Solution Structure of a Consensus Stem-Loop D RNA Domain that Plays Important Roles in Regulating Translation and Replication in Enteroviruses and Rhinoviruses[†,‡]

Zhihua Du,[§] Jinghua Yu,[§] Nikolai B. Ulyanov,[§] Raul Andino,[#] and Thomas L. James[§,*]

Departments of Pharmaceutical Chemistry, and Microbiology and Immunology, University of California, San Francisco, California 94143-2280

Received May 20, 2004; Revised Manuscript Received July 20, 2004

ABSTRACT: Stem-loop D from the cloverleaf RNA is a highly conserved domain within the 5′-UTR of enteroviruses and rhinoviruses. Interaction between the stem-loop D RNA and the viral 3C or 3CD proteins constitutes an essential feature of a ribonucleoprotein complex that plays a critical role in regulating viral translation and replication. Here we report the solution NMR structure of a 38-nucleotide RNA with a sequence that encompasses the entire stem-loop D domain and corresponds to the consensus sequence found in enteroviruses and rhinoviruses. Sequence variants corresponding to *Poliovirus* type 1 and *Coxsackievirus* B3 have virtually the same structure, based on small differences in chemical shifts. A substantial number (136) of $^{1}H-^{13}C$ one-bond residual dipolar coupling (RDC) values were used in the structure determination in addition to conventional distance and torsion angle restraints. Inclusion of the RDC restraints was essential for achieving well-defined structures, both globally and locally. The structure of the consensus stem-loop D is an elongated A-type helical stem capped by a UACG tetraloop with a wobble UG closing base pair. Three consecutive pyrimidine base pairs (two UU and one CU pair) are present in the middle of the helical stem, creating distinctive local structural features such as a dramatically widened major groove. A dinucleotide bulge is located near the base of the stem. The bulge itself is flexible and not as well defined as the other parts of the molecule, but the flanking base pairs are intact. The peculiar spatial arrangement of the distinctive structural elements implies that they may work synergistically to achieve optimal binding affinity and specificity toward the viral 3C or 3CD proteins.

Enteroviruses, such as *Poliovirus* type 1 (PV1)[1] and *Coxsackievirus* B3 (CVB3), and rhinoviruses, such as *Human rhinovirus* 14 (HRV14), are members of the *Picornaviridae* family. These positive-strand RNA viruses carry their genetic information on a small (<10 000 bases) single-stranded genomic RNA molecule. To fulfill multiple functional roles of this molecule, the viruses evolved certain regulatory mechanisms to balance the usage of the genomic RNA in different stages of the viral life cycle. *Poliovirus* has been used as a paradigm to investigate the translational and replicational regulatory processes in picornaviruses. The 5′-untranslated region (5′UTR) of *Poliovirus* genomic RNA harbors many cis-acting RNA elements that play critical roles in both of these processes. The secondary structure of poliovirus 5′UTR is predicted to form six distinct domains, I−VI (Figure 1A). The first domain, which folds into a cloverleaf-like structure, is shown to be an essential component of a ribonucleoprotein complex that provides a switching control between protein synthesis and RNA replication (*1*). Domains II−VI constitute the internal ribosomal entry site (IRES). The IRES element mediates the cap-independent translation by allowing ribosomes to enter the mRNA internally without scanning from a 5′-cap structure.

The cloverleaf RNA at the 5′-UTR can be divided into four subdomains: stem A and stem-loops B, C, and D (Figure 1B). Although the stem-loop B domain serves as the binding site for various cellular proteins called poly(rC)-binding proteins (PCBP), the stem-loop D domain is recognized by the viral protein 3C (the viral protease), or 3CD (fusion of 3C with 3D, the viral RNA-dependent RNA polymerase) (*2*, *3*). Formation of the high affinity cloverleaf-PCBP-3CD ternary complex provides a switching mechanism to dictate the use of the viral genomic RNA as either a mRNA for protein synthesis or a template for RNA replication. Although interaction of PCBP with the cloverleaf RNA stimulates viral translation, addition of viral protein 3CD to the complex inhibits translation and promotes RNA replication. This provides a positive feedback mechanism to ensure

[†] This work was partially supported by the National Institutes of Health grant AI46967 (T. L. J.).

[‡] The coordinates of the NMR structures have been deposited in the RCSB Protein Data Bank as entry 1TXS.

* To whom correspondence should be addressed. E-mail: james@picasso.ucsf.edu. Telephone: (415) 476-1916. Fax: (415) 502-8298.

[§] Department of Pharmaceutical Chemistry.

[#] Department of Microbiology and Immunology.

[1] Abbreviations: 2D, two-dimensional; 3D, three-dimensional; 5′UTR, 5′-untranslated region; BEV1, *Bovine enterovirus* 1; COSY, correlated spectroscopy; CVB3, *Coxsackievirus* B3; HRV14, *Human rhinovirus* 14; HSQC, heteronuclear single-quantum correlated spectroscopy; IRES, internal ribosomal entry site; mRNA, messenger RNA.; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; NOESY, NOE spectroscopy; PCBP, poly(rC)-binding protein; PV1, *Poliovirus* type 1; RDC, residual dipolar coupling; RMSD, root-mean square deviation; RNA, ribonucleic acid; TOCSY, total correlation spectroscopy.

10.1021/bi048973p CCC: $27.50
Published on Web 09/02/2004

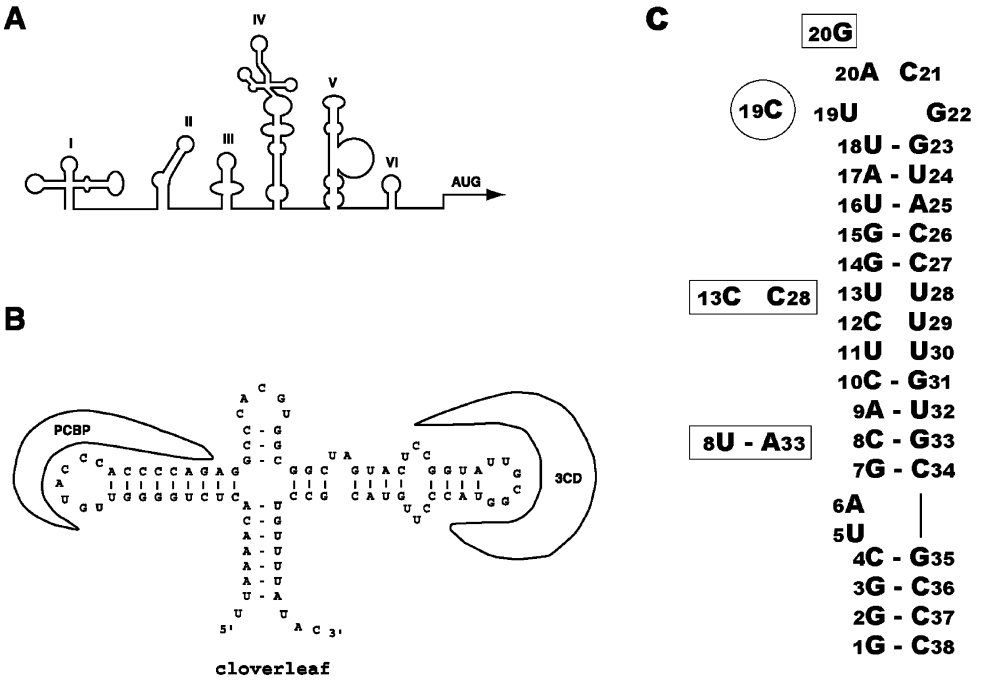

FIGURE 1: Putative secondary structures within the 5′-noncoding region of enteroviral and rhinoviral RNAs. (A) Schematic representation of the six predicted structural domains. Domain I is also known as the cloverleaf structure; domains II to VI form the IRES. (B) Schematic representation and sequence of the PV1 cloverleaf RNA and its interaction with the proteins PCBP and 3CD. Starting from the bottom stem of the cloverleaf and counting clock-wise, the four structural domains arranged around the four-way junction are known as stem A, stem-loop B, stem-loop C, and stem-loop D, respectively. The stem-loop D is the recognition site for the PV1 3CD protein. (C) The three 38-nucleotide RNA molecules investigated by NMR in this study. These sequences correspond to the entire stem-loop D within the cloverleaf RNA, with one additional GC pair (G1−C38) added for the in vitro transcription by T7 RNA polymerase. The main sequence represents the consensus stem-loop D sequence found in sixty enteroviruses and rhinoviruses (*5*). The CVB3 sequence differs from the consensus sequence only in position 19 with a C residue (circled); about one-third of the enteroviral and rhinoviral sequences have C in this position (*5*). The PV1 sequence differs from the consensus sequence in five positions, which are indicated by residues inside the boxes.

the proper balance between protein expression and RNA replication.

Since the cloverleaf RNA structure is highly conserved in enteroviruses and rhinoviruses, it is very likely that the interplay among the cloverleaf RNA and the PCBP and 3CD proteins is a common mechanism utilized by these viruses. To support this hypothesis, it was shown that protein 3C from HRV14 binds to an RNA corresponding to the HRV14 stem-loop D (*4*). More recently, it was shown that protein 3C from CVB3 recognizes the stem-loop D RNA but not the stem-loop B RNA, and a PCBP variant, PCBP2, recognizes the stem-loop B but not the stem-loop D RNA. Each protein specifically recognizes its target RNA in CVB3 (*5*). As shown in Figure 1B, stem-loop D has a mismatch region in the middle of the stem with three pyrimidines in each strand, and a two-nucleotide asymmetric internal loop near the base of the stem, which will be called a bulge in this paper. Comparison of 60 enteroviral and rhinoviral sequences indicated that the stem region (including the mismatch and the bulge) is highly conserved. Although the apical loop is less conserved in sequence, most often a four-nucleotide loop is found (*5*).

How the stem-loop D RNA is recognized by 3C or 3CD protein is not known. In HRV14, which has a three-nucleotide apical loop, it was found that the stem region harbors the most important determinants for 3C binding (*4*). However, in both PV1 and CVB3, the four-nucleotide apical loop appears to be the major determinant for specific recognition by 3C or 3CD protein (*2*, *5*). Besides binding to its cognate stem-loop D RNA, 3C protein from CVB3 also binds to the stem-loop D RNA from a number of other viruses, such as PV1, *Human rhinovirus* 2 and *Bovine enterovirus* 1 (BEV1, which has two cloverleaf RNAs in its 5′UTR; CVB3 binds to the first cloverleaf with a four-nucleotide apical loop). Although all of these stem-loop D RNA variants have a four-nucleotide apical loop, the loop sequences vary considerably. More interestingly, although the CVB3 3C protein recognizes the stem-loop D RNA nonspecies-specifically, it is not able to bind the cloverleaf from HRV14 with a three-nucleotide apical loop nor the second cloverleaf from BEV1 with a five-nucleotide apical loop (*5*). Shortening the CVB3 apical loop by one nucleotide disrupted its interaction with the CVB3 3C. In contrast, insertion of one nucleotide into the apical loop sequence of HRV14 enabled the mutant RNA to bind CVB3 3C (*5*). These experimental data, coupled with the phylogenetic observation that most enteroviral and rhinoviral apical loops have four nucleotides, suggest that interaction between the apical loop and viral 3C or 3CD is probably a common mechanism utilized by most picornaviruses.

