Solution Structure of the SL1 RNA of the M1 Double-Stranded RNA Virus of *Saccharomyces cerevisiae*

Jun-Seok Yoo,*† Hae-Kap Cheong,* Bong Jin Lee,† Yang-Bae Kim,† and Chaejoon Cheong*
*Magnetic Resonance Team, Korea Basic Science Institute, Taejon 305-333, and †Department of Pharmacy, Seoul National University, Seoul 151-742, Korea

ABSTRACT The 20-nucleotide SL1 VBS RNA, 5′-GGAGACGC[GAUUC]GCGCUCC (bulged A underlined and loop bases in brackets), plays a crucial role in viral particle binding to the plus strand and packaging of the RNA. Its structure was determined by NMR spectroscopy. Structure calculations gave a precisely defined structure, with an average pairwise root mean square deviation (RMSD) of 1.28 Å for the entire molecule, 0.57 Å for the loop region (C8-G14), and 0.46 Å for the bulge region (G4-G7, C15-C17). Base stacking continues for three nucleotides on the 5′ side of the loop. The final structure contains a single hydrogen bond involving the guanine imino proton and the carbonyl O_2 of the cytosine between the nucleotides on the 5′ and 3′ ends of the loop, although they do not form a Watson-Crick base pair. All three pyrimidine bases in the loop point toward the major groove, which implies that Cap-Pol protein may recognize the major groove of the SL1 loop region. The bulged A5 residue is stacked in the stem, but nuclear Overhauser enhancements (NOEs) suggest that A5 spends part of the time in the bulged-out conformation. The rigid conformation of the upper stem and loop regions may allow the SL1 VBS RNA to interact with Cap-Pol protein without drastically changing its own conformation.

INTRODUCTION

Hairpin loops and bulges are important RNA structural motifs that serve as sites for protein recognition and binding (Witherell et al., 1991; Eguchi and Tomizawa, 1991; Jaeger et al., 1994) and for initiating RNA folding (Uhlenbeck, 1990; Woese et al., 1990; Tang and Draper, 1990). NMR spectroscopy has helped to determine the three-dimensional structures of a variety of RNA molecules containing small secondary structural elements, such as hairpin loops, bulges, internal loops, and pseudoknots (for reviews see Moore, 1993; Feigon et al., 1996). Although Watson-Crick base-paired stem regions generally adopt an A-form structure, local perturbations due to sequence or structural context are common. Individual loop conformations in hairpins are highly variable. Some small loops have surprisingly well defined structures (Cheong et al., 1990; Heus and Pardi, 1991; Laing and Hall, 1996), often containing noncanonical base pairs, and base-phosphate, base-sugar, and sugar-phosphate interactions (for reviews see Wyatt and Tinoco, 1993; Shen et al., 1995). Larger loops tend to exhibit poorly defined structural features, due to the presence of multiple conformations or significant dynamics (Jaeger and Tinoco, 1993; Colvin et al., 1993). However, it is not always correct to correlate the size of the loop with the conformational flexibility. The conformations of bulged bases also vary widely. For single bulges in RNA oligonucleotides, the conformation usually depends on the nature of the flanking base pairs (Wimberly et al., 1993; Borer et al., 1995; Greenbaum et al., 1996).

The Cap-Pol fusion protein of *Saccharomyces cerevisiae* virus L1 (ScVL1) recognizes a short sequence of the viral plus strand RNA. This viral binding site (VBS) is responsible for viral particle binding to the plus strand and packaging the RNA (for a review see Wickner, 1996). This region, which is a minimum of 20 bases long, consists of a hairpin with a five-nucleotide loop and a stem interrupted by an internal bulge-loop (Fujimura et al., 1990; Shen and Bruenn, 1993). ScVL1 has a single VBS, whereas ScVM1, a satellite virus of ScVL1, has two VBSs, SL1 and SL2 (Fig. 1) (Shen and Bruenn, 1993). It has been suggested that the structure, not the sequence, of the stem, the sequences of the loop, and the bulged A residue are important for binding the Cap-Pol fusion protein (Shen and Bruenn, 1993; Yao et al., 1997). ScVM1, which is present in some strains and is separately packaged in ScV particles provided by the L1 virus, encodes only two proteins: a secreted protein toxin (the killer toxin) and one conferring immunity to that toxin. The phenotype of this toxin has been used in the genetic analysis of the L1 virus system.

The secondary structures of the VBSs of ScVL1 and ScVM1 (SL1) are similar to that of the coat protein binding site for the single-stranded RNA bacteriophage R17 (Fig. 1 *D*) (Romaniuk et al., 1987; Schneider et al., 1992; Stockley et al., 1995). The crystal structures of the MS2 coat protein-R17 operator complex have been reported (Valegård et al., 1994, 1997; Grahn et al., 1999) as well as the solution structure of the R17 operator fragment (Borer et al., 1995). Extensive biochemical data on the ScV system have been reported (Fujimura et al., 1990; Shen and Bruenn, 1993; Wickner, 1996; Yao et al., 1997). In this study, we report

Received for publication 16 August 2000 and in final form 18 January 2001.

Address reprint requests to Dr. Chaejoon Cheong, Magnetic Resonance Team, Korea Basic Science Institute, 52 Eoun-dong, Yusung-gu, Taejon 305-333, South Korea. Tel.: 82-42-865-3431; Fax: 82-42-865-3419; E-mail: cheong@comp.kbsi.re.kr or Prof. Yang-Bae Kim, Department of Pharmacy, Seoul National University, San 56-1 Shinlim-dong, Kwanak-gu, Seoul 151-742, South Korea. Tel.: 82-2-880-7866; Fax: 82-2-872-3632; E-mail: ybkim@plaza.snu.ac.kr.

0006-3495/01/04/1957/10 $2.00

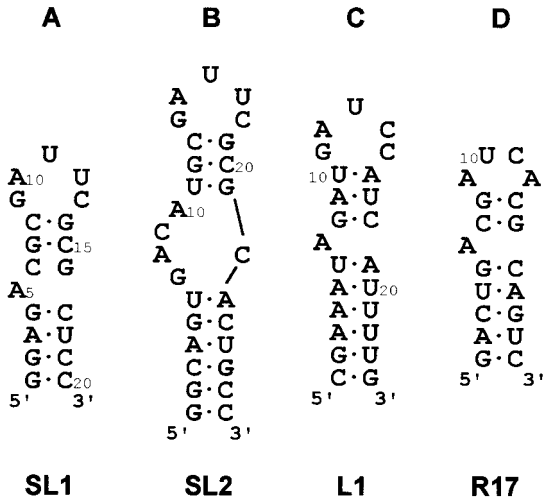

FIGURE 1 Sequences and secondary structures of the ScV VBS's, SL1, SL2, and L1 RNA (*A–C*) and a variant R17 operator fragment (*D*). (*A*) The sequence of the 20-nucleotide RNA used for NMR spectroscopic study was altered from C1-G20 in the wild type to G1-C20 to promote efficient transcription, and the secondary structures determined by NMR are shown. (*D*) A variant of the wild-type sequence studied by NMR is shown.

the conformation of SL1 VBS in solution determined by NMR spectroscopy and restrained molecular dynamics.

