NMR Structures of Loop B RNAs from the Stem−Loop IV Domain of the *Enterovirus* Internal Ribosome Entry Site: A Single C to U Substitution Drastically Changes the Shape and Flexibility of RNA[†,‡]

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

Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143-2280, and Department of Microbiology and Immunology, University of California, San Francisco, California 94143-0414

Received December 26, 2003; Revised Manuscript Received March 10, 2004

ABSTRACT: The 5′-untranslated region of positive-strand RNA viruses harbors many *cis*-acting RNA structural elements that are important for various viral processes such as replication, translation, and packaging of new virions. Among these is loop B RNA of the stem−loop IV domain within the internal ribosomal entry site (IRES) of enteroviruses, including *Poliovirus* type 1 (PV1). Studies on PV1 have shown that specific recognition of loop B by the first KH (hnRNP K homology) domain of cellular poly(rC)-binding protein 2 (PCBP2) is essential for efficient translation of the viral mRNA. Here we report the NMR solution structures of two representative sequence variants of enteroviral loop B RNA. The two RNA variants differ at only one position (C vs U) within a six-nucleotide asymmetric internal loop sequence that is the binding site for the PCBP2 KH1 domain. Surprisingly, the two RNAs are drastically different in the overall shape and local dynamics of the bulge region. The RNA with the 5′-AUCCCU bulge sequence adopts an overall L shape. Its bulge nucleotides, especially the last four, are highly flexible and not very well defined by NMR. The RNA with the 5′-AUUCCU bulge sequence adopts an overall U shape, and its bulge sequence exhibits only limited flexibility. A detailed analysis of the two RNA structures and their dynamic properties, as well as available sequence data and known KH domain−RNA complex structures, not only provides insights into how loop B RNA might be recognized by the PCBP2 KH1 domain but also suggests a possible correlation between structural flexibility and pre-existing structural features for protein recognition.

Positive-strand RNA viruses in the Picornaviridae family utilize their genomic RNAs for at least three important functions: as an mRNA for viral protein synthesis, as a template for negative-strand RNA replication, and for packaging of new virions. To fulfill the multiple functional roles of the genomic RNAs, the viruses have evolved certain regulatory mechanisms for balancing the usage of the genomic RNAs in different stages of the viral life cycle. Research efforts to unveil these mechanisms have thus far been focused on *Poliovirus* type 1 (PV1),[1] a prototypic member of the Picornaviridae family.

The 5′-untranslated region (5′UTR) of PV1 genomic RNA harbors many *cis*-acting RNA elements that play critical roles in both viral translation and RNA replication. The secondary structure of the PV1 5′UTR is predicted to form six distinct stem−loops domains (I−VI, Figure 1A). Functionally, the first stem−loop domain, which folds into a cloverleaf-like structure, is shown to provide a switching control between protein synthesis and RNA replication (*1*). Stem−loops II−VI form another important functional domain known as the internal ribosomal entry site (IRES) (see ref *2* for a review). The IRES element mediates cap-independent translation by allowing ribosomes to enter the viral mRNA internally without scanning from a 7-methylguanosine cap structure at the 5′ end.

In addition to the *cis*-acting RNA elements, a number of *trans*-acting cellular and viral proteins have also been identified as essential players in the regulatory processes. Among these is a host cell protein called poly(rC)-binding protein [PCBP, which exists in two closely related (83% identical) variants, PCBP1 and PCBP2, also known as hnRNP E1 and E2 or αCP1 and αCP2, respectively]. PCBP2 regulates viral translation through its direct interaction with both the cloverleaf RNA and stem−loop IV within IRES. It has been shown that PCBP2 alone binds to the cloverleaf RNA with low affinity. In the presence of viral protein 3CD (the precursor of viral protease 3C and viral polymerase 3D), a high-affinity ternary ribonucleoprotein complex involving the cloverleaf RNA, PCBP2, and 3CD is formed. Formation

[†] This work was partially supported by the National Institutes of Health.

[‡] The coordinates of the NMR structures and the structural restraints for the WT34 and 10U RNA have been deposited in the RCSB Protein Data Bank as entries 1R7W and 1R7Z, respectively.

* 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: 5′UTR, 5′-untranslated region; PV1, *Poliovirus* type 1; IRES, internal ribosome entry site; PCBP, poly(rC)-binding protein; NMR, nuclear magnetic resonance; FID, free induction decay; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser effect spectroscopy; DQF-COSY, double-quantum-filtered correlation spectroscopy; TOCSY, total correlation spectroscopy; HMQC, heteronuclear multiple-quantum coherence.

10.1021/bi0363228 CCC: $27.50
Published on Web 04/24/2004

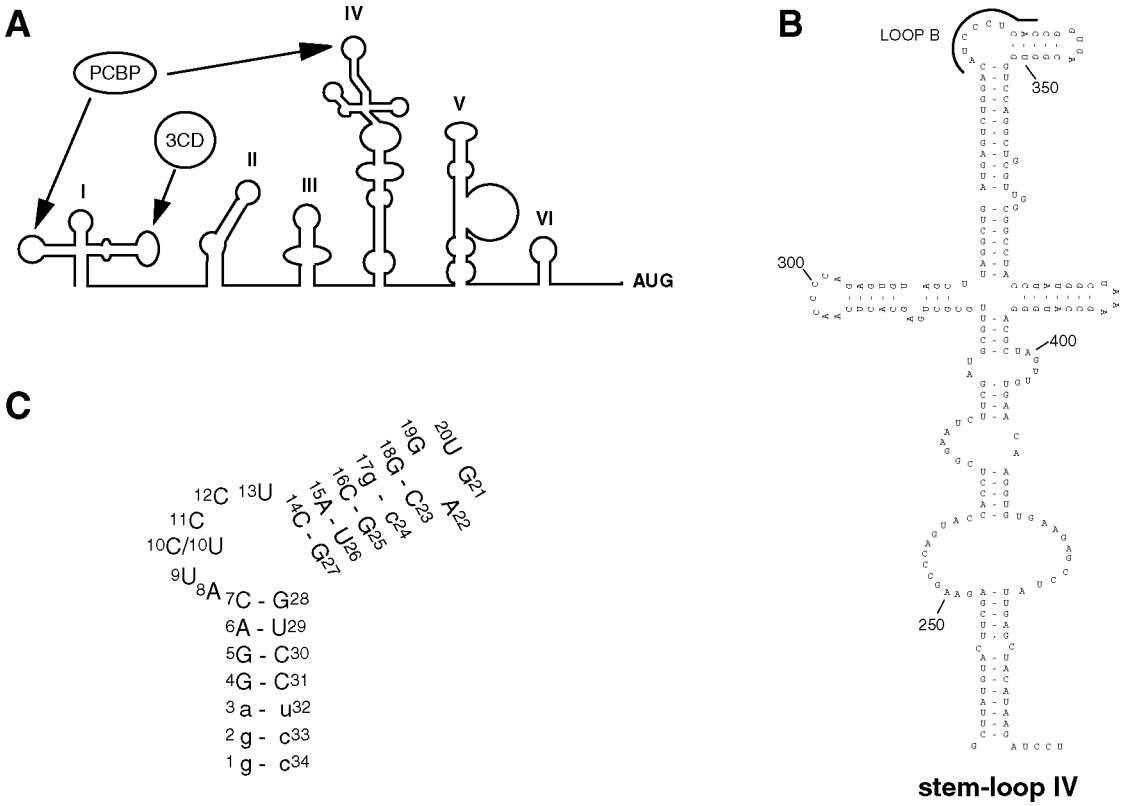

FIGURE 1: (A) Schematic representation of the secondary structure of the *Poliovirus* 5′UTR. Predicted stem−loop domains I−VI are shown. AUG represents the initiation codon for translation. Black arrows denote the protein binding targets within the 5′UTR. These are as follows (from 5′ to 3′): stem−loop B and stem−loop D RNAs within the cloverleaf (domain I) and loop B RNA within domain IV. Both cloverleaf stem−loop B and domain IV loop B bind to cellular protein PCBP2, while stem−loop D RNA binds to viral protein 3CD. (B) Secondary structure of domain IV of *Poliovirus* (Mahoney strain) IRES. The functionally important interacting sequence for PCBP binding is denoted with the black line in loop B. (C) Schematic diagram of the two 34-nucleotide RNA constructs used in this NMR study. These sequences correspond to the loop B region of stem−loop IV for several viruses of the *Enterovirus* genus. Nucleotides in lowercase letters are non-natural. Among these, the G17•C24 base pair is switched from the wild-type pair to prevent formation of a duplex dimer; the first three base pairs also differ from those in the wild type for the purpose of optimizing transcription using T7 polymerase. The construct with a 5′-AUCCCU bulge sequence is termed WT34 RNA in this paper. The other construct with a C to U substitution is called 10U RNA.

of this ternary complex provides a switching mechanism for dictating usage of the viral genomic RNA as either an mRNA for protein synthesis or a template for RNA replication. While interaction of PCBP2 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 that proper balance between protein expression and RNA replication is maintained. Another site of action for PCBP2 is located in stem−loop IV of the IRES element. Binding of PCBP2 to stem−loop IV promotes the IRES-mediated translation. Mutations within stem−loop IV RNA that interrupt its interaction with PCBP impair viral translation.

The specific sites of binding of PCBP2 to the cloverleaf and stem−loop IV RNA have been mapped. Within the cloverleaf structure, a stretch of three cytosine residues in stem−loop B was shown to be essential for protein recognition as well as for viral viability (*3*−*5*). Within the large stem−loop IV domain, a C-rich asymmetric internal loop that also contains a stretch of three cytosine residues (loop B, 1B) is found to be important for PCBP2 protein binding and crucial for *Poliovirus* translation. Interestingly, although C-rich loop sequences are the major determinants for specific binding of PCBP2 to the cloverleaf and stem−loop IV RNAs, the stem−loop IV RNA exhibits a much higher affinity for PCBP2 than the cloverleaf RNA. The dissociation constant for the PCBP2−stem−loop IV RNA complex and the PCBP2−cloverleaf complex was estimated to be ∼15 and ∼95 nM, respectively (*6*). In the presence of viral protein 3CD, the PCBP2−cloverleaf interaction becomes stronger (estimated dissociation constant of ∼1 nM) due to the formation of a high-affinity ternary complex. The PCBP2−stem−loop IV RNA interaction is not altered by the presence of 3CD. Therefore, viral protein 3CD modulates formation of a complex between PCBP2 and the two RNAs within the 5′UTR of the *Poliovirus* genome.

PCBP2 contains three copies of a conserved RNA-binding domain known as the KH (K homology) motif, first identified in hnRNP K (*7*). The arrangement of these three motifs within PCBP2 is similar to a number of other KH domain-containing proteins such as hnRNP K and Nova protein: two consecutive KH domains at the amino terminus and a third KH motif at the carboxy terminus with an intervening sequence variable in length. Except for the three KH domains, no other known RNA-binding motif is found within PCBP2. The nucleic acid binding ability of the three KH domains was first established by expressing each domain individually and showing that the isolated KH1 and KH3 (but not KH2) domains are able to bind strongly to poly(rC) homopolymers (*8*). Interestingly, only the KH1 domain was

shown to interact specifically with the C-rich sequences of the cloverleaf and the stem−loop IV structures within the *Poliovirus* 5′UTR (*9*), despite KH1 and KH3 being 95% identical. The isolated KH1 domain could form a high-affinity ternary complex with the cloverleaf RNA and the viral 3CD protein. It could also inhibit viral IRES-dependent translation, presumably as a dominant-negative competitor against PCBP function. Therefore, the KH1 domain is the major determinant for specific interaction between PCBP2 and its viral RNA targets within the 5′UTR.