Recognizing the importance of the apical loop does not rule out the possibility that other parts of the stem-loop D domain may also play a role. Sequence and secondary structure of the stem region (including the symmetric mismatch region and the asymmetric bulge) are highly conserved. Most noticeably, the three pyrimidine mismatches in the middle of the stem are present in all 60 enteroviral and rhinoviral sequences investigated. Although the U11−U30 mismatch (Figure 1C, numbered according to the NMR construct) becomes an AU pair and the U13−U28 mismatch

becomes a CU or CC mismatch in some viruses, the middle C12−U29 mismatch is invariant (*5*). The high degree of conservation by itself is a good argument for functional importance. It is quite tempting to speculate that the tandem mismatches somehow induce a distinctive local structure to permit protein interaction; and structural features from the stem region and the apical loop synergistically define the affinity and specificity of the interaction between stem-loop D RNA and viral 3C or 3CD protein.

In this paper, we report the NMR solution structure of a 38-nucleotide RNA representing the entire stem-loop D domain with the consensus sequence for enteroviruses and rhinoviruses (Figure 1C). We have also analyzed two other stem-loop D RNAs corresponding to the PV1 and CVB3 sequences (Figure 1C). Our results indicate that, although the primary sequences of the apical loops in these three RNAs are different, they adopt very similar three-dimensional structures that conform to the previously established UNCG-type tetraloop motif. The stem region of the stem-loop D domain adopts a largely A-form structure, although the presence of the pyrimidine mismatches and the bulge does induce distinctive structural features such as a change in groove width.

MATERIALS AND METHODS

Preparation of RNA Samples. Three different unlabeled RNA samples with sequences shown in Figure 1C were prepared. These RNAs correspond to the entire stem-loop D domain of the consensus, the CVB3, and the PV1 sequence, respectively. A uniformly ^{13}C,^{15}N-labeled sample of the consensus construct was also prepared. All samples were obtained by in vitro transcription using T7 RNA polymerase and a synthetic DNA template (*6*). T7 RNA polymerase was overexpressed in *Escherichia coli* using the plasmid pAR1219 and purified using an S-sepharose column followed by a DEAE column. The DNA template consisted of a double-stranded 18-base pair T7 promoter sequence and a single-stranded coding sequence. The transcription buffer typically contained 80 mM Tris (pH 8.1), 10 mM DTT, 1 mM spermidine, 300 nM of template DNA, and various amounts of $MgCl_2$, NTPs, and T7 RNA polymerase for optimal yields. The transcription reaction was incubated for 4 h at 37 °C. The RNAs were precipitated with three volumes of ethanol and dissolved in 8 M urea. The desired products were separated on 20% (w/v) 19:1 polyacrylamide gels containing 8 M urea. The RNA band was visualized by UV shadowing and excised from the gel; 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 the RNA were combined and lyophilized. The purified RNAs were finally resuspended in an NMR buffer that typically contained 25 mM sodium phosphate (pH 6.5) 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. Liquid crystalline media for measurements of RDC constants was achieved at 25 °C by stepwise addition with vigorous vortexing of microliter aliquots of hexanol to the RNA solution containing 2.5% (w/v) of C12E6 (hexaethylene glycol monododecyl ether) until the solution became translucent.

NMR Spectroscopy. All NMR experiments were performed on Varian Inova spectrometers operating at 600 MHz for protons. Spectra were processed with NMRPipe/NMRDraw (*7*) and analyzed with SPARKY (*8*). In 2D and 3D homonuclear and heteronuclear experiments, the States-TPPI method (*9*) was used to achieve quadrature detection in the indirect dimensions. Homonuclear 2D NOESY spectra in 90% H_2O/10% D_2O were recorded at 5 and 10 °C with a symmetrically shifted shaped pulse for solvent suppression (*10*). Homonuclear 2D NOESY, DQF−COSY, and TOCSY spectra in the D_2O buffer were recorded at 30 °C, with a very low-power presaturation for suppression of the residual HDO signal. NOESY spectra were recorded with a relaxation delay of 2.5 s and mixing times of 50, 100, 200, and 300 ms to monitor the NOE cross-peak intensity build-ups. One-bond correlated $^1H-^{15}N$ HSQC spectra were obtained at 5 and 10 °C in the 90% H_2O/10% D_2O buffer. Acquisition parameters were optimized for detection of imino or amino protons in different experiments. All multidimensional heteronuclear experiments in D_2O (except HSQC experiments for RDC measurements) were performed at 30 °C. 2D HMQC experiments with a very low-power presaturation during relaxation delay for water suppression were performed to obtain the natural-abundance one-bond $^1H-^{13}C$ correlation of the unlabeled RNAs (CVB3 and PV1 constructs). 2D constant-time HSQC experiments for one-bond $^1H-^{13}C$ correlations were recorded on the fully labeled consensus construct sample with ^{13}C and ^{15}N decoupling during acquisition using GARP (*11*). Acquisition parameters such as spectral widths, carrier frequencies, constant time delay, and $^1H-^{13}C$ one-bond coupling constants were optimized for detection of aromatic or ribose moiety in different experiments. A set of 3D experiments in D_2O was performed to establish the ribose spin systems. These are 3D ^{13}C-edited HCCH−COSY, HCCH−COSY−RELAY (*12*), and HCCH-TOCSY (*13*). The spin lock period for the HCCH-TOCSY experiment was 12 ms with a DIPSI-3 isotropic mixing scheme (*14*). A 2D version ($^1H-^1H$) of the HCCH-TOCSY experiment with 65 ms isotropic mixing time was also performed for correlating adenine H2 and H8 protons. Correlations of base and ribose protons were established by 2D HCN experiments optimized for ribose H1′ or aromatic H6/H8 protons, respectively. 3D ^{13}C-edited NOESY−HMQC experiments were recorded in D_2O with a mixing time of 150 ms and a relaxation delay of 1.2 s. A 2D HNN−COSY experiment (*15*) was used to detect directly NH···N hydrogen bonds in base pairs. A 2D H(N)CO experiment was used to connect imino protons to the C2 and C4 nuclei within uridine residues for chemical shift-based analysis of the hydrogen bonding patterns in the U11−U30 and U13−U28 mismatch pairs. RDC constants for one-bond C−H vectors were measured by comparing splitting of HSQC peaks in spectra acquired without ^{13}C decoupling in isotropic and liquid crystalline solutions. Each component of the HSQC peaks was line-fitted with the Gaussian shape to determine more accurately the position of each peak. RDC measurements were performed at 25 °C.

NMR-Derived Restraints. Interproton distance restraints between nonexchangeable protons, obtained from the intensities (peak volumes) of cross-peaks in the 2D NOE build-up

series and the 3D ^{13}C-edited NOESY spectra, were assigned to four categories as follows: strong (1.8−3.0 Å), medium (1.8−4.0 Å), weak (1.8−5.0 Å), and very weak (1.8−6.0 Å). A shorter upper bound of 2.2 Å was given to the distance between G22 H1′ and H8 due to the extreme intensity of the corresponding NOE. Distance restraints involving exchangeable protons were assigned to two categories with upper bounds of 4.0 or 5.5 Å, respectively. The very strong NOEs from adenine H2 to uridine NH, and cytosine NH_2 to guanosine NH within Watson−Crick base pairs were assigned an upper bound of 4.0 Å. All other NOEs in water were assigned an upper bound of 5.5 Å. Hydrogen bond restraints (two distance restraints per hydrogen bond) were imposed on standard Watson−Crick GC and AU base pairs established by the 2D HNN−COSY experiment, observation of imino proton resonances, and internucleotide NOEs characteristic of base pair formation. These restraints were based on standard base pair geometry of nucleic acids (*16*) and given a range of ±0.2 Å. Hydrogen bond restraints were also imposed for the U18-G23 wobble base pair, the two U−U mismatch pairs (U11−U30 and U13−U28), and the C12−U29 base pair. Hydrogen bonding of the two U−U mismatch pairs were inferred from C2 and C4 chemical shifts of the residues obtained from a 2D H(N)CO experiment (Supporting Information). Hydrogen bonding between the C12−U29 base pair was revealed by the ^{1}H and ^{15}N chemical shifts of the imino group of U29. Torsion angle restraints for sugar rings were based on analysis of the 2D DQF−COSY and TOCSY spectra. Sugars showing a strong H1′−H2′ cross-peak (U5, A20, and C21) were restrained to C2′-*endo*. Those with no COSY and TOCSY cross-peaks were interpreted as C3′-*endo*. All other nucleotides were left unconstrained (G1, A6, and C38). Backbone torsion angles for residues G1-C4, G7-C10, G14-A17, U24−C27, G31-C34, and G35-C38 were restricted to the A-form helical values within a range of ±45°. No backbone torsion angle restraints were imposed for the U5A6 bulge, the U11C12U13−U28U29U30 mismatches, and the tetraloop with the closing G−U base pair (U18-G23). A total of 136 RDC values for one-bond C−H vectors with a flat well of ±1.5 Hz were used in the structure determination.

Structure Calculation and Refinement. Structure calculation and refinement were performed using the programs DYANA 1.5 (*17*) and Xplor-NIH (*18*). Using DYANA, 800 random initial structures were generated and subjected to simulated annealing in the torsion angle space, followed by variable target function minimization. The 200 best DYANA structures with the lowest target function values and correct chirality (checked by an in-house-developed program CHIranO) were chosen for further refinement using Xplor-NIH. The final values of the constants in Xplor-NIH refinement were 1 kcal mol^{-1} Hz^{-2} for RDC restraints, 50 kcal mol^{-1} Å^{-2} for the distance restraints, and 200 kcal mol^{-1} rad^{-2} for the torsion angle restraints. The 25 structures with the lowest total energy were chosen to represent the final structure ensemble. To assess the impact of RDC restraints on the refined structures, two refinement schemes were carried out: one with RDC values and the other without. To estimate the magnitude of the molecular alignment, *Da* −40 to +40 Hz and *R* was varied from 0 to 0.7 in steps of 0.5 Hz and 0.05, respectively. All molecular graphics was generated using MidasPlus (*19*) or Chimera (*20*) from the Computer Graphics Laboratory, University of California, San Francisco. Analysis of helical parameters was performed using the programs 3DNA (*21*) and CURVES (*22*).

RESULTS

Choice of Sequences for NMR Study. Since most studies of the cloverleaf-PCBP-3C/3CD ribonucleoprotein complex have entailed PV1, we initially prepared a 38-nucleotide RNA corresponding to the PV1 stem-loop D domain (Figure 1C, deviations of the PV1 sequence from the consensus sequence are shown inside the boxes). Unfortunately, the NMR construct appeared to exist as a mixture of monomeric and dimeric forms (estimated molar ratio 3:1) in solution as judged by nondenaturing polyacrylamide gel electrophoresis. Inspection of the pyrimidine H5−H6 region in a DQF−COSY spectrum (data not shown) also revealed more cross-peaks than expected from the sequence, indicating the presence of more than one conformation. A possible reason for formation of the dimeric form is the self-complementary nature of the apical loop sequence UGCG. Although valuable insights into the PV1 apical loop structure could still be obtained from analysis of the NMR spectra due primarily to the characteristic and unusual chemical shifts of a number of the protons belonging to the apical loop residues (vide infra), it appeared inappropriate to pursue a full-blown structure determination using the PV1 construct.

