Extending the Family of UNCG-like Tetraloop Motifs: NMR Structure of a CACG Tetraloop from Coxsackievirus B3†

Zhihua Du,‡ Jinghua Yu,‡ Raul Andino,§ and Thomas L. James*,‡

Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143-0446, and Department of Microbiology and Immunology, University of California, San Francisco, California 94143-0414

Received December 5, 2002; Revised Manuscript Received February 20, 2003

ABSTRACT: Stable RNA tetraloop motifs are found frequently in biologically active RNAs. These motifs carry out a wide variety of functions in RNA folding, in RNA−RNA and RNA−protein interactions. A great deal of knowledge about the structures and functions of tetraloop motifs has accumulated largely due to intensive theoretical, biochemical, and biophysical studies on three most frequently occurring families of tetraloop sequences, namely, the cUNCGg, the cGNRAg, and the gCUUGc sequences. Our knowledge surely is not exhaustive, and efforts are still being made to gain a better understanding. Here we report the NMR structure of a uCACGg tetraloop that occurs naturally within the cloverleaf RNA structure of the 5′-UTR of coxsackievirus B3. This tetraloop is the major determinant for interaction between the cloverleaf RNA and viral 3C protease, which is an essential part of a ribonucleoprotein complex that plays a critical role in the regulation of viral translation and replication. Our structure shows that the CACG tetraloop is closed by a wobble U·G base pair. The structure of the CACG tetraloop is stabilized by extensive base stacking and hydrogen bonding interactions strikingly similar to those previously reported for the cUUCGg tetraloop. Identification of these hallmark structural features strongly supports the existence of an extended YNCG tetraloop family. The U·G base pair closing the stem and the A residue in the loop introduce some small structural and themodynamic distinctions from the canonical cUUCGg tetraloop that may be important for recognition by the viral 3C protease.

It has long been known that tetraloop structural motifs occur frequently in biologically active RNAs (*1*, *2*). Tetraloops carry out a variety of important functions, such as directing correct folding of RNAs with complicated structures (*3*, *4*), mediating RNA−RNA tertiary interactions (*5*−*9*), and serving as a recognition site for RNA−protein interactions (*10*, *11*). So far, three major families of tetraloops have been recognized, based primarily on their frequent appearance: the cUNCGg (N can be any type of nucleotide), the cGNRAg (R is a purine), and the gCUUGc tetraloop families (lowercase letters represent the closing base pairs, which are very conserved phylogenetically along with the tetraloop sequences, in uppercase letters). Subsequent studies showed that these three families of tetraloop-closing base pair combinations are thermodynamically very stable (*12*−*15*).

High-resolution structures of several tetraloops are now available as determined by NMR[1] or X-ray crystallography: UNCG tetraloops (*16*−*20*), GNRA tetraloops (*21*−*27*), and CUUG tetraloops (*28*). While sequences and structural details vary, some common structural features have emerged. It appears that all three tetraloop families are stabilized by a dense network of base stacking and hydrogen bonding interactions. In the NMR structure of a gCUUGc motif, it was shown that the second loop nucleotide interacts directly with the closing base pair of the stem by folding into the minor groove (*28*), providing a good explanation for the phylogenetically conserved g·c closing base pair.

Given the large number of possible sequence variations for tetraloop-closing base pair combinations, a number of questions regarding stable tetraloop motifs remain to be answered. Do other stable tetraloop-closing base pair combinations exist that do not conform, in sequence or in structure, to the three previously recognized families? Can the three tetraloop families tolerate certain sequence deviations without losing their key structural features that define them as distinctive structural motifs? If sequence deviations are tolerated in terms of maintaining structural integrity, how do they affect the thermostability and flexibility of the tetraloop motifs? Are changes in the thermostability and flexibility of the tetraloop RNAs induced by sequence variations related to biological functions? While high stability is a characteristic feature common in the three established classes of tetraloop motifs, it is not difficult to envisage that in certain cases reduced thermostability and enhanced

† This work was largely supported by the National Institutes of Health.

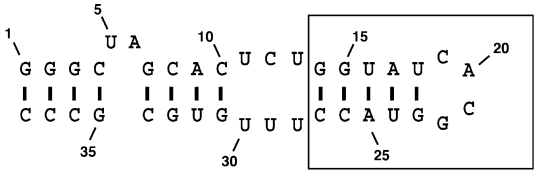
* To whom correspondence should be addressed. E-mail: james@picasso.ucsf.edu. Telephone: (415) 476-1916. Fax: (415) 502-4690.

‡ Department of Pharmaceutical Chemistry.

§ Department of Microbiology and Immunology.

[1] Abbreviations: 5′-UTR, 5′-untranslated region; CVB3, coxsackievirus B3; PV1, poliovirus type 1; HRV14, human rhinovirus type 14; 3C, protease 3C from picornavirus; 3D, RNA-dependent RNA polymerase from picornavirus; 3CD, fusion of 3C and 3D proteins; NMR, nuclear magnetic resonance; FID, free induction decay; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser effect spectroscopy; DQF-COSY, double-quantum-filtered correlation spectroscopy; TOCSY, total correlation spectroscopy; HMQC, heteronuclear multiple-quantum coherence; UV, ultraviolet.

10.1021/bi027314e CCC: $25.00
Published on Web 03/25/2003



CVB3 stem-loop D

FIGURE 1: Thirty-eight-nucleotide RNA investigated by NMR in this study. This sequence corresponds to the entire stem−loop D domain within the 5′-UTR cloverleaf RNA of coxsackievirus B3, with two additional G·C base pairs (G1·C38 and G2·C37) added for in vitro transcription by T7 RNA polymerase. High-resolution structure determination was carried out for the 14-nucleotide hairpin within the rectangle (from residue G14 to C27; numbering is arbitrary starting from the first residue of the 38-nucleotide construct). Another 38-nucleotide RNA that differs from the CVB3 construct by only one nucleotide (U19 vs C19 in CVB3) was also studied. This corresponds to the most consensus stem−loop D RNA sequence in enteroviruses and rhinoviruses.

flexibility (presumably induced by sequence variations) would actually be required for formation of tertiary RNA−RNA interactions or RNA−protein recognition. A very recently published study has advanced our knowledge along these directions (*29*).

To deal with the large possibility of sequence variations, Proctor and co-workers used in vitro selection to identify stable tetraloop motifs from a six-nucleotide RNA hairpin loop library (*29*). Besides the established cUNCGg and gCUUGc motifs (the GNRA motif was not found presumably due to inefficient PCR), previously unknown motifs cUNAGg, cCNAGg, and cCNCGg were found to be almost as stable as the cUNCGg motif. Furthermore, structural similarity between the newly identified motifs and the established cUNCGg motif was inferred from similarities in circular dichroism and NMR spectra. A phylogenetic analysis of rRNA sequences revealed that many of the newly identified motifs are at least as common as the gCUUGc motif. The authors therefore proposed the existence of an extended family of cUNCGg-like tetraloop motifs with a sequence of cYNMGg (Y is either a C or a U, N is any type of nucleotide, and M is either a C or an A). Although this proposal was strongly supported by the data presented, no high-resolution structure of any of the newly identified sequences was determined. In the CACG and CAAG sequences, NMR analysis was complicated by the presence of a duplex (dimer) conformation (*29*).

Here we report the NMR structure of a 14-nucleotide RNA hairpin with a CACG tetraloop sequence and an uncommon U·G closing base pair. This RNA sequence occurs as the apical D-loop within the so-called stem−loop D domain of the cloverleaf RNA found at the 5′-untranslated region (5′-UTR) of coxsackievirus B3 (CVB3) genomic RNA (Figure 1). This D-loop RNA is the most important determinant for a regulatory interaction between the stem−loop RNA D domain and viral protein 3C (the viral protease) (*30*). A corresponding D-loop with the uUGCGg sequence in poliovirus type 1 (PV1) was also found to be important for interaction between the PV1 stem−loop D RNA and the cognate viral protein 3CD (fusion of viral protease 3C and the viral RNA-dependent RNA polymerase 3D) (*31*).

The structure of the 14-nucleotide RNA CVB3 hairpin, determined in the context of the entire stem−loop D domain (Figure 1), shows that although uCACGg differs from the cUNCGg sequences at the first nucleotide of the loop and the closing base pair, its structure is nonetheless strikingly similar to those of the cUNCGg tetraloops. Our structure therefore not only provides unequivocal and timely support for the extended YNMG tetraloop motif but also indicates that the identity of the closing base pair is not restricted to a Watson−Crick C·G pair as far as structural integrity is concerned. The uCACGg tetraloop structural details also suggest features that are important for its recognition by the viral 3C protease.