Although substantial progress has been made in defining important elements of the primary sequence, important aspects of secondary structure, many of the regulatory proteins that are involved, and the mechanisms choreographing interactions of the various structures within the 5′UTR and with regulatory proteins are still not fully elucidated. In particular, there is very little three-dimensional structural information about IRES RNA of the enterovirus-rhinovirus class (a subfamily of picornaviruses which includes PV1), indeed any picornaviruses. To the best of our knowledge, only a few hairpin RNA fragments from domains III and VI of PV1 have been studied by NMR (*10*). There is also some structural information about other distantly related viruses, notably, hepatitis C virus (*11*−*15*).

We have recently initiated structural work on the PCBP2−RNA interaction that is important for viral regulation. Here, we present the RNA structures determined by NMR for two representative sequence variants of the portion of loop B within stem−loop IV of the IRES of the *Enterovirus* genus, including the bulge sequence critical for PCBP2 binding.

MATERIALS AND METHODS

Preparation of RNA Samples. Four different RNA samples of the 34-nucleotide construct (WT34 RNA, as shown in Figure 1C) of IRES stem−loop IV, loop B, were prepared via *in vitro* transcription using T7 RNA polymerase and a synthetic DNA template using published procedures (*16*, *17*): unlabeled, uniformly ^{13}C- and ^{15}N-labeled, or type-specifically ^{13}C- and ^{15}N-labeled in A/U or C residues, respectively. Two samples were prepared for the 10U RNA construct: unlabeled and type-specifically ^{13}C- and ^{15}N-labeled at A/U residues. One mutant RNA with C replacing U (compared to WT34 RNA) at position 12 was also prepared to assist with the assignment of the three C residues in the internal bulge. The RNA samples typically contained 1−2 mM RNA in 25 mM sodium phosphate (pH 6.5) and 25 mM sodium chloride. The liquid crystalline medium for measurements of residual dipolar coupling constants was achieved by stepwise addition with vigorous vortexing of microliter aliquots of hexanol to the RNA solution containing 5% (w/v) of C12E6 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 (*18*) and analyzed with SPARKY (*19*). In the two- and three-dimensional (3D) homonuclear and heteronuclear experiments, the States−TPPI method was used to achieve quadrature detection in the indirect dimension(s). Homonuclear two-dimensional (2D) NOESY spectra in a 90% H_2O/10% D_2O mixture were recorded at 10 °C with a symmetrically shifted shaped pulse for solvent suppression (*20*). The spectral width was set to 12 500 Hz in both dimensions. A mixing time of 150 ms and a relaxation delay of 1.5 s were applied. A total of 600 FIDs of 2048 complex data points were collected. Homonuclear 2D NOESY, DQF-COSY, and TOCSY spectra in D_2O were recorded at 25 or 35 °C, with low-power presaturation for the suppression of the residual HDO signal. The spectral width was set to 5500 Hz in both dimensions. Typically, the COSY and TOCSY spectra were recorded with 512 FIDs of 2048 complex data points, and a relaxation delay of 1.5 s. A 60 ms clean-MLEV-17 isotropic mixing sequence was used in the TOCSY experiments. The 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 buildups. Typically, 600 or 700 FIDs of 2048 complex data points were recorded for each NOESY spectrum.

One-bond correlated ^{1}H−^{15}N HMQC spectra were obtained at 10 °C in a 90% H_2O/10% D_2O mixture using the jump-and-return solvent suppression scheme. The spectral width was set to 12 500 and 2500 Hz for the proton and nitrogen dimensions, respectively. Acquisition parameters were optimized for detection of imino or amino protons in different experiments. All multidimensional heteronuclear experiments in D_2O were performed at 25 °C. 2D HMQC experiments with very low power presaturation during the relaxation delay for water suppression were performed to obtain the natural abundance one-bond ^{1}H−^{13}C correlation of the unlabeled RNAs. The spectral widths were 5500 and 3500 Hz for the ^{1}H and ^{13}C dimensions, respectively, and the ^{13}C carrier frequency was set at 120 ppm. The one-bond ^{1}H−^{13}C coupling constant was set to 165 Hz as a compromise between aromatic and ribose moieties. 2D constant-time HSQC experiments for one-bond ^{1}H−^{13}C correlation were recorded on the fully labeled and type-specifically labeled samples with ^{13}C and ^{15}N decoupling during acquisition using GARP. Acquisition parameters such as the spectral widths, carrier frequencies, the constant time delay, and ^{1}H−^{13}C one-bond coupling constants were optimized for detection of aromatic or ribose moieties 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-RELAY, and HCCH-TOCSY. The spectral widths were set to 2800 Hz in the proton dimensions and 4000 Hz in the carbon dimensions. The ^{13}C carrier frequency was set to 79 ppm, and the 3D data sets typically contained 512 (^{1}H) × 128 (^{1}H) × 64 (^{13}C) complex points. The spin lock period for the HCCH-TOCSY experiment was 12 ms with a DIPSI-3 isotropic mixing scheme. A 2D version (^{1}H−^{1}H) of the HCCH-TOCSY experiment with an isotropic mixing time of 65 ms was also performed for correlating adenine H2 and H8 protons. Correlations of base and ribose protons were established by 2D HCN experiments. 3D ^{13}C-edited NOESY-HMQC experiments (*21*) were carried out in D_2O with a mixing time of 150 ms and a relaxation delay of 1.2 s. The spectral widths in both proton dimensions were 4800 Hz. For the ^{13}C dimension, the spectral width was 3800 or 2500 Hz for the aromatic- or ribose-optimized experiment, respectively; the ^{13}C carrier frequency was set at 79 and 140 ppm accordingly. The 3D sets contained 512 (^{1}H) × 128 (^{1}H) × 64 (^{13}C) complex points or 512 (^{1}H) × 128 (^{1}H) × 48 (^{13}C) complex points for the two experiments. 2D ω_1-

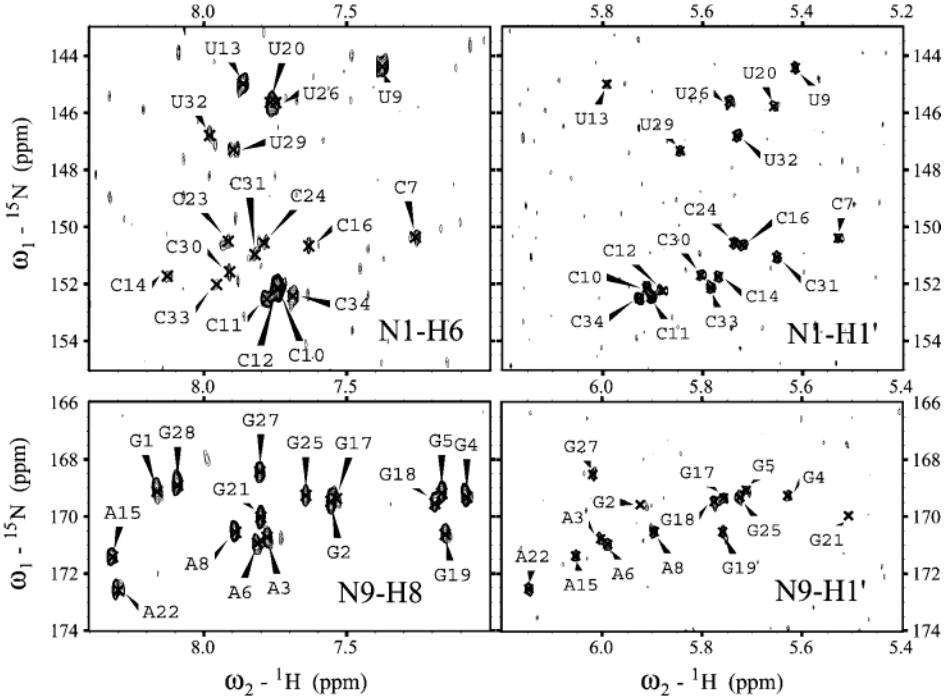

FIGURE 2: Portions of the 2D versions of triple-resonance mqHCN spectra of the 34-mer WT34 (*Poliovirus* IRES stem−loop IV-B) in D_2O, 25 mM phosphate buffer, and 25 mM NaCl at pH 6.5 and 25 °C. It was beneficial to acquire two spectra, one optimized for the base N1/N9−[C6/C8]−H6/H8 correlations (left) and another for the ribose N1/N9−[C1′]−H1′ correlations (right). Top panels show pyrimidine N1−H6 and N1−H1′ cross-peaks; bottom panels show purine N9−H8 and N9−H1′ cross-peaks.

^{13}C-filtered NOESY experiments (*22*) were performed on the two type-specifically labeled samples with a mixing time of 100 or 200 ms.

Resonance Assignments. Four samples, namely, unlabeled, uniformly ^{13}C- and ^{15}N-labeled, type-specifically ^{13}C- and ^{15}N-labeled at A/U or C residues, respectively, of the WT34 RNA were used to obtain nearly complete resonance assignments of this RNA. Assignments of exchangeable imino protons and cytosine amino protons in base pairs were straightforward based on homonuclear water NOESY spectra. Assignments of the nonexchangeable protons were achieved by analysis of a set of standard NMR experiments performed in D_2O buffer. Briefly, ribose spin system identification was achieved by analysis of the 3D (^{1}H, ^{13}C, ^{1}H) HCCH-COSY, HCCH-RELAY, and HCCH-TOCSY experiments. Intranucleotide ribose H1′−base H6/H8 connections were established by 2D H(C)N experiments (Figure 2). Intranucleotide adenine H2 and H8 protons were connected by a 2D version (^{1}H−^{1}H) of the HCCH-TOCSY experiment. Sequence-specific assignments were made using 2D homonuclear TOCSY, DQF-COSY, and NOESY spectra acquired in D_2O solutions, as well as 3D (^{1}H, ^{13}C, ^{1}H) NOESY-HMQC spectra of the uniformly labeled and type-specifically labeled samples and 2D filtered spectra of the type-specifically labeled samples. Assignments of the stem and tetraloop regions, as well as A8, U9, and U13 from the bulge, were easily obtained using the standard sequential assignment procedures. Assignments of the three bulge cytidines (C10−C12) were assisted by comparison with spectra of the 10U RNA and another mutant RNA with a C to U substitution at position 12.

Proton assignments for the stem regions and for the apical tetraloop could be easily transferred from WT34 RNA to 10U RNA. Assignments for the six-nucleotide internal loop of 10U RNA were done with the help of a ^{1}H−^{13}C constant-time HSQC spectrum of the A/U-labeled 10U RNA sample, natural abundance ^{1}H−^{13}C HMQC spectrum of unlabeled 10U RNA sample, and the usual set of homonuclear proton spectra.

Structural Restraints. For structure calculation, intensities of reasonably resolved cross-peaks in the D_2O homonuclear 2D NOESY spectra were integrated by Gaussian line fitting with SPARKY for each of the two RNA molecules, WT34 and 10U. Quantitative interproton distance restraints between nonexchangeable protons were calculated on the basis of these intensities, using a complete relaxation matrix method, MARDIGRAS (*23*), with the RANDMARDI analysis of experimental errors in the intensities (*24*). Using this procedure for RNA structure determination has been described by us previously (*25*, *26*). The input WT34 and 10U structures for MARDIGRAS calculations were initially model-built and energy-minimized with miniCarlo (*27*). The effective rotational correlation time τ_c for MARDIGRAS calculations was estimated by running MARDIGRAS at various τ_c values and selecting values that best reproduced fixed or nearly fixed interproton distances. This produced unexpectedly different results for the two RNA molecules: 15 ns for WT34 and 8 ns for 10U. The effective correlation time could increase due to nonspecific microaggregation. However, sample conditions were the same for both molecules; the lower τ_c value for the 10U RNA indicates a significantly more compact shape for this molecule. Distance restraints calculated with RANDMARDI together with other restraints were used as a flat-well potential for structural refinement (*vide infra*). These steps (distance restraint determination and structure refinement) were iterated several

times to reduce the residual dependence of MARDIGRAS-calculated distances on initial structures.