MATERIALS AND METHODS

Preparation of RNA samples

The SL1 VBS oligoribonucleotide, 5′-GGAGACGCGAUUCGCGCUCC-3′, was enzymatically synthesized in vitro using T7 RNA polymerase and a chemically synthesized oligodeoxyribonucleotide template (Milligan et al., 1987; Wyatt et al., 1991). The T7 RNA polymerase was obtained from a culture of overexpressed *Escherichia coli* strain BL21 containing the plasmid pAR1219 carrying the gene for the enzyme (Wyatt et al., 1991). The transcription product was purified using denaturing 20% polyacrylamide gel electrophoresis. The purified oligonucleotide was dialyzed extensively against 10 mM sodium phosphate, pH 6.5, 0.01 mM EDTA, using a Centricon-3 concentrator (Amicon). Following dialysis, the sample was lyophilized to dryness. For experiments involving the nonexchangeable protons, the sample was exchanged several times with 99.9% D_2O and then resuspended in 400 μl of 99.96% D_2O. For experiments involving the exchangeable protons, the sample was resuspended in 400 μl of 90% H_2O/10% D_2O. The RNA concentration was 2 mM.

NMR spectroscopy

All experiments were carried out using a Bruker DMX600 NMR spectrometer operating at 600.13 MHz proton frequency. Spectra were acquired at 20°C unless otherwise stated. NMR data were processed and displayed using the programs FELIX (Biosym/MSI) and UXNMR (Bruker). One-dimensional (1D) exchangeable proton spectra were obtained using the jump-return method (Hore, 1983).

All two-dimensional (2D) NMR spectra were recorded in the phase-sensitive mode using the TPPI method (Marion and Wüthrich, 1983). Nuclear Overhauser enhancement spectroscopy (NOESY) spectra of the sample in H_2O were acquired at 5°C, 10°C, and 20°C. Water suppression was achieved using a jump-return pulse with a 150-μs delay between the observed pulses (Sklenar and Bax, 1987). Forty-eight scans were averaged for each free induction decay (FID), with a 2.0-s relaxation delay between scans. Approximately 512 FIDs of 2000 complex points were collected. The mixing times used were 100 and 400 ms. The spectral width in both dimensions was 12,000 Hz. The spectra were zero-filled to 2000 complex points in the t_1 dimension and apodized with a 45–60° phase-shifted squared sine bell function in both dimensions. NOESY spectra of the sample in D_2O were acquired with mixing times of 50, 100, 150, 200, and 400 ms at primarily 20°C, and at 10°C and 30°C to resolve overlapping resonances. A relaxation delay of 1.8 s was used for all experiments.

A double-quantum-filtered COSY (DQF-COSY) spectrum was acquired with a spectral width of 4800 Hz, using the standard pulse sequence (Müller et al., 1986). A high-resolution, phosphorus-decoupled DQF-COSY spectrum was also acquired with a spectral width of 1800 Hz in both dimensions, 2000 complex points in t_2, and 512 points (FIDs) in t_1. Phosphorus was decoupled using the GARP1 decoupling sequence (Shaka et al., 1985) during acquisition. A total correlation spectroscopy (TOCSY) spectrum (Griesinger et al., 1988) was acquired with a compensated MLEV17 sequence for mixing (Bax and Davis, 1985). A mixing time of 100 ms was used.

A natural abundance heteronuclear ^{1}H-^{13}C multiple quantum coherence (HMQC) spectrum was acquired using a standard pulse sequence (Bax et al., 1983; Varani and Tinoco, 1991a) and the GARP1 sequence for carbon decoupling during acquisition. A total of 112 scans were collected for each FID, and 256 FIDs of 2000 complex points were collected. The spectral widths were 4800 Hz (8 ppm) in the ^{1}H dimension and 15,000 Hz (100 ppm) in the ^{13}C dimension. Proton-detected ^{1}H-^{31}P heteronuclear correlation spectroscopy (HETCOR) was acquired as previously proposed (Sklenar et al., 1986). A total of 256 FIDs of 2000 complex points and 16 scans each were recorded. The spectral widths were 1800 Hz (3 ppm) in the ^{1}H dimension and 580 Hz (2.4 ppm) in the ^{31}P dimension.

Interproton distance constraints

Semiquantitative distance constraints between nonexchangeable protons were estimated from cross-peak intensities in 2D NOESY spectra acquired at different mixing times (50, 100, and 150 ms). Using the covalently fixed pyrimidine H5-H6 distance (2.45 Å) as a reference, peak intensities were classified as strong, medium, weak, very weak, and for a few unambiguous cases, absent. The corresponding distances were 1.8−3, 2−4, 3−5, 4−7, and >4.5 Å, respectively. Peaks that were not present in any of the above three NOESY spectra, but were present in the 400-ms NOESY, were classified as very weak. Distance constraints involving exchangeable protons were estimated from 2D NOESY spectra and were classified as either weak or very weak, except for the intra-base-pair distances A·U H2−NH and G·C NH−NH2, which were classified as strong constraints. Inter-base-pair imino to imino cross-peaks were assigned a distance range of 2.0–5.0 Å. Intra-residue sugar-to-sugar constraints were not included in the calculations.

Dihedral angle constraints

Constraints on the ribose ring and backbone dihedral angles were derived from semiquantitative measurements of $^3J_{H\text{-}H}$, $^3J_{H\text{-}P}$ couplings (Williamson and Bax, 1988; Legault et al., 1995). Sugar pucker conformations were determined from $^3J_{H1'\text{-}H2'}$ couplings in the high-resolution DQF-COSY spectrum. Residues for which no H1′-H2′ coupling was observed were constrained to the C3′-*endo* conformation. Residues having 3 Hz $< {}^3J_{H1'\text{-}H2'} <$ 7 Hz were not constrained. Residues with a large (>7 Hz) H1′-H2′ coupling were constrained to the C2′-*endo* conformation. All the glycosidic torsion angles, χ, were constrained to the *anti* conformation ($-150 \pm 40°$) based on the lack of a strong intranucleotide H1′-H8/6 NOE.

Dihedral angle constraints for the β- and ϵ-torsion angles were derived from $^3J_{P\text{-}H5'}$, $^3J_{P\text{-}H5''}$, and $^3J_{P\text{-}H3'}$ couplings measured in the HETCOR spectrum. β was constrained to the *trans* conformation ($180 \pm 40°$) for

residues in which P-H5′ and P-H5″ peaks were clearly absent, representing couplings of <5 Hz (Varani et al., 1996). ϵ was constrained to include both the *trans* and the *gauche*$^-$ conformation ($-130 \pm 60°$) for residues with $^3J_{P\text{-}H3'} > 5$ Hz. Dihedral angle constraints for the γ-torsion angle were derived from $^3J_{H4'\text{-}H5'}$ and $^3J_{H4'\text{-}H5''}$ couplings in the DQF-COSY spectrum. For residues in which H4′-H5′ and H4′-H5″ peaks were clearly absent, representing couplings of <5 Hz, γ was constrained to the *gauche*$^+$ conformation ($55 \pm 30°$) (Varani et al., 1996). No dihedral angle constraints were used for the α- and ζ-torsion angles.

