Published in final edited form as:

Biochemistry. 2017 May 23; 56(20): 2549–2557. doi:10.1021/acs.biochem.7b00141.

Structure of RNA Stem Loop B from the Picornavirus Replication Platform

Meghan S. Warden[†], Marco Tonelli[‡], Gabriel Cornilescu[‡], Dong Liu[†], Lorelei J. Hopersberger[†], Komala Ponniah[†], and Steven M. Pascal^{*,†}

[†]Department of Chemistry and Biochemistry, Old Dominion University, Norfolk, Virginia 23529, United States

[‡]National Magnetic Resonance Facility at Madison (NMRFAM), University of Wisconsin–Madison, Madison, Wisconsin 53706, United States

Abstract

The presumptive RNA cloverleaf at the start of the 5′-untranslated region of the picornavirus genome is an essential element in replication. Stem loop B (SLB) of the cloverleaf is a recognition site for the host polyC-binding protein, which initiates a switch from translation to replication. Here we present the solution structure of human rhinovirus isotype 14 SLB using nuclear magnetic resonance spectroscopy. SLB adopts a predominantly A-form helical structure. The stem contains five Watson–Crick base pairs and one wobble base pair and is capped by an eight-nucleotide loop. The wobble base pair introduces perturbations into the helical parameters but does not appear to introduce flexibility. However, the helix major groove appears to be accessible. Flexibility is seen throughout the loop and in the terminal nucleotides. The pyrimidine-rich region of the loop, the apparent recognition site for the polyC-binding protein, is the most disordered region of the structure.

Graphical abstract

***Corresponding Author:** Department of Chemistry and Biochemistry, Old Dominion University, 4541 Hampton Blvd., Norfolk, VA 23529. Telephone: 757-683-6763. spascal@odu.edu.

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.biochem.7b00141. Plot comparing experimental and calculated RDCs, comparison of SLB and 5′CL ¹H–¹H NOESY spectra, the aromatic region of the ¹H–¹³C HSQC spectrum, and the secondary structure of SLB (PDF)

ORCID

Gabriel Cornilescu: 0000-0002-1204-8904
Steven M. Pascal: 0000-0002-9492-6167

Notes The authors declare no competing financial interest.

Human rhinoviruses (HRVs) are members of the Picornaviridae family that consists of more than 28 species grouped into 12 genera.¹ Picornaviruses are responsible for a diverse number of human and animal diseases, including poliomyelitis, hepatitis A, foot and mouth disease, and the common cold.² Genetic information is carried as a small single-stranded RNA molecule (~7–8.5 kb) that is replicated via a highly conserved mechanism.³

The rhinovirus genomic RNA consists of three distinct regions: the 5′ and 3′ untranslated regions (UTRs) flanking a gene encoding a single ~250 kDa polyprotein. After translation, the polyprotein is cleaved into several structural and nonstructural proteins by virus proteases that are part of the polyprotein. Most of these cleavages are performed by the 3C protease, which also performs a second task: some 3C protease remains covalently joined to the virus-encoded RNA-dependent RNA polymerase (3D) to form the 3CD fusion protein. The 3CD protein directly interacts with the 5′UTR as an essential step in replication.^{4–8} The 5′UTR itself contains two regions: the large (typically 400–500 bases) internal ribosome entry site (IRES) that directs translation and a smaller (<85 bases) region, predicted to form a cloverleaf [5′CL (see Figure 1)] in most picornaviruses, that serves as a platform for RNA replication.^{9–11} Stem loop D (SLD) from the 5′CL attracts the 3CD fusion protein,⁵ effectively delivering the 3D RNA polymerase to the required site to begin replication. The C-rich region of stem loop B (SLB) attracts the host polyC-binding protein (PCBP) that helps to circularize the genomic RNA.^{12,13}

The interactions of 5′CL with 3CD and with PCBP mentioned above are not independent. PCBP facilitates both translation and replication, by binding C-rich loops in both the IRES and the 5′CL.^{5,14–17} However, PCBP can be cleaved by 3C, and the product subsequently functions only in replication, effecting a switch from translation to replication.^{18,19} Furthermore, interaction of PCBP with 5′CL SLB facilitates binding of 3CD to SLD. A better understanding of the structural basis for these interactions and their interplay could lead to ways to inhibit virus replication.

In this study, the nuclear magnetic resonance (NMR) solution structure of a 24-nucleotide RNA hairpin representing 5′CL SLB (Figure 1) from isotype 14 of human rhinovirus

(HRV-14) was determined. The structure contains five stable Watson–Crick base pairs and one wobble base pair with an eight-nucleotide C-rich hairpin loop. The loop is highly flexible, with the C-rich region exposed, while the stem region presents an accessible major groove that may be important for protein interactions. This study represents an important step in the process of building a high-resolution structural understanding of picornavirus replication.

MATERIALS AND METHODS

RNA Preparation

A natural abundance RNA sample, 5'GCGGAUGGGUAUCCCACCAUUCGA^{3'}, corresponding to HRV-14 SLB, was produced by *in vitro* transcription using T7 RNA polymerase (expressed and purified in house) with RNAPoly reaction buffer [New England Biolabs; 40 mM Tris-HCl, 6 mM MgCl₂, 10 mM DTT, and 2 mM spermidine (pH 7.9)], each rNTP at 5 mM (Sigma), 1 μ M double-stranded DNA (as a template), and 0.001 unit/mL pyrophosphatase (Fisher). A 20 mL transcription reaction mixture was incubated at 37 °C for 3 h; then 2.5 volumes of 100% ethanol and 0.1 volume of 3 M sodium acetate were added to the reaction mixture, and the mixture was incubated overnight at –20 °C. The solution was spun down at 9200g and 4 °C. The resulting pellet was washed in 70% ethanol and spun down for an additional 10 min at 9200g. The pellet was dissolved in 8 M urea loading buffer [8 M urea, 25 mM EDTA, 500 mM Tris-HCl, 4.6 mM xylene cyanol, and 3.7 mM bromophenol blue (pH 7.9)] and heated for 3 min at 98 °C prior to gel purification. The RNA was resolved on a 15% denaturing polyacrylamide gel containing 8 M urea and visualized via ultraviolet (UV) shadowing at 254 nm, and the corresponding band was excised from the gel. The RNA was recovered from the gel pieces using an Elutrap device (Whatman) in 1× Tris-borate-EDTA (TBE) buffer (pH 7.9). The concentration of the eluted RNA was determined using a Biodrop UV–visible spectrophotometer. The RNA was concentrated and buffer exchanged using a Vivaspinn spin concentrator (3 kDa cutoff) to a concentration of 0.5 mM in NMR buffer [10 mM sodium phosphate and 0.1 mM EDTA (pH 7.0)].