In addition, upper bounds of 6.0 Å were used for the WT34 RNA for peaks observed in the 3D ^{13}C-edited NOESY spectrum, in cases when the quantitative restraints could not be calculated from 2D NOE intensities due to spectral overlap. Distance restraints involving exchangeable protons were assigned to two categories with upper bounds of 4.0 and 6.0 Å, based on the intensities of the corresponding cross-peaks in the water SSNOESY spectra. Idealized hydrogen bond restraints were imposed on Watson−Crick base pairs in stems established by the observation of imino proton resonances. Also, GNRA tetraloop-specific hydrogen bond restraints were used: G19 N2−A22 N7, G19 N2−A22 O2P, G19 N1−A22 O2P, and G19 O2′−G21 N7. Torsion angle restraints for the ribose sugar conformation were based on analysis of the 2D DQF-COSY spectra. Sugars showing a strong H1′−H2′ cross-peak were restrained to C2′-endo (residues 9−13 and 27 for both WT34 and 10U); other residues were interpreted as C3′-endo. NOE intensities for the stem residues showed patterns typical of an A conformation (in particular, H1′−H2′ sequential peaks and very strong H2′−H6/H8 sequential peaks). Backbone torsion angle restraints consistent with a generic A conformation were also used for the stem residues. In addition, "non-NOE" distance restraints (lower bounds of 5 Å) were added for proton pairs that did not have any intensity in any of the 2D or 3D NOESY data sets, but tended to have short distance in preliminary rounds of structure calculation.

Residual dipolar coupling constants (RDCs) for C−H vectors were measured for partially oriented ^{13}C- and ^{15}N-labeled WT34 and A/U ^{13}C- and ^{15}N-labeled 10U RNA constructs. For this purpose, we acquired ^{1}H−^{13}C constant-time HSQC spectra of each RNA molecule in a dilute liquid crystalline solution. To create liquid crystalline conditions, we used a mixture of *n*-alkyl poly(ethylene glycol) C12E6 with *n*-hexanol. C12E6 is uncharged, insensitive to pH, and barely sensitive to salt, and it forms planar bilayers in a suitable temperature range. As described previously (*28*), we used a C12E6:*n*-hexanol molar ratio of 0.64 and a ratio of surfactant to D_2O solvent [25 mM sodium phosphate and 25 mM NaCl (pH 6.5)] of 5 wt % at 25 °C. The liquid crystalline solution was readily formed, as monitored by the quadrupolar splitting of the solvent ^{2}H signal (of 28 Hz). The C−H couplings were directly measured as the splitting in the ^{13}C dimension in the non-decoupled HSQC peaks, acquired separately for aromatic and ribose correlations. RDC values were measured as the difference in the coupling under isotropic and liquid crystalline conditions. Each component of the HSQC peaks was line-fitted with the Gaussian shape to determine more accurately the position of each peak. The RDC is defined as isotropic splitting minus splitting observed for partially oriented samples.

Structure Refinement. The NMR structures for WT34 and 10U RNA were refined in three stages, using DYANA version 1.5 (*29*) and AMBER version 7 (*30*). In the first stage, DYANA was used to carry out a simulated annealing with 10 000 steps of torsion angle dynamics starting with 100 randomized initial structures. The weights used for NOE-derived distance restraints, non-NOE restraints, and hydrogen bond restraints were 1.0, 10.0, and 15.0, respectively. The weight for the sugar pucker-related torsion and bond angle restraints was 10.0, and the weight for backbone torsion angle restraints in stems was 20.0. In addition, the C4′−O4′ bond length was constrained in the range of 1.41−1.42 Å with a weight of 100.0, and the O4′−H4′ distance was constrained in the range of 2.10−2.20 Å with a weight of 20.0. These constraints are usually sufficient to ensure proper closure of the five-member furanose rings in DYANA and the correct stereochemistry for the prochiral centers, which was checked with a program developed in-house (CHIRANO). DYANA calculations are fast and easy to set up. However, it uses a simplified force field, so these structures were considered preliminary and are not presented here. The DYANA stage was used mainly to speed the iterations between determination of distance restraints with MARDIGRAS, structure calculations, and search for non-NOE restraints, possible errors in NOE cross-peak integration, etc.

For the final two stages, the 50 best DYANA structures were passed to the SANDER module of the AMBER version 7 suite of programs and refined with the same type of structural restraints. It is not necessary to use constraints for sugar ring closure in AMBER. However, local geometries may be violated when running AMBER simulations with experimental restraints; the planarity of aromatic bases and positions of hydrogen atoms are especially vulnerable (*31*, *32*). To prevent this, a large number of bond length, bond angle, and improper torsion constraints were generated, based on geometric parameters observed in high-resolution crystal structures of nucleosides and nucleotides (*33*, *34*). These parameters are somewhat different for C2′-endo and C3′-endo sugar conformations; the constraints were generated accordingly. Altogether, several thousand such local geometry constraints were generated, with force constants of 1000 kcal mol^{-1} Å^{-2} and 1000 kcal mol^{-1} rad^{-2} for distances and angles, respectively. A Python program generating local geometry restraints for RNA molecules is available from the authors upon request. For the experimental NOE distances, non-NOE restraints, and hydrogen bond restraints, a force constant of 10 kcal mol^{-1} Å^{-2} was used. For sugar pucker restraints and backbone torsion restraints, a force constant of 100 kcal mol^{-1} rad^{-2} was used.

The second stage of refinement consisted of equilibration, high-temperature simulated annealing, and restrained minimization. The equilibration included restrained minimization, cold-temperature restrained molecular dynamics (rMD), and another restrained minimization. In each case, restrained minimization included 500 steps of steepest descent and 1000 steps of conjugate gradient minimization. Cold-temperature rMD included 1000 steps (0.5 fs) of simulation, when the weight of restraints was increased from 0.1 to 1.0, and the target temperature was gradually changed from 5 to 50 K, with a thermal bath coupling constant τ_{TP} of 0.2 ps. The high-temperature simulated annealing included 20 000 rMD 0.5 fs steps, when the target temperature changed from 5 to 1000 K during the first 1000 steps, stayed at 1000 K during the next 4000 steps, then decreased to 50 K during the next 5000 steps, and was kept at 50 K for the last 10 000 steps. The weight of restraints increased from 0.1 to 1.5 during the high-temperature period, then decreased to 1.0, and stayed at 1.0 during the last 5000 steps. The thermal coupling constant gradually decreased during simulations from 0.2 to 0.01 ps, at which it was kept for the last 5000 steps of rMD. All AMBER calculations, rMD and restrained minimization,

Table 1: Bulge Sequence of IRES Stem−Loop IV-B

bulge sequence	size	viruses where this sequence occurs[a]
AUCCCU	6	PV1 (Mahoney strain); EV30
AUUCCU	6	PV1 and -2; EV6, -11, -30; CXA6, -A15
ACUCCU	6	PV1; EV18; unclassified *Enterovirus*
AUUCCC[b]	6	PV1 and -3; EV1, -5, -6, -11, -13, -19, -20, -26, -29, -32; CXA7, -A12, -A13, -B2, -B3, -B4, -B6
AUCCCC	6	PV1 and -2; EV11, -16, -32, -33; CXA10; EN69
ACUCCC	6	PV2; CXA15, -A21
GUUCCU	6	PV2 and -3; EV7, -18; CXA9, -A10, -B4
GUCCCU	6	PV1 and -3; EV30
GUUCCC	6	PV1; EV4, -12, -15, -19, -24, -25, -30; CXA5, -A8, -A17, -A19, -A24, -B1, -B4
GUCCCC	6	PV1
AAUCCU	6	CXA24
AAUCCC	6	EV3, -19; EN71
AACCCC	6	EV14; EN71
AGCCCU	6	CXA11
AGUCCC	6	PV3
GAUCCU	6	PV3
GAUCCC	6	EV27; CXA2, -A16; EN71
GACCCC	6	EN71
AUACCC	6	CXA1, -A5, -A22; EV33
ACGCCC	6	PV3
GUACCU	6	PV1
GUACCC	6	EN68
GCACCC	6	EN68
AAACCU	6	PV2 and -3
AAACCC	6	CXA2, -A4, -A14
GAACCU	6	PV3
UUCCCU	6	unclassified *Enterovirus*
UUUCCC	6	EV14, -17; A-2 plaque virus
UUCCCC	6	EV6
UAUCCC	6	EN71
GUUCC	5	EV9
UUCCC	5	EV2
AUAUCCC	7	unclassified *Enterovirus*
ACUUCCC	7	CXA3
AUUCCUC	7	CXB5
AUUCUCC	7	CXB5
AUUUCCC	7	EV6

[a] Abbreviations: PV, *Poliovirus*; EV, *Human echovirus*; EN, *Human enterovirus*; CX, *Human coxsackievirus*. [b] Consensus for the *Enterovirus* genus.

were carried out with the parm94 force field (*35*) in a vacuum, without counterions but with 80% neutralized charges on phosphate groups, with a dielectric constant of 4.0. No SHAKE option was necessary because of the large number of local geometry constraints that were used.

The third and last stage of refinement included high-temperature simulated annealing and restrained minimization, similar to the previous stage, using distance and angle restraints, but also with the RDC restraints included. The weight of the RDC restraints was 0.2 kcal mol^{-1} Hz^{-2} for WT34 RNA and 0.5 kcal mol^{-1} Hz^{-2} for 10U RNA. In contrast to XPLOR refinement with RDC restraints (*36*, *37*), the alignment tensor is floating during AMBER refinement. For each input structure, the initial RNA orientation was calculated with the singular-value decomposition method (best fitting of experimental RDC) (*38*) using the PALES program (*39*). Saupe matrix components S_{zz}, S_{xx_yy}, S_{xy}, S_{xz}, and S_{yz} (PALES output) were converted to AMBER-style components S_{11}, S_{12}, S_{13}, S_{22}, and S_{23}, respectively, of the initial alignment tensor according to the expressions $S_{11} = 10^5(S_{xx\_yy} - S_{zz})/2$, $S_{12} = 10^5 S_{xy}$, $S_{13} = 10^5 S_{xz}$, $S_{22} = -10^5(S_{xx\_yy} + S_{zz})/2$, and $S_{23} = 10^5 S_{yz}$, respectively.

One additional refinement was carried out for WT34 RNA using distance and angle restraints as described above, but with RDC restraints included for the stem residues and not for the bulge residues. The goal of this refinement was to determine more accurately the orientation of the two stems in the molecule.

Calculated structures were analyzed using FITPARAM (*40*) and CHIRANO and, interactively, using MidasPlus (*41*) and Chimera (*42*). Figures of merit comparing experimental and calculated NOE intensities were calculated with CORMA (*43*) and RDC with AMBER and PALES.