Structure calculations

Structure computation and graphic display were achieved using the InsightII/Discover software package with an AMBER force field (Molecular Simulations). Forty initial A-form structures were generated in InsightII. The structures were first minimized using 2000 cycles of energy minimization; 313 distance constraints and 98 dihedral angle constraints were used. After the minimization, restrained molecular dynamics simulation was initiated at 1000 K in which the NOE and dihedral angle constraints were gradually introduced, with a step size of 1.0 fs for a period of 30 ps. Then, the temperature was gradually lowered to 300 K over 20 ps followed by 2000 cycles of energy minimization. After the simulated annealing, the structures were equilibrated at 300 K for 20 ps with the van der Waals interactions turned on. Energy minimization of 2000 cycles then followed. The procedure produced 22 converged structures.

RESULTS

Assignment of exchangeable protons

The imino and amino proton spectra were assigned by standard methods using one- and two-dimensional NMR spectroscopy in H_2O (Table 1). The SL1 RNA molecule contains seven guanosine and three uridine residues; the 1D imino proton spectrum contains 5 sharp resonances out of the possible 10 (data not shown). The resonances were assigned to the imino protons of the stem. These assignments were confirmed from the imino-imino cross-peaks seen between G2 and U18, G4 and U18, and G7 and G14 in 2D NOESY. The 5′-terminal guanosine G1 is not observed due to the fraying effect. G16 next to the bulge A5 does not show imino proton resonance at 1°C or higher. U18 imino proton was identified by its strong NOE to the H2 proton of base-paired A3. The amino protons of the cytosines in the stem were assigned by NOEs to their own H5 protons and to the cross-strand guanine imino protons.

Assignment of nonexchangeable protons

Pyrimidine H5 and H6 resonances were identified by their strong cross-peaks in the DQF-COSY spectrum. The number of H5-H6 cross-peaks observed in the spectrum is consistent with a single conformation of a molecule that contains ten pyrimidines. Cytosines were distinguished from uridines by the chemical shifts of their C5 carbons in the natural abundance ^{1}H-^{13}C HMQC. This information provides a starting point for the assignment of nonexchangeable protons followed by the standard procedure based on the sequential NOE connectivities and through-bond correlations (Varani and Tinoco, 1991b). Fig. 2 shows the H8/H6/H2-H5/H1′ region of the 400-ms NOESY spectrum. The H8/H6-H1′ connectivities are continuous from G1 to A10 and from U11 to C20. No cross-peak was observed between A10 H1′ and U11 H6 in the 400-ms NOESY.

The H1′-H2′ region of the 50- and 100-ms NOESY and the DQF-COSY spectra was used to assign all the H2′ resonances. Alternate H8/H6-H2′ connectivities confirmed the sequential assignments and resolved some ambiguities in the H8/H6-H1′ connectivities. The H8/H6-H2′ connectivities are continuous for the entire molecule, except for the A10-U11 step. The remaining sugar protons were assigned by identifying the sugar spin systems using high-resolution, ^{31}P-decoupled DQF-COSY. This assignment was confirmed by the sequential H8/H6-H3′ NOE connectivities and the inter- and intra-H1′-sugar proton NOEs. Adenosine H2 resonances were identified by the chemical shifts of the bound carbons in ^{1}H-^{13}C HMQC.

Sequence-specific carbon assignments for the base and C1′ resonances were obtained from the natural abundance ^{1}H-^{13}C HMQC spectrum using the available proton assignments. All three adenine C2s and 20 of the aromatic resonances were unambiguously assigned. All C1′ carbon resonances were assigned, with the exception of G1 and C13 whose H1′ resonances overlapped and were indistinguishable.

The ^{1}H-^{31}P HETCOR determines the backbone sequence along the backbone. This spectrum was assigned using the previously determined H3′ and H5′/H5″ assignments. Strong sequential H3′(i)-^{31}P($i+1$) cross-peaks and weak intranucleotide H5′/H5″-^{31}P cross-peaks are seen at the ^{31}P resonances. These scalar connectivities unambiguously confirmed the nucleotide sequence, especially in the loop region, where structural assumptions cannot be made.

The proton, carbon, and phosphorus assignments are summarized in Table 1.

Structural calculations

The final 22 structures have low NOE violations and total energies. None of the 22 final structures violates any of the experimentally derived NOE constraints by more than 0.1 Å. The average pairwise root mean square deviation (RMSD) is 1.28 Å for the entire molecule. Local structural features, such as loop and bulge regions, are precisely defined by NMR data. Local superposition of all 22 converged structures for the loop (C8 to G14) and the bulge (G4 to G7 and C15 to C17) regions gives average pairwise RMSD values of 0.57 Å and 0.46 Å, respectively. The structure calculation statistics involving distance and dihedral angle constraints are reported in Table 2.

TABLE 1 Proton, carbon, and phosphorus chemical shifts (ppm) for the SL1 VBS RNA in 10 mM sodium phosphate and 10 μM EDTA, pH 6.5

Residue	H8/H6	H2/H5	HI′	H2′	H3′	H4′	H5′/H5″
G1	8.11	na	5.81	4.93	4.69	4.54	4.26/4.41
G2	7.52	na	5.90	4.73	4.57	4.55	4.26/4.47
A3	7.63	7.38	5.91	4.54	4.58	4.50	4.13/4.52
G4	7.28	na	5.68	4.20	4.61	4.47	4.09/4.49
A5	8.25	8.05	6.07	4.65	4.77	4.49	4.16/4.32
C6	7.51	5.31	5.47	4.53	4.40	4.46	4.14/4.32
G7	7.67	na	5.73	4.53	4.53	4.46	4.16/4.44
C8	7.56	5.13	5.50	4.46	4.47	4.35	4.07/4.53
G9	7.45	na	5.67	4.56	4.49	4.45	4.08/4.36
A10	7.92	7.86	5.85	4.56	4.36	4.43	4.04/4.28
U11	7.51	5.46	5.58	4.10	4.46	4.25	3.91/4.05
U12	7.77	5.84	5.97	4.33	4.68	4.52	4.05/4.05
C13	7.83	6.05	5.81	4.62	4.53	4.50	4.19/4.23
G14	7.78	na	5.72	4.57	4.34	4.45	4.10/4.42
C15	7.66	5.20	5.55	4.63	4.49	4.41	4.14/4.54
G16	7.47	na	5.76	4.56	4.41	4.48	4.09/4.49
C17	7.60	5.30	5.39	4.41	4.45	4.40	4.05/4.51
U18	7.98	5.49	5.44	4.44	4.53	4.40	4.05/4.55
C19	7.88	5.62	5.54	4.16	4.46	4.34	4.02/4.53
C20	7.62	5.45	5.71	3.96	4.14	4.12	3.98/4.45

Residue	Imino	Amino	C8/C6	C2/C5	C1′	P
G1			139.2	na	*	na
G2	12.36		137.2	na	93.1	−1.76
A3	na		139.4	153.4	93.7	−1.97
G4	12.99		136.9	na	91.0	−2.43
A5	na		141.7	155.6	90.9	−1.89
C6	na		142.0	98.4	93.8	−1.86
G7	12.77		137.0	na	93.2	−2.22
C8	na	8.27/6.59	140.7	97.7	94.0	−2.52
G9			136.7	na	92.4	−2.04
A10	na		140.7	155.3	91.5	−2.08
U11		na	143.2	104.9	89.8	−2.08
U12		na	143.9	106.0	90.1	−1.56
C13	na		143.0	99.7	*	−2.00
G14	12.73		137.5	na	93.1	−1.63
C15	na	8.37/6.58	140.9	97.7	93.9	−2.51
G16			136.0	na	93.5	−1.91
C17	na	8.59/6.92	141.0	97.3	94.5	−2.84
U18	14.40	na	142.4	103.4	94.0	−2.72
C19	na	8.27/6.91	142.1	97.5	94.5	−2.50
C20	na	8.18/6.97	142.1	98.3	92.9	−2.43

All proton and carbon shifts were referenced to TSP, and phosphorus chemical shifts were referenced to phosphates. Imino and amino resonances were measured at 10°C. All other proton, carbon, and phosphorus resonances were measured at 20°C. na, not applicable.

