7814–7821 *Nucleic Acids Research, 2010, Vol. 38, No. 21* *Published online 26 July 2010*
doi:10.1093/nar/gkq640

NMR structure of the *let-7* miRNA interacting with the site LCS1 of *lin-41* mRNA from *Caenorhabditis elegans*

Mirko Cevec[1], Christophe Thibaudeau[1] and Janez Plavec[1,2,3,*]

[1]Slovenian NMR Center, National Institute of Chemistry, Hajdrihova 19, [2]Faculty of Chemistry and Chemical Technology, University of Ljubljana, Aškerčeva cesta 5 and [3]EN-FIST Center of Excellence, SI-1000 Ljubljana, Slovenia

Received May 5, 2010; Revised July 2, 2010; Accepted July 5, 2010

ABSTRACT

We have determined the 3D structure of a 34-nt RNA construct, herein named LCS1co, which mimics the interaction of *let-7* microRNA (miRNA) to one of its complementary binding sites, LCS1, in the 3′-untranslated region of *lin-41* mRNA by solution-state NMR spectroscopy. *let-7* miRNAs control the timing of development of the nematode *Caenorhabditis elegans* and are highly conserved in mammals. The sequence and structure of the two conserved *let-7* complementary sites, LCS1 and LCS2, in the 3′-untranslated region of *lin-41* mRNA are important for a proper downregulation of *lin-41*. The high-resolution NMR structure reveals details of the binding of *let-7* miRNA to *lin-41* mRNA which involves formation of a complex with non-canonical structural elements within the seed region. LCS1co exhibits a stem-loop structure with two stems, an asymmetric internal loop and an adenine bulge. Comparison with the NMR solution-state structure of the *let-7*:*lin-41* complex involving the LCS2-binding site shows that conformational freedom of the asymmetric internal loop of LCS1co correlates with a smaller bend between the upper and lower stems in comparison to the well-defined asymmetric loop of LCS2co.

INTRODUCTION

let-7 microRNA (miRNA), a 21-nt RNA oligonucleotide, has been shown to control the timing of development of nematode *Caenorhabditis elegans* from the last larval stage to adulthood (1,2). The enzyme Dicer produces *let-7* miRNA by fragmentation of an endogenous hairpin RNA precursor (3). The mature *let-7* miRNA assembles into the Argonaute protein of the miRNA-induced silencing complex (miRISC) which uses *let-7* miRNA to identify target *lin-41* mRNA and to repress protein synthesis. In *C. elegans*, the structure of the dsRNA influences the selection of an Argonaute family member. RDE-1 is involved in RNA interference induced by perfectly paired dsRNA, while ALG-1 and ALG-2 are necessary for the miRNA function carried out by dsRNA containing nucleotide mismatches (4).

Efficient regulation of *lin-41* requires the presence of the two conserved *let-7* complementary sites, called LCS1 and LCS2, within the 3′-untranslated region (3′-UTR) of *lin-41* mRNA. miRNA target sites are typically composed of 7- to 8-nt seed-matched regions (5). However, in the case of *let-7* miRNA both interaction sites are composed of a seed-mismatched region, an asymmetric internal loop and a 3′-compensatory Watson–Crick pairing. This family of target sites named 3′-compensatory miRNA target sites is rare and represents only 1% of all conserved target sites in mammals (5). A tentative model was proposed in which miRNA is bound by Argonaute at residues 2–8 (the seed region) in a preorganized form to support efficient pairing (5). After nucleation of the seed residues central residues do not base pair and pairing continues between residues 13 and 16. Apparently, this is energetically favorable, because there is no need for the miRNA to wrap around the mRNA or for the Argonaute protein to change its conformation, which is the usual event for typical target sites after forming central base pairs (5). Structural investigation of *Thermus thermophilus* Argonaute protein bound to a 5′-phosphorylated 21-nt guide DNA and a 20-nt target RNA showed that the guide strand was anchored within the Argonaute scaffold with extensive intermolecular interactions (6). On the other hand, the target strand (i.e. mRNA) which is complementary to the seed segment makes no contacts with the Argonaute scaffold and can accommodate a single bulge when sufficient pairing is retained (6).

The *let-7* sequence is well conserved among diverse organisms (2). It was proposed that *let-7* miRNA

*To whom correspondence should be addressed. Tel: +386 1 47 60 353; Fax: +386 1 47 60 300; Email: janez.plavec@ki.si

regulates cell differentiation and proliferation in animals. *let-7* miRNA has been linked to multiple oncogenes and could thus represent a good diagnostic marker and target in potential cancer therapy. Low levels of *let-7* miRNA have been correlated with high protein levels of RAS, which is a known oncogene in lung cancers (2).

In the present study, we focus on the first *let-7* complementary site LCS1 in order to delineate structural features of the *let-7*:*lin-41* complex that is being recognized and interacts with enzymes of the miRNA machinery. The LCS1 interaction site complex consists of two stem regions composed of a G·U wobble, nine AU and seven GC Watson–Crick base pairs. We designed a unimolecular 34-nt RNA construct (herein named LCS1co) with a GAAA tetraloop connecting the strands originating from *let-7* miRNA and *lin-41* mRNA and retained the asymmetric internal loop and the adenine bulge of the original complex (Figure 1). It has been shown, that the nucleotide sequence of *let-7* target sites and the linker region are important for a proper downregulation of *lin-41* (7). The adenine bulge was predicted to be a binding site for proteins specific to the *let-7*:*lin-41* interaction, while the imperfect base pairing in the seed region prevented other *let-7* family members to regulate *lin-41* (8). Besides *lsy-6* miRNA target sites within the 3′-UTR of *cog-1* mRNA, there are two additional AU-rich sequences in *C. elegans* which are located 3′ to the two *lsy-6* sites and play a role in regulation of protein expression (9). Structural configurations of these sequences are expected to influence RISC access to *lsy-6* sites.