MATERIALS AND METHODS

RNA Synthesis and Purification. The 38-nucleotide RNA molecule with a sequence corresponding to the entire stem−loop D domain of CVB3 (Figure 1) was transcribed using T7 RNA polymerase and a synthetic DNA template, essentially as described by Milligan et al. (*32*). The DNA template consisted of a double-stranded 18 bp T7 promoter sequence and a single-stranded coding sequence. T7 RNA polymerase was overexpressed in *Escherichia coli* strain pAR1219 and purified by using an S-Sepharose column followed by a DEAE column. The transcription buffer typically contained 80 mM Tris (pH 8.1), 10 mM DTT, 1 mM spermidine, 300 nM template DNA, and various amounts of $MgCl_2$, NTPs, and T7 RNA polymerase for optimum yields. The transcription reaction mixture was incubated for 4 h at 37 °C. The RNAs were precipitated with 3 volumes of ethanol and dissolved in 8 M urea. The desired products were separated on 20% (w/v) 19:1 polyacryamide gels containing 8 M urea. The RNA band was visualized by UV shadowing and excised from the gel, and the RNA was extracted by electroelution using an Elutrap apparatus from Schleicher & Schuell. The RNA was further subjected to repeated ethanol precipitation and then passed through a Sephadex G25 gel filtration column. The fractions containing RNA were combined and lyophilized. The purified RNAs were finally resuspended in NMR buffer that typically contains 25 mM sodium phosphate (pH 6.0) and 25 mM sodium chloride. The RNA solution was heated to 90 °C and slowly cooled to room temperature before being transferred to the NMR tube. Final RNA concentrations were 1−2 mM. Unlabeled and doubly ^{13}C- and ^{15}N-labeled RNA samples for another construct that differs from the CVB3 sequence only at one position (U19 vs C19) were also prepared in the same way as outlined above. This RNA corresponds to the consensus sequence found in enteroviruses and rhinoviruses.

NMR Spectroscopy. All NMR experiments were performed on Varian Inova spectrometers operating at 600 MHz for protons. Spectra were processed with NMRPipe/NMRDraw (*33*) and analyzed with SPARKY (Goddard and Kneller, UCSF). The two-dimensional (2D) and three-dimensional (3D) spectra utilized the States−TPPI method (*34*) for quadrature detection. Homonuclear 2D NOE spectra in a 90% H_2O/10% D_2O buffer were recorded at 5 or 10 °C using SSnoesy (*35*). The spectral width was 12 500 Hz in both dimensions; the mixing time was 150 ms, and the relaxation delay was 1.5 s, with 600 FIDs of 2048 complex data points being collected. Homonuclear 2D NOE, DQF-COSY, and TOCSY spectra in 100% D_2O were recorded at 30 °C, with low power presaturation for suppression of the residual HDO

signal. The spectral width was set to 6000 Hz in both dimensions. Typically, COSY and TOCSY spectra were recorded with 512 FIDs of 2048 complex data points, and a relaxation delay of 1.5 s. A 60 ms clean-MLEV-17 isotropic mixing sequence was used for TOCSY. 2D NOE spectra were recorded with a relaxation delay of 3.0 s and mixing times of 50, 100, 200, and 300 ms, with 600 or 700 FIDs of 2048 complex data points being recorded. 2D HMQC experiments, with very low power water presaturation during the relaxation delay, were performed to obtain the natural-abundance one-bond $^{1}H-^{13}C$ correlation of the unlabeled RNAs (CVB3 construct). Spectral widths were 5500 and 3500 Hz for the ^{1}H dimension and ^{13}C dimension, respectively, and the ^{13}C carrier frequency was set at 120 ppm. The one-bond $^{1}H-^{13}C$ coupling constant was set to 165 Hz as a compromise between aromatic and ribose moieties. Extensive heteronuclear NMR spectra were acquired with the doubly labeled consensus RNA.

UV Melting Studies. UV absorbance melting studies were performed using a Cary 3E UV−visible spectrophotometer with a heating rate of 1.0 °C/min. The RNA concentration was varied from 3 to 300 μM.

NMR-Derived Restraints. Distance restraints between nonexchangeable protons, obtained from 2D NOE cross-peak intensities (50, 100, 200, and 300 ms mixing times) recorded in D_2O buffer, were assigned to four categories: very strong (1.8−3.0 Å), strong (1.8−4.0 Å), medium (1.8−5.0 Å), and weak (1.8−6.0 Å). NOE cross-peaks appearing only in the 300 ms mixing time spectrum were classified as very weak with an upper bound of 7 Å. Distance restraints involving exchangeable protons were assigned to three conservative categories with upper bounds of 3.5 (strong), 5.0 (medium), and 6.0 Å (weak). Hydrogen bond restraints were imposed on base pairs established by the observation of imino proton resonances and internucleotide NOEs characteristic of base pair formation. These restraints were based on the standard base pair geometry of nucleic acids and given a range of ±0.2 Å. Dihedral angle restraints for the ribose sugar conformation were based on analysis of the 2D DQF-COSY spectra. Sugars showing a strong H1′−H2′ cross-peak were restrained to C2′-endo. Those with no COSY and TOCSY cross-peaks were interpreted as C3′-endo, with the exception of G23, which was left unconstrained. Although the glycosidic torsion angle (χ) of G22 is established as being in a *syn* conformation on the basis of NMR data, it is not constrained. Instead, distance constraints between the H1′ and H8 protons of G22 were given an upper bound of 2.6 Å due to the extremely intense NOE, sufficient to keep the glycosidic torsion angle in a *syn* conformation. The backbone dihedral angles α, β, δ, ϵ, and ζ for all of the tetraloop residues and the U18·G23 closing base pair were not restrained. Loose A-form backbone dihedral angles constrains (±10°) were applied to residues in the stem regions (G14−A17 and U24−C27).

Structure Calculations. Structure calculation and refinement were performed using DYANA 1.5 (*36*) and CNS 1.1 (*37*). Using DYANA, 200 random initial structures were generated and subjected to simulated annealing in torsion angle space, followed by variable target function minimization. Twenty DYANA structures with the lowest target function values and correct chirality [checked by the program CHIRANO by N. Ulyanov, University of California, San Francisco (UCSF)] were chosen for further refinement via restrained molecular dynamics using CNS. The 10 structures with the lowest total energy comprise the final structure ensemble. All molecular graphics were generated using MidasPlus (*38*) or Chimera (*39*) from the Computer Graphics Laboratory (UCSF).

RESULTS

Choice of Sequence for Structure Determination by NMR. The structure of the 14-nucleotide CVB3 apical uCACGg tetraloop was determined in the context of a longer (38-nucleotide) NMR construct corresponding to the entire stem−loop D domain in CVB3 (Figure 1). Our long-term goal is to study the structure and interaction with 3C/3CD protein of the stem−loop D RNA domain of enteroviruses and rhinoviruses. Consequently, the 38-nucleotide sequence shown in Figure 1 was chosen. In view of the recently published observation that short sequences (12-nucleotides) containing cCACGg and cCAAGg sequences exist as a mixture of monomeric and dimeric duplex forms in solution (*29*), we decided to study the D-loop tetraloop motif structure within a longer NMR construct. In this context, the 38-nucleotide CVB3 stem−loop D RNA is mostly monomeric in solution as judged by nondenaturing electrophoresis, UV melting, and NMR spectroscopy.

Besides the CVB3 constructs, another 38-nucleotide RNA with a uUACGg D-loop sequence was also made. This construct, which differs from the CVB3 construct only at the first nucleotide of the D-loop, represents the phylogenetically most conserved stem−loop D sequence found in enteroviruses and rhinoviruses (*30*) (we refer to this sequence as the consensus construct). The CVB3 and consensus constructs both yielded excellent NMR spectra. A doubly ^{13}C- and ^{15}N-labeled sample was prepared for the consensus construct, and nearly complete ^{1}H, ^{13}C, and ^{15}N chemical shift assignments were obtained. While the structure of the entire stem−loop D domain will be reported elsewhere at a later time on the consensus sequence, we focus our presentation here on the structure of the apical uCACGg tetraloop of the CVB3 sequence. Two sets of high-quality NMR spectra of closely related sequences has greatly facilitated the assignment processes by resolving spectral overlaps and cross-validating assignments.