*The H1's of these nucleotides resonate at the same frequency; hence, it is not possible to assign the C1′ chemical shifts of these residues uniquely. Resonances are seen at 91.4 and 91.8 ppm.

Structural features determined by NMR

NOE connectivities, sugar puckers, and other NMR results reveal a great deal of structural information about the RNA molecule. The schematic structure and the overlap of 11 of the 22 final structures of the SL1 VBS RNA determined by NMR are shown in Figs. 3 and 4, respectively.

Loop region

The loop region of the representative structure is shown in Fig. 5 *A*. The abundance of distance constraints in the loop region precisely defines the loop conformation. NOEs from A10 H2 to U11 H1′ and across to C13 H1′ indicate that the A10 base stacks on top of the G9 base. Moreover, NOEs from A10 H2 to U12 H1′, U12 H5′ and H5″, C13 H5, and to G14 H1′ are also observed. The presence of these NOEs, which play a meaningful role structurally in the loop, is supported by the T1 relaxation property of the A10 H2. This proton has a relatively short T1 of 4.84 ± 0.02 s, which is about the same as most of the H6 and H8 resonances (data not shown), suggesting that A10 H2 has more proton neighbors than other AH2 protons (Szewczak and Moore, 1995).

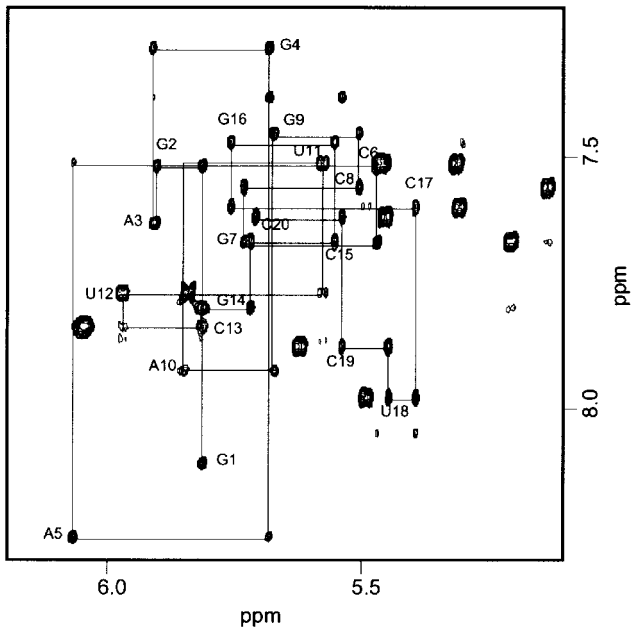

FIGURE 2 Fingerprint (aromatic to anomeric) region of a 400-ms NOESY spectrum of the SL1 VBS RNA. The H8/H6(i)-H1′(i)-H8/H6(i+1) walk is shown in solid lines. The nucleotides near the cross-peak that correspond to the intranucleotide H8/H6-H1′ NOEs have been labeled.

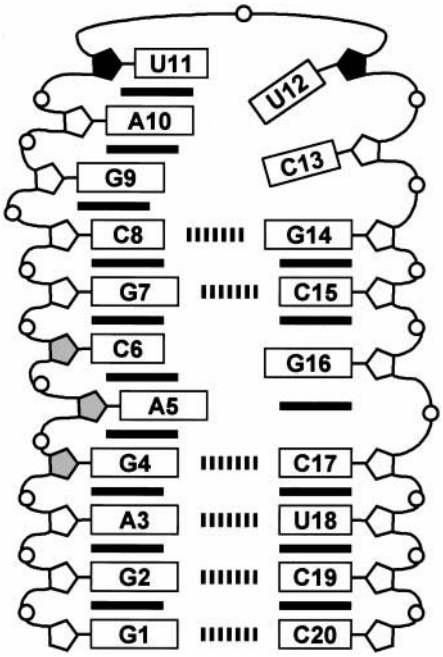

FIGURE 3 Schematic summary of the NMR data for the SL1 VBS RNA. Base pairing and base stacking are represented by hatched bars and solid bars, respectively. Sugars are represented by pentagons, the colors of which indicate the sugar puckers: white for C3′-*endo*, black for C2′-*endo*, and gray for an equilibrium between C2′- and C3′-*endo*. Phosphate groups are represented by small circles.

An unusual inter-residue sugar-to-sugar NOE is observed between C13 H1′ and G14 H3′.

Several nucleotides in the loop have unusual sugar-phosphate backbone conformations. The large H1′-H2′ couplings of U11 and U12 indicate that their ribose sugar rings adopt the C2′-*endo* conformation. As can be seen from Fig. 6, although the γ-angle of U11 has not been constrained in the structure calculation, it adopts the nonstandard *trans* conformation for all the converged structures. The γ-angle of C13 has been also left unconstrained. This angle tends to cluster into the *gauche*$^+$ conformation, although the *trans* conformation exists in a minority. The ϵ-angles of U11 and U12 adopt the *gauche*$^-$ conformation. The α- and ζ-torsion angles, which have been left unconstrained, tend to cluster into a few distinct conformations in the converged structures (Fig. 6). All scalar coupling measurements indicate a standard A-form geometry for the sugar rings and the β-, γ-, and ϵ-backbone angles of residues C8, G9, A10, and G14.

TABLE 2 Summary of experimental distance and dihedral angle constraints

Constraint	
NOE distance constraints	
Intra-residue	122
Inter-residue	185
Mean number per residue	15
NOE constraints by category	
Strong (1.8–3 Å)	17
Medium (2–4 Å)	36
Weak (3–5 Å)	128
Very weak (4–7 Å)	114
Absent (>4.5 Å)	12
Base pair constraints	
Total	36
Dihedral angle constraints	
Ribose ring	14
Backbone	69
Mean number per residue	4.2

At the base of the loop, the bases G9 and C13 do not form a Watson-Crick base pair. However, the G9 imino proton forms a hydrogen bond with C13 O_2 in all 22 final structures. Therefore, base C13 points in the direction of the major groove. The helical base stack continues up the 5′ side of the loop, with A10 stacking on top of G9, which in turn stacks on top of the C8 · G14 base pair. The U11 base partially stacks on top of A10. Residues U12 and C13, however, do not participate in either the 5′- or 3′-base stack. Base U12 clearly points in the direction of the major groove. A break in NOE connectivity was observed between A10 and U11. The two residues are only connected by a weak NOE from A10 H3′ to U11 H6; the NOEs from A10 H1′ and H2′ to U11 H6 are absent. This might be correlated with all nonstandard backbone angles of U11 and reflects the extended conformation of the backbone in the A10−U11 step.