A uniformly ¹³C- and ⁵N-labeled RNA sample of SLB was similarly produced using each ¹³C- and ⁵N-labeled rNTP at 2.5 mM (Cambridge Isotope Laboratories) and 0.00075 unit/mL pyrophosphatase.

NMR Spectroscopy

NMR spectra of the unlabeled SLB sample were a one-dimensional (1D) ¹H spectrum, ¹H–¹H NOESY in H₂O and D₂O, and ¹H–¹H TOCSY spectra. The spectra obtained with the labeled sample were ¹H–¹³C and ¹H–¹⁵N HSQC, ¹H–¹³C ARTSY, ¹H–¹⁵N ARTSY, and ¹H–¹H–¹³C-edited three-dimensional NOESY-HSQC spectra. Two copies of each spectrum measuring ¹³C shifts were obtained, with the ¹³C transmitter centered on the C6/C8/C2 region or the C1'/C5 region.

A 0.5 mM unlabeled RNA sample in 10 mM sodium phosphate and 0.1 mM EDTA (pH 7.0) containing a 90% H₂O/10% D₂O mixture was used for the analysis of exchangeable protons.

For the analysis of nonexchangeable protons, the 0.5 mM sample was vacuum concentrated using a Thermo Scientific SpeedVac concentrator and resuspended in 99.9% D₂O. NMR experiments were performed at 750, 800, and 900 MHz (Varian) at the National Magnetic Resonance Facility at Madison (NMRFAM) equipped with a cryoprobe. For the assignment of nonexchangeable base proton resonances, ¹H–¹H NOESY spectra in D₂O with mixing times of 125 and 250 ms were recorded at 26.5 °C with 256 increments, 4096 data points, 32 scans per fid, and a 3 s recycle time with presaturation for suppression of the water signal. A ¹H–¹H NOESY spectrum in H₂O with a mixing time of 250 ms was recorded at 5 °C using a 1–1 echo NOESY pulse sequence with 256 increments, 4096 data points, and 96 scans for the assignment of exchangeable imino proton resonances. Two ¹H–¹H TOCSY spectra with mixing times of 50 and 200 ms with a 5 kHz spin-lock field were recorded at 26.5 °C to obtain sugar pucker conformations. Two ¹H–¹³C HSQC spectra were recorded for C6/C8/C2 and C1'/C5 with spectral widths of 13020 and 14534 Hz for ¹H and 4000 and 5200 Hz for ¹³C, respectively. Residual dipolar coupling (RDC) experiments were performed on a 750 MHz Bruker spectrometer. The samples contained 18 mg/mL Pf1 filamentous bacteriophage, and one-bond ¹D_{NH} and ¹D_{CH} couplings were measured from ¹H–¹³C and ¹H–¹⁵N ARTSY spectra.²⁰ Data for all spectra were processed using NMRPipe²¹ and visualized and analyzed using NMRViewJ.²²

Structural Constraints

NOE distance constraints were calculated using ¹H–¹H NOESY and three-dimensional ¹³C-edited NOESY-HSQC spectra. The interproton distances calculated for samples in D₂O were calibrated according to the intranucleotide H5–H6 intensities (distance of 2.4 Å) using the equation $I = kr^{-6}$, where r is the distance between two protons. The NOE distances corresponding to assignments in the ¹H–¹H NOESY spectrum in H₂O were calibrated using the intranucleotide H5–H3 distance of uracil (4.1 Å). A 33% upper limit was added to all NOE distances up to a maximal upper limit of 6 Å. Because of the high degree of flexibility associated with the loop and terminal regions, NOE distances corresponding to loop nucleotides (G9–A16) and terminal nucleotides (G1, C2, G23, and A24) were treated more conservatively with a further 1 Å added to the upper limit. The lower limit of all NOE distances was set to 1.8 Å.

Sugar puckers were determined using ¹H–¹H TOCSY spectra. Nucleotides for which strong H1'–H2' cross peaks were observed were constrained to the C2'-endo sugar pucker conformation with ν_0 – ν_4 set to $-4.2 \pm 15.0^\circ$, $24.9 \pm 15.0^\circ$, $-34.9 \pm 15.0^\circ$, $33.0 \pm 15.0^\circ$, and $-18.3 \pm 15.0^\circ$, respectively.²³ Nucleotides without strong H1'–H2' cross peaks were constrained to C3'-endo sugar conformations with ν_0 – ν_4 set to $5.8 \pm 15.0^\circ$, $-26.2 \pm 15.0^\circ$, $36.5 \pm 15.0^\circ$, $-33.0 \pm 15.0^\circ$, and $16.8 \pm 15.0^\circ$, respectively.²³ Nucleotides with intermediate H1'–H2' cross peaks were left unconstrained. On the basis of preliminary structure calculations showing the stem region is consistent with A-form geometry, backbone dihedral angle constraints corresponding to the A-form were applied to nucleotides G3–G8 and C17–U22 with α , β , γ , δ , ϵ , and ζ set to $292 \pm 15.0^\circ$, $178 \pm 15.0^\circ$, $54 \pm 15.0^\circ$, $82 \pm 15.0^\circ$, $207 \pm 15.0^\circ$, and $289 \pm 15.0^\circ$, respectively.²³ CURVES analysis of the initial calculated structures revealed that G4 of the wobble base pair (G4–U21) exhibited atypical α and γ backbone dihedral angles that were correlated with dihedral angle violations in the XPLOR

calculations.^{24,25} Typically, α and γ form a gauche⁻/ gauche⁺ conformation in A-form geometry; however, the α and γ backbone dihedral angles of G4 formed a trans/trans conformation. This conformation is typical of wobble base pairs embedded in the helix of RNA.²⁶ Therefore, α and γ backbone dihedral angles of G4 were left unconstrained in the final structure calculations, and U21 backbone dihedral angles were given an additional tolerance of $\pm 30.0^\circ$ to alleviate any possible strain experienced due to unusual dihedral angles of G4.