Assignment of Exchangeable Proton Resonances. Assignment of exchangeable proton resonances was started from analysis of the imino−imino proton region of a 2D NOE spectrum recorded at 5 °C in a 90% H_2O/10% D_2O buffer (Figure 2). Except G1 and G2, all of the G and U residues predicted to form a standard Watson−Crick G·C or A·U base pair in the helical stem yield observable imino proton resonances and sequential NOE cross-peaks in a region from ∼12.4 to ∼13.6 ppm (within the typical chemical shift range for imino protons involved in standard Watson−Crick-type hydrogen bonding). Seven upfield-shifted imino proton resonances were also observed between ∼9.8 and ∼11.8 ppm, indicating that there are seven imino protons involved in non-Watson−Crick-type hydrogen bonding. These imino resonances were assigned to U30, U13, G22, U28, U11, U18, and G23 (from upfield to downfield). Strong cross-peaks between two pairs of U residues indicate formation of two U·U base pairs (U11·U30 and U13·U28) in the pyrimidine-

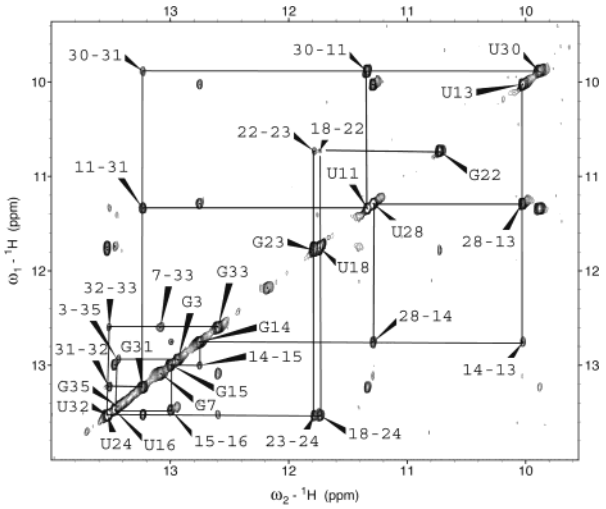


FIGURE 2: Section of the 2D NOE spectrum (150 ms mixing time) showing NOE cross-peaks between imino protons of the 38-nucleotide CVB3 stem−loop D RNA in 90% H_2O/10% D_2O solvent and 25 mM sodium chloride, 25 mM sodium phosphate buffer (pH 6.0) at 5 °C. Three sequential NOE walks were established: from G22 to the U13·U28 base pair, from G7 to the U11·U30 base pair, and between G35 and G3. Of particular interest for the CACG tetraloop structure are NOEs arising among G22 H1, G23 H1, and U18 H3. These NOEs indicate the existence of a U18·G23 closing base pair and the involvement of the G22 H1 proton in a stable non-Watson-Crick-type hydrogen bond.

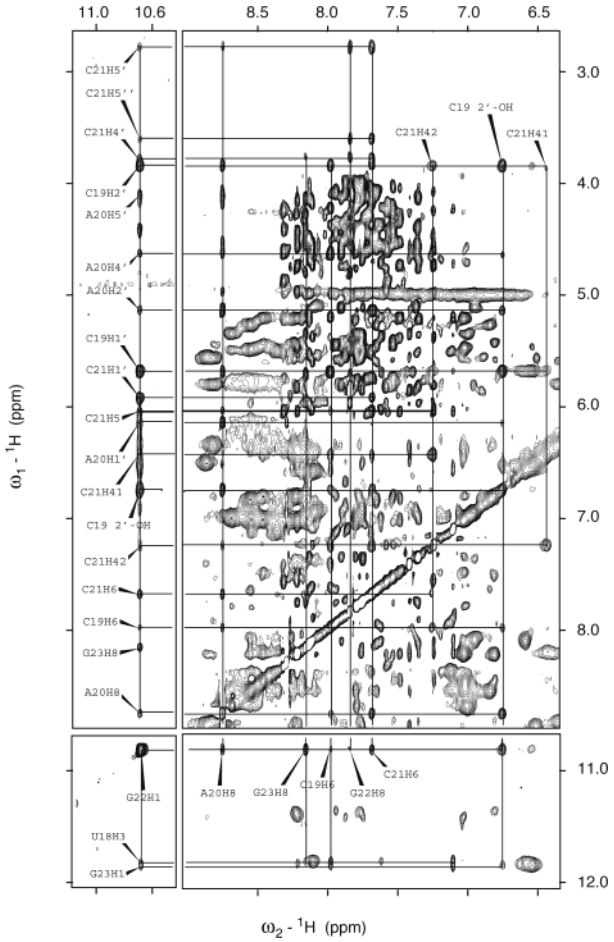


FIGURE 3: Dense network of NOEs from residues in the tetraloop region observed in other sections of the same 2D NOE spectrum as that in Figure 2. NOEs are indicated by crossings of the horizontal and vertical grid lines, which represent chemical shift values of the resonances. Assignments for the horizontal lines are indicated by the column of atom names near the left edge of the figure, with arrows pointing to the right. Assignments for the vertical lines are indicated in two places. The three exchangeable protons (C19 2′-OH and C21 H41 and H42) are indicated at the upper-right corner with arrows pointing downward. The five aromatic H6 and H8 protons (C19 H6, A20 H8, C21 H6, G22 H8, and G23 H8), as well as the imino proton of G22 (G22 H1), are indicated at the bottom-left corner with arrows pointing upward. The observation of G22 H1, C19 2′-OH, and C21 H41 and H42 resonances indicates that these protons are involved in stable hydrogen bonds. Note that a large number of inter-residue NOEs are detected from G22 H1 and C19 2′-OH to various partners in the tetraloop, suggesting that hydrogen bonds involving these two protons play a central role in the structure of the tetraloop.

rich symmetric mismatched region of the stem. Chemical shifts of U18 and G23 imino resonances are too close to each other, preventing unambiguous identification of a cross-peak between them. However, in the consensus construct, U18 and G23 imino resonances are well-separated, so a strong cross-peak was observed between them, indicating the existence of a U18·G23 wobble base pair.

Assignment of the imino proton resonances proves the formation of the predicted base pairs in this construct. Of particular interest to the apical D-loop structure is the observation of slowly exchanging imino resonances from U18, G23, and G22. While observation of U18 and G23 imino resonances can be explained by the formation of a U18·G23 wobble base pair (later confirmed by structure calculation), a Watson−Crick C19·G22 base pair is unlikely to be responsible for observation of the G22 imino resonance. First of all, the G22 imino resonance is shifted way upfield from other imino resonances involved in standard Watson−Crick base pairs. Second, unlike other imino resonances involved in standard Watson−Crick base pairs, the G22 imino resonance gives no strong cross-peaks to a pair of amino resonances. These NMR data clearly indicate that G22 does not form a Watson−Crick base pair with C19 in the CVB3 D-loop structure. This is in contrast to the structure of a gCUUGc tetraloop in which the first and last loop nucleotides form a standard Watson−Crick C·G base pair (*28*).

With the imino resonances assigned, it was straightforward to assign cytosine amino proton resonances on the basis of their strong NOEs to the base-paired guanidine imino protons. No irregularity was seen for residues in the stem region. What drew our special attention was the observation of a number of slowly exchanging resonances between ∼6.3 and ∼7.3 ppm (Figure 3). On the basis of similarities to previous studies on the cUUCGg tetraloop (*18*), we assigned the resonance at 6.72 ppm to the 2′-hydroxyl (2′-OH) proton of C19, and we assigned a pair of resonances at 6.40 and 7.21 ppm to the amino (H41 and H42, respectively) protons of C21. As noted previously (*18*), the small chemical shift difference between C21 H41 and C21 H42 resonances is a good indicator of their involvement in novel (non-Watson−Crick) hydrogen bonding. Assignment of the C19 2′-OH was primarily based on its very intense NOEs to H1′ and H2′ protons of C19 (Figure 3). A number of other NOEs to H6 of C19, H1′, H2′, H4′, and H8 of A20, H6 of C21, and imino protons of G22 and G23 were also observed. It is relatively rare to detect 2′-hydroxyl protons in NMR due to their rapid exchange with solvent. The presence of a C19 2′-hydroxyl resonance and the many associated NOEs suggest that this