RESULTS

Sequence Analysis of Loop B RNA. Binding of PCBP2 to loop B of stem−loop IV (Figure 1A,B) promotes the IRES-mediated translation. Stem−loop IV has an apical GNRA tetraloop and a bulge separated from the tetraloop by 5 bp. The sequence of this bulge is essential for PCBP2 binding and regulation of *Poliovirus* type 1 translation (*6*). The secondary structure of the 63-nucleotide stem−loop IV-B is highly conserved in the *Enterovirus* genus (our unpublished data). Of 516 sequences of this region available from the GenBank as of February 2003, 173 are unique, and the vast majority of them have a six-nucleotide bulge (Table 1). In general, the sequence of this bulge is pyrimidine-rich, and the fourth and fifth residues are always cytidines. For our NMR study, we chose two sequence variants of the portion of stem−loop IV-B, including the bulge that is critical for PCBP2 binding (Figure 1C). One variant (WT34 RNA) has

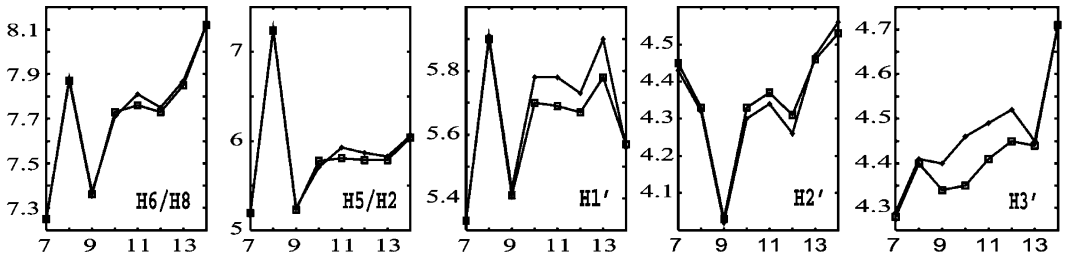

FIGURE 3: Comparison of proton chemical shifts for the WT34 RNA (□) and 10U RNA (◆). Only bulge residues (8−13) and flanking cytidines C7 and C14 are shown. The *x*-axis is the residue number, and the *y*-axis is the chemical shift (parts per million).

the six-nucleotide bulge sequence corresponding to the *Poliovirus* 1 Mahoney strain, a "reference" strain for *Poliovirus* type 1. The bulge sequence, AUCCCU, occurs also in *Human echovirus* 30 (Table 1). Another variant (10U RNA) has a single substitution in the bulge (Figure 1C). The bulge sequence, AUUCCU, naturally occurs in various isolates of *Poliovirus* types 1 and 2, *Human echoviruses* 6, 11, and 30, and *Human coxsackieviruses* A6 and A15 (Table 1).

Resonance Assignments. Resonances of protons and heteronuclei were assigned for WT34 and 10U RNA as described in Materials and Methods. Imino proton resonances were observed at 10 °C for all expected stem base pairs, except the G1·C34 terminal pair. No imino or amino protons were observed for the bulge residues. The H1′−H6/H8 connectivities are observed for both RNA sequences for all residues with the exceptions of bulge cytidines and the tetraloop, where the G19 H1′−U20 H6 cross-peak is missing, and instead the G19 H1′−G21 H8 cross-peak is observed. C23 H1′ has a chemical shift ca. 3.9 ppm, characteristic of GNRA tetraloops. In general, the C to U substitution in the 10U RNA practically does not affect chemical shifts of any stem residues, but it perturbs chemical shifts of all internal loop residues with the exception of A8 (summarized in Figure 3). This gave an early indication that the conformation and/or dynamics of WT34 and 10U sequences may be, in fact, different. This was also confirmed by comparing distributions of nontrivial NOE connectivities in the bulge of the two molecules. Some examples illustrating the differences between the two molecules are shown in Figure 4. In the step opposite from the bulge, six NOE cross-peaks were observed between G27 and G28 for the U10 RNA: G27 H1′−G28 H8 (Figure 4A), G27 H3′−G28 H8 (Figure 4D), G27 H2′−G28 H8, G27 H8−G28 H8, G27 H1′−G28 H5′, and G27 H1′−G28 H5″. However, only an extremely weak G27 H2′−G28 H8 peak was observed for the WT34 in this step. For the internal loop residues, connectivity between U9 and C14 (Figure 4B) and between A8 and C11 (Figure 4C) was observed for WT34 but not for 10U RNA. In many other instances, cross-peaks between a pair of bulge residues were present in both molecules, but the type of connectivity was different. For example, there are a number of cross-peaks between C12 and C14 in both molecules, but only a weak C14 H5−C12 H6 cross-peak was observed for the 10U RNA (Figure 4A, dashed box).

Structural Restraints. The H1′−H2′ cross-peaks were observed for both molecules in the DQF-COSY spectra for G27 and for all bulge residues except A8. This implies a predominantly C2′-endo conformation for these residues. Also, weak H1′−H2′ cross-peaks were observed for C34 (3′ terminus) and U20 (tetraloop residue), which is consistent with flexible sugar puckers.

Quantitative distance restraints were calculated as described in Materials and Methods on the basis of D_2O 2D NOESY intensities using complete relaxation matrix analysis and supplemented with qualitative restraints from water 2D NOESY and (for WT34) from 3D NOESY. In total, we used 575 distance restraints for WT34 RNA (16.9 per residue) and 614 distance restraints for 10U (18.1 per residue). The statistics are very good (Table 2), especially taking into account that approximately half of the restraints are quantitative restraints calculated with RANDMARDI and that these numbers do not include any fixed distances or distances with low variation (such as H1′−H2′), which do not have much informational content. The larger number of distance restraints for the 10U RNA (despite the fact that we used the 3D ^{13}C-edited NOESY data for WT34 and not for 10U) is due to better spectral resolution for the internal loop pyrimidines in the case of the 10U sequence.

Residual dipolar coupling constants (RDCs) were measured for partially oriented ^{13}C- and ^{15}N-labeled WT34 and A/U ^{13}C- and ^{15}N-labeled 10U RNA constructs as described in Materials and Methods. In total, 86 RDC values for WT34 RNA and 26 RDC values for 10U RNA were measured. Following the analysis of Al-Hashimi *et al*. for the free HIV-1 TAR RNA (*44*), we plotted the observed RDC values against the secondary structure of WT34 and 10U RNAs (Figure 5). RDC values are significantly attenuated for the bulge residues in WT34 RNA compared to both stems, which is indicative of significant dynamics for the loop (*44*). There is no such attenuation in the 10U RNA, which is consistent with a single conformation or a few alternative conformations.

Structure Calculations. Several structural ensembles were calculated for each RNA sequence as described in Materials and Methods. The final stage of refinement consisted of AMBER high-temperature simulated annealing and restrained minimization, either with distance and torsion angle restraints or with distance, torsion angle, and RDC restraints. Both rMD and restrained minimization with RDC restraints initially presented serious problems. It turned out that it was very easy to satisfy all RDC restraints even during a short restrained minimization, without significantly changing the RNA conformation. Inspection of resulting structures showed that it was achieved by violating local geometries, mostly by distorting bond angles involving protons with RDC restraints (data not shown). We have observed previously similar, although not as severe, problems during refinement using distance restraints for a DNA duplex (*32*). Tsui *et al*. (*31*) used increased bond angle force constants in the

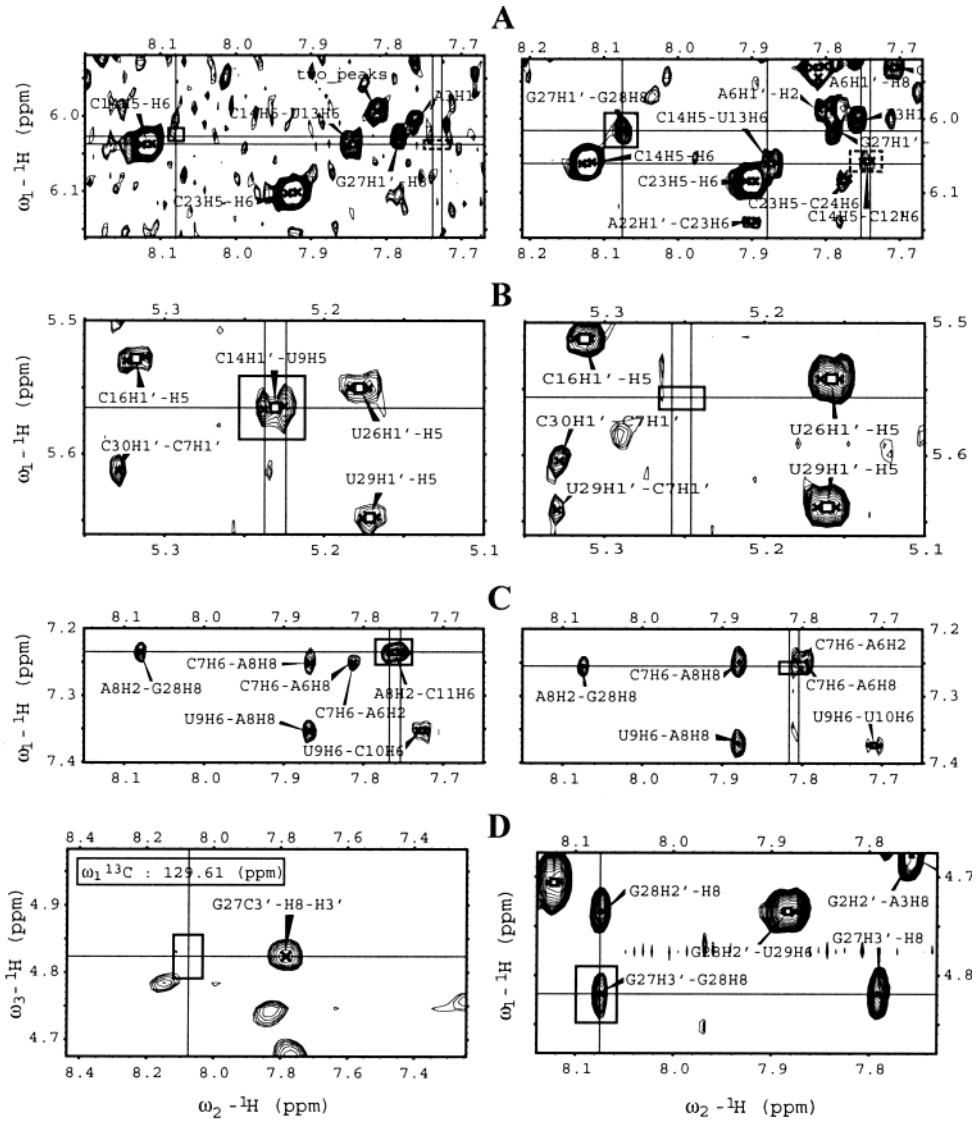

FIGURE 4: Differences in the NOE connectivities in the internal loop region of WT34 RNA (left) and U10 RNA (right). All panels show portions of D_2O 2D NOESY data sets at 200 ms, except for the left panel of part D, which shows a portion of the G27 C3′ plane of the 3D ^{13}C-edited NOESY-HMQC data set for WT34. Cross-peaks observed for only one of the molecules are boxed, as are the spectral regions where these cross-peaks are expected for the other molecule. Grid lines are drawn through the corresponding frequencies. Note that the H5 and H6 resonances of pyrimidines are split in the ω_2 dimension due to the H5−H6 scalar coupling. In such cases, the grid lines are drawn through each component of the doublet. (A) The G27 H1′−G28 H8 (solid box) and C14 H5−C12 H6 (dashed box) cross-peaks are observed for 10U (right) but not for WT34 RNA. (B) The C14 H1′−U9 H5 cross-peak is observed for only WT34 RNA. (C) The A8 H2−C11 H6 cross-peak is observed for only WT34 RNA. Note that the expected position of the A8 H2−C11 H6 doublet is partially overlapped with the C7 H6−A6 H2 and C7 H6−A6 H8 cross-peaks for the 10U RNA (right); however, one component of this doublet is clear of the overlap. (D) The C27 H3′−G28 H8 cross-peak is observed for 10U, and the G27 C3′−G28 H8−G27 H3′ cross-peak is not observed for WT34 RNA.