Two other constructs, corresponding to the CVB3 and the consensus sequences, were then investigated (Figure 1C; the CVB3 sequence differs from the consensus sequence by only one nucleotide, indicated inside the circle). Nondenaturing electrophoresis showed that both constructs were mostly monomeric in solution, with the CVB3 construct having a faint minor band (<5%) at the dimeric form position. Within the pyrimidine H5−H6 region in the DQF−COSY and TOCSY spectra, the correct number of cross-peaks was detected (C24 and C36 are overlapped), indicating that both constructs were sufficiently pure in monomeric conformation for detailed structure determination. A uniformly ^{13}C,^{15}N-labeled sample of the consensus construct was prepared. Nearly complete ^{1}H, ^{13}C and ^{15}N chemical shift assignments were obtained for that construct (Supporting Information). Since a high-resolution structure of a CACG tetraloop, such as that in CVB3, had not yet been reported at the time, we also assigned ^{1}H resonances and solved the structure for the apical stem-loop (Figure 1C, from residue G14 to C27) in the context of the 38-nucleotide CVB3 construct (*23*). Here, we focus on the structure determination of the 38-nucleotide consensus stem-loop D RNA, with description extended to the other two RNAs in terms of spectral and structural similarities.

Assignment of Imino Group Resonances and Establishment of Secondary Structure. Imino proton resonances were assigned first. These were the easiest to analyze and yet most indicative of the existence of hydrogen bonds within base pairs in an RNA molecule. As shown in Figure 2A, the U18-G23, U13−U28, and U11−U30 base pairs were readily identified for the consensus sequence since each residue gives rise to a sharp imino resonance and a strong cross-peak is seen for each pair of resonances. The UG imino protons resonate at an upfield region compared to the imino resonances from Watson−Crick GC and AU base pairs.

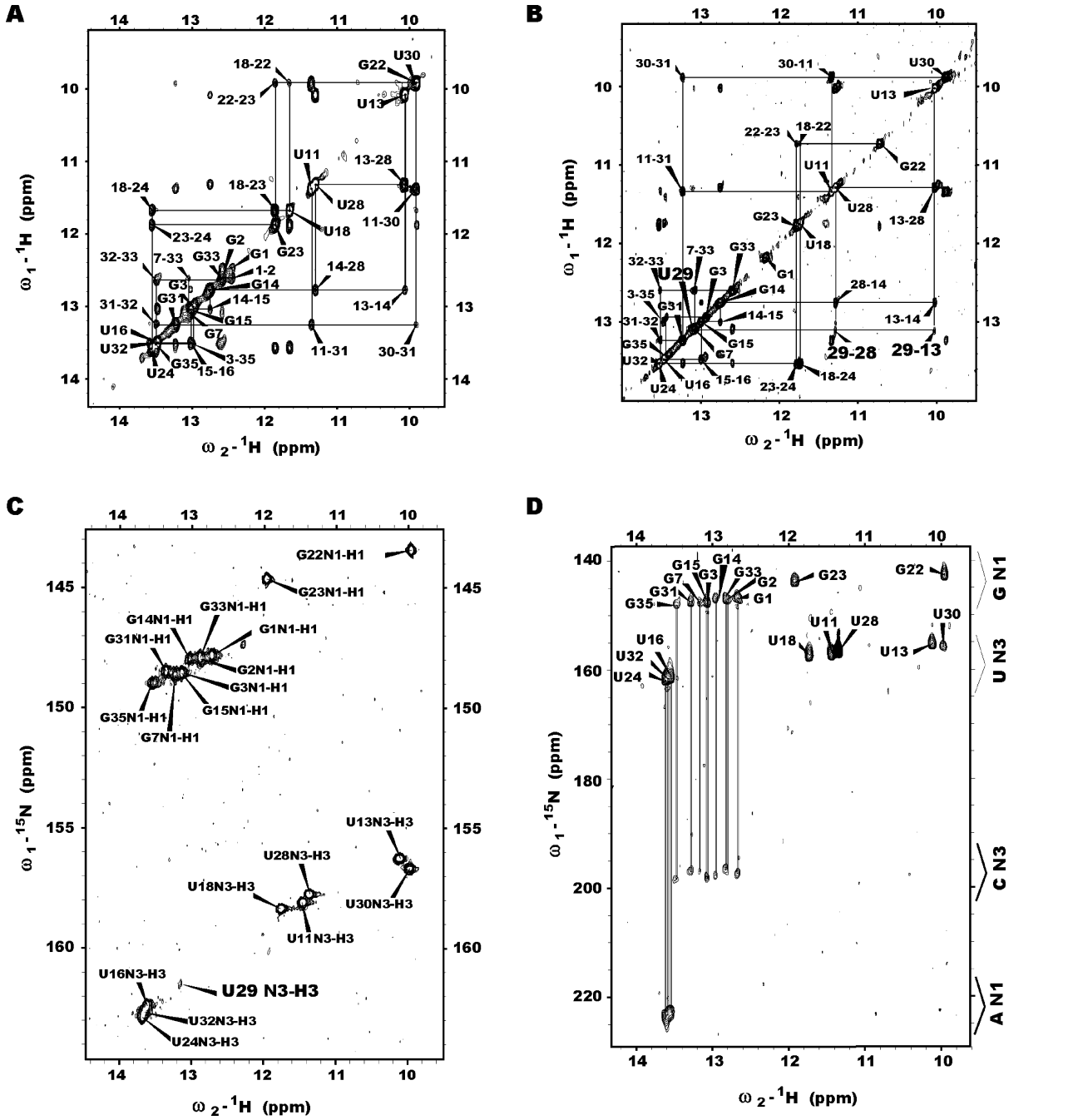

FIGURE 2: (A) Section of a 2D NOESY spectrum (150 ms mixing time) showing NOEs between imino protons of the 38-nucleotide consensus stem-loop D RNA in the 90% H_2O/10% D_2O solvent and 25 mM sodium chloride, 25 mM sodium phosphate buffer, pH 6.5, at 5 °C. Assignments of the diagonal peaks are indicated by residue type and number; cross-peaks are indicated by the residue numbers involved. Note the resonances shifted upfield between 9.5 and 11.5 ppm, corresponding to two UU base pairs in the mismatch region, a wobble GU base pair and hydrogen bonding involving G22 in the tetraloop region. (B) Corresponding spectrum from CVB3 stem-loop D RNA at 5 °C. Note that the U29 imino resonance is observed in this construct, which has two NOEs to the imino protons of U13 and U28 (emphasized labels). (C) The imino-optimized $^{15}N-HSQC$ spectrum of the consensus stem-loop D RNA at 10 °C. Imino resonances from G and U residues are distinguished by their ^{15}N chemical shifts. ^{15}N chemical shift differences within each group are also indicative of the nature of the hydrogen bonding interaction; note the N1 chemical shift difference between G22, G23 and guanines in Watson−Crick base pairs. The U29 peak (emphasized label) is very weak compared to the others. (D) The 2D HNN−COSY spectrum of the consensus stem-loop D at 10 °C for the direct detection of hydrogen bonds in Watson−Crick base pairs. Labeled peaks in the upper part of the spectrum arise from the one-bond correlations of the NH imino groups (comparable to the spectrum in C). Cross-peaks in the lower part of the spectrum (not labeled) arise from the $^{2h}J_{NN}$ couplings across NH···N hydrogen bonds; they have the opposite phase compared to the NH peaks. Note that imino groups participating in NH···O hydrogen bonds do not have such cross-peaks (U11, U13, U18, U28, U30, G22, G23).

Similar patterns are observed for the CVB3 construct, however, U18 and G23 imino resonances overlap in this case (Figure 2B). The residue type (U vs G) for these imino protons was further confirmed by a 2D $^{15}N-HSQC$ spectrum optimized for imino resonances (Figure 2C).

G22 from the tetraloop also gives an upfield-shifted imino resonance, overlapped with the U30 imino. The two resonances are readily resolved in the 2D $^{15}N-HSQC$ spectrum into a G imino resonance and a U imino resonance (Figure 2C). Although the G22 imino proton resonance is clearly

present, the U19 imino proton resonance is not observed. G22 N1 resonance is also shifted upfield by more than 1 ppm from G23 N1, from the classical U18-G23 wobble pair (Figure 2C). Therefore, G22 is unlikely to form a wobble base pair with U19. In the CVB3 construct with the CACG tetraloop, G22 still gives rise to a sharp upfield-shifted imino proton resonance (Figure 2B). Moreover, the G22 imino resonance shows a virtually identical set of NOEs in the CVB3 construct and the consensus construct, suggesting that residue G22 is involved in similar hydrogen bonding interactions in the two RNA molecules.

In the consensus construct, we did not observe the imino resonance of U29 or the amino resonance of C12 in the NOESY spectrum; however, a weak cross-peak for U29 imino proton is detected in the ^{15}N−HSQC spectrum (Figure 2C). The imino resonance of U29 is clearly observed in the CVB3 construct, and it yields two NOEs to the imino protons of U13 and U28 (Figure 2B). These data indicate the existence of the C12−U29 base pair in both constructs; a weaker intensity of the U29 imino resonance may indicate increased proton exchange with solvent in the central mismatch.

NH···N hydrogen bonds in Watson−Crick AU and GC base paired are directly confirmed by the 2D HNN−COSY spectrum (Figure 2D). The top part of the spectrum shows peaks (labeled in the Figure) due to the one-bond $^{1}J_{NH}$ couplings, and the bottom part shows the opposite-phase cross-peaks (unlabeled) due to the $^{2h}J_{NN}$ couplings across the NH···N hydrogen bonds (*15*). Imino groups participating in the NH···O hydrogen bonds in UG and UU pairs (U11, U13, U18, G22, G23, U28, U30) have only peaks due to the covalent bond $^{1}J_{NH}$ couplings, because of the lack of NH···N hydrogen bonds. The imino groups of U19 and U29 are not observed in this spectrum (however, compare with the weak U29 HSQC peak in Figure 2C).

The geometry of the two UU base pairs was inferred from the ^{13}C chemical shifts (obtained from a 2D H(N)CO experiment, see Supporting Information) of the carbonyl C2 and C4 nuclei of the uridine residues involved (*24*−*26*). C2 chemical shifts of U11 and U28 are similar to that of U18 in the U18-G23 wobble base pair, indicating O2 of U11 and U28 is the hydrogen bond acceptor as in U18. For U13 and U30, both the C2 and C4 chemical shifts are comparable to those of the uridines in the AU base pairs, indicating that O4 is the hydrogen-bond acceptor, whereas O2 is not involved in hydrogen bond. This conclusion is also corroborated by the ^{1}H and ^{15}N chemical shifts: U11 and U28 are similar to U18 from the wobble UG pair, which is using O2 as the hydrogen bond acceptor, and all three are different from U13 and U30 with the O4 acceptor (Figure 2C).

