6 Å. The U21 base is disordered. Both the bulged regions in the Sp-free form have the same intramolecular hydrogen bond patterns as in **B**.

### Ligand binding: Spermine and $Mg^{2+}$

A spermine molecule (in the Sp-form) is located between the base pairs A4•U25 and U11•A18, and it is extensively hydrogen bonded in the major groove of the duplex. The terminal nitrogen N1 makes a hydrogen bond with O4(U25) and is also connected to an O2P(G24) atom through a bridging water molecule. The other terminal nitrogen, N14, is hydrogen bonded to the O4 atom of U11. The remaining nitrogen atoms, N5 and N10, are involved in bifurcated hydrogen bonds, N5 with O6 atoms of G23 and G24 and N10 with O6 and O4 atoms of G10 and U19, respectively. CH–O interactions are also found between C7 and O6(G23) and C12 and O2 (U6) with an average distance of 3.08 Å.

Superposition of the two structures shows that one of the magnesium ion sites (Mg1) in the two forms is within 1.4 Å distance, while the other (Mg2) in the Sp-free form is not there. Mg2 is within 1.0 Å distance of the C12 atom of the spermine molecule. This position of the spermine eliminates the possibility of the second Mg ion in the Sp-form although it is observed in the other. The two crystal forms reveal that both spermine and Mg ions can contribute to the stabilization of the bulges. The magnesium ion in the Sp-form is coordinated with five water molecules (Fig. 7), three of which are further connected to O1P(U7), O6(G16), and bifurcated hydrogen bonding to O2′ (U6), O4 (U17) atoms. Mg1 in the Sp-free form also has five directly coordinated water molecules, four of which are further associated to U6, G15, and G16 residues. Two of these water molecules form bridged hydrogen bonds to N7(G15) and N7(G16). The third water molecule forms bifurcated hydrogen bonding with O6(G15) and O6(G16). The fourth water of this magnesium is connected to O2′(U6) through a bridging water molecule in the major groove. Only three water molecules are found