Table 2: Experimental Restraints for WT34 and 10U RNA

type of restraint	WT34	10U
MARDIGRAS distances	281[a]	376[b]
water NOESY	97	128
3D NOESY	72	0
non-NOEs	88	73
hydrogen bonds	37	37
total distances	575	614
distances per residue	16.9	18.1
RDC	86	26

[a] Quantitative distance bounds calculated with the RANDMARDI procedure; average flat well width of 1.73 Å. [b] Quantitative distance bounds calculated with the RANDMARDI procedure; average flat well width of 1.37 Å.

AMBER force field to preserve local geometries of a protein during refinement with RDC restraints. Trying a similar approach for RNA, however, led to severe flattening of sugar rings, because equilibrium bond angles within the sugar moieties have tetrahedral values in the AMBER force field (data not shown). To overcome this problem, we generated, as described in Materials and Methods, a set of sugar pucker-dependent strong local geometry constraints for each sugar (C2′-endo for residues 9−13 and 27 and C3′-endo for the rest). These constraints (bond lengths and bond angles) are based on a survey of high-resolution crystal structures (*33*, *34*); they are consistent with a particular sugar pucker, but do not enforce it, because they do not include torsion angles. In addition, we included similar constraints for the nitrogenous bases, but using AMBER equilibrium values (essentially, increasing AMBER force constants), and a large number of improper torsion angles. Because of these local

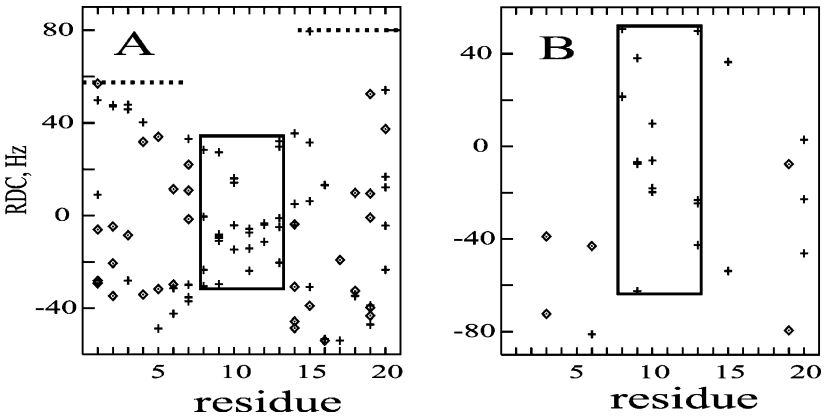

FIGURE 5: Experimental RDC values plotted against secondary structure for WT34 (A) and 10U (B) RNA. For example, "residue 5" represents both G5 (+) and base-paired C30 (◇). Solid boxes show regions corresponding to the internal loop (residues 8−13). Dashed lines for WT34 show maximum RDC values for the lower and upper stems. Note the attenuation of RDC values for loop residues in WT34.

geometry constraints, we gained flexibility in choosing the weight of RDC restraints without compromising the quality of structures. Using a large number (several thousand for an RNA 34-mer) of strong constraints required a tighter coupling to the thermal bath during rMD simulations to keep the temperature under control (see Materials and Methods). During the high-temperature period of simulated annealing, the target temperature was 1000 K, but the actual temperature typically rose to 20000−30000 K, very efficiently randomizing initial structures. During the cooling period, the target temperature was kept at 50 K with the coupling constant of 0.01 ps for the last 10 000 steps; the actual temperature decreased to 50−100 K and stayed in this range for the last 5000 steps of rMD.

Refined Structures. Calculated structures are presented in Figures 5−7, and comparison with experimental data is shown in Table 3; this table also shows the magnitude of alignment and rhombicity of the alignment tensor as calculated by PALES. WT34 RNA refined with distance and torsion angle restraints (Figure 6A) presents an ensemble of structures typical of RNA with internal loops or bulges in solution (*26*, *44*). Each stem is defined relatively tightly, to an rmsd of 1.3−1.6 Å (Table 4), but the bulge conformation and the relative position of stems are not defined well (overall rmsd of 7.1 Å and rmsd for both stems of 5.6 Å). Although the residual distance deviation R_{dev} is low [0.07 (Table 3)], this is due to some extent to the fact that many distance restraints have wide bounds of 1.8−6.0 Å (restraints from water NOESY and 3D NOESY), and the quantitative MARDIGRAS restraints have a relatively wide flat well width of 1.7 Å (Table 2). Indeed, the NOE-based sixth-root weighted R^x factor is relatively high (Table 3), and there are individual distances violated by more than 1 Å. And, not surprisingly, calculated RDC constants agree poorly with the experimental values (Table 3 and Figure 10). In strict terms, an ensemble of structures calculated in such a manner (with all restraints enforced simultaneously) is not meant to represent the true dynamics of the molecule; it rather shows the extent to which the conformation is underdetermined by the restraints. Nevertheless, the presence of individual distance violations due to internal conflicts in experimental data often indicates conformational averaging (*45*). Attenuation of the RDC constants for the bulge residues (Figure 5) also confirms the localized flexibility of WT34 RNA.

Including RDC restraints in the refinement of WT34 RNA tightens up considerably the envelope of calculated conformations (Figure 6B) to an average rmsd of 4.1 Å. However, it comes at the expense of a noticeable increase in residual distance deviation R_{dev} to 0.11 Å (Table 3). Ignoring RDC restraints for the bulge residues during refinement makes the envelope even tighter (Figure 6C), with an overall rmsd of 3.7 Å, and an rmsd for the two stems excluding the bulge of 2.4 Å. This also relieves somewhat the distortion of distance restraints (R_{dev} decreases to 0.09 Å), but the overall average deviation between the calculated and observed RDC increases to 9.3 Hz because of the bulge residues. It is clear that there are internal conflicts not only between NOE-derived distance restraints but also between distances and RDC data, which have very different motional averaging properties. These conflicts could be exploited by such methods as MDtar (*45*, *46*) or multiple-copy refinement (*47*) to determine individual conformations contributing to the NMR signals. At present, we consider the structural envelope calculated with the most complete set of experimental data the best representation of the average conformation for WT34 RNA (Figure 6B).

Very surprisingly, the situation is dramatically different for 10U RNA, which has a single C to U substitution compared to WT34 RNA. Refinement with distance and torsion restraints produced a relatively tight envelope of structures (Figure 6D) with an overall rmsd of 1.6 Å. The residual distance deviation R_{dev} (0.10 Å) is higher for this molecule than for WT34 RNA refined with a similar protocol (Table 3). This is explained by a greater overall number of distance restraints for 10U RNA, a greater number of quantitative restraints calculated with MARDIGRAS, and a tighter flat well width for these restraints (Table 2). The 10U RNA structure also has some individual distance violations greater than 1 Å (fewer than WT34) indicative of some conformational averaging for the bulge residues. Nevertheless, it is clear that the degree of flexibility is significantly reduced for the 10U RNA. This is reflected in a significantly lower value of the NOE-based R^x factor, 0.077, and in the apparent lack of attenuation of RDC constants for the bulge residues (Figure 5). However, agreement between calculated and observed RDC constants is still poor for this structure (Table 3).

Refining 10U RNA with RDC restraints significantly improves the agreement between the calculated and observed

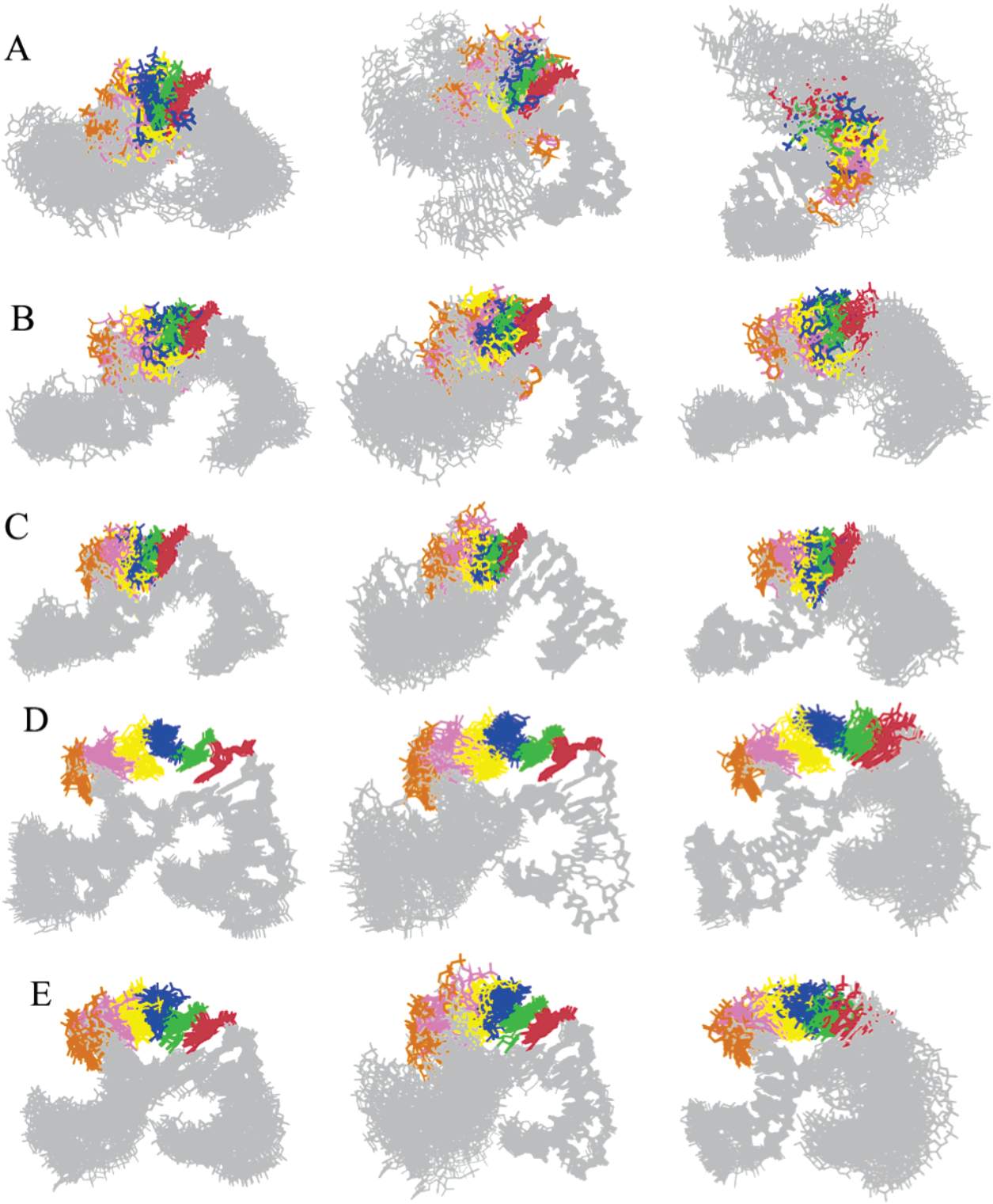

FIGURE 6: Superposition of the 20 best structures refined with AMBER for the WT34 (A−C) and U10 (D and E) RNA. (A and D) Structures refined without RDC restraints. (B and E) Structures refined with a full set of restraints, including the RDC. (C) WT34 structures refined with RDC restraints for the bulge residues excluded from the set of restraints. In all cases, the left, middle, and right panels show structures superimposed using all residues, the first stem, and the second stem, respectively. Residue A8 is shown in red, U9 in green, C10/U10 in blue, C11 in yellow, C12 in magenta, and U13 in orange.