In the C12−U29 base pair, the imino proton of U29 has two possible acceptors: O2 or N3 of C12. It is clear from Figure 2C that the N3 chemical shifts of base-paired uridines have distinctive values depending on the atom type of the acceptor. U16, U24, and U32 are involved in standard Watson−Crick AU base pairs with nitrogen as acceptor. U11, U13, U18, U28, and U30 are involved in UU or UG base pairs with carbonyl oxygen as acceptor. The N3 chemical shifts of these two groups of uridines differ by more than 4 ppm. The similarity of the N3 chemical shift of U29 and the U16/24/32 group (Figure 2C) indicates that N3 of C12 is the hydrogen bond acceptor for the U29 imino proton.

All residues from the mismatch region are in the anti conformation with respect to the glycosidic bonds (vide infra), so each pyrimidine−pyrimidine pair assumes a Watson−Crick-type configuration (as opposed to reverse Watson−Crick). Assignments of imino group resonances confirm the predicted stem-loop nature of the D domain RNA. Base pairing of all three mismatches is directly established by NMR.

Observation of 2′-Hydroxyl and Amino Protons Involved in Hydrogen-Bonding within the Tetraloop. We identified a number of slowly exchanging resonances (presumably protected from exchange with solvent by involvement in hydrogen bonds) in the region between 6.5 and 7.5 ppm. By analogy to previous studies on the cUUCGg tetraloop (*27*), we assigned the resonance at 6.54 ppm to the 2′-OH proton of U19, and a pair of resonances at 6.53 and 7.31 ppm to the amino protons H41 and H42 of C21, respectively. The identities of the C21 H41 and H42 protons were confirmed by a 2D ^{15}N−HSQC spectrum showing that their attached ^{15}N nuclei resonate at 94.70 ppm, within the region of cytosine N4 resonances. As noted previously (*27*), this ^{15}N chemical shift is distinct from other N4 resonances (>97 ppm) that are involved in standard Watson−Crick GC base pairs. The smaller difference between the H41 and H42 chemical shifts of C21 also distinguishes this residue from cytosines in GC base pairs. Most of the NOEs involving the U19 2′-OH and C21 H41/H42 protons are identical (in terms of the two protons involved) in all three RNAs investigated, so hydrogen bonding interactions in the three apical loops should be very similar.

Assignment of the Nonexchangeable Proton Resonances. Nonexchangeable protons were assigned using a set of standard NMR experiments performed in D_2O buffer. Ribose spin systems were identified by analysis of 3D HCCH−COSY, 3D HCCH−RELAY, and 3D HCCH-TOCSY spectra. Intra-nucleotide ribose H1′-to-base H6/H8 connections were established by 2D H(C)N experiments. Adenine H2 and H8 protons were connected by a 2D version (^{1}H−^{1}H) of the HCCH-TOCSY experiment. The relatively large size of the RNA constructs necessitated NOE-based sequential walks for sequence-specific assignments. Although this through-space strategy is potentially troublesome, several favorable factors made it less so in our case. First, the resonances are well dispersed (see Figure 3, for example). Second, the helical nature of the stem region makes sequential walks relatively straightforward. Finally, each residue of the tetraloop belongs to a specific residue type; and the tetraloop residues yield a number of characteristically shifted resonances, which made their assignments unambiguous.

Figure 3 shows sequential walks in the H1′/H5 to H2/H6/H8 region of a 2D NOESY spectrum with 200 ms mixing time. Four continuous walks were identified, connecting most residues of the molecule: from G1 to U5, A6 to C21, G23 to C34, and G35 to C38. Assignments for overlapping peaks were resolved and confirmed by a 3D ^{13}C-edited NOESY spectrum. Assignments of H1′ protons were readily extended to other ribose protons by the three HCCH-type experiments (Supporting Information).

Several unusual proton chemical shifts were observed in stem-loop D RNA, nearly all associated with the apical loop residues. Some of these protons resonate in regions typical of other classes of protons. For examples, G23 H1′ (4.46

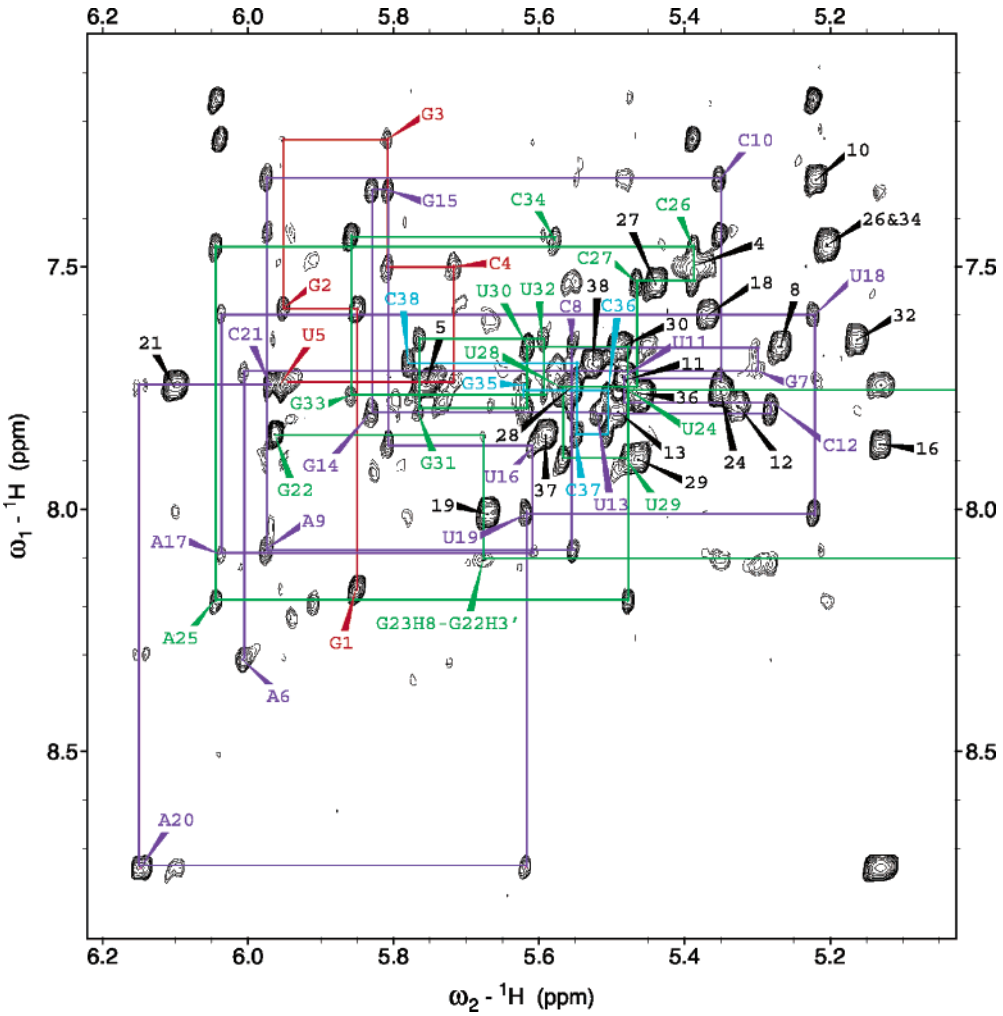

FIGURE 3: Section of a 2D NOESY spectrum (200 ms mixing time) of the 38-nucleotide consensus stem-loop D RNA in the D_2O solvent, 25 mM sodium chloride, 25 mM sodium phosphate buffer, pH 6.5, at 30 °C. Essentially all observed peaks in this region of the spectrum have been assigned. Despite the relatively large size of the RNA molecule, excellent dispersion of the resonances enables the establishment of four sequential H6/H8−H1′ walks connecting all of the residues except G22, which is connected to G23 through NOEs to its unusually shifted H3′ proton (5.65 ppm, green). Intra-nucleotide H6/H8−H1′ NOEs are labeled by type and number, for example, G22. Pyrimidine intra-nucleotide H5−H6 NOE peaks are labeled by the number of the nucleotide, for example, 18. Sequential walk from residue G1 to U5 is traced in red, A6 to C21 in purple, G23 to C34 in green, and G35 to C38 in cyan. Since G23 H1′ is shifted upfield to 4.47 ppm, its NOEs to G23 H8 and U24 H6 are not shown in the figure but are indicated by two green lines pointing out of the graph. NOEs in the overlapped regions were confirmed by 3D ^{13}C-separated NOESY.

ppm) is found in the region normal for H2′-H5′/H5″; A20 H2′ (5.13 ppm) and G22 H3′ (5.68 ppm) are found in the normal H1′/H5 region. A number of protons, including U19 H2′ (3.81 ppm), C21 H4′ (3.74 ppm), and H5′/H5″ (3.62/2.80 ppm, not stereospecifically assigned), were shifted upfield to a region where usually no RNA resonances are found. The real identities of these unusually shifted protons are revealed by the chemical shifts of their attached ^{13}C nuclei and by correlations in the HCCH-type spectra. All of these unusual chemical shifts are virtually identical in all three tetraloops studied here: uUACGg (consensus), uCACGg (CVB3), and uUGCGg (PV1) (Supporting Information). Moreover, these chemical shifts are almost identical to those first reported in the paradigmatic study of the unusually stable cUUCGg tetraloop (*27*−*29*). Since chemical shifts are very sensitive to local structure, it strongly argues for the structural similarity of this class of tetraloops.

No unusual chemical shifts were detected for residues in the three pyrimidine mismatches, again consistent with these mismatches being in a largely A-form conformation.