Because of the absence of imino resonances for nucleotide G23 in the ^1H - ^1H NOESY spectra, no hydrogen bond constraints for C2-G23 were included in the calculation. The properties of C2 and G23 resemble those observed for the loop nucleotides: C2'-endo sugar conformation assigned to C2 and partial C2'-endo character for G23 with no stable hydrogen bonding. They display a degree of flexibility similar to those of nucleotides G1, G9-A16, and A24 (see Results). Therefore, for NOE constraints involving C2 and G23, an additional 1 Å was added to the upper distance limit, and no backbone dihedral angle constraints were included in the structural calculation for C2 and G23.

Residual dipolar couplings were measured from partial alignment of the labeled SLB sample via 18 mg/mL Pf1 filamentous bacteriophage. One-bond $1D_{\text{NH}}$ and $1D_{\text{CH}}$ couplings were obtained from two-dimensional ARTSY spectra. RDC errors were estimated via comparison between isotropic peak heights, aligned sample peak heights, and the average spectral noise. RDC values for nucleotides G3-G7 and C17-U22 were entered as constraints in the simulated annealing protocol. The RDCs for the loop and terminal regions display values significantly smaller than those of the stem (see Results), most likely as a result of flexibility associated with the loop and terminal nucleotides. Therefore, loop and terminal RDC values were not used as constraints, to prevent incorrect alignment of the associated bonds in these disordered regions.

Structural Calculation and Refinement

A total of 400 structures were calculated using a simulated annealing protocol in XPLOR-NIH. The initial annealing temperature was set to 3000 K with a final temperature of 25 K. The force constants for NOE and hydrogen bond constraints were ramped from 2 to 50 kcal mol⁻¹ Å⁻² and from 0.02 to 5 kcal mol⁻¹ rad⁻² for RDC constraints. The structural calculations contained 257 NOE distance constraints, 115 sugar puckers, 73 backbone dihedral angles, 30 hydrogen bond constraints, 12 planarity constraints, and 37 ^1H - ^{15}N and ^1H - ^{13}C residual dipolar couplings. The 10 lowest-energy structures (see Results) containing no NOE violations of >0.5 Å and no dihedral angle violations of $>5^\circ$ were selected to represent the structural ensemble. The structures were visualized with PyMol and analyzed using CURVES.^{24,27}

RESULTS

Assignment of Imino Protons

The number of observable imino resonances in a 1D ^1H NMR spectrum of RNA is correlated to the number of base pairs. Resonances of imino protons in standard Watson-

Crick base pairs typically appear in the spectral range of ~13–15 ppm (uracil) and ~11–13.5 ppm (guanine). The absence of an imino resonance suggests that the nucleotide is not base-paired, allowing the imino proton to rapidly exchange with the solvent. Imino resonances outside of the spectral ranges mentioned above indicate the presence of non-Watson–Crick base pairs, such as G–U wobble pairs (10–12 ppm).

Cross peaks in the NOESY spectrum should be seen between imino protons from adjacent base pairs, providing a means to walk through the helix and obtain sequential imino assignments. HRV-14 SLB contains 12 imino protons, seven of which were observed in the 1D ^1H spectrum (top of Figure 2) and in the ^1H – ^1H NOESY spectrum in H_2O (Figure 2). The strongest five imino resonances were sequentially assigned to G3, G4, U6, U20, and U21 on the basis of the imino–imino NOE cross peaks labeled in Figure 2. The two weaker resonances were assigned to G7 and G8. These two assignments were made via cross-strand NOEs to C18 and C17 amino protons, respectively. As a result, the moderately weak imino peak (~11.7 ppm) is assigned to G7, and the weakest observed imino is assigned to G8, consistent with the fact that the G8–C17 base pair is the closing base pair. In fact, the G8 imino resonance produced no observable diagonal peak.

We considered the possibility that the broad resonance assigned to the G8 imino proton was instead the G9 imino proton, and that the G8 base is not base-paired to an extent allowing detection of its imino proton. If G8 is extruded from the helix, this could allow the formation of a G9–C17 base pair and an additional base pair could then potentially form between U10 and A16 to extend the helix (Figure S4). However, the base–sugar NOE walk (Figure 3) is continuous through the sequence as currently assigned. Such continuity would not be expected if the G8 base were extruded. Furthermore, the U10 and A16 peak characteristics (see Figure 5) are consistent with disorder, not with stable base pair formation. Also, the U10 imino peak is not observed. For these reasons, we maintained the assignment of the broadest observed imino proton to G8.

Therefore, the structural calculations included hydrogen bond constraints for the following base pairs: G3–C22, G4–U21, A5–U20, G7–C18, and G8–C17. To account for broadening of the G8 imino proton, the hydrogen bond parameters for this pair were weakened as described by Baba et al.²⁸ The absence of the imino resonances for G1, G9, U10, U12, and G23 indicates that these nucleotides are not involved in stable hydrogen bonding.

Assignment of Sugar and Nonexchangeable Base Protons

Intranucleotide and sequential $\text{H1}'$ – $\text{H6}/\text{H8}$ NOEs provide a means to “walk” down a polynucleotide chain to obtain sequential chemical shift assignments. The walk region of the SLB 250 ms mixing time ^1H – ^1H NOESY spectrum in D_2O (Figure 3) was used to assign the $\text{H1}'$, H2, H5, H6, and H8 resonances of nucleotides C2–A24. The corresponding ^{13}C resonances were assigned using ^1H – ^{13}C HSQC spectra. NOESY peak overlap in the U12–C15 region was resolved, and all assignments were confirmed, using ^{13}C -edited three-dimensional NOESY–HSQC spectra. The $\text{H1}'$, H8, $\text{C1}'$, and C8 resonances of G1 were assigned via NOEs to A24, because no G1–C2 NOEs were detected. In all, every nonexchangeable base resonance was assigned along with all sugar $\text{H1}'$ and $\text{H2}'$

resonances. Of the remaining sugar resonances, 19 H3' and 13 H4' resonances were assigned.