RDC (Figure 10). However, the residual distance deviation R_{dev} increases by 30%, and the NOE-based R^x factor increases by 10% (Table 3). Counterintuitively at first glance, the conformational envelope becomes less tight, when 10U RNA is refined with RDC restraints (Figure 6E), with an overall rmsd of 2.5 Å (Table 4). The explanation for this result is similar to that for WT34 RNA. It suggests that there is limited flexibility in the bulge of 10U; different averaging properties of RDC and NOE would lead to internal conflicts in experimental restraints. One possible source of flexibility is residue U13. In most calculated conformations, it is positioned in the major groove of stem 2 (Figure 8), but in some structures, it is partially stacked between C12 and C14 (Figure 9). It is interesting that the difference between the

Table 3: Figures of Merit for the WT34 and 10U Refined NMR Structures[a]

	energy[e]	R_{dev}[f]	R^{xg}	RDC[h]	RDC[i]	magnitude[j]	rhombicity[k]
WT34[b]	−535 (54)	0.074 (0.005)	0.106 (0.002)	22.2 (2.0)	22.9 (3.4)	18.8 (3.8)	0.62 (0.21)
WT34[c]	−451 (67)	0.110 (0.007)	0.109 (0.003)	2.2 (0.3)	1.9 (0.2)	40.0 (1.6)	0.35 (0.11)
WT34[d]	−500 (69)	0.090 (0.003)	0.108 (0.003)	9.3 (1.3)	1.9 (0.2)	28.6 (3.2)	0.43 (0.17)
10U[b]	−407 (48)	0.104 (0.003)	0.077 (0.002)	23.3 (1.1)	n/a[l]	27.7 (1.5)	0.45 (0.18)
10U[c]	−502 (64)	0.132 (0.003)	0.084 (0.003)	1.2 (0.3)	n/a[l]	41.3 (1.4)	0.74 (0.12)

[a] Average and standard deviations (in parentheses) for the 20 best structures. [b] Structures refined with distance and torsion angle restraints but without RDCs. [c] Structures refined with a full set of restraints, including all RDCs. [d] Structures refined with distance and torsion angle restraints and with RDCs for the stem residues (excluding the bulge residues). [e] Amber conformational energy in kilocalories per mole. [f] Average distance deviation in angstroms. [g] Sixth-root weighted R^x factor. [h] Average deviation between the calculated and observed RDC (hertz), calculated for all residues. [i] Average deviation between the calculated and observed RDC (hertz), calculated for only the stem residues. [j] Absolute value of the magnitude of the alignment of refined structures, calculated from PALES output, assuming a CH distance of 1.085 Å (average value between aromatic and aliphatic CH vectors). [k] Rhombicity of the alignment tensor for the refined structures recalculated from the Eigen values reported by PALES as $(A_{xx} - A_{yy})/A_{zz}$. [l] Not available.

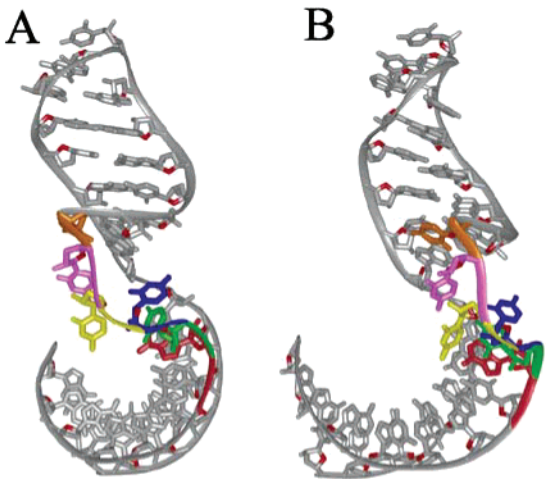

FIGURE 7: Ribbon representation of the best WT34 structure refined with the full set of restraints. Panels A and B show two orthogonal views. Residues are colored as described in the legend of Figure 6. Note the L shape of the molecule.

Table 4: Pairwise Atomic rmsds within the Ensembles of Refined WT34 and 10U Structures[a]

	all[e]	both stems[f]	stem 1	stem 2[g]	Figure[h]
WT34[b]	7.1 (2.4)	5.6 (2.2)	1.3 (0.2)	1.6 (0.6)	6A
WT34[c]	4.1 (1.3)	2.9 (1.3)	0.9 (0.2)	1.5 (0.6)	6B
WT34[d]	3.7 (1.1)	2.4 (1.0)	0.9 (0.3)	1.0 (0.3)	6C
10U[b]	1.6 (0.4)	1.3 (0.4)	0.4 (0.2)	0.8 (0.2)	6D
10U[c]	2.5 (0.9)	1.9 (0.7)	0.8 (0.5)	1.2 (0.4)	6E

[a] Average and standard deviations (in parentheses) of rmsds calculated using heavy atoms among the 20 best structures. [b] Structures refined with distance and torsion angle restraints but without RDCs. [c] Structures refined with a full set of restraints, including all RDCs. [d] Structures refined with distance and torsion angle restraints and with RDCs for the stem residues (excluding the bulge residues). [e] All residues. [f] Excluding bulge residues. [g] Stem 2, including the tetraloop. [h] Figure showing the superposition of the corresponding ensemble.

10U structures calculated with and without RDC is not that great, 3.1 Å on average (Table 5), but it is greater than the average pairwise rmsd for each of the two individual envelopes [2.5 Å for the structures refined with the RDC and 1.6 Å without the RDC (Table 4)]. As in the case with WT34 RNA, we consider the structures refined with the complete set of experimental data the best representation of the average conformation of 10U RNA (Figure 6E).

In addition to very different dynamic properties, the two RNAs have drastically different overall shapes. 10U RNA is significantly more compact, with two stems making a U turn at the bulge (Figure 8). WT34 RNA is more extended, with an overall L shape (Figure 7). This shape is consistent with WT34 RNA having a higher overall correlation time τ_c (see Materials and Methods), although different dynamics may also contribute to differences in the apparent τ_c.

Within the bulge region of the 10U RNA, A helix-like stacking extends from C7 to the first two bulge nucleotides (A8 and U9) in most structures (Figure 8). A8 H2 has a strong sequential NOE to U9 H1′ and a strong cross-strand NOE to G28 H1′, characteristic of A helical structure. The position of U10 is more varied than that of A8 and U9. In the best structure and some other structures, the base of U10 is partially stacked on U9. At the other end of the bulge, two major arrangements of C12 and U13 could be identified among the ensembles (Figure 9). In one of them, U13 is positioned in the major groove of the second stem and C12 is most often partially stacked on C14. In another, U13 is partially stacked between C14 and C12. Viewed from an orientation as shown in Figure 8, the three bulge nucleotides U10, C11, and C12 are juxtaposed at the top of the structure, exposed to solvent. In WT34 RNA, as in 10U RNA, helix-like stacking of A8 and U9 on C7 is also seen in most structures within the ensembles (Figure 7). The rest of the bulge nucleotides (C10−U13) assume a wide variety of conformations, reflecting the more dynamic nature of these residues.

DISCUSSION

Since detection of the KH motif as one of the most common RNA-binding modules, more than 100 KH-containing proteins have been identified (*48*, *49*), most of which contain multiple copies of the KH motif. Intriguingly, only a very small number of *in vivo* RNA targets of the KH domains have been identified to date; no atomic structure of any *bona fide* target RNA has been reported in its unbound state. The two RNA structures we describe here provide the first detailed structural and dynamic information of a KH target RNA free in solution. These structures are also quite different from the majority of reported solution RNA structures in that they have a large asymmetric internal loop sequence (six nucleotides vs two and three nucleotides in HIV-2 and HIV-1 TAR RNAs, respectively) present in only one of the strands, instead of being hairpin loops or bulged at both strands. In fact, asymmetric internal loops of this size are underrepresented even in a large database of crystal structures of ribosomal RNA (our unpublished data).

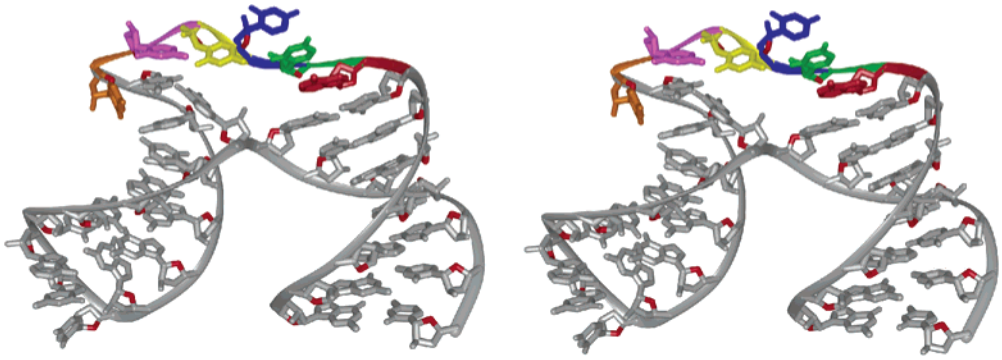

FIGURE 8: Stereoview of the best 10U structure refined with the full set of restraints. Residues are colored as described in the legend of Figure 6. Note the U shape of the molecule.

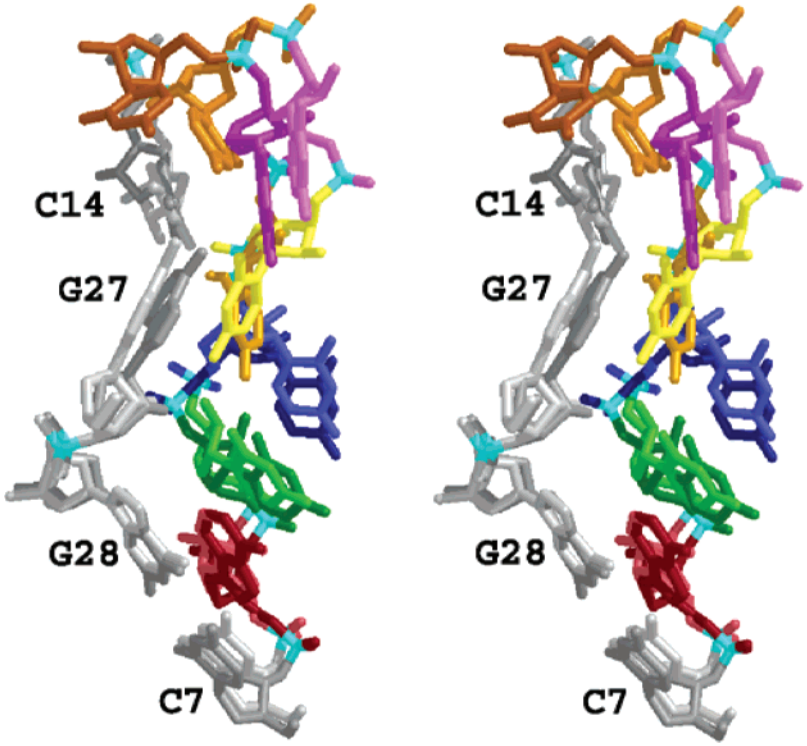

FIGURE 9: Stereoview of a superposition of two conformations of the 10U bulge, refined with the full set of restraints. In one of them, the U13 residue (orange) is positioned in the major groove of the second stem (the same as shown in Figure 8). In another, U13 is partially stacked between C14 and C12. Residues are colored as described in the legend of Figure 6; however, darker colors are used for the most frequent conformation.