Ribose Conformations and Glycosidic Torsion Angles. In the H1′−H2′ region of the 2D DQF−COSY spectrum, only U5, A20, and C21 had strong multiplet cross-peaks corresponding to $^3J_{H1'-H2'} = 6-8$ Hz. C38 exhibited a weak H1′−H2′ COSY peak. In the 2D TOCSY spectrum, H1′−H2′ cross-peaks in G1 and A6 were also detected. These data indicate that ribose rings of U5, A20, and C21 adopt predominantly the C2′-*endo* conformation. G1, A6, and C38 riboses probably exist in equilibrium between C3′-*endo* and C2′-*endo* conformations. The absence of COSY and TOCSY H1′−H2′ cross-peaks for the rest of the residues except G23 suggests that their riboses are predominantly C3′-*endo*. Noticeably, a strong COSY peak was unambiguously identified for H3′−H4′ of G22 due to a very unusual chemical shift of the H3′ proton at 5.68 ppm. The presence of this peak and the absence of a H1′−H2′ COSY peak unequivocally indicate the C3′-*endo* sugar pucker for G22. Because the H1′ resonance of G23 is shifted upfield to 4.52 ppm and overlapped with the H2′ resonance, the ribose conformation of G23 could not be determined solely

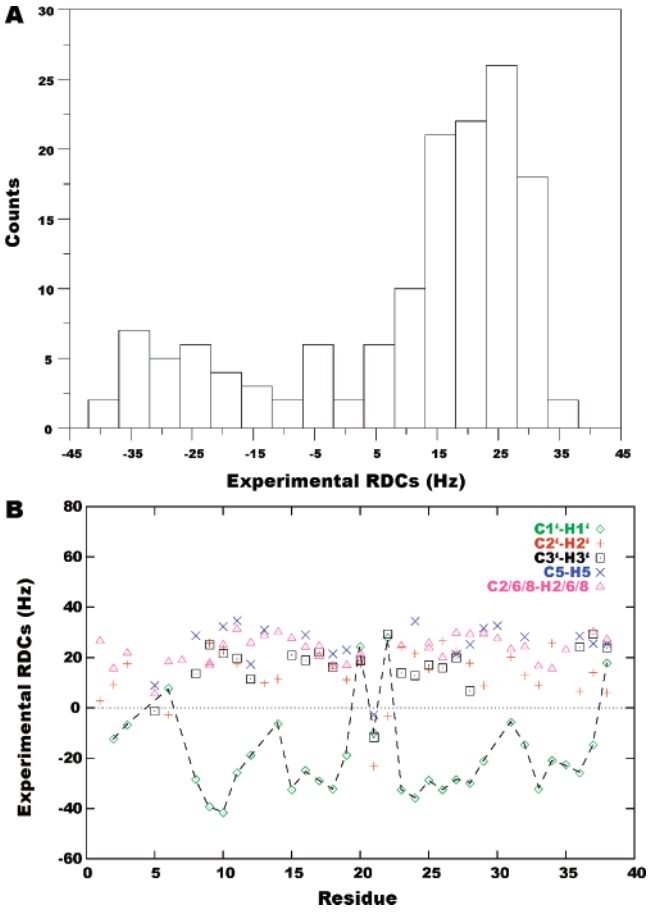

FIGURE 4: (A) Histogram of the 146 experimental $^{1}H-^{13}C$ one-bond RDCs plotted with a bin size of 5 Hz, showing the nonrandom distribution of the RDCs. (B) A graph showing the 146 experimental $^{1}H-^{13}C$ one-bond RDCs plotted against the residue numbers. C1′−H1′ RDC points are connected with a dashed line. Note attenuation of RDC values for the bulge residues, U5 and A6.

on analysis of COSY and TOCSY spectra. However, chemical shift values for ribose carbons (C1′−C4′) of G23 suggest that the ribose conformation is C3′-*endo*. None of the six riboses of the three pyrimidine mismatches (U11 to U13 and U28 to U30) exhibited any cross-peaks in the COSY and TOCSY spectra, consistent with NOE data indicating that these residues are part of a largely A-helical stem structure.

Inspection of the intraresidue H1′−H6/H8 NOE cross-peaks showed that only G22 has an NOE far stronger than any of the cytosine and uridine intraresidue H5−H6 cross-peaks, indicating that G22 is in the syn conformation with respect to its glycosidic torsion angle, while all other residues are in the anti conformation.

Measurements of Residual Dipolar Couplings. A substantial number of $^{13}C-^{1}H$ one-bond RDCs (46 base and 98 ribose) were measured in D_2O buffer for the consensus stem-loop D RNA, ranging from −42 to +35 Hz. Weak alignment of RNA was achieved using a C12E6-hexanol mixture (*30*). In Figure 4A, the distribution of measured RDC values is shown. A strong bias toward the right edge of the distribution is observed, which is typical for predominantly helical A-form RNA. In Figure 4B, the RDCs are plotted against residue numbers based on categories (H1′−C1′, H2′−C2′, H3′−C3′, H5−C5, and H2/6/8−C2/6/8) of the $^{13}C-^{1}H$ one-bond vectors. RDC values within each category generally

Table 1: Structural Statistics for the Consensus Cloverleaf Stem-Loop D RNA

NOE-derived distance restraints	541
intra-nucleotide	122
inter-nucleotide sequential	348
inter-nucleotide nonsequential	71
residual deviation from experimental restraints[a,b]	
distance restraints (Å)[c]	0.075 ± 0.003
torsion angle restraints (°)[d]	0.29 ± 0.04
RDC restraints (Hz)	1.30 ± 0.04
deviation from the idealized covalent geometry[a,b]	
bond lengths (Å)	0.005 ± 0.000
bond angles (°)	1.02 ± 0.02
improper torsion angles (°)	0.25 ± 0.01
heavy atoms RMSD from the mean structure (Å)[a]	
all residues	1.09
all except U5, A6	0.94
tetraloop (U18-G23)	0.67
mismatch region (U11−U13 & U28−U30)	0.84
bulge with the flanking base pairs (G4-G7 & C34-G35)	1.28

[a] Averaged over 25 structures. [b] Plus/minus standard deviation. [c] No violations >0.5 Å. [d] No violations >5°.

have the same sign except for apical loop residues (A20, C21, and G22), the dinucleotide bulge (U5 and A6), and the terminal C38. All measured RDCs for the pyrimidine mismatches (U11−U13 and U28−U30) show no abnormality in this regard. This might reflect a largely A-like helical structure of these mismatch pairs, although a deviation from standard geometry is also suggested by careful analysis of RDC data. A dashed curve connecting RDCs for the C1′−H1′ vectors in Figure 4B illustrates this point. The curve has two pronounced peaks at residues G14 and G31 (disregarding peaks associated with bulge and tetraloop residues), flanking the mismatch pairs. Moreover, these two peaks have similar shapes, which might indicate local structural similarities between the U13G14-C27U28 and U30G31-C10U11 base pair steps. Such structural similarities are also suggested by similar chemical shifts between U13 and U30, and between U28 and U11 (e.g., see Figure 2, parts A and C). Another feature of the RDC plot is an attenuation of RDC values in the U5A6 bulge region, especially for U5. This attenuation is most likely due to increased dynamics in the bulge (*31*). In comparison, the tetraloop and the mismatch region do not show similar attenuation of RDC values, indicating less flexible structures in these regions.

Structure Calculation of the Consensus Stem-Loop D RNA. The structure of the consensus stem-loop D RNA was calculated as outlined in MATERIALS AND METHODS; structural statistics are summarized in Table 1. To assess the influence of RDCs on NMR structures, two sets of structures, with or without RDCs, were calculated. For the refinement with RDC restraints, 136 one-bond $^{13}C-^{1}H$ RDCs were used; these did not include RDCs for residues U5 and A6. A test run with all RDCs including those for U5 and A6 did not result in a better-defined structure of the two bulged nucleotides. The axial and rhombic components of the molecular alignment tensor D_a and R were estimated with a grid search within Xplor-NIH. For our data set, D_a is −31.0 Hz and R is 0.10. Refinements using higher or lower values did not produce structures with lower total energy.

Impact of RDC Restraints on Calculated Structures. The superimpositions of the 25 lowest-energy structures calculated with and without RDCs are shown in Figure 5, parts

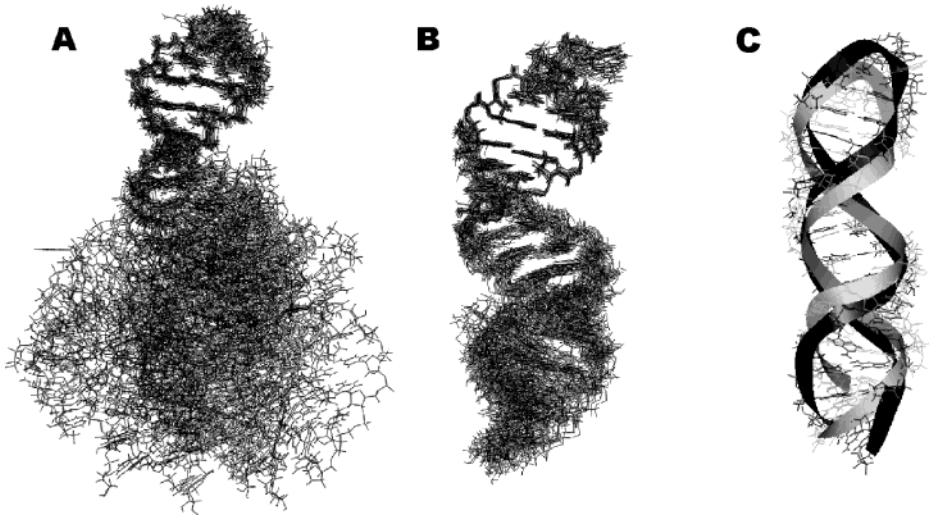

FIGURE 5: NMR-determined structures of the consensus stem-loop D RNA. (A) The 25 lowest energy structures refined with the conventional distance and torsion angle restraints (without RDC restraints). Superimposition is done using the upper helix and tetraloop (residues G14 to C27). (B) The 25 lowest energy structures refined with the conventional distance and torsion angle restraints plus 136 $^{1}H-^{13}C$ one-bond RDC restraints (excluding RDC restraints for the bulge residues, U5 and A6). Superimposition is done the same way as in panel A. (C) Superimposition of the average structures refined with (gray) or without (black) RDC restraints; the ribbons trace the RNA backbone.

A and B, respectively. Without RDC restraints, the overall structure of the molecule is ill defined. Most noticeably, the two fragments of the molecule above and below the mismatch region could not be superimposed simultaneously, although each individual fragment was relatively well defined (Figure 5A). Inclusion of RDCs in the refinement led to better defined overall structures (Table 1 and Figure 5B). Superimposition of the structures refined with and without RDCs further revealed that structures refined without RDCs were more elongated along the helical axis compared to structures refined with RDCs (Figure 5C). Besides the global structure, definition of local structural features was also significantly improved by the inclusion of RDCs. For example, in the absence of RDC restraints, helical parameters of the mismatch region and the base pairs adjacent to the U5A6 bulge could not be defined due to lack of convergence of the structures. These ill-defined local structures actually are the origin of the ill-defined global structures derived in the absence of the RDC restraints. There is no doubt that RDC restraints played an essential role in defining both global and local structures more precisely and presumably more accurately. We will mainly discuss the structures refined with RDCs in the following sections.

Global Structure of the Consensus Stem-Loop D Domain. The overall structure of the consensus stem-loop D RNA is largely an A-form helical stem capped by a stable tetraloop with a wobble UG closing base pair. Three pyrimidine mismatch base pairs are present in the middle of the helical stem, which causes some deviation from standard A-form geometry. The two-nucleotide bulge also has distinctive features such as disruption of the A-form helical stacking on the bulged strand and changes in the groove widths (vide infra). All three junctions within the stem (one at the bulge and two at each end of the pyrimidine mismatch region) are associated with some bending of the helical axis (Figure 6A). The irregular helical axis for each structure was calculated with the program CURVES and then approximated with segments of a straight line. For the structures refined with RDCs, each of the three junctions has a bend angle of 30−35° on average, with the range of ca. 20−50°. However, because the junctions are separated by 3−4 base pairs, individual bends cancel each other, so that the overall axis of the stem-loop can be approximated with a straight line (Figure 6A).