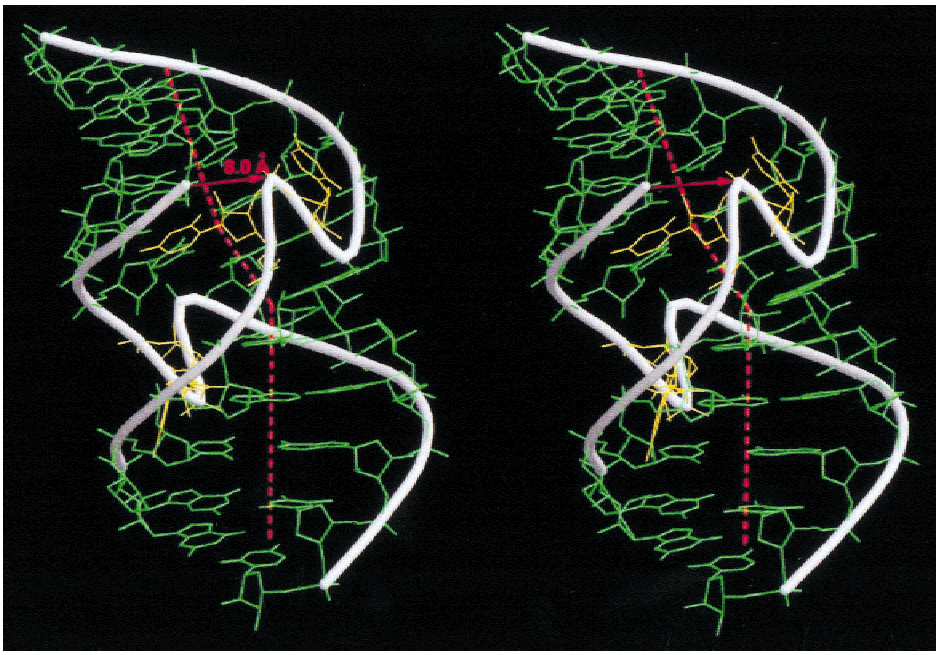


**FIGURE 5.** The kink in the duplex. The RNA duplex is bent, and the major groove is considerably narrowed (2.2 Å).

to be associated with the second magnesium ion in the Sp-free form, all of which are further hydrogen bonded to the RNA atoms O1P(G9), N7(G10), and O2(U6). The nonbridging phosphate groups and N7 atoms of guanines are generally seen as potential sites for cation binding. In this structure, the short distances between phosphate oxygen atoms are observed close to the bulges, but their phosphate oxygen atoms are facing in opposite directions, which does not allow their coordination to metal ions.

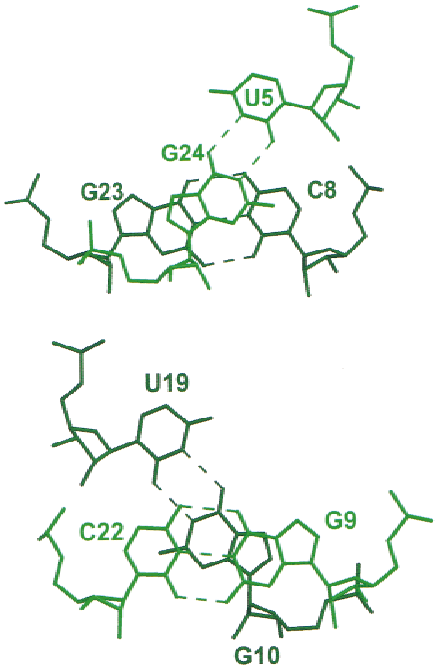


**FIGURE 6.** Significantly high twist angles between the flanking base pairs of the bulged uridines (55°).

## MATERIALS AND METHODS

### Synthesis and crystallization

The 14mer chimer (gguauuucggua$^{Br}$Cc)$_2$ were synthesized by solid-phase phosphoramidite chemistry and cleaved from the solid support using 3:1 (v/v) ammonia/ethanol and incubated overnight in the same solution at 55 °C for deprotection. Crystals were grown in a few days by the hanging drop

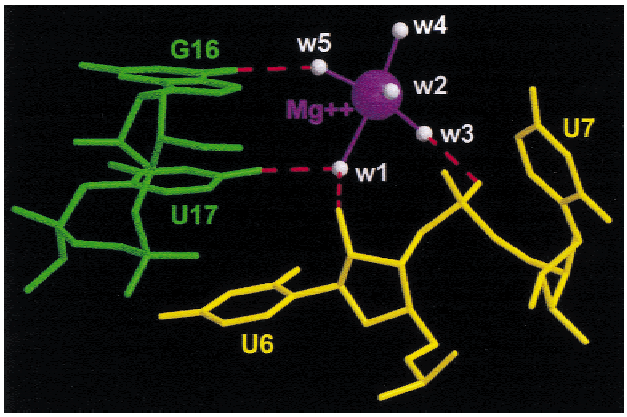


**FIGURE 7.** The geometry of the hydrated $Mg^{2+}$ ion (purple ball) in the Sp-form. Five water molecules (white balls) are found to be coordinated with the $Mg^{2+}$ ion.

vapor diffusion method from a drop containing 1 mM double-stranded RNA, 10 mM spermine tetrachloride, 40 mM sodium cacodylate buffer, pH 7.0, and 10% (v/v) methyl-2,4-pentanediol (MPD), 20 mM magnesium chloride, equilibrated against a reservoir of 1 mL 40% MPD at room temperature (290 K). Crystals of dimension 0.3 mm × 0.3 mm × 0.3 mm were obtained in two weeks.