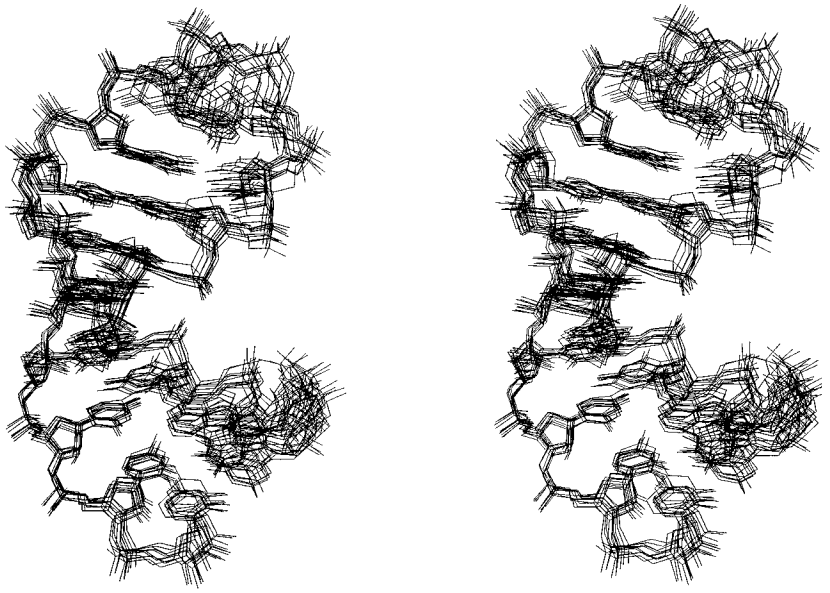

FIGURE 4 Stereoview superposition of 11 of the 22 converged structures of the SL1 VBS RNA. The average pairwise RMSD was 1.28 Å.

Bulge region

The bulge region of the representative structure is shown in Fig. 5 *B*. The bulge A5 base is stacked into the helical stem as is evidenced by the normal stacking pattern for the G4-A5-C6 NOE connectivity and the A5 H2-C17 H1′ cross-peak. However, the G4 H1′-A5 H8 and A5 H1′-C6 H6 NOEs are weak, and some G4-C6 NOEs are observed. The observed NOEs are G4 H1′-C6 H5, G4 H2′-C6 H5, G4 H3′-C6 H5, and G4 H2′-C6 H6. This suggests that A5 spends part of the time in the bulged-out conformation.

Exchangeable resonances in the bulge region, including the G16 NH and the A5 and C6 NH_2, are not observed. This may be due to the line broadening caused by mobility or exchange with solvent. A previous nuclease sensitivity experiment (Shen and Bruenn, 1993) suggested that the C6 and G16 bases form a hydrogen bond at least part of the time. In the study of Borer et al. (1995), the corresponding CG base pair in the variant R17 RNA fragment, which has similar sequence to that of SL1 VBS (see Fig. 1), shows broad imino proton resonance. This base pair is slightly opened to the major groove. In our case, G16 imino proton resonance could not be observed. The structures of the CG pairs of the SL1 VBS RNA and the variant R17 RNA are similar, but that of SL1 VBS is more staggered and does not form a stable Watson-Crick base pair.

The H1′−H2′ coupling constants of G4, A5, and C6 are intermediate between the values characteristic of C3′-*endo* and C2′-*endo* sugar conformations. Sugar puckers of these residues are left unconstrained to encompass the C2′-*endo*, C3′-*endo*, and O4′-*endo* sugar conformations. The dynamic sugars of G4, A5, and C6 might accommodate unusual inter-proton distances in the bulge region or reflect the averaging between stacked and bulged-out conformation of the A5 base. On the strand opposite the bulge, the C15-G16-C17 ribose rings adopt the C3′-*endo* conformation of the standard A-family duplex.

DISCUSSION

The crucial process in the assembly of an L1 and M1 double-stranded RNA virus is the recognition of its plus single-stranded RNA by the Cap-Pol protein (Wickner, 1996). The Pol region of this protein binds RNA hairpins composed of five-nucleotide loops and stems containing a single bulged adenine base (Fujimura et al., 1992; Ribas et al., 1994), although M1 has another binding site, with a three-base bulge in place of the single A bulge (Shen and Bruenn, 1993). The high-resolution structure of the SL1 VBS RNA proposed in this paper provides helpful insight into the pentaloop and single-bulge conformations and the significance of this RNA structure for Cap-Pol protein-RNA complex formation in genomic RNA packaging.

Loop conformation

The nuclease sensitivity mapping of the SL1 VBS shows that a pentaloop with the sequence GAUUC is formed (Shen and Bruenn, 1993). Although the G9 and C13 residues at the bottom of the loop could potentially undergo Watson-Crick base pairing, our structure rules this out. The position of C13 is well defined by the NOEs to nearby protons (see below), and the resulting structure makes formation of a Watson-Crick base pair impossible. However, the structure reveals an unusual interaction between G9 and C13. In all 22 final structures, the imino proton of G9 forms a hydrogen bond with the O_2 of C13 (Fig. 7). The average distance between these atoms is 2.13 ± 0.08 Å. There are enough inter- and intra-residue distance constraints to position the

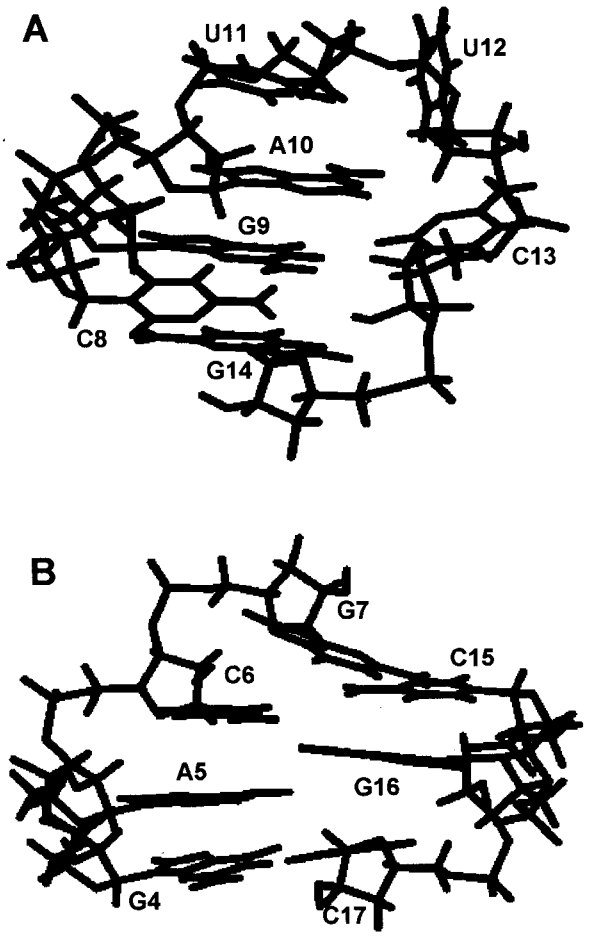

FIGURE 5 Representative structures for the loop (*A*) and bulge (*B*) regions of the SL1 VBS RNA. (*A*) Base stacking of the G9-A10-U11 residues is shown. (*B*) Intrahelical stacking conformation of A5 residue is shown.