Structure of the uUACGg Tetraloop. The uUACGg tetraloop in the consensus stem-loop D RNA is quite similar to the uCACGg tetraloop in the CVB3 stem-loop D RNA we previously reported (*23*). All of our detailed description and analysis of the CVB3 tetraloop seem applicable to the consensus tetraloop. We will therefore only summarize key structural features here. The tetraloop is closed by a UG wobble base pair. The first (U19) and last (G22) nucleotides of the tetraloop form an unusual U19-G22 base pair, with U19 in anti and G22 in syn glycosidic conformations. The 2′-hydroxyl group of U19 forms a hydrogen bond with O6 of G22, and the imino group of G22 makes another hydrogen bond to O2 of U19. The sugars of the second and third nucleotides of the tetraloop, A20 and C21, are in the C2′-*endo* conformation. The base moieties of A20 and C21 are on the minor and major groove side, respectively, with their hydrophobic edges (H8 in A20, and H5 and H6 in C21) facing each other across the backbone. The hydrophilic edges of A20 and C21 (the two Watson−Crick positions N6 amino group and N1 in A20, and the three Watson−Crick positions O2, N3, and the N4 amino group in C21) are fully exposed to solvent (Figure 6B). The amino group of C21 is hydrogen-bonded to the pro(R) phosphate oxygen of A20. Bases of C21, U19 and U18 form a continuous stack. Hydrogen bonds are also possible between the hydroxyl group of A20 and N7 of G22, and between the hydroxyl group and O2 of C21 (Figure 6B). With these characteristic structural features, there is no doubt that the uUACGg tetraloop of the consensus stem-loop D RNA is structurally a member of the UNCG tetraloop family.

Structure of the Pyrimidine Mismatches. Formation of two UU base pairs (U11−U30 and U13−U28) results in protection of the four imino protons from rapid exchange with solvent (Figures 2A, 2C, and 2D). Each of the base pairs is

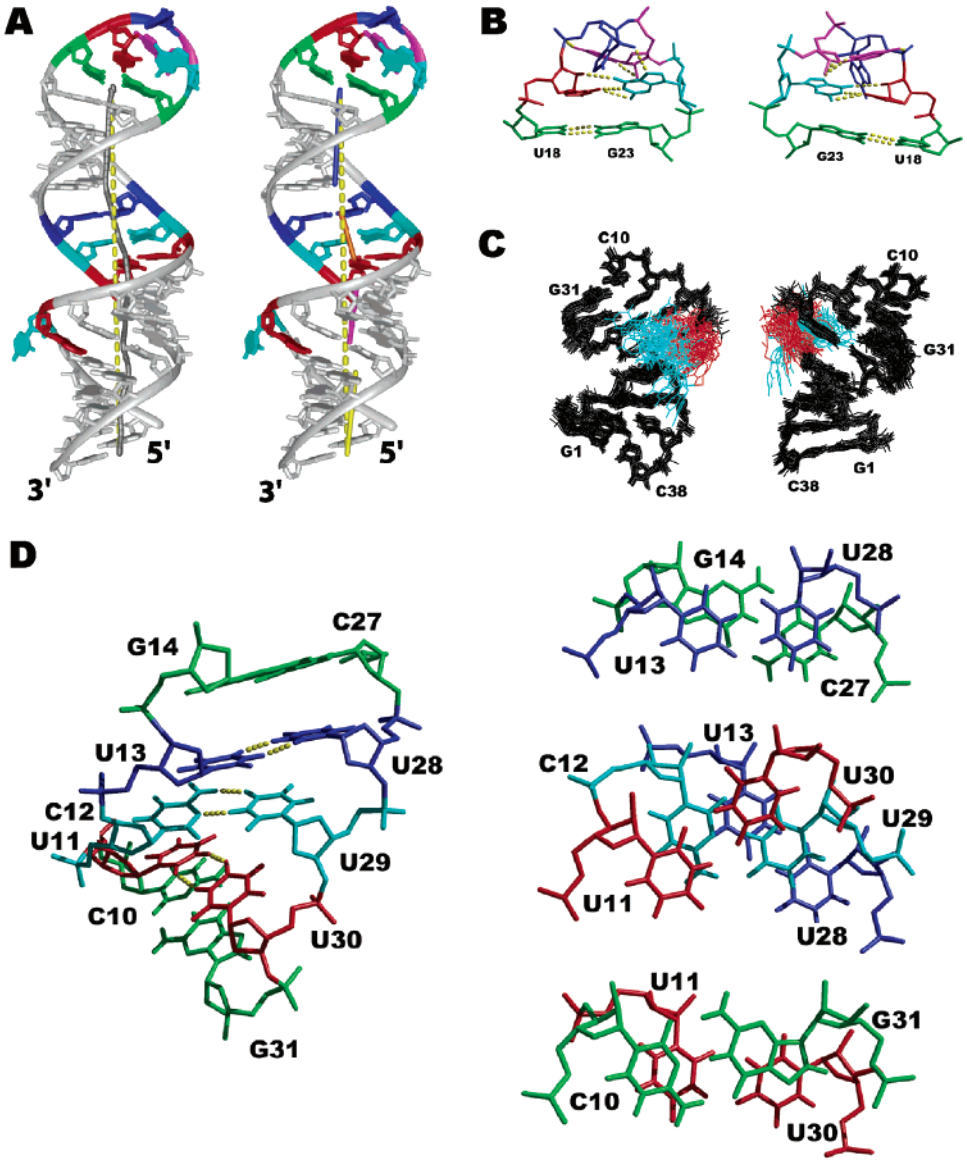

FIGURE 6: (A) Stereoview of the ribbon representation of the lowest-energy structure refined with the RDC restraints. Here, and in panels B and C, the closing wobble U18-G23 pair is shown in green, residues U19, A20, C21, and G22 are shown in red, blue, magenta, and cyan, respectively. Bulge residues U5 and A6 are cyan and red, respectively; all other residues are gray. The helical axes for the four helical regions, as well as the overall axis (yellow dashed line), are also shown. These axes were calculated with CURVES (see text). (B) The apical tetraloop uUACGg including the U18-G23 closing base pair in the lowest-energy structure, showing the minor (left) and major (right) grooves. Only heavy atoms are shown. (C) Structure of the lower stem and the bulge showing the superimposition of the lower stem and bulge (residues G1−C10 and G31−C38) of the 25 lowest-energy structures of the RDC-refined ensemble. All residues shown, except U5 and A6, were used for superimposition. U5 and A6 are colored cyan and red, respectively, and other residues are black; only heavy atoms are shown. (D) The mismatch region in the lowest-energy structure. (Left) The three pyrimidine-pyrimidine base pairs and the two flanking GC base pairs, view to the minor groove. The hydrogen bonds in the mismatch pairs are indicated as yellow dotted lines. (Right) Base stacking in the mismatch region. Top panel, base stacking between G14-C27 and U13−U28. Middle panel, three pyrimidine-pyrimidine pairs; note the absence of the intra-strand base stacking. Bottom panel, base stacking between U11−U30 and C10-G31. The two flanking base pairs G14−C27 and C10−G31 are in green; U11−U30, C12−U29 and U13−U28 are in red, cyan and blue, respectively.

stabilized by two hydrogen bonds: from H3 of U11/U28 to O4 of U30/U13, and from H3 of U30/U13 to O2 of U11/U28 (Figure 6D). The C12−U29 base pair also involves two hydrogen bonds: from the amino group of C12 to O4 of U29, and from H3 of U29 to N3 of C12 (Figure 6D). This base pair is identical to the corresponding base pair in a 30-nucleotide CVB3 stem-loop D RNA in which the C12N3−U29H3 hydrogen bond was directly detected by a 2D HNN−COSY experiment (*26*). A similar UC base pair has also been observed in a human telomerase RNA hairpin (*24*).

The geometry of the mismatch region belongs to the A-form family, as judged by values of several characteristic local and global helical parameters including negative slide and X-displacement, and positive inclination and roll. As expected, deviations from standard A-form do exist. Each of the three pyrimidine base pairs exhibits an opening of ~13−14° toward the major groove, while the standard Watson−Crick base pairs have an opening of 2−3°. Second, unlike the classical A-form structure, the mismatch region shows relatively little intra-strand stacking (Figure 6D). Third, the UU and UC pairs are smaller: the C1′−C1′

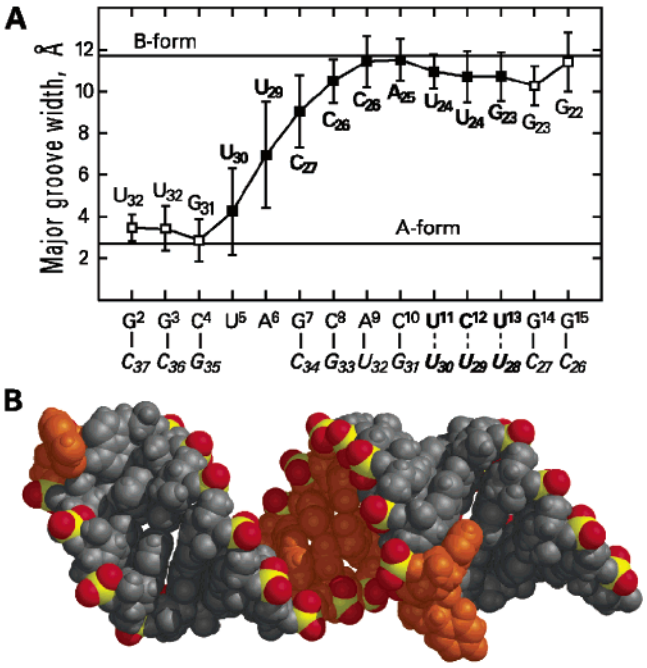

FIGURE 7: (A) Major groove width in the NMR structure of the consensus stem-loop D RNA, measured as the closest inter-phosphorus separation minus 5.6 Å. For each inter-phosphorus distance plotted, the first residue is shown at the *x*-axis (with the base-paired residue below in italic), and the second residue is shown inside the plot panel. Average values with the error bars representing the standard deviations among the 25 refined structures are plotted. Filled boxes correspond to the inter-phosphorus distances spanning the major groove of the mismatch region. Major groove width in canonical A- and B-forms is shown as straight line for comparison. Note that the phosphates of bulge residues U5 and A6 also participate in forming the major groove, although the corresponding distances have larger error bars since residue positions are less defined. (B) Space-filling representation of consensus stem-loop D showing the wide major groove of the mismatch region (in the middle). The bulge residues, pyrimidine-pyrimidine base pairs and A20 of the apical loop are colored orange; all other residues are gray. Phosphate oxygens are red, and phosphorus atoms are yellow.

distance for each mismatch base pair (~8.6 Å) is about 2 Å shorter than the distance for a standard Watson−Crick base pair (~10.7 Å); the C6−C6 distance is also about 1 Å shorter than the C8−C6 distance of standard Watson−Crick base pairs. The minor groove width is about 2.5 Å narrower and noticeably deeper than in the ideal A-form. The most remarkable feature of the pyrimidine mismatch region, however, is a dramatic widening of the major groove, which attains a width more typical of B-conformations (Figure 7A). In addition, the mismatch region is strongly under-twisted, with the average local helical twist 23° and 13° for the U11pC12 and C12pU13 steps, respectively (calculated with the 3DNA program).