In both structures, the GUGA tetraloop and the two helical stems are very well defined (Figure 6). Formation of the Watson−Crick C7·G28 and C14·G27 base pairs is evident in the observation of sharp imino resonances from G27 and G28 and their characteristic NOEs to the amino resonances of C14 and C7 (data not shown). Clearly, the predicted secondary structure of loop B RNA (Figure 1B,C) is correct; a six-nucleotide asymmetric internal loop resides on the 5′ side of an otherwise double-stranded RNA. Given the relatively large bulge, it is not surprising to see that the structures exhibit certain degrees of flexibility around this region (Figure 6). What is quite unexpected is the fact that the two RNA molecules exhibit such a drastic difference in both overall shape and dynamics (Figures 5−8), all caused by a single C to U substitution at the third position of the bulge sequence. How a single substitution causes these changes is not readily understood from the structures, and it warrants further investigation. A similar effect has been reported for the influenza A virus promoter (*50*), but in that case, the substitution caused a local rearrangement in the secondary structure of RNA. In our case, both C10 in WT34 RNA and U10 in 10U RNA are relatively exposed to the solvent and not involved in hydrogen bonding (Figures 6 and 7). A substitution of a carbonyl oxygen for an amino group must change to some extent the hydration properties of the bulge. Realistic MD simulations in water might show if this change in hydration would be sufficient to bring about such dramatic changes in both the dynamics and overall shape of this RNA. Another possible explanation of the differences between the two molecules is a transient formation of intramolecular hydrogen bonds (not observed by NMR). If this is the case, such transient hydrogen bonds are more plausible for the more dynamic internal loop of WT34.

The bulge sequences of WT34 and 10U RNA we investigate here are two representative sequences of the *Enterovirus* genus. Our sequence analysis (summarized in Table 1) shows that the vast majority of the sequences have six nucleotides. The identity of the first nucleotide is A, G,

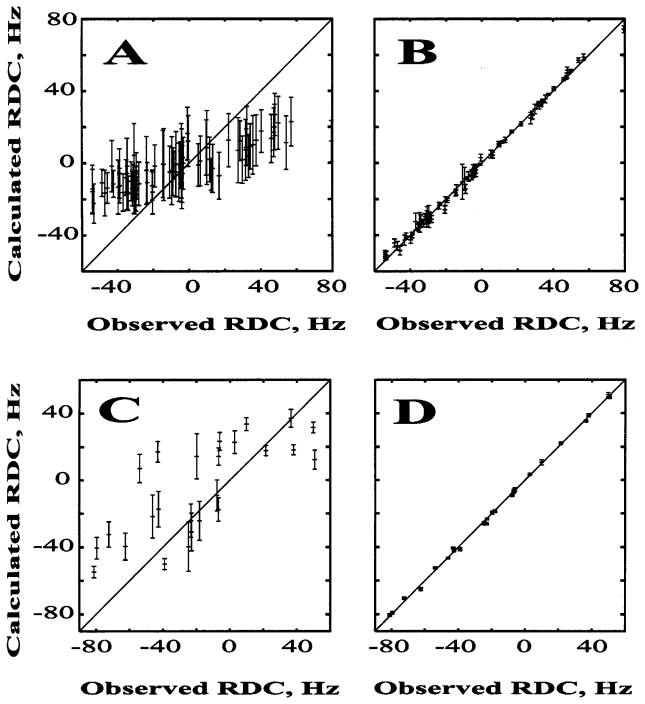

FIGURE 10: Comparison of the experimental and calculated RDC for the refined WT34 (A and B) and 10U (C and D) sequences. Experimental RDC were measured as described in the text. Theoretical RDCs were calculated with the PALES program for the structures refined using distance and torsion angle restraints but no RDC (A and C) and for the structures refined with the full set of restraints, including the RDC (B and D). In each case, error bars show standard deviations of the RDC calculated for the 20 best structures.

Table 5: Pairwise Atomic rmsds between Various Ensembles of Refined WT34 and 10U Structures[a]

	WT34[c]	WT34[d]	10U[b]	10U[c]
WT34[b]	6.1 (1.6)	6.2 (1.5)	8.0 (1.0)	7.9 (1.3)
WT34[c]		4.0 (1.3)	6.8 (0.7)	6.9 (0.9)
WT34[d]			6.4 (0.6)	6.7 (1.0)
10U[b]				3.1 (0.6)

[a] Average and standard deviations (in parentheses) of rmsds calculated using all heavy atoms among 20 pairs of structures. [b] Structures refined with distance and torsion angle restraints but without RDCs. [c] Structures refined with a full set of restraints, including all RDCs. [d] Structures refined with distance and torsion angle restraints and with RDCs for the stem residues.

or U, with A and G being in more than 90% of the sequences. The second nucleotide can be any type, although pyrimidine (especially U) appears more frequently than purine. The third nucleotide can also be any type, but the tendency to have a pyrimidine is even stronger (>90%) than at the second position. The fourth and fifth nucleotides are exclusively cytidines. The last nucleotide is either C or U. In general, the bulge sequence is pyrimidine-rich, and the 5′ half of the sequence exhibits more variation than the 3′ half. Interestingly, in the structures of the WT34 and 10U RNAs, the 5′ half of the bulge sequence is apparently more structured (less flexible) than the 3′ half. Notably, continuous stacking of the first two bulge nucleotides (A8 and U9) on C7 seems to be a common feature among most structures within both WT34 and 10U ensembles. Partial stacking of U10 on U9 is predominant in the 10U structures, and it is also seen in some structures of the WT34 ensemble. Structures of the 3′ half of the bulge are less converged, even in the relatively well-defined 10U ensemble. These interesting observations about sequences and structures prompt us to hypothesize that different residues of the bulge assume distinct roles in protein recognition. Sequence variation and structure conservation of the first two bulge nucleotides may suggest that their role is mainly structural in presenting the rest of the bulge sequence in an appropriate orientation or conformation for interacting with PCBP2. While sequence conservation of the 3′ half of the bulge is probably dictated by the need for sequence-specific recognition, adequate structural flexibility may be required to ensure that an induced fit of the RNA target can take place upon protein contact.

So far, two structures exhibiting the KH domain−RNA interaction have been published. One is the Nova2 KH3 domain bound to a stem−loop RNA (*51*), and another is the KH-QUA2 domain of SF1 (splicing factor 1) bound to a single-stranded intron branch site RNA (*52*). Two more structures have been determined for KH domain−DNA interactions: KH3 and KH4 of FUSE-binding protein bound to a single-stranded DNA (*53*) and hnRNP K-KH3 bound to a single-stranded DNA (*54*). In all of these complexes, the overall KH domain structures are similar. In the cases of Nova2 KH3 and hnRNP K-KH3, the KH structures in their unbound states were also known (*49*, *55*), showing that the RNA and DNA binding platforms are largely preformed (no substantial changes in protein structures upon binding of RNA or DNA). The nucleic acid-binding platform is similar in all complexes, regardless of whether the target is RNA or DNA. Dictated by the size of the platforms, the RNA and DNA targets in general assume an extended single-stranded conformation with only four or five core recognition residues for each KH domain. Base-specific recognition is mainly achieved by molecular mimicry of Watson−Crick base pairing from the protein functional groups.

We have used NMR to study the structure and RNA binding properties of the KH1 domain of PCBP2, which is the major determinant for interaction with both loop B of stem−loop IV within the IRES and stem−loop B RNA within the cloverleaf (Figure 1A), and its RNA binding properties (unpublished data). The overall structure of the free PCBP2-KH1 domain in solution closely resembles other KH protein structures determined to date (*49*, *51*−*57*). Given the emerging general features of KH domain−nucleic acids recognition (see above) and the striking similarity of the KH domains, there is little reason to doubt that the PCBP2-KH1 domain would use the same platform to bind the viral RNA targets. Assuming this is the case, we would expect the bound form of the target RNA has four or five residues presented in the binding interface with an extended conformation, by analogy to the bound RNA and DNA structures in the known complexes. Inspection of the 10U RNA structure ensemble reveals that the six-nucleotide bulge sequence (especially the central portion from U9 to C12) in general adopts a rather extended conformation (Figures 7 and 8). Residues U9−C12 are juxtaposed on one edge of the molecule, with the functional groups of the bases exposed to solvent and very accessible for interacting with the protein partner. When the conformation of the bulge sequence of 10U RNA is compared to reported RNA structures bound to KH domains (*51*, *52*), it is not hard to envisage that substantial structural rearrangement in the bulge region is probably not required

for a similar KH domain−RNA interaction to occur with enteroviruses. If we assume the central four nucleotides (U9−C12) represent the core recognition sequence by KH1, superposition of the backbone atoms (P, O3′, O5′, C3′, C4′, and C5′) of these residues with those of the core recognition sequence within the stem−loop RNA recognized by Nova2-KH3 gives an rmsd of 1.8 ± 0.4 Å averaged over the 20 best 10U structures. In comparison, this value is 3.0 ± 0.7 Å for the 20 best WT34 structures, suggesting that more substantial conformational changes are required for the WT34 bulge to interact with the KH1 domain in a manner similar to that for the Nova2-KH3−stem−loop RNA complex. In fact, the bulge sequence in more than half of the structures within the WT34 ensemble has a quite twisted backbone, which, in conjunction with the different overall shape of the molecule, often makes the bulge residues less accessible to the solvent than in the 10U structures. For this subset of structures, it is hard to imagine KH domain−RNA interaction conforming to the emerging common features would take place without incurring substantial structural rearrangements. On the basis of this analysis and the difference in dynamic properties of the WT34 and 10U RNA, we speculate that less RNA structural flexibility correlates with pre-existing conformational features favoring protein recognition. Of course, that simplistically suggests that 10U RNA binding should have an entropic advantage.

ACKNOWLEDGMENT

We acknowledge use of the University of California at San Francisco Computer Graphics Laboratory, which is supported by an NIH grant.