Sugar Conformations

Standard A-form RNA contains predominantly C3'-endo sugar conformations that help to create a short sequential phosphate-phosphate distance (~5.8 Å). As a result, A-form RNA exhibits a helical structure more compact than that of the B-form helix that is typical of DNA. The SLB ¹H-¹H TOCSY spectra revealed strong intra-nucleotide H1'-H2' cross peaks for nucleotides G1, C2, G9-A16 and A24, consistent with the C2'-endo sugar conformation at these sites. The absence of intranucleotide H1'-H2' cross peaks for nucleotides G3-G8 and C17-C22 is consistent with the C3'-endo sugar conformation expected in an A-form helix. Nucleotide G23 exhibits weak but detectable H1'-H2' cross peaks typical of an intermediate sugar pucker or equilibrium between C3'-endo and C2'-endo conformations. Therefore, the sugars of nucleotides G3-G8 and C17-C22 were constrained to C3'-endo.²³ Nucleotides G1, C2, G9-A16, and A24 were constrained to C2'-endo.²³ The sugar of G23 was left unconstrained.

Structural Calculations

Structures were calculated using a total of 524 constraints (Table 1). The calculation included 257 NOE-derived distances along with 30 hydrogen bond-derived constraints. The latter represent 15 hydrogen bonds in the 6 bp identified above via imino proton analysis. Twelve planarity constraints, 37 ¹H-¹⁵N and ¹H-¹³C residual dipolar coupling-based constraints, and 188 dihedral angle constraints were applied. The majority of the dihedral constraints specify sugar puckers, but backbone dihedral angles of Watson-Crick base-paired nucleotides determined to have C3'-endo sugar puckers (G3, A5-G8, C17-U20, and C22), were constrained to typical A-form helix ranges.²³ Backbone dihedral angle constraints were not added for G23 because of the observed intermediate H1'-H2' *J* couplings. Initially, nucleotide G4 from the wobble base pair was constrained to A-form backbone dihedral angles, but NOE violations and dihedral angle violations were observed. Therefore, the α and γ G4 backbone dihedral angles were left unconstrained in the final structure calculations. The structures were calculated using a standard simulated annealing protocol in XPLOR-NIH.²⁵ A total of 400 structures were calculated. The 10 lowest-energy structures containing no NOE violations of >0.5 Å, no RDC violations of >3 Hz, and no dihedral angle violations of >5° were selected to represent the structural ensemble (Figure 4).

Solution Structure of SLB

HRV-14 SLB forms a 6 bp helical stem (including G3-G8 and C17-C22) that is largely consistent with A-form geometry, and an eight-nucleotide hairpin loop (G9-A16). The helical region is well-defined, with a pairwise root-mean-square deviation (RMSD) of 0.415 Å, and contains a wide and shallow minor groove (~10.5 Å wide) typical of A-form RNA. However, its major groove (~8.7 Å wide) shares some characteristics with the major groove of B-form DNA, as will be discussed further below. The overall SLB RMSD of 2.249 Å reflects the high degree of conformational variability in the regions outside of the helix.

Each of the loop and terminal nucleotides adopts a C2'-endo sugar conformation, except G23, which is C2'-endo in eight structures and C3'-endo in two structures. This result accurately reflects the observed intermediate G23 H1'-H2' J coupling and the observed strong couplings for the other loop and terminal nucleotides. The helix nucleotides adopt exclusively C3'-endo sugar pucker conformations. This is consistent with the weak H1'-H2' couplings detected for each helix nucleotide, G3-G8 and C17-C22.

Evidence of Disorder in HRV-14 SLB Loop and Termini

Flexible regions within an RNA structure can be identified via analysis of chemical shift, resonance intensity, and residual dipolar couplings. Chemical shifts from highly flexible regions tend toward the chemical shifts of NMPs.^{29,30} The differences between the carbon (¹³C) chemical shifts of SLB and NMP values³¹ are shown in Figure 5A. As expected, the loop region displays chemical shifts more similar to those of the NMPs. The four terminal residues (G1, C2, G23, and A24) were not constrained to A-form geometry because of the absence of observable imino peaks and the presence of strong H1'-H2' couplings. The 3'-terminus does, however, have chemical shifts more closely resembling those of the helical region. This suggests that the terminal residues may have some helical or other partially ordered character.

Flexible regions are also expected to produce narrow, intense resonances.^{29,30,32,33} Figure 5B displays intensities from a ¹H-¹³C HSQC spectrum. Resonance intensities for the loop region and terminal region are noticeably higher than those of the helix. There is a gradual decrease in intensity proceeding down the helix toward the loop, suggesting increased rigidity near the loop. Within the loop, U11-A16 are the most intense, suggesting that this region is the most dynamic on the nanosecond to picosecond time scale. However, the C8 resonance of G9 was not observed. This suggests millisecond to microsecond time scale dynamics causing exchange broadening at the 5' end of the loop.

Disordered residues typically have RDC values lower than those of ordered regions.³⁴⁻³⁷ In Figure 5C, the RDC values for C5, C6, C8, and C2 are presented. The low RDC values in the loop and termini are again consistent with mobility in these regions. Some relatively low RDC values within the helix, such as at G3 and C17, are also consistent with mobility. However, rigid regions may be oriented in such a way that the RDC value is small. That appears to be the case within the helix, because inclusion of these RDCs as orientational constraints resulted in a well-ordered ensemble that is largely consistent with A-form geometry.

In summary, chemical shift, resonance intensity, and RDC values are each consistent with the calculated disorder in the loop and termini. There may be a slight trend, particularly in the 5' strand, showing more order at the end of the helix nearer the loop (see Figure 5). However, all regions of the helix appear to be far more ordered than the loop and termini are.