Structure of the Lower Stem and the Two-Nucleotide Bulge. As initially implicated by the RDC data (Figure 4B), the structure of the two-nucleotide bulge is not very well defined, most likely due to significant internal motion. In agreement with the attenuation in measured RDCs, residue U5 is the least defined in the molecule (Figure 6C). The base moiety of U5 samples a wide range of conformational space from inside of the major groove to the surface of the minor groove; most often, it is hovering over the backbone or pointing toward the minor groove side (Figures 6A and 6C). Compared to U5, A6 is defined relatively better. The base moiety of A6 is most often positioned on the minor groove side and partially stacked underneath the base of G7 (Figures 6A and 6C). Although the bulged nucleotides are not well defined, the flanking base pairs are intact. The existence of the C4-G35 and G7-C34 base pairs is unambiguously confirmed by the observation of imino proton resonances of G35 and G7 (Figure 2, parts A and C) and direct detection of their hydrogen bonds in the 2D HNN−COSY spectrum (Figure 2D). The presence of the U5A6 dinucleotide bulge results in higher values for twist (59 ± 7°), rise (4.1 ± 0.5 Å), tilt (7 ± 4°), and roll (22 ± 3°) at the C4G7/C34G35 step (calculated with the 3DNA program). While continuous stacking is still present on the nonbulged strand between C34 and G35, no base stacking is observed between C4 and G7 on the bulged strand (Figure 6C). The two helical segments separated by the bulge assume a bend angle of about 30−35°.

DISCUSSION

Structural Features of the Cloverleaf Stem-Loop D RNA. Several studies conclude that using RDC data significantly improves solution structures of RNA (*32*−*34*), especially for elongated RNA with helical domains separated by irregular elements such as bulges or internal loops. The presence of the pyrimidine mismatch region and dinucleotide bulge in the stem-loop D made it impossible to define the relative orientation of the upper and lower stems without RDC restraints. Including RDCs in the refinement resulted in a significantly better definition of the global conformation, even though no artificial backbone torsion angle restraints were used for the mismatch and bulge regions (Figure 5). Interestingly, the overall shapes of the *average* structures refined with and without RDCs are quite similar, apart from a difference in the overall length (Figure 5C). Apparently, structures refined without RDC restraints assume a variety of fully randomized orientations with no preference for a particular bending direction.

Less expected, inclusion of RDC restraints also had a significant impact on the definition of local structures, in particular, for the pyrimidine mismatch region, which is absolutely conserved in stem-loop D in enteroviruses and rhinoviruses (*5*). The hallmark of this structure is an overall A-like conformation, but with a dramatically widened major groove, due to the smaller size of pyrimidine−pyrimidine base pairs. Because of the Watson−Crick-type configuration of these mismatches, all six O2 carbonyl atoms of pyrimidines are positioned in the minor groove, and all five O4 carbonyl atoms of uridines and the N4 amino group of C12 are positioned in the major groove. Both grooves thus have potential to be involved in electrostatic and/or hydrogen bonding interactions (Figure 8). However, the major groove is the most striking feature, which is comparable in width with B-conformations (Figure 7) and thus is capable of accommodating even a protein α-helix.

The apical loop is structurally conserved and belongs to the expanded UNCG-like tetraloop family. Together with the structure reported here, we have determined by NMR three structures of this family (*23*, *35*). Despite sequence variations, the structures are very similar; they are stabilized by the same set of hydrogen bonds and stacking-interactions as found in previously determined UNCG-like tetraloops

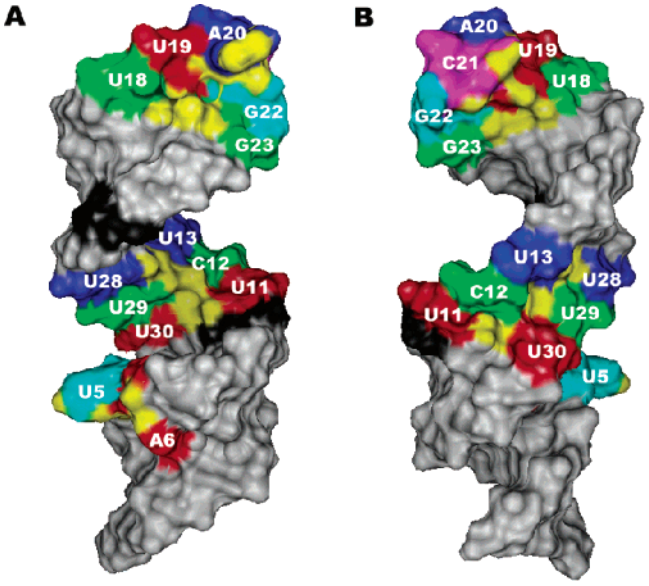

FIGURE 8: Surface representation of the consensus stem-loop D structure. (A) View of the minor groove of the apical loop and the major groove of the pyrimidine mismatch region (orientation similar to Figure 7B). (B) View of the major groove of the apical loop and the minor groove of the mismatch region (~180° rotation about the helical axis from A). Coloring schemes in both panels are: U5 cyan, A6 red, U11−U30 base pair red, C12−U29 base pair green, U13−U28 base pair blue, U18-G23 wobble base pair green, U19 red, A20 blue, C21 magenta, and G22 cyan. Potential hydrogen bond donor and acceptor atoms of these bases are colored yellow. C10 and C26 are black; all other residues are gray.

(*27*−*29*, *36*). The apical loop RNA from PV1 must also have a similar structure based on the characteristic set of unusual chemical shifts of the loop nucleotides. Structural conservation of these tetraloops provides evidence for the expansion of the UNCG tetraloop family to the more extensive YNMG family (Y is U or C, N is any residue, and M is A or C). This extension was deduced from a set of RNA tetraloops selected in vitro based on their thermal stability (*37*), although no tetraloop structure has been solved so far with an adenine in the third position. In addition, structures solved for the variants of the enteroviral stem-loop D show that this family can also accommodate the closing wobble UG pair, instead of a more common Watson−Crick CG (this work; (*23*, *26*)). The identity of the second tetraloop position (A20 in CVB3 and in the consensus construct) is not structurally important. Nevertheless, the cloverleaf stem-loop D has a strong preference for purines (52% A, 23% G) at this position (*5*). Partial conservation of a structurally insignificant position suggests that this residue may contribute to recognition of stem-loop D by the 3C/3CD proteins.

The presence of a bulge near the base of the stem, but not its sequence, is also a highly conserved feature of the stem-loop D RNA domain (*5*). The structure of the bulge with a quite variable U5 does not provide a clear structural explanation for the high degree of conservation. Nevertheless, the bulge creates a distinctive local structure in this region. The bulge marks the starting point of a dramatic increase in major groove width (Figure 7A), and its residues may be used to cap a major groove cavity for a putative interaction with protein secondary structure elements. It is interesting, that although the bulge has an irregular structure, the bulge residues seem to continue the helical arrangement of phosphate groups, which are part of the wide major groove encompassing the mismatch region (Figure 7).

A structure of a shorter 30-nucleotide fragment of the CVB3 cloverleaf stem-loop D has been solved recently by NMR without RDC restraints but with an extraordinarily high number of NOE restraints (*26*); this fragment lacked the dinucleotide bulge and the lower stem. Similar to the structure reported here, the 30-nucleotide CVB3 fragment had a wide major groove in the pyrimidine mismatch region.

Possible Effects of Conformational Averaging. When multiple conformations contribute to the observed NMR signal, the derived structural restraints represent population-weighted averaged values. In the worst-case, these values may not correspond to a single physically meaningful structure, so attempts to impose simultaneously all restraints may lead to structural artifacts. In stem-loop D, in addition to the terminal base pair, an apparently conformationally flexible site is the dinucleotide bulge: U5 has a variable orientation, judging from the attenuation of its RDC values, and A6 has a flexible sugar ring. To reduce averaging artifacts, we did not use RDC restraints for the bulge and sugar pucker restraints for A6, although we used NOE restraints. In the final ensemble of 25 structures, indeed, U5 adopts a variety of orientations; in 37% of the structures, A6 has C2′-*endo* sugar pucker. Although this structural variability agrees with experimental RDC and TOCSY data qualitatively, it may or may not reflect the true flexibility of the bulge quantitatively. Rigorously, variability observed in the 25 refined structures only reflects the degree to which the structure is underdetermined by the experimental restraints used. Determination of true solution conformers of the bulge would involve using methods such as molecular dynamics with time averaging of restraints or multiple-copy refinement (see, e.g., (*38*, *39*)).

Possible Implications for Binding Viral Protein 3C/3CD. Biochemical data have established that the cloverleaf stem-loop D domain is the major determinant for viral 3C proteinase binding. Using a mammalian cell-based RNA binding assay (*40*), it was shown that specific RNA binding by the *Poliovirus* protein 3C depends on the integrity of stem-loop D of the *Poliovirus* genomic RNA. Similarly, CVB3 stem-loop D RNA is the recognition element for CVB3 3C protein binding. Furthermore, a 38-nucleotide RNA construct (nearly identical to our CVB3 construct shown in Figure 1C) was sufficient for binding the 3C protein, with a K_d of 4.6 μM (determined by in vitro filter binding assay), similar to that of the entire cloverleaf RNA (*5*). Three distinctive structural elements within the stem-loop D domain reported here may serve as sites for specific interactions with the 3C/3CD protein. These are the apical loop, the pyrimidine mismatch region, and the dinucleotide bulge. As discussed above, the mismatch region, and presence of the tetraloop and bulge are highly conserved among enteroviral and rhinoviral cloverleaf RNA; in addition, the second position of the tetraloop (A20) is partially conserved (*5*). In a 30-nucleotide CVB3 construct lacking the bulge and lower stem, the tetraloop with the closing UG pair and two residues flanking the mismatch region are affected by addition of the 3C protein, based on NMR line-broadening and chemical shift changes (*26*). Based on these data, Ohlenschlager et al. suggested that the minor groove side of the tetraloop and the major groove side of the

mismatch region form a bipartite recognition site for 3C (*26*). The residues implicated correspond to residues U18 through G23, C10 and C26 in our construct. Because of the specific sequence spacing between these elements, the major groove of the mismatch region, the bulge, and the A20 base are positioned approximately on the same side of stem-loop D (colored orange in Figure 7B). It is possible that the bulge residues are also part of this extended recognition site. The major groove in typical A-form helices is too deep and narrow to allow for interaction with protein secondary structures. However, junctions, bulges and mismatches are often used in RNA as a strategy to widen the major groove and make it accessible for a protein (see, e.g., (*41*)), with a possible induced fit of an RNA and/or protein structure (*42*). In this case, three pyrimidine-pyrimidine base pairs are largely responsible for increasing the major groove width (Figure 7A). The PV1 protease 3C residues implicated in the stem-loop D recognition include the conserved KFRDIR sequence of the domain linker, and amino acids from helix C and two hairpin turns (*3*, *43*). These amino acids form a continuous cluster on the surface of the crystal structure of the PV1 3C protein (*43*), and they can potentially make contacts with available phosphate groups (Figure 7B) and/or functional groups of RNA bases (Figure 8). The longest span of implicated amino acids, ca. 30 Å, is, however, significantly shorter than the longest span of implicated RNA residues, ca. 40 Å (data not shown). Nevertheless, the role of some RNA residues may be structural rather than providing direct interaction with amino acids. For example, tetraloop residues and the closing wobble pair are responsible for extending the wide major groove beyond the mismatch region (Figure 7A). Alternatively, some residues, e.g., bulge residues and the partially conserved A20, may be involved in the high-affinity PCBP-2/3CD/cloverleaf ternary complex, which is the one considered biologically relevant (*44*). Of course, additional studies of RNA-3C/3CD interactions are required; the stem-loop D RNA structure reported here can be used as model to design such studies.