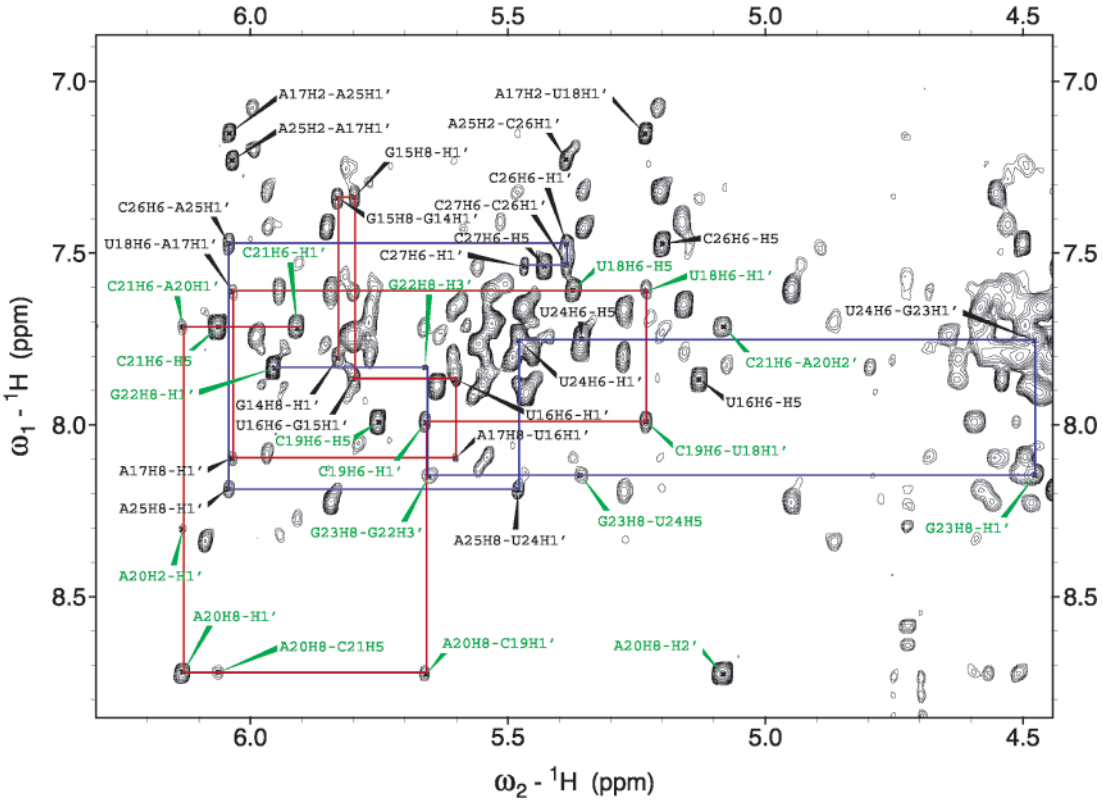


FIGURE 4: Section of a 2D NOE spectrum (200 ms mixing time) of the 38-nucleotide CVB3 stem−loop D RNA in D_2O solvent and 25 mM sodium chloride, 25 mM sodium phosphate buffer (pH 6.0) at 30 °C. Although very extensive assignments were obtained for the whole molecule, only cross-peaks arising from the apical D-loop hairpin (residues G14−C27) are labeled. Two sequential NOE walks are clearly established: one from G14 to C21 (red) and the other from G23 to C27 (blue). NOE cross-peaks from the CACG tetraloop and the U18·G23 closing base pair are labeled in green. Note the very unusual chemical shifts of G23 H1′ (4.47 ppm) and G22 H3′ (5.65 ppm). The exceptional spectral dispersion, especially for resonances from the tetraloop region, enabled the assignment of a large number of NOEs without ambiguity.

hydroxyl group is involved in a stable hydrogen bond that is of key importance to the tetraloop structure.

Assignment of Nonexchangeable Proton Resonances. Assignment of nonexchangeable proton resonances was greatly facilitated by the excellent spectral dispersion despite the relatively large size (by NMR standard) of the molecule. Figure 4 shows the base (H2/H6/H8) to H1′ and H5 region of a 200 ms mixing time 2D NOE spectrum recorded at 30 °C in 100% D_2O buffer. Two H1′−H6/H8 NOE sequential walks are shown for residues within the apical D-loop (from G14 to C27), one from G14 to C21 and the other from G23 to C27. No NOE between C21 and G22 was observed in this region. A NOE between G22 H3′ and G23 H8 was detected thanks to the very unusual chemical shift of the G22 H3′ resonance (5.65 ppm). Assisted by assignments obtained for the consensus construct in which heteronuclear NMR data were available, we were able to assign almost all of the nonexchangeable protons in the tetraloop hairpin (Table 1 of the Supporting Information).

Besides G22 H3′, many other protons yielded unusually shifted resonances. These include C19 H2′ (3.84 ppm), A20 H2′ (5.09 ppm) and H8 (8.72 ppm), C21 H4′ (3.78 ppm) and H5′/H5″ (3.62/2.84 ppm, assignments are not stereospecific), and G23 H1′ (4.47 ppm). Some of these resonances shifted to regions typically reserved for other classes of protons. Real identities of these protons were revealed by the chemical shifts of the attached ^{13}C obtained in a 2D natural-abundance ^{13}C HMQC experiment. All of these unusual chemical shifts are associated with protons belonging to residues in the tetraloop and closing base pair. They neatly match those unusual chemical shifts we observed in the uUACGg tetraloop of the consensus construct, and those previously reported for a gUUCGc tetraloop (*16*−*18*) and a gUACGg tetraloop (*20*). Since unusual chemical shifts are very sensitive indicators for unusual local structure, observation of a similar set of unusual chemical shifts in each one of the uCACGg, uUACGg, gUUCGc, and gUACGc tetraloops suggests that these tetraloops should be structurally very similar.

Glycosidic Torsion Angles and Sugar Conformations. In all of the 2D NOE spectra with different mixing times (50−300 ms), the intraresidue H1′−H8 NOE cross-peak from G22 has a stronger intensity than any of the pyrimidine H5−H6 NOE cross-peaks, indicating that the glycosidic torsion angle of G22 is in the *syn* conformation. All other glycosidic torsion angles are in the *anti* conformation.

Within the apical D-loop, only A20 and C21 showed strong H1′−H2′ COSY cross-peaks, indicating that the sugar puckers of A20 and C21 adopt predominantly the C2′-*endo* conformation. No other residue in the D-loop hairpin (G14−C27) yielded an observable H1′−H2′ COSY cross-peak. Overlapping of G23 H1′ and H2′ proton resonances made it impossible to determine the sugar conformation of this residue on the basis of analysis of COSY and TOCSY spectra. Other than that, the absence of H1′−H2′ COSY and TOCSY cross-peaks indicates predominantly C3′-*endo* sugar conformations. For residue G22, a strong COSY cross-peak was unambiguously identified between H3′ and H4′ due to

Table 1: Structure Determination Statistics for the uCACGg Tetraloop

no. of NOE distance constraints	
total	279
intranucleotide	104
internucleotide sequential	127
internucleotide nonsequential	48
no. of NOE constraints by category	
very strong (1.8−3.0 Å)	12
strong (1.8−4.0 Å)	42
medium (1.8−5.0 Å)	123
weak (1.8−6.0 Å)	91
very weak (1.8−7.0 Å)	11
no. of mean NOE constraints per nucleotide	
all (G14−C27)	20
tetraloop region (U18−G23)	27.5
no. of hydrogen bond constraints	30
no. of dihedral angle constraints	42
no. of violations	
NOE violation >0.2 Å	0
angle violations >5 °C	0
mean deviation from covalent geometry	
bond lengths (Å)	0.0013 ± 0.0005
bond angles (deg)	0.580 ± 0.007
impropers (deg)	0.311 ± 0.015
rmsd from mean structure (for heavy atoms)	
all (G14−C27)	0.84 ± 0.38
tetraloop (U18−G23)	0.55 ± 0.23

the unusual chemical shift of the H3′ proton at 5.68 ppm. The presence of this H3′−H4′ COSY peak and the absence of a H1′−H2′ COSY peak make it particularly unequivocal that the sugar conformation of G22 is C3′-*endo*.

The glycosidic torsion angles and sugar conformations we observed in our uCACGg tetraloop are virtually identical to those observed in the cUUCGg tetraloop (*16*−*20*) comparing residue by residue.