DISCUSSION

Helical Parameters of SLB

Rigid body parameters, including shear, rise, and twist, can be used to describe the geometry of a base pair and the sequential base pair steps. Typical A-form RNA parameters include the following parameters: base pair shear of ~ 0.0 Å and helical twist of $\sim 31^\circ$.^{38,39} Deviation from these parameters is often observed when a non-Watson–Crick base pair is incorporated into a helix. The most prevalent non-Watson–Crick base pair is the G-U wobble base pair.⁴⁰

The rigid body parameters of the final 10 structures of SLB were analyzed using CURVES.²⁴ In each of the 10 structures, the shear, twist, and rise parameters fall close to standard A-form values, with the following exceptions. (i) The shear values of G4-U21 and A5-U20 base pairs are approximately -1.9 Å and approximately 1.5 Å, respectively, and (ii) the twist values of interbase pair steps G4/A5 and A5/U6 are $\sim 41.9^\circ$ and $\sim 17.0^\circ$, respectively.

Shear is an intrabase pair parameter that is directly related to the hydrogen bonds and indicates the sliding of one base with respect to the other. Therefore, the shear value is correlated with the twist value, the latter of which is measured via the relative position of sequential C1' atoms.⁴¹ The average observed shear value for the G4-U21 wobble base pair (approximately -1.9 Å) is a feature consistent with wobble base pairs presented in the literature, as are unusual twist parameters near a G-U pair.^{42–47} The observed twists can be characterized as an overtwisting of the helix at the G4/A5 step and an undertwisting at the next step, A5/U6. Ananth et al. report that overtwisting ($\sim 47^\circ$) is common at Watson–Crick (WC)/GU base pair steps while undertwisting is often observed in the next base pair step (GU/WC $\sim 19^\circ$).⁴² However, the overtwisting and undertwisting near a G-U pair is often reported at different positions. For instance, Joli et al. reported the WC/GU base pair step experiences an under-twisting while the GU/WC step is slightly overtwisted.⁴⁴ Here, SLB contains a pattern similar to that reported by Ananth et al., an overtwist followed by an undertwist. However, this pattern appears one base pair further down the chain relative to the wobble pair.

To ensure the overtwisting or undertwisting of the helix was not influenced by artificially restricted backbone dihedral angles, 400 additional structures were calculated excluding backbone dihedral angle constraints for nucleotides G3, G4, and A5. In the 10 lowest-energy structures, the helical parameters, including the position of the overtwisting and undertwisting, were similar to those discussed above, and no other noticeable changes were observed.

Major and Minor Groove Width

Typical A-form RNA contains a narrow (~ 3.0 Å) but deep major groove and a wide (~ 11 Å) but shallow minor groove. Therefore, protein–RNA base interactions are largely restricted to the wider minor groove side. In contrast, B-form DNA has a wide (~ 11 Å) major groove that often accommodates protein α helices. The groove width is defined via the shortest cross-strand P–P distances on the relevant side of the double strand.⁴⁸ For A-form RNA, these shortest distances are from P_i to P_{i-6} (major groove) and from P_i to P_{i+3} (minor groove).

In HRV-14 SLB, the minor groove width, as measured from cross-strand P_i-P_{i+3} distances, averages 10.7 Å, which is close to the typical A-form range. However, the shortest cross-strand phosphorus distances in the SLB major groove are from P_i to P_{i-5} and yield an average major groove width of 8.7 Å (Figure 6), atypically large for A-form RNA, despite the fact that other helical parameters are consistent with A-form geometry.

The reason for this discrepancy is the length of the helix. Typical A-form RNA contains 11 bp per turn of the helix, with 7 bp being the minimum required to produce a P_i-P_{i-6} contact to close off the major groove. SLB forms only 6 bp, so the stem, though it follows the pattern of the standard A form, is not long enough to close off access to the major groove. Thus, the SLB major groove, as calculated, is more accessible than is the major groove in a longer A-form stem. This accessibility could play a role in interactions with PCBP or other unidentified proteins and molecules.

Interestingly, the major groove of SLD of the HRV-14 5'CL has previously been shown to contain a similarly accessible major groove,^{49,50} also with the potential to enhance interactions with host or virus proteins. It remains to be seen whether these two major grooves remain as accessible in the context of the intact 5'CL RNA. Also to be determined is their relative orientation, which could influence the juxtaposition of proteins that bind to each of these stem loops, including potentially to their major grooves.

Role of SLB and PCBP

Picornaviruses are positive-sense single-stranded RNA viruses. The known role for SLB from the picornavirus 5'CL in virus replication is to attract the host PCBP.^{12,13} The PCBP then coordinates with host poly-A-binding protein (PABP) that binds to the A-rich 3'UTR of the virus RNA. This interaction effectively circularizes the virus RNA and assists negative strand RNA synthesis, using the positive sense genomic RNA strand as a template.^{15,51} The nascent negative strand is then used as a template to create multiple positive strands, effecting virus replication. Thus, interaction of SLB with PCBP is a critical step in the replication process, and detailed structural knowledge of this interaction and its constituents may lead to ways to block replication.

Human PCBP contains three K-homology (KH) domains, the first and third of which are capable of binding C-rich single-stranded RNA, while the second KH domain is important for dimerization of the protein.⁵² The first KH domain (KH1) is the major determinant of 5'CL binding.⁵³ The presumptive site on SLB for PCBP is the UCCCA sequence (nucleotides 12–16) located in the loop (G9–A16) (Figure 1).

Two crystal structures of the KH1 domain from PCBP, in complex with a single-stranded C-rich DNA strand representing a human telomere, have been determined,^{54,55} along with a third structure in which the DNA is replaced with its RNA equivalent.⁵⁴ In each case, the RNA-binding cleft in PCBP is narrow; therefore, direct contact is limited to a single-stranded nucleic acid sequence of four nucleotides in length. The crystal structures, though similar in many ways, show an interesting difference in the register of the interacting polynucleotide: KH1 contacts the ACCC,⁵⁵ CCCT, and CCCU⁵⁴ sequences in the three different structures. Each crystal structure shows essentially identical interactions between

the KH1 domain and the two central (cytosine) nucleotides from the interacting tetrad. These two cytosine nucleotides hydrogen bond to two guanidinium groups (from KH1 R57 and R40, respectively) via the cytosine hydrogen bond acceptors O2 and N3. Specific interactions were not observed between KH1 and the first nucleotide (A or C) of the tetranucleotide sequence. The last nucleotide hydrogen bonds to KH1 via side chains of E51 (if the nucleotide is C) or via the carbonyl group of I49 (if the nucleotide is T or U). The reason for this apparent difference in register of the CCC sequence is not completely clear at this time.