ACKNOWLEDGMENT

We acknowledge use of the UCSF Computer Graphics Laboratory (supported by NIH Grant P41 RR-01081).

SUPPORTING INFORMATION AVAILABLE

Figures showing a portion of the amino-optimized 2D ^{15}N−HSQC spectrum, an example of identification of the ribose spin system with the help of the 3D HCCH−COSY, HCCH−RELAY and HCCH-TOCSY spectra, correlation of ribose and base nuclei in the triple-resonance 2D HCN spectra, 2D H(N)CO spectrum for establishing C2 and C4 chemical shifts in uracils, 2D ^{13}C−HSQC spectra for the RDC measurements, plot of predicted vs experimental RDC values, and tables with chemical shift assignments. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Gamarnik, A. V., and Andino, R. (1998) Switch from translation to RNA replication in a positive-stranded RNA virus, *Gen. Dev. 12*, 2293−304.
2. Andino, R., Rieckhof, G. E., Trono, D., and Baltimore, D. (1990) Substitutions in the protease (3Cpro) gene of poliovirus can suppress a mutation in the 5′ noncoding region, *J. Virol. 64*, 607−12.
3. Andino, R., Rieckhof, G. E., Achacoso, P. L., and Baltimore, D. (1993) Poliovirus RNA synthesis utilizes an RNP complex formed around the 5′-end of viral RNA, *EMBO J. 12*, 3587−98.
4. Walker, P. A., Leong, L. E., and Porter, A. G. (1995) Sequence and structural determimants of the interaction between the 5′-noncoding region of picornavirus RNA and rhinovirus protease 3C, *J. Biol. Chem. 270*, 14510−6.
5. Zell, R., Sidigi, K., Bucci, E., Stelzner, A., and Görlach, M. (2002) Determinants of the recognition of enteroviral cloverleaf RNA by coxsackievirus B3 proteinase 3C, *RNA 8*, 188−201.
6. Milligan, J. F., and Uhlenbeck, O. C. (1989) Synthesis of small RNAs using T7 RNA polymerase, *Methods Enzymol. 180*, 51−62.
7. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes, *J. Biomol. NMR 6*, 277−93.
8. Goddard, T. D., and Kneller, D. G. (1998) *SPARKY*, version 3.0, University of California, San Francisco, CA.
9. Marion, D., Ikura, M., Tschudin, R., and Bax, A. (1989) Rapid recording of 2D NMR spectra without phase cycling. Application to the study of hydrogen exchange in proteins, *J. Magn. Reson. 85*, 393−9.
10. Smallcombe, S. H. (1993) Solvent suppression with symmetrically shifted pulses, *J. Am. Chem. Soc. 115*, 4776−85.
11. Shaka, A. J., and Freeman, R. (1985) A composite 180° pulse for signal localization, *J. Magn. Reson. 63*, 596−600.
12. Bax, A., Clore, G. M., and Gronenborn, A. M. (1990) ^{1}H-^{1}H correlation via isotropic mixing of ^{13}C magnetization, a new three-dimensional approach for assigning ^{1}H and ^{13}C spectra of ^{13}C-enriched proteins, *J. Magn. Reson. 88*, 425−31.
13. Clore, G. M., Bax, A., Driscoll, P. C., Wingfield, P. T., and Gronenborn, A. M. (1990) Assignment of the side-chain ^{1}H and ^{13}C resonances of interleukin-1β using double- and triple-resonance heteronuclear three-dimensional NMR spectroscopy, *Biochemistry 29*, 8172−84.
14. Shaka, A. J., and Freeman, R. (1983) Composite pulses with dual compensation, *J. Magn. Reson. 55*, 487−93.
15. Dingley, A. J., Masse, J. E., Peterson, R. D., Barfield, M., Feigon, J., and Grzesiek, S. (1999) Internucleotide scalar couplings across hydrogen bonds in Watson−Crick and Hoogsteen base pairs of a DNA triplex, *J. Am. Chem. Soc. 121*, 6019−27.
16. Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.
17. Güntert, P., Mumenthaler, C., and Wüthrich, K. (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA, *J. Mol. Biol. 273*, 283−98.
18. Schwieters, C. D., Kuszewski, J. J., Tjandra, N., and Clore, G. M. (2003) The Xplor-NIH NMR molecular structure determination package, *J. Magn. Reson. 160*, 66−74.
19. Ferrin, T. E., Huang, C. C., Jarvis, L. E., and Langridge, R. (1988) The MIDAS display system, *Mol. Graph. 6*, 13−27.
20. Huang, C. C., Couch, G. S., Pettersen, E. F., and Ferrin, T. E. (1996) Chimera: An extensible molecular modeling application constructed using standard components, *Pacific Symp. Biocomput. 1*, 724.
21. Lu, X., and Olson, W. K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures, *Nucleic Acids Res. 31*, 5108−21.
22. Lavery, R., and Sklenar, H. (1990) *CURVES*, version 5.1, Laboratory for Theoretical Biology, CNRS, Paris.
23. Du, Z., Yu, J., Andino, R., and James, T. L. (2003) Extending the family of UNCG-like tetraloop motifs: NMR structure of a CACG tetraloop from coxsackievirus B3, *Biochemistry 42*, 4373−83.
24. Theimer, C. A., Finger, L. D., Trantirek, L., and Feigon, J. (2003) Mutations linked to dyskeratosis congenita cause changes in the structural equilibrium in telomerase RNA, *Proc. Natl. Acad. Sci. U.S.A. 100*, 449−54.
25. Fürtig, B., Richter, C., Wöhnert, J., and Schwalbe, H. (2003) NMR Spectroscopy of RNA, *Chem. Biol. Chem. 4*, 936−62.
26. Ohlenschläger, O., Wöhnert, J., Bucci, E., Seitz, S., Häfner, S., Ramachandran, R., Zell, R., and Görlach, M. (2004) The structure of the stemloop D subdomain of coxsackievirus B3 cloverleaf RNA and its interaction with the proteinase 3C, *Structure 12*, 237−48.
27. Allain, F. H., and Varani, G. (1995) Structure of the P1 helix from group I self-splicing introns, *J. Mol. Biol. 250*, 333−53.

28. Cheong, C., Varani, G., and Tinoco, I., Jr. (1990) Solution structure of an unusually stable RNA hairpin, 5′GGAC(UUCG)GUCC, *Nature 346*, 680−2.
29. Varani, G., Cheong, C., and Tinoco, I., Jr. (1991) Structure of an unusually stable RNA hairpin, *Biochemistry 30*, 3280−9.
30. Rückert, M., and Otting, G. (2000) Alignment of biological molecules in novel nonionic liquid crystalline media for NMR experiments, *J. Am. Chem. Soc. 122*, 7793−7.
31. Al-Hashimi, H. M., Gosser, Y., Gorin, A., Hu, W. D., Majumdar, A., and Patel, D. J. (2002) Concerted motions in HIV-1 TAR RNA may allow access to bound state conformations: RNA dynamics from NMR residual dipolar couplings, *J. Mol. Biol. 315*, 95−102.
32. McCallum, S. A., and Pardi, A. (2003) Refined solution structure of the iron-responsive element RNA using residual dipolar couplings, *J. Mol. Biol. 326*, 1037−50.
33. Warren, J. J., and Moore, P. B. (2001) Application of dipolar coupling data to the refinement of the solution structure of the Sarcin-Ricin loop RNA, *J. Biomol. NMR 20*, 311−23.
34. Sibille, N., Pardi, A., Simorre, J. P., and Blackledge, M. (2001) Refinement of local and long-range structural order in theophylline-binding RNA using (13)C−(1)H residual dipolar couplings and restrained molecular dynamics, *J. Am. Chem. Soc. 123*, 12135−46.
35. Comolli, L. R., Ulyanov, N. B., Soto, A. M., Marky, L. A., James, T. L., and Gmeiner, W. H. (2002) NMR structure of the 3′ stem-loop from human U4 snRNA, *Nucleic Acids Res. 30*, 4371−9.
36. Ennifar, E., Nikulin, A., Tishchenko, S., Serganov, A., Nevskaya, N., Garber, M., Ehresmann, B., Ehresmann, C., Nikonov, S., and Dumas, P. (2000) The crystal structure of UUCG tetraloop, *J. Mol. Biol. 304*, 35−42.
37. Proctor, D. J., Schaak, J. E., Bevilacqua, J. M., Falzone, C. J., and Bevilacqua, P. C. (2002) Isolation and characterization of a family of stable RNA tetraloops with the motif YNMG that participate in tertiary interactions, *Biochemistry 41*, 12062−75.
38. Yao, L. J., James, T. L., Kealey, J. T., Santi, D. V., and Schmitz, U. (1997) The dynamic NMR structure of the TΨC-loop: Implications for the specificity of tRNA methylation, *J. Biomol. NMR 9*, 229−44.
39. Görler, A., Ulyanov, N. B., and James, T. L. (2000) Determination of the populations and structures of multiple conformers in an ensemble from NMR data: Multiple-copy refinement of nucleic acid structures using floating weights, *J. Biomol. NMR 16*, 147−64.
40. Blair, W. S., Parsley, T. B., Bogerd, H. P., Towner, J. S., Semler, B. L., and Cullen, B. R. (1998) Utilization of a mammalian cell-based RNA binding assay to characterize the RNA binding properties of picornavirus 3C proteinases, *RNA 4*, 215−25.
41. Draper, D. E. (1999) Themes in RNA-protein recognition, *J. Mol. Biol. 293*, 255−70.
42. Williamson, J. R. (2000) Induced fit in RNA-protein recognition, *Nat. Struct. Biol. 7*, 834−7.
43. Mosimann, S. C., Cherney, M. M., Sia, S., Plotch, S., and James, M. N. (1997) Refined X-ray crystallographic structure of the poliovirus 3C gene product, *J. Mol. Biol. 273*, 1032−47.
44. Andino, R., Rieckhof, G. E., and Baltimore, D. (1990) A functional ribonucleoprotein complex forms around the 5′ End of poliovirus RNA, *Cell 63*, 369−80.

BI048973P