Structure of the uCACGg Tetraloop. The solution structure of the uCACGg tetraloop hairpin was calculated from NMR-derived distance and dihedral angle constraints using the protocol described in Materials and Methods. Statistics for the structure calculation and refinement are summarized in Table 1. Thanks to the exceptional quality of the NMR spectra (see Figures 2−4), several NOEs were identified, especially in the tetraloop region (Table 1). These lead to very well-defined structures. During DYANA calculations, generic hydrogen bond distance restraints were applied only to the canonical Watson−Crick G·C and A·U base pairs in the helical stem. No such restraints were used for the U18·G23 base pair or for the tetraloop. Analysis of the converged best-scoring DYANA structures revealed that the orientation and distances of the hydrogen bond-forming groups of U18 and G23 are consistent with the wobble pair formation in all structures. Therefore, hydrogen bond distance restraints for wobble pair formation and dihedral restraints for base pair planarity were introduced during subsequent CNS refinement. Protection of the C19 2′-hydroxyl proton (2′-OH), the C21 amino protons (H41 and H42), and the G22 imino proton (H1) from rapid exchange (Figure 3) indicates that these protons are involved in stable hydrogen bonds. The hydrogen bond acceptor group for each of these protons was clearly identified by the initial DYANA calculation, and hydrogen bonding distance restraints were consequently introduced during CNS refinement. Observation of a strong NOE cross-peak between C19 2′-OH and G22 H1 (Figure 3) suggests that G22 O6 is a good candidate for the acceptor of C19 2′-OH. Analysis of the DYANA structures confirms that G22 O6 is indeed the only reasonable choice. The distance between C19 O2′ and G22 O6 is no more than 2.90 Å in all 20 of the best-scoring DYANA structures. Analysis of DYANA structures also identified C19 O2 as the only reasonable acceptor for G22 H1; the distance between C19 O2 and G22 N1 is no more than 3.16 Å (average distance of 2.86 Å) in all 20 of the best-scoring DYANA structures. Since the distance between C19 O2 and G22 N2 is also reasonably short (average distance of 3.18 Å), it is not impossible that C19 O2 is shared by G22 imino and amino groups for hydrogen bond formation, as seen in the previously reported UUCG tetraloop structures (*18*, *19*). Although the pro-R phosphate oxygen of A20 (A20 OP2) is slightly outside the hydrogen bonding distance range of the C21 amino group (∼3.6 Å between A20 OP2 and C21 N4), it seems to be the only possibility. The other two groups within 5 Å of C21 N4 are O2 and N3 of C19, but they are unlikely to be the acceptor group because base stacking between C21 and C19 is clearly seen in all initial DYANA structures.

One structure calculation was carried out with hydrogen bonding restraints involving C19 2′-OH, C21 H41/H42, and G22 H1 removed from the final restraint set. The average structure of the ensemble calculated without these hydrogen bond restraints is very close to the average structure of the ensemble calculated with the hydrogen bond restraints (pairwise rmsd of 0.22 Å).

Panels A and B of Figure 5 show the superimposition of a set of 10 structures with lowest energy from CNS and the minimized average structure of the CACG tetraloop hairpin. Views of the tetraloop and the closing U18·G23 base pair region are shown in panels A and B of Figure 6. In panels C and D of Figure 6, the structure of the uCACGg tetraloop is superimposed with the crystal structure of a cUUCGg tetraloop (*19*) and the NMR structure of a cUACGg tetraloop (*20*), respectively.

Clearly, a U18·G23 wobble base pair is formed at the base of the tetraloop. The possibility of a reverse wobble U18·G23 base pair is basically ruled out by the absence of NOE cross-peaks between the G23 imino proton and the aromatic (H5 and H6) protons of U18. Although a U·G wobble base pair is not commonly found in stable tetraloop-closing base pair combinations, the presence of such a base pair as the closing base pair in the CVB3 CACG tetraloop does not seem to prevent the CACG tetraloop from adopting a stable tetraloop-like structure.

The first (C19) and last (G22) nucleotides of the tetraloop form an unusual C19·G22 base pair, with C19 in an *anti* and G22 in a *syn* glycosidic conformation. The 2′-hydroxyl group of C19 is involved in a hydrogen bond with O6 of G22, and the imino group of G22 makes another hydrogen bond with O2 of C19. Formation of these two hydrogen bonds satisfactorily explains the observation of the G22 imino and C19 hydroxyl proton resonances and the many NOEs they involve. While one of the amino protons of G22 could possibly form a hydrogen bond with O2 of C19 (shared with the G22 imino proton), the amino group of C19 is exposed to solvent in the major groove side. Neither the amino protons of G22 nor the amino protons of C19 produced observable resonances. Base stacking is seen between C19 and U18, and between G23 and G22.

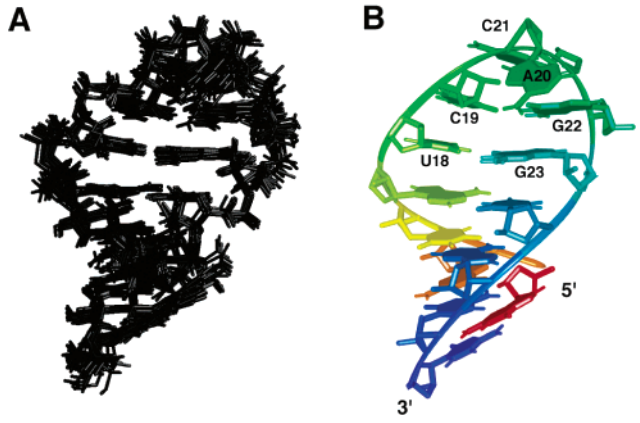


FIGURE 5: NMR-determined structure of the D-loop hairpin RNA (residues G14−C27). (A) Superimposition of the 10 lowest-energy structures. (B) Minimized average structure, with color ramping from the 5′-end (red) to the 3′-end (blue). The ribbon traces the backbone of the hairpin. The stem region (including the U18·G23 base pair) adopts an overall α-helical structure. An extensive network of hydrogen bonding and base stacking interactions provides the stabilizing force for the CACG tetraloop motif (see the text for details).

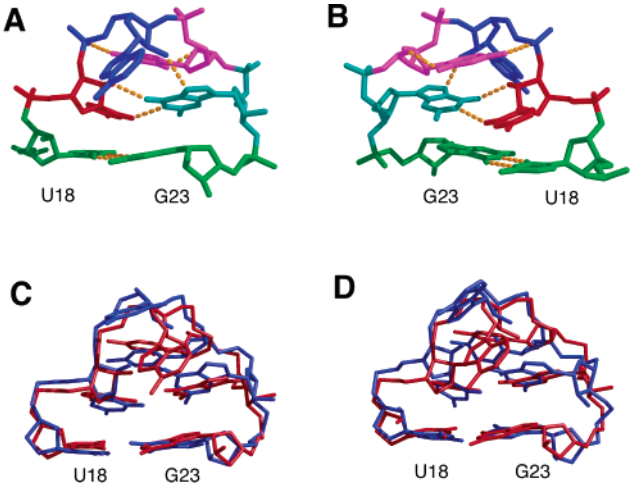


FIGURE 6: Structure of the CACG tetraloop region, including the U18·G23 closing base pair, (A) looking into the minor groove and (B) looking into the major groove (~180° rotation from panel A). The color scheme in panels A and B is as follows: U18·G23 wobble base pair in green, C19 in red, A20 in blue, C21 in magenta, and G22 in cyan. Hydrogen bonds other than Watson−Crick base pairs are shown as yellow dotted lines. (C) Superimposition of the uCACGg tetraloop structure (minimized average structure of the 10 best structures) from the study presented here (in red) with the crystal structure of the cUUCGg tetraloop (in blue; PDB entry 1F7Y). Superimposition is performed on all of the residues shown except the second residue in the loop. The pairwise atomic root-mean-square deviation (rmsd) calculated for the superimposed residues is 1.10 Å. (D) Superimposition of the uCACGg tetraloop structure with the recent NMR structure of a cUACGg tetraloop (in blue; PDB entry 1MFJ). Superimposition was performed with all the residues that are shown. The calculated pairwise rmsd is 1.52 Å. For clarity, ribose protons are not shown in panels A and B, and only heavy atoms are shown in panels C and D.

The second nucleotide of the tetraloop, A20, is the most flexible residue in the structures (Figure 5A). Its base is flipped out into solution on the minor groove side. The orientation of the base is such that the H8 proton edge is closer to the minor groove surface, and the H2 proton edge is fully exposed to solvent. While A20 H8 gives rise to several internucleotide NOEs (Figure 3), the H2 proton sees only one very weak NOE to its own H1′ proton (Figure 4). Both Watson−Crick positions (N6 amino group and N1) of A20 are readily available for hydrogen bond formation with proteins or other RNAs that might interact with the CACG tetraloop. As for the crystal structure of the UUCG tetraloop (*19*), it is possible that A20 is stabilized by a hydrogen bond between its 2′-hydroxyl group and N7 of G22. The distance between O2′ of A20 and N7 of G22 is short in all refined structures (~2.9 ± 0.1 Å). Slightly rotating the 2′-hydroxyl group around the C2′−O2′ bond would suffice to put it into a hydrogen bonding orientation with N7 of G22.