## Data collection, structure solution, and refinement

### *Sp-form*

A set of three wavelength MAD data were obtained from a single crystal at X-4A beam line at Brookhaven National Laboratory after fluorescence scanning of the same crystal for the determination of the bromine absorption edge wavelength. The intensity data were integrated and scaled with DENZO and SCALEPACK (Otwinowski & Minor, 1997). Resolution data of 1.85 Å were collected at both the absorption edge and at the remote sites of bromine. A total of 16,802 independent reflections were collected to 1.70 Å resolution at the bromine peak site (Freidel pair unmerged), corresponding to 98.6% of the theoretically possible data, with a $R_{merge}$ of 5.2%. The data completeness for the last resolution bin was 88.5%. The crystal belonged to the monoclinic space group C2 with the cell constants of a = 60.06 Å, b = 29.10 Å, c = 52.57 Å, and $\beta = 120.69°$. The asymmetric unit contains one duplex with a volume/bp of 1,410 Å$^3$. We clearly solved the two bromine sites in the asymmetric unit using the SnB program (Weeks & Miller, 1999). We obtained the same results from the anomalous data collected at both the bromine peak and the remote sites.

Phase refinement was successfully done by MLPHARE/CCP4 (Bailey, 1994). Density modifications including non-crystallographic symmetry (NCS) averaging were performed with DM/CCP4 (Cowtan, 1994). The MAD phased electron density map could be clearly traced. The data set collected at the bromine peak site was used in the refinement. After simulated annealing, iterative positional and B-refinement and rebuilding of the model were carried out with CNS (Brunger et al., 1998) and the graphics package O (Jones & Kjeldgaard, 1995). The final $R_{work}/R_{free}$ for the Sp-form is 19.8%/22.7% using 8,549 reflections ranging from 10 Å to 1.7 Å resolution. The final Sp-form model contains 586 RNA atoms, 1 magnesium ion, 1 spermine molecule, and 68 water molecules. The atomic coordinates and the structure factors have been deposited in the Nucleic Acid Database (Berman et al., 1992) with accession code AR0029. Information about the crystal, the MAD phasing, and the structure refinement have been summarized in Table 1.

**TABLE 1**. Crystal data, structure refinement, and MAD phasing statistics.

| | Sp-form | | | Sp-free |
|---|---|---|---|---|
| | Inflection (0.92021 Å) | Peak (0.9196 Å) | Remote (0.90998 Å) | Cu-K$\alpha$ (1.5418 Å) |
| Crystal data | | | | |
| Space group | | C2 | | C2 |
| Cell constants | | | | |
| a (Å) | | 60.06 | | 60.06 |
| b (Å) | | 29.10 | | 29.58 |
| c (Å) | | 52.57 | | 52.50 |
| $\beta$ (°) | | 120.69 | | 120.85 |
| Resolution (Å) | 1.85 | 1.7 | 1.85 | 1.80 |
| Unique reflections | 13,118* | 16,802* | 13,148* | 7,100 |
| Completeness | 100% | 98.6% | 100% | 91% |
| Rsym | 5.1% | 5.2% | 5.3% | 6.0% |
| MAD phasing statistics at 1.85 Å | | | | |
| Phasing power | | | | |
| Isomorphous centric | — | 0.73 | 0.46 | |
| Isomorphous acentric | — | 0.91 | 0.55 | |
| Rcull_a(Rcull_c) | 1.42 (1.0) | 1.0 (1.0) | 0.97 (0.93) | |
| Rcull_ano | 0.76 | 0.75 | 0.81 | |
| FOM | | | | |
| Before DM (density modification) | | 0.65 | | |
| After DM | | 0.74 | | |
| Refinement | | | | |
| Resolution range (Å) | | 10–1.7 | | 10–1.8 |
| Reflection used | | 8,549 | | 6,285 |
| $R/R_{free}$ (%) | | 19.8/22.7 | | 20.8/23.2 |
| RMSD | | | | |
| Bond (Å) | | 0.0092 | | 0.0084 |
| Angle (°) | | 1.91 | | 1.44 |

*Before merging freidel pairs.