Mutational analysis has been performed to help identify the site on SLB for PCBP recognition.^{12–14,56,57} PCBP binding is inhibited when the CCC nucleotide sequence (C13–C15) is mutated to CAC, identifying C14 as being important.^{14,56,57} Also, deletion of the CCCA sequence (C13–A16) significantly decreases the level of binding of PCBP to SLB.¹² These studies do not, however, specify whether PCBP contacts the UCCC (nucleotides 12–15) or CCCA (nucleotides 13–16) tetrad in SLB. Some insight can be gained by analysis of the SLB structure. The four consecutive pyrimidine bases located in the loop region (nucleotides 12–15) are solvent-exposed in most of the final calculated structures. The increased flexibility observed within this region of the loop (Figure 5) suggests all or some of these nucleotides would be immediately accessible for protein interactions. Therefore, UCCC may be the tetra-recognition site for PCBP binding. This would place C13 and C14 central in the recognition tetrad, with conserved contacts to the arginine as discussed above for the crystal structures. If this is correct, then C15 would likely interact with the PCBP glutamic acid group, as in the first structure discussed above,⁵⁵ with perhaps nonspecific interactions involving U12. However, structural analysis of SLB with PCBP is needed to confirm the tetra-recognition sequence.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study made use of NMR spectrometers at the National Magnetic Resonance Facility at Madison, which is supported by National Institutes of Health (NIH) Grant P41GM103399 (National Institute of General Medical Sciences). NMR equipment was purchased with funds from the University of Wisconsin–Madison, the NIH (P41GM103399, S10RR02781, S10RR08438, S10RR023438, S10RR025062, and S10RR029220), the National Science Foundation (DMB-8415048, OIA-9977486, and BIR-9214394), and the U.S. Department of Agriculture. The authors thank Samuel Butcher and John Markley.

Funding This research was supported by startup funds and undergraduate research grants from Old Dominion University.

References

1. Knowles, N., Hovi, T., King, A., Stanway, G. The picornaviruses. American Society for Microbiology Press; Washington, DC: 2010. Overview of taxonomy; p. 19–32.
2. Greenberg SB. Respiratory consequences of rhinovirus infection. Arch Intern Med. 2003; 163:278–284. [PubMed: 12578507]
3. Goodfellow IG, Kerrigan D, Evans DJ. Structure and function analysis of the poliovirus cis-acting replication element (CRE). RNA. 2003; 9:124–137. [PubMed: 12554882]

4. Agol VI, Paul AV, Wimmer E. Paradoxes of the replication of picornaviral genomes. *Virus Res.* 1999; 62:129–147. [PubMed: 10507323]
5. Andino R, Rieckhof GE, Baltimore D. A functional ribonucleoprotein complex forms around the 5' end of poliovirus RNA. *Cell.* 1990; 63:369–380. [PubMed: 2170027]
6. Goodfellow I, Chaudhry Y, Richardson A, Meredith J, Almond JW, Barclay W, Evans DJ. Identification of a cis-acting replication element within the poliovirus coding region. *J Virol.* 2000; 74:4590–4600. [PubMed: 10775595]
7. Jacobson SJ, Konings DA, Sarnow P. Biochemical and genetic evidence for a pseudoknot structure at the 3' terminus of the poliovirus RNA genome and its role in viral RNA amplification. *J Virol.* 1993; 67:2961–2971. [PubMed: 8388482]
8. Paul AV, Rieder E, Kim DW, van Boom JH, Wimmer E. Identification of an RNA hairpin in poliovirus RNA that serves as the primary template in the in vitro uridylylation of VPg. *J Virol.* 2000; 74:10359–10370. [PubMed: 11044080]
9. Le SY, Zuker M. Common structures of the 5' non-coding RNA in enteroviruses and rhinoviruses: Thermodynamical stability and statistical significance. *J Mol Biol.* 1990; 216:729–741. [PubMed: 2175364]
10. Pilipenko EV, Blinov VM, Romanova LI, Sinyakov AN, Maslova SV, Agol VI. Conserved structural domains in the 5'-untranslated region of picornaviral genomes: an analysis of the segment controlling translation and neurovirulence. *Virology.* 1989; 168:201–209. [PubMed: 2536978]
11. Rivera VM, Welsh JD, Maizel JV Jr. Comparative sequence analysis of the 5' noncoding region of the enteroviruses and rhinoviruses. *Virology.* 1988; 165:42–50. [PubMed: 2838971]
12. Gamarnik AV, Andino R. Two functional complexes formed by KH domain containing proteins with the 5' noncoding region of poliovirus RNA. *RNA.* 1997; 3:882–892. [PubMed: 9257647]
13. Parsley TB, Towner JS, Blyn LB, Ehrenfeld E, Semler BL. Poly (rC) binding protein 2 forms a ternary complex with the 5'-terminal sequences of poliovirus RNA and the viral 3CD proteinase. *RNA.* 1997; 3:1124–1134. [PubMed: 9326487]
14. Andino R, Rieckhof GE, Achacoso PL, Baltimore D. Poliovirus RNA synthesis utilizes an RNP complex formed around the 5'-end of viral RNA. *EMBO J.* 1993; 12:3587–3598. [PubMed: 8253083]
15. Barton DJ, O'Donnell BJ, Flanagan JB. 5' cloverleaf in poliovirus RNA is a cis-acting replication element required for negative-strand synthesis. *EMBO J.* 2001; 20:1439–1448. [PubMed: 11250909]
16. Blyn LB, Swiderek KM, Richards O, Stahl DC, Semler BL, Ehrenfeld E. Poly(rC) binding protein 2 binds to stem-loop IV of the poliovirus RNA 5' noncoding region: identification by automated liquid chromatography-tandem mass spectrometry. *Proc Natl Acad Sci U S A.* 1996; 93:11115–11120. [PubMed: 8855318]
17. Blyn LB, Towner JS, Semler BL, Ehrenfeld E. Requirement of poly(rC) binding protein 2 for translation of poliovirus RNA. *J Virol.* 1997; 71:6243–6246. [PubMed: 9223526]
18. Back SH, Kim YK, Kim WJ, Cho S, Oh HR, Kim JE, Jang SK. Translation of polioviral mRNA is inhibited by cleavage of polypyrimidine tract-binding proteins executed by polioviral 3C(pro). *J Virol.* 2002; 76:2529–2542. [PubMed: 11836431]
19. Perera R, Daijogo S, Walter BL, Nguyen JH, Semler BL. Cellular protein modification by poliovirus: the two faces of poly(rC)-binding protein. *J Virol.* 2007; 81:8919–8932. [PubMed: 17581994]
20. Ying J, Wang J, Grishaev A, Yu P, Wang YX, Bax A. Measurement of (1)H-(15)N and (1)H-(13)C residual dipolar couplings in nucleic acids from TROSY intensities. *J Biomol NMR.* 2011; 51:89–103. [PubMed: 21947918]
21. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR.* 1995; 6:277–293. [PubMed: 8520220]
22. Norris M, Fetler B, Marchant J, Johnson BA. NMRfX Processor: a cross-platform NMR data processing program. *J Biomol NMR.* 2016; 65:205–216. [PubMed: 27457481]