The third nucleotide of the tetraloop, C21, is quite well defined in our structures. It is positioned on the major groove side, opposite A20. The sugar puckers of C21 and A20 are both C2′-*endo*. Strikingly similar to the crystal structure of the UUCG tetraloop, three different interactions stabilize C21. First, the C21 base is clearly stacked on the C19 base. Second, the amino group of C21 is hydrogen bonded to the pro-R phosphate oxygen of A20, which explains the protection of C21 amino protons from rapid exchange with solvent. Finally, formation of a hydrogen bond between the C21 2′-hydroxyl group and O2 of C21 is very likely based on the short distance (~3.0 ± 0.2 Å) between O2′ and O2 of C21 found in all refined structures. Besides these three factors, we also noticed that the hydrophobic side of C21 (H5 and H6 protons) is pointing toward the major groove, leaving the hydrophilic side [the three Watson−Crick positions (O2, N3, and the amino group)] fully exposed to solvent (Figure 6B). Even after formation of the two hydrogen bonds involving O2 and the amino group described above, all three Watson−Crick positions of C21 are still available for hydrogen bonding with proteins or RNAs that might recognize the D-loop.

One similarity between A20 and C21 is implied by the above description of the base orientations in A20 and C21. Each has its hydrophobic side embedded in either the major groove or the minor groove, and its hydrophilic side with the Watson−Crick functional groups pointed into solution. This interesting similarity prompted us to look more carefully at these two residues. It is very clear from our NMR data that A20 and C21 are the only two residues within the hairpin D-loop that adopt the C2′-*endo* sugar conformation. The extended conformation of these two residues accounts for reversal of the backbone direction. The bases of C21 and A20 are placed in the major and minor groove, respectively, with their hydrophobic sides facing each other across the backbone. NOEs from A20 H8 to H5 and H6 of C21 (Figures 3 and 4) reflect the proximity of A20 H8 to C21 H5/H6. A20 H8−C21 H5 and A20 H8−C21 H6 distances are 4.4 ± 0.2 and 5.0 ± 0.1 Å, respectively, in our CACG tetraloops (corresponding distances in the UUCG tetraloop crystal structure of 5.25 and 6.08 Å, respectively). Considering the spatial and orientational relationship of A20 and C21, we surmise that hydrophobic interaction assists in retaining the two residues on top of the tetraloop in position, and thereby contributes to the overall stability of the tetraloop.

DISCUSSION

The CACG Tetraloop Is Structurally a Member of the UNCG Tetraloop Family. Here we have reported the solution

structure of the apical D-loop RNA within the cloverleaf structure of coxsackievirus B3. Establishing the existence of a U·G wobble base pair at the junction between the A-form helical stem and the loop confirms that the apical D-loop is indeed a tetraloop. The tetraloop is stabilized by an extensive network of base stacking and hydrogen bonding interactions, as well as hydrophobic contacts. Essentially all key stabilizing interactions found in the uCACGg tetraloop were basically identical to those previously identified in the cUUCGg tetraloop. The structural similarity between the uCACGg and cUUCGg tetraloops is at the atomic level; e.g., the identities of the characteristic set of hydrogen bond donor and acceptor groups are virtually identical. This contrasts somewhat with the cGNRAg tetraloop family, in which a network of heterogeneous hydrogen bonds was identified as an important feature of the GNRA motif (*26*). The uCACGg and cUUCGg tetraloop structures apparently exhibit a higher degree of similarity than structures within the GNRA tetraloop family.

The uCACGg tetraloop structure clearly adopts a UNCG-like tetraloop structure in solution, although it does not conform to the UNCG sequence in either identities of the first loop nucleotide or the closing base pair. We should consider the uCACGg tetraloop structure as a member of the UNCG tetraloop family, or perhaps more appropriately, the UNCG tetraloop family should be extended, as recently proposed (*29*).

Structural Basis for the YNMG Motif. Establishing the CACG tetraloop as a structural member of the UNCG tetraloop family no doubt provides the strongest possible support for the recently proposed YNMG tetraloop motif. As demonstrated in the CUUG stable tetraloop structure (*28*), it is quite possible for a standard Watson−Crick C·G base pair to form between the first and last residues of a stable tetraloop, leaving only two unpaired nucleotides (in this sense, the CUUG tetraloop is actually a biloop). Interestingly, this option is not taken by the uCACGg D-loop sequence of CVB3. Instead, an unusual noncanonical base pair is formed. Since the groups of the first loop nucleotide that participate in the novel hydrogen bonding interactions are the 2′-hydroxyl and the O2 groups, either a C or U residue is fully functional.

The second loop nucleotide is the least restrictive. According to previously reported structures of UUCG-type tetraloops and the CACG tetraloop structure we present here, two different interactions might be involved in stabilizing the second loop nucleotide: a hydrogen bond between the 2′-hydroxyl group of this nucleotide and the N7 (or O6) group of the G residue at the fourth position and a hydrophobic interaction between the bases of this nucleotide and the third loop nucleotide. Neither of these interactions requires a specific type of nucleotide, so there is no nucleotide preference for the second loop position.

Preference for either a C or an A residue at the third position of the putative YNMG tetraloop motif is not as readily explained by available structures as are the first and second positions. To find a feasible explanation, we performed a modeling study in which we substituted the cytosine residue at the third position of the CACG tetraloop with an adenine. The NOE data set was modified accordingly to reflect the change (NOEs to C21 H6 and H41/H42 were changed to NOEs to A21 H8 and H61/H62; NOEs to C21 H5 were deleted). Structure calculation was carried out using the same protocol with the mock NOE data set. The outcome was surprisingly good: a very decent structure was determined without distance violations bigger than 0.2 Å and dihedral angle violations larger than 5°. Taking a quick look at the CAAG tetraloop structure determined this way immediately provides a satisfactory explanation for the nucleotide type preference at the third position. As shown in Figure 7, all of the interactions that involved C21 in the CACG tetraloop structure find their equivalence in the CAAG tetraloop structure, without compromising any key structural feature in other parts of the tetraloop. Base stacking between A21 and C19 is seen. Hydrophobic interaction between the bases of A21 and A20 is present. The amino group of A21 is within hydrogen bonding distance of the pro-R phosphate oxygen of A20. An intraresidue hydrogen bond is feasible between the 2′-hydroxyl and N3 groups of A21 (N3−O2′ and N3−HO2′ distances of ∼2.8 and ∼2.0 Å, respectively). A G or a U residue cannot produce all of these different interactions simultaneously in the third position of the tetraloop, given the availability and relative orientation of the functional groups in these two types of nucleotides.

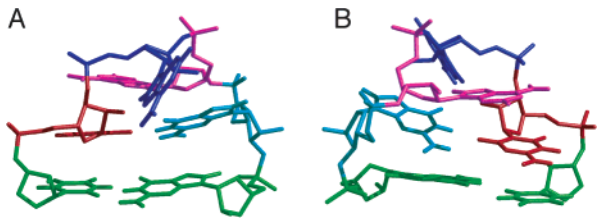


FIGURE 7: Structural model of a uCAAGg tetraloop calculated from a mock NMR data set (A) looking into the minor groove and (B) looking into the major groove (∼180° rotation from panel A). The color schemes are as follows: U18·G23 wobble base pair in green, C19 in red, A20 in blue, A21 in magenta, and G22 in cyan. Ribose protons are not shown for clarity.

The requirement for a G residue at the forth position of the YNMG motif can be enforced by the formation of the three hydrogen bonds involving the imino proton, O6, and N7 of G22 in the CACG tetraloop, and extensive purine−purine stacking between G22 and G23.

Our NMR structure determination and modeling studies reveal, at atomic resolution, the structural basis for the recently proposed YNMG tetraloop motif. Our analyses demonstrate that extension of the UNCG stable tetraloop motif to the YNMG motif is strongly justified. Moreover, we have shown that a wobble U·G closing base pair is a viable alternative to the previously observed C·G closing base pair for maintaining the distinctive structural features of the tetraloop, although this deviation does cause distortion in the stem region and may also reduce the stability of the tetraloop motif (*vide infra*). It is possible that the structural irregularity and decreased stability produced by the presence of a U·G closing base pair is necessary for optimum biological function in some systems such as the D-loop RNAs of CVB3 and other related viruses. The stage is now set for further investigations into the sequence-dependent variations in structural, dynamic, and thermodynamic properties of the YNMG tetraloop motif, and their biological relevance.