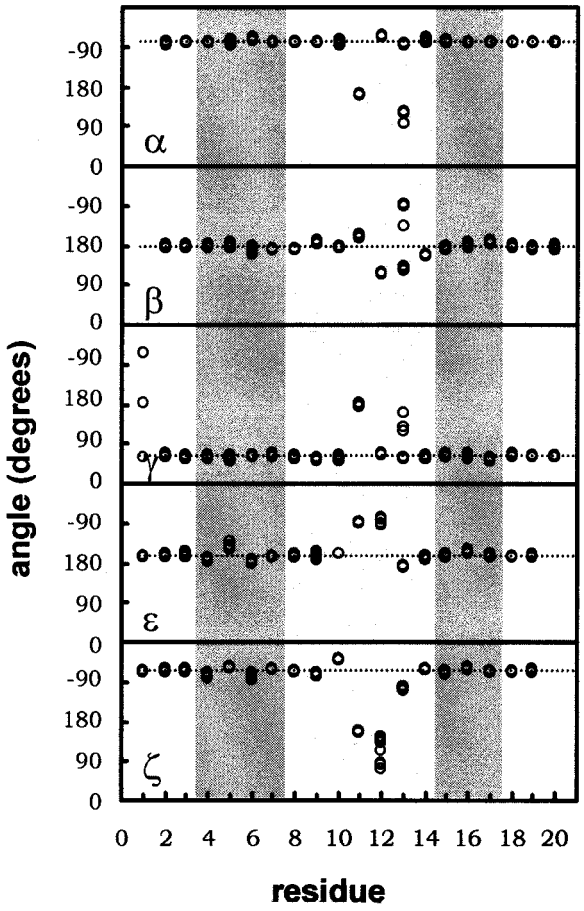

FIGURE 6 Phosphate backbone angles for the 22 converged structures. The loop region is shaded light gray and the bulge region is dark gray. Dotted lines indicate values for standard A-form helices.

G9 and C13 residues precisely. In particular, C13 has 8 intra-residue and 19 inter-residue constraints, which is markedly more than the average number of constraints shown in Table 2. Superposition of the two residues, G9 and C13, gives an average pairwise RMSD value of only 0.33 Å. Although the relatively high precision by which the loop structure in this region is defined adds confidence to the cross-strand hydrogen bond interaction involving G9, no direct NMR evidence has been found to support the formation of this hydrogen bond. The absence of sharp G9 imino resonance indicates that the G9 imino proton is not strongly protected from the solvent; however, this does not rule out the possibility that the G9 imino proton forms a hydrogen bond.

Hairpin motifs with related loop sequences such as 5′-UUUCUGA-3′ (Puglisi et al., 1990), 5′-GCUUUGC-3′ (Davis et al., 1993), and 5′-GGUCUCC-3′ (Grüne et al., 1996) have triloop conformations (loop sequences underlined) closed by U·G, C·G, and G·C base pairs, respectively. Whereas, the hairpin sequence 5′-CGUUUCG-3′ (Sich et al., 1997) assumes a pentaloop conformation like the sequence 5′-CGAUUCG-3′ presented here. The previous results for hairpins containing a UUU sequence (Davis et al., 1993; Sich et al., 1997) emphasize the context dependence and the importance of the closing base pair for the stability of the triloop structure; the UUU triloop can be closed by a C·G, but not by a G·C base pair. In our case, the AUU sequence could not be closed by the G·C base pair either.

According to an in vitro selection experiment, the loop sequences for Cap-Pol protein recognition are completely conserved, except A10 and U12 (Yao et al., 1997). The loop sequences of SL1 and SL2 are the same, and that of L1 differs only at the 4th residue (Fig. 1). The conservation seems to reflect the essential role of the unique pentaloop structure. The pentaloops 5′-CGAUUCG-3′ and 5′-CGUUUCG-3′ have quite different structural features. In the 5′-GAUUC-3′ loop studied here, base stacking extends to the 3rd residue on the 5′ side, whereas the 4th and 5th residues do not stack on the 3′ side. The 3rd and 4th residues adopt C2′-*endo* sugar puckers. Although the 1st

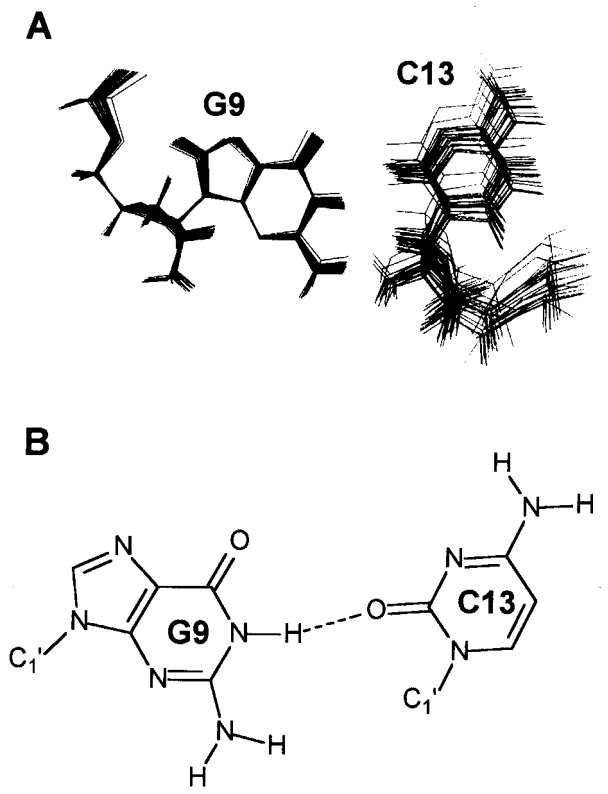

FIGURE 7 Proposed hydrogen bond found in the converged structures generated by restrained molecular dynamics. (*A*) Relative orientation of G9 and C13. The 22 converged structures are superimposed relative to G9. No constraints were used between G9 and C13 in the structural calculation. (*B*) The hydrogen bond formed between G9 H1 and C13 O_2 is indicated by a dashed line and is found in all 22 final structures.

residue, G, and the 5th residue, C, do not form a Watson-Crick base pair, these two bases are connected by a hydrogen bond between G9 H1 and C13 O_2. The 5th residue, C, points somewhat in the direction of the major groove. In the loop with 5′-GUUUC-3′ (Sich et al., 1997), two interconverting conformations with alternate stacking interactions are suggested. The 2nd and 5th residues are markedly exposed to solvent. The 3rd, 4th, and 5th residues adopt C2′-*endo* sugar puckers. The 1st residue, G, and the 5th residue, C, do not form a Watson-Crick base pair, but instead point in the direction of the minor and major grooves, respectively.