SUPPORTING INFORMATION AVAILABLE

Fingerprint portions of 600 MHz NOESY spectra, portions of 600 MHz constant time $^{1}H-^{13}C$ HSQC spectra of the ^{13}C- and ^{15}N-labeled WT34 RNA, portion of a 600 MHz natural abundance $^{1}H-^{13}C$ HMQC spectrum, overlaid portions of homonuclear TOCSY spectra of WT34 and 10U RNA, portions of the DQF-COSY spectra of WT34 and 10U RNA, ^{2}H spectrum showing quadrupolar splitting of the solvent peak under dilute liquid crystalline conditions, and an example illustrating measurements of residual couplings. 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, *Gene Dev. 12*, 2293−2304.
2. Ehrenfeld, E., and Semler, B. L. (1995) Anatomy of the poliovirus internal ribosome entry site, *Curr. Top. Microbiol. Immunol. 203*, 65−83.
3. Andino, R., Rieckhoff, G. E., and Baltimore, D. (1990) A functional ribonucleoprotein complex forms around the 5′-end of poliovirus RNA, *Cell 63*, 369−380.
4. Andino, R., Rieckhoff, 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*, 3587−3598.
5. Harris, K. S., Xiang, W., Alexander, L., Lane, W. S., Paul, A. V., and Wimmer, E. (1994) Interaction of poliovirus polypeptide 3CDpro with the 5′ and 3′ termini of the poliovirus genome. Identification of viral and cellular cofactors needed for efficient binding, *J. Biol. Chem. 269*, 27004−27014.
6. Gamarnik, A. V., and Andino, R. (2000) Interactions of viral protein 3CD and poly(rC) binding protein with the 5′ untranslated region of the poliovirus genome, *J. Virol. 74*, 2219−2226.
7. Siomi, H., Siomi, M. C., Nussbaum, R. L., and Dreyfuss, G. (1993) The protein product of the fragile X gene, FMR1, has characteristics of an RNA-binding protein, *Cell 74*, 291−298.
8. Dejgaard, K., and Leffers, H. (1996) Characterisation of the nucleic-acid binding activity of KH domains. Different properties of different domains, *Eur. J. Biochem. 241*, 425−431.
9. Silvera, D., Gamarnik, A. V., and Andino, R. (1999) 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. 274*, 38163−38170.
10. Klinck, R., Sprules, T., and Gehring, K. (1997) Structural characterization of three RNA hexanucleotide loops from the internal ribosome entry site of polioviruses, *Nucleic Acids Res. 25*, 2129−2137.
11. Collier, A. J., Gallego, J., Klinck, R., Cole, P. T., Harris, S. J., Harrison, G. P., Aboul-Ela, F., Varani, G., and Walker, S. (2002) A conserved RNA structure within the HCV IRES eIF3-binding site, *Nat. Struct. Biol. 9*, 375−380.
12. Kieft, J. S., Zhou, K. H., Jubin, R., Murray, M. G., Lau, J. Y. N., and Doudna, J. A. (1999) The hepatitis C virus internal ribosome entry site adopts an ion-dependent tertiary fold, *J. Mol. Biol. 292*, 513−529.
13. Kieft, J. S., Zhou, K. H., Grech, A., Jubin, R., and Doudna, J. A. (2002) Crystal structure of an RNA tertiary domain essential to HCV IRES-mediated translation initiation, *Nat. Struct. Biol. 9*, 370−374.
14. Klinck, R., Westhof, E., Walker, S., Afshar, M., Collier, A., and Aboul-Ela, F. (2000) A potential RNA drug target in the hepatitis C virus internal ribosomal entry site, *RNA 6*, 1423−1431.
15. Lukavsky, P. J., Otto, G. A., Lancaster, A. M., Sarnow, P., and Puglisi, J. D. (2000) Structures of two RNA domains essential for hepatitis C virus internal ribosome entry site function, *Nat. Struct. Biol. 7*, 1105−1110.
16. Batey, R. T., Battiste, J. L., and Williamson, J. R. (1995) Preparation of isotopically enriched RNAs for heteronuclear NMR, in *Methods in Enzymology* (James, T. L., Ed.) Vol. 261, pp 300−322, Academic Press, New York.
17. Milligan, J. F., and Uhlenbeck, O. C. (1989) Synthesis of small RNAs using T7 RNA polymerase, in *Methods in Enzymology* (Dahlberg, J. E., and Abelson, J. N., Eds.) Vol. 180, pp 51−62, Academic Press, New York.
18. 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*, 277−293.
19. Goddard, T. D., and Kneller, D. G. (1998) *SPARKY*, version 3.0, University of California, San Francisco.
20. Smallcombe, S. H. (1993) Solvent suppression with symmetrically-shifted pulses, *J. Am. Chem. Soc. 115*, 4776−4785.
21. Zuiderweg, E. R. P., McIntosh, L. P., Dahlquist, F. W., and Fesik, S. W. (1990) Three-dimensional C-13-resolved proton NOE spectroscopy of uniformly C-13-labeled proteins for the NMR assignments and structure determination of larger molecules, *J. Magn. Reson. 86*, 210−216.
22. Otting, G., and Wüthrich, K. (1990) Heteronuclear filters in 2D [^{1}H,^{1}H]-NMR spectroscopy: Combined use with isotope labeling for studies of macromolecular conformation and intermolecular interaction, *Q. Rev. Biophys. 32*, 39−96.
23. Borgias, B. A., and James, T. L. (1990) MARDIGRAS: Procedure for Matrix Analysis of Relaxation for DIscerning Geometry of an Aqueous Structure, *J. Magn. Reson. 87*, 475−487.
24. Liu, H., Spielmann, H. P., Ulyanov, N. B., Wemmer, D. E., and James, T. L. (1995) Interproton distance bounds from 2D-NOE intensities: Effect of experimental noise and peak integration errors, *J. Biomol. NMR 6*, 390−402.
25. 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*, 4371−4379.
26. Schmitz, U., Behrens, S., Freymann, D. M., Keenan, R. J., Lukavsky, P., Walter, P., and James, T. L. (1999) Structure of the phylogenetically most conserved domain of SRP RNA, *RNA 5*, 1419−1429.
27. Zhurkin, V. B., Ulyanov, N. B., Gorin, A. A., and Jernigan, R. L. (1991) Static and statistical bending of DNA evaluated by Monte Carlo simulations, *Proc. Natl. Acad. Sci. U.S.A. 88*, 7046−7050.

28. Ruckert, M., and Otting, G. (2000) Alignment of biological molecules in novel nonionic crystalline media for NMR experiments, *J. Am. Chem. Soc. 122*, 7793−7797.
29. Güntert, P., Mumenthaler, C., and Wüthrich, K. (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA, *J. Mol. Biol. 273*, 283−298.
30. Case, D. A., Pearlman, D. A., Caldwell, J. W., Cheatham, I. T. E., Wang, J., Ross, W. S., Simmerling, C. L., Darden, T. A., Merz, K. M., Stanton, R. V., Cheng, A. L., Vincent, J. J., Crowley, M., Tsui, V., Gohlke, H., Radmer, R. J., Duan, Y., Pitera, J., Massova, I., Seibel, G. L., Singh, U. C., Weiner, P. K., and Kollman, P. A. (2002) *AMBER*, version 7, University of California, San Francisco.
31. Tsui, V., Zhu, L., Huang, T.-H., Wright, P. E., and Case, D. A. (2000) Assessment of zinc finger orientations by residual dipolar coupling constants, *J. Biomol. NMR 16*, 9−21.
32. Ulyanov, N. B., Schmitz, U., and James, T. L. (1993) Metropolis Monte Carlo calculations of DNA structure using internal coordinates and NMR distance restraints: An alternative method for generating high-resolution solution structure, *J. Biomol. NMR 3*, 547−568.
33. Clowney, L., Jain, S. C., Srinivasan, A. R., Westbrook, J., Olson, W. K., and Berman, H. M. (1996) Geometric parameters in nucleic acids: Nitrogenous bases, *J. Am. Chem. Soc. 118*, 509−518.
34. Gelbin, A., Schneider, B., Clowney, L., Hsieh, S.-H., Olson, W. K., and Berman, H. M. (1996) Geometric parameters in nucleic acids: Sugar and phosphate constituents, *J. Am. Chem. Soc. 118*, 519−529.
35. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules, *J. Am. Chem. Soc. 117*, 5179−5197.
36. Tjandra, N., Tate, S., Ono, A., Kainosho, M., and Bax, A. (2000) The NMR structure of a DNA dodecamer in an aqueous dilute liquid crystalline phase, *J. Am. Chem. Soc. 122*, 6190−6200.
37. Wu, Z., Delaglio, F., Tjandra, N., Zhurkin, V. B., and Bax, A. (2003) Overall structure and sugar dynamics of a DNA dodecamer from homo- and heteronuclear dipolar couplings and ^{31}P chemical shift anisotropy, *J. Biomol. NMR 26*, 297−315.
38. Losonczi, J. A., Andrec, M., Fischer, M. W., and Prestegard, J. H. (1999) Order matrix analysis of residual dipolar couplings using singular value decomposition, *J. Magn. Reson. 138*, 334−342.
39. Zweckstetter, M., and Bax, A. (2000) Prediction of sterically induced alignment in a dilute liquid crystalline phase: Aid to protein structure determination by NMR, *J. Am. Chem. Soc. 122*, 3791−3792.
40. Ulyanov, N. B., and James, T. L. (1995) Statistical analysis of DNA duplex structural features, in *Methods in Enzymology* (James, T. L., Ed.) Vol. 261, pp 90−120, Academic Press, New York.
41. Ferrin, T. E., Huang, C. C., Jarvis, L. E., and Langridge, R. (1988) The MIDAS display system, *Mol. Graphics 6*, 13−27.
42. Huang, C. C., Couch, G. S., Pettersen, E. F., and Ferrin, T. E. (1996) Chimera: An extensible molecular modeling application constructed using standard components, *Pac. Symp. Biocomput. '96 1*, 724.
43. Keepers, J. W., and James, T. L. (1984) A theoretical study of distance determinations from NMR. Two-dimensional nuclear Overhauser effect spectra, *J. Magn. Reson. 57*, 404−426.
44. Al-Hashimi, H. M., Gosser, Y., Gorin, A., Hu, W., 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.
45. 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*, 229−244.
46. Hess, B., and Scheek, R. M. (2003) Orientation restraints in molecular dynamics simulations using time and ensemble averaging, *J. Magn. Reson. 164*, 19−27.
47. 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*, 147−164.
48. Burd, C. G., and Dreyfuss, G. (1994) Conserved structures and diversity of functions of RNA-binding proteins, *Science 265*, 615−621.
49. Lewis, H. A., Chen, H., Edo, C., Buckanovich, R. J., Yang, Y. Y., Musunuru, K., Zhong, R., Darnell, R. B., and Burley, S. K. (1999) Crystal structures of Nova-1 and Nova-2 K-homology RNA-binding domains, *Struct. Folding Des. 7*, 191−203.
50. Lee, M. K., Bae, S. H., Park, C. J., Cheong, H. K., Cheong, C., and Choi, B. S. (2003) A single-nucleotide natural variation (U4 to C4) in an influenza A virus promoter exhibits a large structural change: implications for differential viral RNA synthesis by RNA-dependent RNA polymerase, *Nucleic Acids Res. 31*, 1216−1223.
51. Lewis, H. A., Musunuru, K., Jensen, K. B., Edo, C., Chen, H., Darnell, R. B., and Burley, S. K. (2000) Sequence-specific RNA binding by a Nova KH domain: implications for paraneoplastic disease and the fragile X syndrome, *Cell 100*, 323−332.
52. Liu, Z., Luyten, I., Bottomley, M. J., Messias, A. C., Houngninou-Molango, S., Sprangers, R., Zanier, K., Kramer, A., and Sattler, M. (2001) Structural basis for recognition of the intron branch site RNA by splicing factor 1, *Science 294*, 1098−1102.
53. Braddock, D. T., Louis, J. M., Baber, J. L., Levens, D., and Clore, G. M. (2002) Structure and dynamics of KH domains from FBP bound to single-stranded DNA, *Nature 415*, 1051−1056.
54. Braddock, D. T., Baber, J. L., Levens, D., and Clore, G. M. (2002) Molecular basis of sequence-specific single-stranded DNA recognition by KH domains: solution structure of a complex between hnRNP K KH3 and single-stranded DNA, *EMBO J. 21*, 3476−3485.
55. Baber, J. L., Libutti, D., Levens, D., and Tjandra, N. (1999) High precision solution structure of the C-terminal KH domain of heterogeneous nuclear ribonucleoprotein K, a c-myc transcription factor, *J. Mol. Biol. 289*, 949−962.
56. Musco, G., Stier, G., Joseph, C., Castiglione Morelli, M. A., Nilges, M., Gibson, T. J., and Pastore, A. (1996) Three-dimensional structure and stability of the KH domain: Molecular insights into the fragile X syndrome, *Cell 85*, 237−245.
57. Musco, G., Kharrat, A., Stier, G., Fraternali, F., Gibson, T. J., Nilges, M., and Pastore, A. (1997) The solution structure of the first KH domain of FMR1, the protein responsible for the fragile X syndrome, *Nat. Struct. Biol. 4*, 712−716.

BI0363228