How Might the D-Loop RNA Be Recognized by 3C Protease? The solution structure of the uCACGg tetraloop

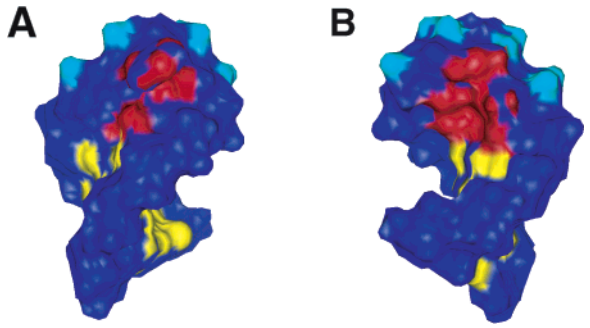


FIGURE 8: Solvent-accessible surface of the CVB3 D-loop RNA hairpin. Hydrogen bond donor and acceptor atoms from the bases of the tetraloop region (U18−G23) and the stem region (G14−A17 and U24−C27) are colored red and yellow, respectively. Phosphate oxygen atoms of the four looped nucleotides (C19−G22) are colored cyan. All other atoms are colored blue. (A) Looking into the minor groove. (B) Looking into the major groove (~180° rotation from panel A). The viewing angles are roughly the same as in panels A and B of Figure 6.

also sheds lights on how the apical D-loop RNA might be recognized by the viral 3C protease. First, since the uCACGg tetraloop has essentially all of the major structural features that define a previously established ultrastable tetraloop motif, it is very likely that some of these preformed features are used by the 3C protease for recognition. The compact and stable UUCG tetraloop resists read-through by reverse transcriptase (*3*). Given the extensive structural similarities between the uCACGg and cUUCGg tetraloops, it seems to be too demanding for the 3C protease to dramatically reorganize the existing interactions within the uCACGg tetraloop upon binding, keeping in mind that the primary biological function of 3C protease is not even as dedicated to RNA processes as those of reverse transcriptase.

Recently, the structure of the apical D-loop from human rhinovirus type 14 (HRV14) was determined by NMR (*40*). Unlike most D-loop hairpins, the HRV14 D-loop is a triloop with the sequence cUAUg. It was shown that the A residue in the loop stacked on top of the closing C·G base pair, leaving both U residues protruding into solution. No hydrogen bonding interaction was confidently identified. Apparently, the structure of the HRV14 triloop bears little similarity to that of the CVB3 tetraloop we present here. 3C protease from CVB3 recognizes several tetraloop-forming D-loop RNAs from various other viruses, including poliovirus type 1 and human rhinovirus type 2, but not the D-loop RNA from HRV14 (*30*). This suggests that some structural features of the D-loop are indeed important for RNA−protein interaction.

In the uCACGg tetraloop, several functional groups in loop residues form stable hydrogen bonds that stabilize the structure. Does this leave the tetraloop structure inept in forming hydrogen bonds with its protein partner without dramatic structural rearrangement? Not at all. Examination of potential hydrogen bond donor and acceptor groups reveals that several are exposed to solvent, ready to form hydrogen bonds with other RNA or protein molecules. On the major groove side, the three Watson−Crick positions (amino, O2, and N3) of C21, the amino and N3 groups of C19, the amino group of G22, and the O4 group of U18 are all fully accessible. The surface of the major groove side of the tetraloop is largely defined by these hydrogen bond-forming groups (Figure 8B). On the minor groove side, the two Watson−Crick positions (amino and N1) of A20 are extended into solvent. The amino and N3 positions of G23 are also ready for hydrogen bonding (Figure 8A). On top of the tetraloop structure, most of the backbone and phosphate oxygen atoms are exposed to solvent on the distinctive roofline of the tetraloop (Figure 8A,B). There is definitely no shortage of readily available hydrogen bond donors and acceptors on the accessible surface of the CVB3 tetraloop. Revelation of these potential hydrogen bond formers also suggests that YNMG tetraloops can not only serve as a stable scaffold in RNA folding but also possess all necessary attributes to be active players in tertiary RNA−RNA and RNA−protein interactions.

So far, we have focused mainly on the structural similarities between the uCACGg tetraloop and the previously determined prototypic cUUCGg tetraloop. Now we can contrast differences in structure and stability, which may pertain to optimum biological functions of the D-loop RNA. To evaluate the functional roles of these deviations, we must account for available phylogenetic data.

Analysis of 60 enteroviral and rhinoviral sequences has shown that the entire stem−loop D RNA enjoys a high degree of sequence conservation (*30*). The consensus sequence for the stem−loop D domain is identical to the naturally occurring sequence of CVB3 (Figure 1) except that at the first position of the D-loop (position 19 in the diagram) a U residue is found more frequently than a C. U and C are the two most frequent nucleotides at this position, accounting for 48 and 27%, respectively. The frequencies of nucleotide usage at other positions of the tetraloop, including the closing base pair, are as follows: 72% U and 27% C at position 18, 52% A and 23% G at position 20, 73% C and 15% U at position 21, 85% G and 10% A at position 22, and 97% G at position 23 (position numbers refer to the arbitrary scheme used in this paper; only the two most frequent nucleotides are listed). While the preference for a YNMG-type tetraloop is evident, it is interesting to see that the second position is strongly biased toward purines and the third position rarely sees an adenosine. Although the cUUCGg tetraloop is the most frequently occurring (*2*) and thermodynamically stable sequence in the YNMG family, it clearly does not enjoy its usual popularity among D-loop RNAs of the enteroviruses and rhinoviruses. This is very intriguing and potentially significant.

Among D-loop RNAs, a U·G base pair is found predominantly, rather than a C·G base pair, as the closing base pair. This unusual tendency certainly deserves more attention. To probe possible structural distortions induced by the presence of the U·G wobble base pair, we used the program CURVES (*41*) to analyze the helical parameters of the stems, including the U·G wobble base pair. The most noticeable deviation from A-helical structure occurred at the twist between U18·G23 and A17·U24 base pairs. This step, which involves the U18·G23 wobble base pair, was overtwisted by ~14° (the interbase pair twist at this step is 45.7 ± 1.8° among the 10 ensemble structures). The interbase pair twist angle between A17·U24 and U16·A25 base pairs is also slightly increased (35.8 ± 1.1°). Rise between A17·U24 and U16·A25 base pairs is 3.6 ± 0.2 Å, ~0.5 Å greater than values observed in other interbase pair steps. Concomitant with these twist angle and rise deviations, different base pair stacking patterns

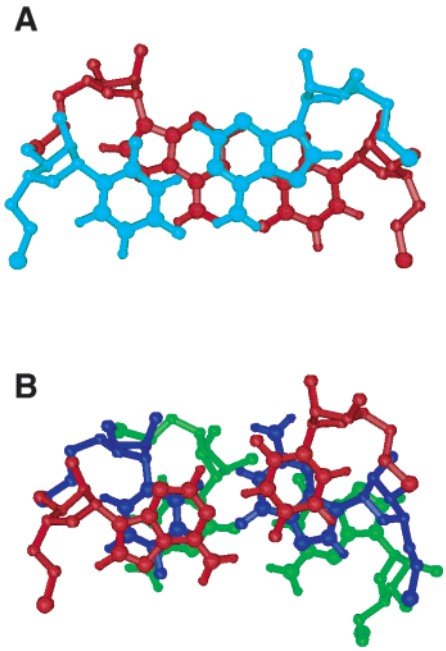


FIGURE 9: Different base stackings seen in the CVB3 D-loop RNA hairpin. (A) Very little overlap of the bases is seen between U16·A25 (cyan) and A17·U24 (red) base pairs. (B) Extensive base stacking is seen with the A17·U24 Watson−Crick base pair (red), the U18·G23 wobble base pair (blue), and the C19·G22 novel base pair (green). For clarity, sugar protons and backbone phosphate oxygen atoms are not shown.

are observed. As shown in Figure 9A, very little overlap of the bases is observed between the U16·A25 and A17·U24 base pairs. This may provide an explanation for the slightly higher value for the rise between these two base pairs. Extensive base stacking is observed in the two successive steps among the Watson−Crick A17·U24 base pair, the wobble U18·G23 base pair, and the noncanonical C19·G22 base pair (Figure 9B). Since the U16·A25 and A17·U24 base pairs are also highly conserved (the U16·A25 pair is invariant; the A17·U24 pair is 90% conserved, with the other 10% of the sequences switching to a U·A pair), it is possible that the peculiar helical features (twist, rise, and base stacking) observed in the three consecutive base pairs adjacent to the CACG tetraloop are defined by the sequence-specific combination of the base pairs and the tetraloop. Because of the limited length of the stem helix, it is difficult to reliably calculate the groove widths with CURVES. Our ongoing structure determination of the entire stem−loop D domain of the most consensus sequence would reveal how the groove widths are affected by the presence of the U·G wobble base pair and the pyrimidine-rich mismatches approximately half a helical turn away in the stem.

Structural deviation from A-form helical conformation in the stem region was also observed in the HRV14 D-loop hairpin (*40*). It is possible that distortion in base pairs adjacent to the loop is a common structural feature among D-loop RNAs of enteroviruses and rhinoviruses that is important for binding proteins.