23. Marino JP, Schwalbe H, Griesinger C. J-Coupling Restraints in RNA Structure Determination. *Acc Chem Res.* 1999; 32:614–623.
24. Lavery R, Moakher M, Maddocks JH, Petkeviciute D, Zakrzewska K. Conformational analysis of nucleic acids revisited: Curves+ *Nucleic Acids Res.* 2009; 37:5917–5929. [PubMed: 19625494]
25. Schwieters CD, Kuszewski JJ, Tjandra N, Marius Clore G. The Xplor-NIH NMR molecular structure determination package. *J Magn Reson.* 2003; 160:65–73. [PubMed: 12565051]
26. Ramos A, Varani G. Structure of the acceptor stem of *Escherichia coli* tRNA Ala: role of the G3.U70 base pair in synthetase recognition. *Nucleic Acids Res.* 1997; 25:2083–2090. [PubMed: 9153306]
27. Delano, WL. The PyMOL Molecular Graphics System. DeLano Scientific; San Carlos, CA: 2002.
28. Baba S, Kajikawa M, Okada N, Kawai G. Solution structure of an RNA stem-loop derived from the 3' conserved region of eel LINE UnaL2. *RNA.* 2004; 10:1380–1387. [PubMed: 15273327]
29. Eichhorn CD, Feng J, Suddala KC, Walter NG, Brooks CL, Al-Hashimi HM. Unraveling the structural complexity in a single-stranded RNA tail: implications for efficient ligand binding in the prequeuosine riboswitch. *Nucleic Acids Res.* 2012; 40:1345–1355. [PubMed: 22009676]
30. Popovic M, Greenbaum NL. Role of helical constraints of the EBS1-IBS1 duplex of a group II intron on demarcation of the 5' splice site. *RNA.* 2014; 20:24–35. [PubMed: 24243113]
31. Kishore AI, Mayer MR, Prestegard JH. Partial ¹³C isotopic enrichment of nucleoside monophosphates: useful reporters for NMR structural studies. *Nucleic Acids Res.* 2005; 33:e164. [PubMed: 16254075]
32. Shajani Z, Drobny G, Varani G. Binding of U1A protein changes RNA dynamics as observed by ¹³C NMR relaxation studies. *Biochemistry.* 2007; 46:5875–5883. [PubMed: 17469848]
33. Zhang Q, Sun X, Watt ED, Al-Hashimi HM. Resolving the Motional Modes That Code for RNA Adaptation. *Science.* 2006; 311:653. [PubMed: 16456078]
34. Eichhorn CD, Al-Hashimi HM. Structural dynamics of a single-stranded RNA-helix junction using NMR. *RNA.* 2014; 20:782–791. [PubMed: 24742933]
35. Getz M, Sun X, Casiano-Negroni A, Zhang Q, Al-Hashimi HM. NMR studies of RNA dynamics and structural plasticity using NMR residual dipolar couplings. *Biopolymers.* 2007; 86:384–402. [PubMed: 17594140]
36. Tolman JR, Flanagan JM, Kennedy MA, Prestegard JH. NMR evidence for slow collective motions in cyanometmyoglobin. *Nat Struct Biol.* 1997; 4:292–297. [PubMed: 9095197]
37. Tolman JR, Ruan K. NMR Residual Dipolar Couplings as Probes of Biomolecular Dynamics. *Chem Rev.* 2006; 106:1720–1736. [PubMed: 16683751]
38. Kailasam S, Bhattacharyya D, Bansal M. Sequence dependent variations in RNA duplex are related to non-canonical hydrogen bond interactions in dinucleotide steps. *BMC Res Notes.* 2014; 7:83. [PubMed: 24502340]
39. Pingali PK, Halder S, Mukherjee D, Basu S, Banerjee R, Choudhury D, Bhattacharyya D. Analysis of stacking overlap in nucleic acid structures: algorithm and application. *J Comput-Aided Mol Des.* 2014; 28:851–867. [PubMed: 24990628]
40. Hermann T, Westhof E. Non-Watson-Crick base pairs in RNA-protein recognition. *Chem Biol.* 1999; 6:R335–343. [PubMed: 10631510]
41. Lu XJ, Olson WK. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.* 2003; 31:5108–5121. [PubMed: 12930962]
42. Ananth P, Goldsmith G, Yathindra N. An innate twist between Crick's wobble and Watson-Crick base pairs. *RNA.* 2013; 19:1038–1053. [PubMed: 23861536]
43. Du Z, Yu J, Andino R, James TL. Extending the family of UNCG-like tetraloop motifs: NMR structure of a CACG tetraloop from coxsackievirus B3. *Biochemistry.* 2003; 42:4373–4383. [PubMed: 12693932]
44. Joli F, Bouchemal N, Laigle A, Hartmann B, Hantz E. Solution structure of a purine rich hexaloop hairpin belonging to PGY/MDR1 mRNA and targeted by antisense oligonucleotides. *Nucleic Acids Res.* 2006; 34:5740–5751. [PubMed: 17041234]