*Sp-free*

The unit cell constants are a = 60.06, b = 29.58, c = 52.50 Å, and $\beta = 120.85°$. The crystal diffracted to 1.8 Å resolution and the data were collected at Cu-Ka radiation ($\lambda = 1.5418$ Å) on an in-house Raxis IIc imaging plate system with a Rigaku rotating anode generator. Crystal data were also processed by the DENZO and SCALEPACK programs. The structure was solved by the molecular replacement method using the program AmoRe (Navaza, 1994) with the Sp-form (without spermine) as the search model. The trial structure was then refined by the CNS program to a final $R_{work}/R_{free}$ of 20.8%/23.2%. A total of 69 water molecules and two magnesium ions were found in the structure. The atomic coordinates and the structure factors have been deposited in the Nucleic Acid Database with accession code AR0033. The data and refinement statistics are summarized in Table 1.

## ACKNOWLEDGMENTS

We thank BNL for providing synchrotron beam time and help in data processing. We acknowledge the support of this work by the National Institute of Health Grant GM-17378 and the Board of Regents of Ohio for an Ohio Eminent Scholar Chair and Endowment to M.S. We also acknowledge the Hays Consortium Investment Fund by the Reagents of Ohio for partial support for purchasing the R-axis IIc imaging plate system.

*Received May 10, 2001; returned for revision*
*May 28, 2001; revised manuscript received July 12, 2001*

## REFERENCES

Berman HM, Olson WK, Beveridge DL, Westbrook J, Gelbin A, Demeny T, Hsiesh SH, Srinivasan AR, Schneider B. 1992. The Nucleic Acid Database: A comprehensive relational data base of three-dimensional structure of nucleic acids. *Biophys J 63*: 751–759.

Bailey S. 1994. The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr D50*:760–763.

Brodsky AS, Williamson JR. 1997. Solution structure of the HIV-2 TAR-Argininamide complex. *J Mol Biol 267*:624–639.

Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jinag J, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. 1998. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D54*:905–921.

Cowtan K. 1994. Joint CCP4 and ESF-EACBM. *Newsl Protein Crystallogr 31*:34–38.

Frank DN, Ellington AD, Pace NR. 1996 In vitro selection of RNase P RNA reveals optimized catalytic activity in highly conserved structural domain. *RNA 2*:1179–1188.

Hermann T, Patel DJ. 2000. RNA bulges as architectural and recognition motifs. *Structure 8*:R47–R54.

Holbrook SR, Cheong C, Tinoco I, Kim SH. 1991. Crystal structure of an RNA double helix incorporating a track of non-Watson–Crick base pairs. *Nature 353*:579–581.

Ippolito JA, Steitz TA. 2000. The structure of the HIV-1 RRE high affinity Rev binding site at 1.6 Å resolution. *J Mol Biol 295*:711–717.

Iwai S, Pritchard C, Mann DA, Karn J, Gait MJ. 1992. Recognition of the high affinity binding site in rev-response element RNA by the human immunodeficiency virus type-1 rev protein. *Nucleic Acids Res 20*:6465–6472.

Jones A, Kjeldgaard M. 1995. The O files. http://origo.imsb.au.dk/~mok/o/.

Navaza J. 1994. AMoRe: An automated package for molecular replacement. *Acta Crystallogr D50*:157–163.

Otwinowski Z, Minor W. 1997. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol 276*:307–326.

Puglisi JD, Tan R, Calnan BJ, Frankel AD, Williamson JR. 1992. Conformations of the TAR RNA-arginine complex by NMR spectroscopy. *Science 257*:76–80.

Schmitz M, Tinoco I Jr. 2000. Solution structure and metal-ion binding of the P4 element from bacterial RNase P RNA. *RNA 6*:1212–1225.

Sudarsanakumar C, Xiong Y, Sundaralingam M. 2000. Crystal structure of an adenine bulge in the RNA chain of a DNA•RNA hybrid, d(CTCCTCTTC)•r(gaagagagag). *J Mol Biol 299*:103–112.

Weeks CM, Miller R. 1999. The design and implementation of *SnB* v2.0. *J Appl Cryst 32*:120–124.

Xiong Y, Sundaralingam M. 2000. Two crystal forms of helix II of *Xenopus laevis* 5S rRNA with a cytosine bulge. *RNA 6*:1316–1324.