Besides structural features, dynamic and thermodynamic properties of the D-loop RNA hairpin may be important in its interaction with the viral 3C protease. The melting temperature of the 38-nucleotide stem−loop D RNA with the CVB3 sequence and the consensus sequence was determined to be 64.5 and 65.9 °C, respectively. Consistent with previously reported results (*29*, *30*), these data indicate that the stability of the CACG tetraloop is comparable to that of the UACG tetraloop. We have not performed systematic studies to investigate how identities of the closing base pairs and sequence variation within the loop influence the flexibility and thermostability of the D-loop RNAs, but qualitative comparison of melting temperatures we measured with those reported for the 5′-ggagCACGcucc (69.6 °C) and 5′-ggagUACGcucc (70.9 °C) tetraloop minihairpins (*29*) allows some conclusions. Since our measurements were obtained for RNAs with much longer helical stems (our ongoing study on the entire consensus stem−loop D RNA revealed that the three mismatches are largely α-helical), the lower melting temperatures suggest that tetraloop hairpins found in most enteroviruses and rhinoviruses are less stable, presumably due to the closing U·G base pair and the two adjacent A·U base pairs. In the loop, viral sequences also maintain a bias for a purine residue at the second position, although U is more favorable for thermostability.

Apparently, high thermostability was not used as the primary criterion by viruses in the evolution of D-loop sequences for protein binding. It is fully possible that the reduced thermostability induced by sequence preference in the loop and the three adjacent base pairs is actually favored by the viruses for achieving optimum RNA−protein interaction. In light of recent studies showing that sequence variations also affect the flexibility of the UNCG-type tetraloop motifs at temperatures well below their melting temperatures (*29*, *42*), it is conceivable that the preferred viral sequences may also be more dynamically suitable for interacting with proteins.

In summary, this study together with all other available structural, thermodynamic, and phylogenetic data leads to a very interesting conclusion that may pertain to the mechanism of molecular recognition between the D-loop RNA and viral 3C/3CD proteins. On one hand, most viral RNAs have evolved to include the YNMG motif, one member of which has one of the most stable RNA structures known. However, this most stable tetraloop (UUCG) has clearly not been chosen, indicating that the maximum stability of the D-loop is not required for optimum biological function. Why the more stable structure is avoided by the viral RNAs is not known conclusively at this point. It is, however, tempting to speculate that reduced thermostability and enhanced dynamic flexibility produced by the preferred D-loop sequences are critical for function. However, particular viral sequences that alter dynamic or thermodynamic properties also produce distinct structural variations in the second position of the loop and the adjacent stem region that may be important for function. It may be that only one of the properties is important for proper function, but it is also possible that the structural, dynamic, and thermodynamic factors together are important.

ACKNOWLEDGMENT

We are grateful to Dr. Nikolai Ulyanov for advice in using the programs DYANA and CURVES and helpful discussions. We acknowledge use of the UCSF Computer Graphics Laboratory, supported by NIH Grant P41 RR-01081.

SUPPORTING INFORMATION AVAILABLE

Proton chemical shift assignments of CBV3 stem−loop D RNA from residue 14 to 27. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Gutell, R. R., Weiser, B., Woese, C. R., and Noller, H. F. (1985) *Prog. Nucleic Acids Res. Mol. Biol. 32*, 155−216.
2. Woese, C. R., Winker, S., and Gutell, R. R. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 8467−8471.
3. Tuerk, C., Gauss, P., Thermes, C., Groebe, D., Gayle, M., Guild, N., Stomo, G., Daubentoncarafa, Y., Uhlenbeck, O. C., Tinoco, I., Brody, E. N., and Gold, L. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 1364−1368.
4. Uhlenbeck, O. C. (1990) *Nature 346*, 613−614.
5. Michel, F., and Westhof, E. (1990) *J. Mol. Biol. 216*, 585−610.
6. Murphy, F. L., and Cech, T. R. (1993) *Biochemistry 32*, 5291−5300.
7. Murphy, F. L., and Cech, T. R. (1994) *J. Mol. Biol. 236*, 49−63.
8. Jaeger, L., Michel, F., and Westhof, E. (1994) *J. Mol. Biol. 236*, 1271−1276.
9. Pley, H. W., Flaherty, K. M., and McKay, D. B. (1994) *Nature 372*, 111−113.
10. Endo, Y., Chan, Y., Lin, A., Tsurugi, K., and Wool, I. G. (1988) *J. Biol. Chem. 263*, 7917−7920.
11. Gluck, A., Endo, Y., and Wool, I. G. (1992) *J. Mol. Biol. 226*, 411−424.
12. Tuerk, C., Gauss, P., Thermes, C., Groebe, D. R., Gayle, M., and Guild, N. (1998) *Proc. Natl. Acad. Sci. U.S.A. 85*, 1364−1368.
13. Antao, V. P., Lai, S. Y., and Tinoco, I. J. (1991) *Nucleic Acids Res. 19*, 5901−5905.
14. Antao, V. P., and Tinoco, I. J. (1992) *Nucleic Acids Res. 20*, 819−824.
15. Santalucia, J., Kierzek, R., and Turner, D. H. (1992) *Science 256*, 217−219.
16. Cheong, C., Varani, G., and Tinoco, I., Jr. (1990) *Nature 346*, 680−682.
17. Varani, G., Cheong, C., and Tinoco, I., Jr. (1991) *Biochemistry 30*, 3280−3289.
18. Allain, F. H., and Varani, G. (1995) *J. Mol. Biol. 250*, 333−353.
19. Ennifar, E., Nikulin, A., Tishchenko, S., Serganov, A., Nevskaya, N., Garber, M., Ehresmann, B., Ehresmann, C., Nikonov, S., and Dumas, P. (2000) *J. Mol. Biol. 304*, 35−42.
20. Comolli, L. R., Ulyanov, N. B., Soto, A. M., Marky, L. A., James, T. L., and Gmeiner, W. H. (2002) *Nucleic Acids Res. 30*, 4371−4379.
21. Orita, M., Nishikawa, F., Shimayama, T., Taira, K., Endo, Y., and Nishikawa, S. (1993) *Nucleic Acids Res. 21*, 5670−5678.
22. Heus, H. A., and Pardi, A. (1991) *Science 253*, 191−194.
23. Pley, H. W., Flaherty, K. M., and McKay, D. B. (1994) *Nature 372*, 68−74.
24. Mueller, L., Legault, P., and Pardi, A. (1995) *J. Am. Chem. Soc. 117*, 11043−11048.
25. Szewczak, A. A., and Moore, P. B. (1995) *J. Mol. Biol. 247*, 81−98.
26. Jucker, F. M., Heus, H. A., Yip, P. F., Moors, E. H. M., and Pardi, A. (1996) *J. Mol. Biol. 264*, 968−980.
27. Kerwood, D. J., Cavaluzzi, M. J., and Borer, P. N. (2001) *Biochemistry 40*, 14518−14529.
28. Jucker, F. M., and Pardi, A. (1995) *Biochemistry 34*, 14416−14427.
29. Proctor, D. J., Schaak, J. E., Bevilacqua, J. M., Falzone, C. J., and Bevilacqua, P. C. (2002) *Biochemistry 41*, 12062−12075.
30. Zell, R., Sidigi, K., Bucci, E., Stelzner, A., and Gorlach, M. (2002) *RNA 8*, 188−201.
31. Andino, R., Rieckhof, G. E., Trono, D., and Baltimore, D. (1990) *J. Virol. 64*, 607−612.
32. Milligan, J. F., Groebe, D. R., Witherell, G. W., and Uhlenbeck, O. C. (1987) *Nucleic Acids Res. 15*, 8783−8798.
33. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) *J. Biomol. NMR 6*, 277−293.
34. Marion, D., and Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967−974.
35. Smallcombe, S. H. (1993) *J. Am. Chem. Soc. 115*, 4776−4785.
36. Guntert, P., Mumenthaler, C., and Wuthrich, K. (1997) *J. Mol. Biol. 273*, 283−298.
37. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., GrosseKunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. D54*, 905−921.
38. Ferrin, T. E., Huang, C. C., Jarvis, L. E., and Langridge, R. (1988) *Mol. Graphics 6*, 13−27.
39. Huang, C. C., Couch, G. S., Pettersen, E. F., and Ferrin, T. E. (1996) *Pac. Symp. Biocomput. '96 1*, 724.
40. Huang, H., Alexandrov, A., Chen, X. Y., Barnes, T. W. I., Zhang, H., Dutta, K., and Pascal, S. M. (2001) *Biochemistry 40*, 8055−8064.
41. Lavery, R., and Sklenar, H. (1996) Laboratoire de Biochimie Theoretique, CNRS, Paris.
42. Williams, D. J., and Hall, K. B. (2000) *J. Mol. Biol. 297*, 1045−1061.

BI027314E