45. Mueller U, Schübel H, Sprinzl M, Heinemann U. Crystal structure of acceptor stem of tRNA(Ala) from *Escherichia coli* shows unique G.U wobble base pair at 1.16 Å resolution. *RNA*. 1999; 5:670–677. [PubMed: 10334337]
46. Trikha J, Filman DJ, Hogle JM. Crystal structure of a 14 bp RNA duplex with non-symmetrical tandem G-U wobble base pairs. *Nucleic Acids Res*. 1999; 27:1728–1739. [PubMed: 10076005]
47. Varani G, McClain WH. The G × U wobble base pair. A fundamental building block of RNA structure crucial to RNA function in diverse biological systems. *EMBO Rep*. 2000; 1:18–23. [PubMed: 11256617]
48. El Hassan MA, Calladine CR. Two distinct modes of protein-induced bending in DNA. *J Mol Biol*. 1998; 282:331–343. [PubMed: 9735291]
49. Headey SJ, Huang H, Claridge JK, Soares GA, Dutta K, Schwalbe M, Yang D, Pascal SM. NMR structure of stem-loop D from human rhinovirus-14. *RNA*. 2007; 13:351–360. [PubMed: 17194719]
50. Huang H, Alexandrov A, Chen X, Barnes TW 3rd, Zhang H, Dutta K, Pascal SM. Structure of an RNA hairpin from HRV-14. *Biochemistry*. 2001; 40:8055–8064. [PubMed: 11434774]
51. Herold J, Andino R. Poliovirus RNA replication requires genome circularization through a protein-protein bridge. *Mol Cell*. 2001; 7:581–591. [PubMed: 11463383]
52. Chase AJ, Daijogo S, Semler BL. Inhibition of poliovirus-induced cleavage of cellular protein PCBP2 reduces the levels of viral RNA replication. *J Virol*. 2014; 88:3192–3201. [PubMed: 24371074]
53. Silvera D, Gamarnik AV, Andino R. The N-terminal K Homology Domain of the Poly(rC)-binding Protein Is a Major Determinant for Binding to the Poliovirus 5'-Untranslated Region and Acts as an Inhibitor of Viral Translation. *J Biol Chem*. 1999; 274:38163–38170. [PubMed: 10608888]
54. Du Z, Lee JK, Fenn S, Tjhen R, Stroud RM, James TL. X-ray crystallographic and NMR studies of protein-protein and protein-nucleic acid interactions involving the KH domains from human poly(C)-binding protein-2. *RNA*. 2007; 13:1043–1051. [PubMed: 17526645]
55. Du Z, Lee JK, Tjhen R, Li S, Pan H, Stroud RM, James TL. Crystal Structure of the First KH Domain of Human Poly(C)-binding Protein-2 in Complex with a C-rich Strand of Human Telomeric DNA at 1.7 Å. *J Biol Chem*. 2005; 280:38823–38830. [PubMed: 16186123]
56. Sharma N, Ogram SA, Morasco BJ, Spear A, Chapman NM, Flanagan JB. Functional role of the 5' terminal cloverleaf in Cocksackievirus RNA replication. *Virology*. 2009; 393:238–249. [PubMed: 19732932]
57. Spear A, Sharma N, Flanagan JB. Protein-RNA tethering: The role of poly(C) binding protein 2 in poliovirus RNA replication. *Virology*. 2008; 374:280–291. [PubMed: 18252259]

Figure 1.

Secondary structure prediction of the HRV-14 5'CL. SLB, the subject of this investigation, is circled. SLB is predicted to contain six Watson–Crick base pairs with one wobble base pair shown in color and an eight-nucleotide hairpin loop. Nucleotides implicated in PCBP binding are located in the loop and are shown in color.

Figure 2.

Assigned imino proton portion of the 250 ms ^1H - ^1H NOESY spectrum at 278 K of HRV-14 SLB in H_2O with the 1D ^1H NMR spectrum shown above. A diagram of the helix is shown at the right.

Figure 3.

Base to H1'/H5 region of a 250 ms ^1H - ^1H NOESY spectrum of SLB in D_2O . Peaks representing intranucleotide H1'-H6/H8 NOEs are labeled with the nucleotide number. The lines connect the labeled peaks through sequential H1'-H6/H8 NOE cross peaks. The H2 resonances not involved in the walk are underlined.

Figure 4.

Superposition of the final 10 structures of HRV-14 SLB determined in this study. The loop and terminal regions are shown as gray ribbons. The bases of hydrogen-bonded base pairs are shown in the following colors: orange for cytosine, yellow for uracil, green for guanine, and blue for adenine.

Figure 5.

(A) Comparison between SLB ^{13}C chemical shifts and chemical shifts of nucleotide monophosphates (NMPs). The chemical shift difference $[\delta(\text{NMP}) - \delta(\text{SLB})]$ is shown in parts per million. (B) Normalized intensity of C6, C8, and C2 resonances in the ^1H - ^{13}C HSQC spectrum. (C) SLB ^1H - ^{13}C RDC values extracted from ^1H - ^{13}C ARSTY spectra. In each section, the regions comprising the helix are shaded gray, and the color/shape coding of the bars is defined to the right of the figure.

Figure 6.

Space-filling view of the average calculated structure of HRV-14 SLB. The helix region is colored gray, and the loop (top) and termini (bottom) are colored blue. The red dashed line shows the shortest distance between phosphorus atoms P_i and P_{i-5} across the major groove (G3-C17).

Table 1**Structural Constraints**

NMR Constraints (number)	
total NOEs	257
intranucleotide	144
internucleotide	73
long-range	40
dihedral angle constraints	188
sugar pucker	115
backbone	73
hydrogen bonds	30
planarity	12
residual dipolar couplings (RDC)	37
Constraint and Geometry Violations	
NOE stem mean (Å)	0.038 ± 0.007
NOE stem max (Å)	0.376
NOE loop mean (Å)	0.063 ± 0.015
NOE loop max (Å)	0.439
dihedral mean (deg)	0.416 ± 0.082
dihedral max (deg)	4.486
hydrogen bond mean (Å)	0.033 ± 0.010
hydrogen bond max (Å)	0.158
RDC mean (Hz)	0.489 ± 0.043
RDC max (Hz)	0.801
bond length mean (Å)	0.006 ± 0.000
bond angle mean (deg)	0.743 ± 0.050
improper mean (deg)	0.612 ± 0.050
Ensemble RMSD	
stem (Å)	0.415
all atoms (Å)	2.249