The bulge conformation

The bulged A5 base is completely conserved in the consensus ScV VBS determined by in vitro selection (Yao et al., 1997), and A5 is absolutely required for VBS activity in mutagenesis (Fujimura et al., 1990). The results suggest that both the conformation of residue A5 itself, and the position of the bulged base relative to the hairpin loop, are important for protein binding. Previous investigations of single bulges in RNA structures proposed that the conformation adopted is dependent on the nature of the flanking base pairs. An adenine single bulge flanked by Watson-Crick base-paired G and C residues adopts an intrahelical stacked conformation (Borer et al., 1995), but an adenine bulge flanked on one side by a non-Watson-Crick base pair adopts an extrahelical conformation (Greenbaum et al., 1996). The A-bulge motif (A-A)·U found in the RNA-binding site of ribosomal protein S8, the RNA motif of the spliceosome branch-point helix, and the RNA-binding site for phage GA coat protein have different structures. The first, which is flanked by G·C and U·G base pairs, has a structure in which the 5′-adenine base is extrahelical and the 3′-adenine participates in the A·U base pair (Kalurachchi et al., 1997). The second and third, which are both flanked by G·C and U·A base pairs, have a structure in which both adenine bases are intrahelical and the 5′-adenine preferentially forms the A·U base pair (Smith and Nikonowicz, 1998). The bulge region of SL1 VBS has the bulged A5 base flanked by Watson-Crick base-paired G and C residues. Based on previous results, this region, which is identical to that of a variant of the RNA-binding site for phage R17 coat protein (Borer et al., 1995), is expected to have an intrahelical stacked conformation. The solution structure of the bulge region of SL1 VBS presented here confirms this expectation; the bulged A5 base is stacked into the helical stem by the NOE cross-peaks indicating the stacking pattern. However, the NOEs between G4 and C6, as in the case of the variant R17 RNA fragment (Borer et al., 1995), suggest the existence of a bulged-out conformation for part of the time.

Significance of the structure of the SL1 VBS RNA for the Cap-Pol protein-RNA complex

Because interaction with the Cap-Pol protein may alter the conformation of the SL1 VBS RNA, the three-dimensional structure of the SL1 alone does not provide evidence for the interaction mechanism of the fusion protein and the SL1 RNA. However, it allows some insight into how the residues in the RNA molecule may interact with the Cap-Pol protein. This is especially true for the loop region, whose NMR structure is quite well defined. All three pyrimidine bases in the loop point in the direction of the major groove. It is likely that the Cap-Pol protein recognizes the major groove of the SL1 loop region.

It is important to note that the bulged A5 spends part of the time in a bulged-out conformation, although it is stacked in the helix most of the time. Any interpretation of the bulge region in the Cap-Pol protein-RNA interaction should be made with caution. The NMR and co-crystal structures of the phage R17 coat protein binding site, which closely resembles the SL1 VBS RNA in the secondary structure, provide a good insight. The NMR structure of the bulge region of the variant R17 coat protein binding site (Borer et al., 1995) is very similar to that of the SL1 VBS RNA, but

the crystal structure in the presence of the protein has a bulged-out conformation (Valegård et al., 1994). Therefore, we cannot exclude the possibility that the bulged A of the SL1 VBS RNA is also extruded from the stem in the complex with Cap-Pol protein. In the loop and upper-stem region other than the bulge region, a notable difference between the SL1 VBS RNA and the R17 coat protein binding site is observed. The variation in the loop sugar conformations of the R17 coat protein binding site and the fact that the two upper-stem iminos are more open to solvent exchange suggest that the hairpin segment of the R17 exhibits fast conformational fluctuations. On the other hand, SL1 VBS RNA has a rigid loop conformation, and the two upper-stem iminos exhibit sharp resonances. Because it is the length of the upper stem of SL1 that is important for protein recognition, not the sequence, the rigid upper stem may play a role in positioning the bulged base at the correct site relative to the hairpin loop. This suggests that the extent of conformational change that an RNA molecule undergoes upon binding to the protein might be quite different in SL1 VBS and R17. SL1 VBS RNA likely interacts with the Cap-Pol protein without drastically changing its own conformation in the loop region. The affinity of the R17 coat protein binding site for the coat protein is one to two orders of magnitude less than that of the VBS for Cap-Pol protein (Romaniuk et al., 1987; Yao et al., 1997). This difference may be related to the structural features of the RNA molecules themselves. The three-dimensional structure of the Cap-Pol protein (or peptide)-SL1 VBS RNA complex remains to be resolved to answer this question precisely.

We thank Prof. I. Tinoco, Jr., for valuable advice.
This work was supported by an international research collaboration grant (99-I-01–03-A-043) from the Ministry of Science and Technology, the Republic of Korea.

REFERENCES

Bax, A., and D. G. Davis. 1985. MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy. *J. Magn. Reson.* 65: 355–360.

Bax, A., R. H. Griffey, and B. L. Hawkins. 1983. Correlation of proton and nitrogen-15 chemical shifts by multiple quantum NMR. *J. Magn. Reson.* 55:301–315.

Borer, P. N., Y. Lin, S. Wang, M. W. Roggenbuck, J. M. Gott, O. C. Uhlenbeck, and I. Pelczer. 1995. Proton NMR and structural features of a 24-nucleotide RNA hairpin. *Biochemistry.* 34:6488–6503.

Cheong, C., G. Varani, and I. Tinoco, Jr. 1990. Solution structure of an unusually stable RNA hairpin, 5′GGAC(UUCG)GUCC. *Nature.* 346: 680–682.

Colvin, R. A., S. W. White, M. A. Garcia-Blanco, and D. W. Hoffman. 1993. Structural features of an RNA containing the CUGGGA loop of the human immunodeficiency virus type 1 trans-activation response element. *Biochemistry.* 32:1105–1112.

Davis, P. W., W. Thurmes, and I. Tinoco, Jr. 1993. Structure of a small RNA hairpin. *Nucleic Acids Res.* 21:537–545.

Eguchi, Y., and J. Tomizawa. 1991. Complexes formed by complementary RNA stem-loops: their formations, structures and interaction with ColE1 Rom protein. *J. Mol. Biol.* 220:831–842.

Feigon, J., T. Dieckmann, and F. W. Smith. 1996. Aptamer structures from A to zeta. *Chem. Biol.* 3:611–617.

Fujimura, T., R. Esteban, L. M. Esteban, and R. B. Wickner. 1990. Portable encapsidation signal of the L-A double-stranded RNA virus of S. cerevisiae. *Cell.* 62:819–828.

Fujimura, T., J. C. Ribas, A. M. Makhov, and R. B. Wickner. 1992. Pol of gag-pol fusion protein required for encapsidation of viral RNA of yeast L-A virus. *Nature.* 359:746–749.

Grahn, E., N. J. Stonehouse, J. B. Murray, S. van den Worm, K. Valegård, K. Fridborg, P. G. Stockley, and L. Liljas. 1999. Crystallographic studies of RNA hairpins in complexes with recombinant MS2 capsids: implications for binding requirements. *RNA.* 5:131–138.

Greenbaum, N. L., I. Radhakrishnan, D. J. Patel, and D. Hirsh. 1996. Solution structure of the donor site of a trans-splicing RNA. *Structure.* 4:725–733.

Griesinger, C., G. Otting, K. Wüthrich, and R. R. Ernst. 1988. Clean TOCSY for ^{1}H spin system identification in macromolecules. *J. Am. Chem. Soc.* 110:7870–7872.

Grüne, M., M. Görlach, V. Soskic, S. Klussmann, R. Bald, J. P. Fürste, V. A. Erdmann, and L. R. Brown. 1996. Initial analysis of 750 MHz NMR spectra of selective ^{15}N-G, U labelled *E. coli* 5S rRNA. *FEBS Lett.* 385:114–118.

Heus, H. A., and A. Pardi. 1991. Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. *Science.* 253:191–194.

Hore, P. J. 1983. Solvent suppression in Fourier transform nuclear magnetic resonance. *J. Magn. Reson.* 55:283–300.

Jaeger, L., F. Michel, and E. Westhof. 1994. Involvement of a GNRA tetraloop in long-range RNA tertiary interactions. *J. Mol. Biol.* 236: 1271–1276.

Jaeger, J. A., and I. Tinoco, Jr. 1993. An NMR study of the HIV-1 TAR element hairpin. *Biochemistry.* 32:12522–12530.

Kalurachchi, K., K. Uma, R. A. Zimmermann, and E. P. Nikonowicz. 1997. Structural features of the binding site for ribosomal protein S8 in *Escherichia coli* 16S rRNA defined using NMR spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.* 94:2139–2144.

Laing, L. G., and K. B. Hall. 1996. A model of the iron responsive element RNA hairpin loop structure determined from NMR and thermodynamic data. *Biochemistry.* 35:13586–13596.

Legault, P., F. M. Jucker, and A. Pardi. 1995. Improved measurement of ^{13}C, ^{31}P J coupling constants in isotopically labeled RNA. *FEBS Lett.* 362:156–160.

Marion, D., and K. Wüthrich. 1983. Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for measurements of ^{1}H-^{1}H spin-spin coupling constants in proteins. *Biochem. Biophys. Res. Commun.* 113:967–974.

Milligan, J. F., D. R. Groebe, G. W. Witherell, and O. C. Uhlenbeck. 1987. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res.* 15:8783–8798.

Moore, P. B. 1993. Recent RNA structures. *Curr. Opin. Struct. Biol.* 3:340–344.

Müller, N., R. R. Ernst, and K. Wüthrich. 1986. Multiple-quantum-filtered two-dimensional correlated NMR spectroscopy of proteins. *J. Am. Chem. Soc.* 108:6482–6492.

Puglisi, J. D., J. R. Wyatt, and I. Tinoco, Jr. 1990. Solution conformation of an RNA hairpin loop. *Biochemistry.* 29:4215–4226.

Ribas, J. C., T. Fujimura, and R. B. Wickner. 1994. Essential RNA binding and packaging domains of the Gag-Pol fusion protein of the L-A double-stranded RNA virus of *Saccharomyces cerevisiae. J. Biol. Chem.* 269:28420–28428.

Romaniuk, P. J., P. Lowary, H. N. Wu, G. Stormo, and O. C. Uhlenbeck. 1987. RNA binding site of R17 coat protein. *Biochemistry.* 26: 1563–1568.

Schneider, D., C. Tuerk, and L. Gold. 1992. Selection of high affinity RNA ligands to the bacteriophage R17 coat protein. *J. Mol. Biol.* 228: 862–869.

Shaka, A. J., P. B. Barker, and R. Freeman. 1985. Computer-optimized decoupling scheme for wideband applications and low-level operation. *J. Magn. Reson.* 64:547–552.

Shen, Y., and J. A. Bruenn. 1993. RNA structural requirements for RNA binding, replication, and packaging in the yeast double-stranded RNA virus. *Virology.* 195:481–491.

Shen, L. X., Z. Cai, and I. Tinoco, Jr. 1995. RNA structure at high resolution. *FASEB J.* 9:1023–1033.

Sich, C., O. Ohlenschlager, R. Ramachandran, M. Gorlach, and L. R. Brown. 1997. Structure of an RNA hairpin loop with a 5′-CGUUUCG-3′ loop motif by heteronuclear NMR spectroscopy and distance geometry. *Biochemistry.* 36:13989–14002.

Sklenar, V., and A. Bax. 1987. A new water suppression technique for generating pure-phase spectra with equal excitation over a wide bandwidth. *J. Magn. Reson.* 75:378–383.

Sklenar, V., H. Miyashiro, G. Zon, H. T. Miles, and A. Bax. 1986. Assignment of the ^{31}P and ^{1}H resonances in oligonucleotides by two-dimensional NMR spectroscopy. *FEBS Lett.* 208:94–98.

Smith, J. S., and E. P. Nikonowicz. 1998. NMR structure and dynamics of an RNA motif common to the spliceosome branch-point helix and the RNA-binding site for phage GA coat protein. *Biochemistry.* 37: 13486–13498.

Stockley, P. G., N. J. Stonehouse, J. B. Murray, S. T. Goodman, S. J. Talbot, C. J. Adams, L. Liljas, and K. Valegård. 1995. Probing sequence-specific RNA recognition by the bacteriophage MS2 coat protein. *Nucleic Acids Res.* 23:2512–2518.

Szewczak, A. A., and P. B. Moore. 1995. The sarcin/ricin loop, a modular RNA. *J. Mol. Biol.* 247:81–98.

Tang, R. S., and D. E. Draper. 1990. Bulge loops used to measure the helical twist of RNA in solution. *Biochemistry.* 29:5232–5237.

Uhlenbeck, O. C. 1990. Tetraloops and RNA folding. *Nature.* 346: 613–614.

Valegård, K., J. B. Murray, N. J. Stonehouse, S. van den Worm, P. G. Stockley, and L. Liljas. 1997. The three-dimensional structures of two complexes between recombinant MS2 capsids and RNA operator fragments reveal sequence-specific protein-RNA interactions. *J. Mol. Biol.* 270:724–738.

Valegård, K., J. B. Murray, P. G. Stockley, N. J. Stonehouse, and L. Liljas. 1994. Crystal structure of an RNA bacteriophage coat protein-operator complex. *Nature.* 371:623–626.

Varani, G., F. Aboul-ela, and F. H.-T. Allain. 1996. NMR investigation of RNA structure. *Prog. NMR Spectrosc.* 29:51–127.

Varani, G., and I. Tinoco, Jr. 1991a. Carbon assignments and heteronuclear coupling constants for an RNA oligonucleotide from natural abundance ^{13}C-^{1}H correlated experiments. *J. Am. Chem. Soc.* 113:9349–9354.

Varani, G., and I. Tinoco, Jr. 1991b. RNA structure and NMR spectroscopy. *Q. Rev. Biophys.* 24:479–532.

Wickner, R. B. 1996. Double-stranded RNA viruses of *Saccharomyces cerevisiae. Microbiol. Rev.* 60:250–265.

Williamson, D., and A. Bax. 1988. Resolution-enhanced correlation of ^{1}H and ^{31}P chemical shifts. *J. Magn. Reson.* 76:174–177.

Wimberly, B., G. Varani, and I. Tinoco, Jr. 1993. The conformation of loop E of eukaryotic 5S ribosomal RNA. *Biochemistry.* 32:1078–1087.

Witherell, G. W., J. M. Gott, and O. C. Uhlenbeck. 1991. Specific interaction between RNA phage coat proteins and RNA. *Prog. Nucleic Acid Res. Mol. Biol.* 40:185–220.

Woese, C. R., S. Winker, and R. R. Gutell. 1990. Architecture of ribosomal RNA: constraints on the sequence of "tetra-loops." *Proc. Natl. Acad. Sci. U.S.A.* 87:8467–8471.

Wyatt, J. R., M. Chastain, and J. D. Puglisi. 1991. Synthesis and purification of large amounts of RNA oligonucleotides. *Biotechniques.* 11: 764–769.

Wyatt, J. R., and I. Tinoco, Jr. 1993. RNA structural elements and RNA function. *In* The RNA World. R. F. Gesteland, and J. F. Atkins, editors. Cold Spring Harbor Laboratory Press, New York. 465–496.

Yao, W., K. Adelman, and J. A. Bruenn. 1997. In vitro selection of packaging sites in a double-stranded RNA virus. *J. Virol.* 71:2157–2162.