The high-resolution 3D structure of the *let-7*:*lin-41* construct based on restraints from a set of standard 2D and 3D NMR experiments was expected to give new insights into a potential role of internal loops and adenine bulges for recognition by proteins of miRISC. In particular, our work aimed at determining the influence of non-canonical structural elements on the stability of the miRNA:mRNA complex. At the outline, a comparison of the structural features of LCS1co and the 33-nt hairpin RNA oligonucleotide (herein named LCS2co) designed to mimic the interaction of *let-7* complementary site LCS2 with its target in the 3′-UTR of *lin-41* mRNA, that was recently determined in our laboratory (10), was expected to provide new insights from a structural perspective into why the two sites on mRNA have been utilized to control expression of a particular gene.

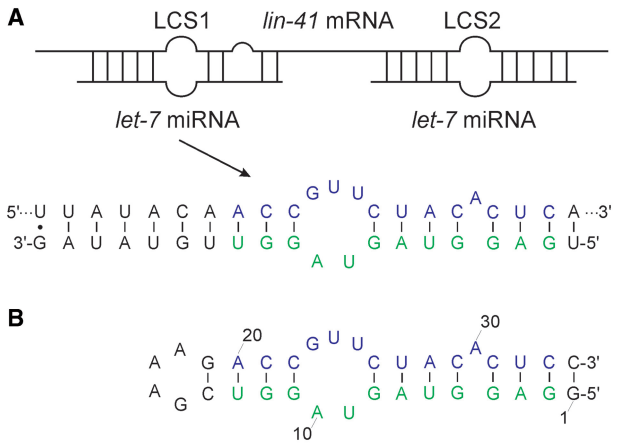

Figure 1. (**A**) Schematic representation of the interaction of *let-7* miRNA with binding sites LCS1 and LCS2 in the 3′-UTR of *lin-41* mRNA. (**B**) 34-nt RNA construct, LCS1co, mimicking the complex involving LCS1. Residues originating from *let-7* and *lin-41* sequences are colored in green and blue, respectively. The numbering scheme used in the current study is indicated.

MATERIALS AND METHODS

RNA preparation

The uniformly isotopically labeled 34-nt RNA was prepared by *in vitro* enzymatic transcription with T7 RNA polymerase (Promega) using a standard published protocol (11), ^{13}C,^{15}N-labeled rNTPs (Silantes) and a DNA template (IDT-DNA) composed of a T7 promoter region and a coding template strand. An additional 5′C-G3′ base pair was included to close the GAAA tetraloop and to increase the construct stability. The terminal 5′U-A3′ base pair was changed into a 5′G-C3′ base pair which allowed us to prepare a ^{13}C,^{15}N-labeled RNA oligonucleotide using T7 RNA polymerase. The last two residues at the 5′ end of the template strand were modified with C2′-methoxy groups for better yield (12). The target oligonucleotide was purified by denaturating 17% PAGE (7 M urea), followed by electroelution (Elutrap system, Schleicher and Schuell BioScience) and extensively dialyzed against an NMR buffer using Centricon YM-3 centrifugation devices (Millipore). The NMR sample was prepared by dissolving RNA into an aqueous solution (10 mM sodium phosphate buffer, pH 6.8, 20 mM NaCl in 95% H_2O, 5% ^2H_2O or 100% ^2H_2O). The sample was heated to 368 K and snap-cooled on ice to ensure proper folding of the RNA hairpin. Sample concentration was determined by measuring UV absorbance at 260 nm on a Perkin-Elmer Lambda Bio 40 UV spectrometer and by assuming a molar extinction coefficient (ε) of 343 600 l/(mol cm). The concentration of the NMR sample used to perform assignments and to extract NOE restraints was 2.0 mM. Another 1.0 mM NMR sample was prepared for measuring residual dipolar couplings (RDCs) by mixing RNA with a filamentous Pf1 phage solution (Asla Biotech) to the total phage concentration of 17 mg/ml (deuterium splitting was 17.3 Hz at 800 MHz). UV melting temperature of RNA sample was 332 K (1.9 μM in oligonucleotide, 10 mM sodium phosphate buffer, pH 6.8, 20 mM NaCl).

NMR spectroscopy and structure calculations

NMR spectra were recorded on a Varian NMR System 600 MHz spectrometer equipped with a 5-mm pulsed field *z*-gradient penta probe and on a Varian NMR System 800 MHz spectrometer with a 5 mm HCN cold probe. Exchangeable proton resonances were assigned using a combination of 2D NOESY spectra recorded in 95% H_2O with mixing times (τ_m) of 75 and 300 ms at 278 K and of ^{15}N-HSQC spectra. The WATERGATE pulse sequence was used for H_2O suppression. The base pairing pattern was further determined by a 2D

HNN-COSY spectrum of LCS1co. 2D NOESY spectra recorded in 100% $^{2}H_2O$ at 303 K with τ_m of 75, 150 and 300 ms and 3D NOESY-^{13}C-HSQC spectra recorded at 303 K with τ_m of 150 ms were used to assign non-exchangeable protons. We have assigned ^{1}H (86% of all ^{1}H), ^{13}C and ^{15}N resonances with the use of 2D TOCSY, 2D DQF-COSY, 2D ^{13}C- and ^{15}N-HSQC, 2D nucleobase specific experiments (2D HCCH-TOCSY, 2D HNCCCH and 2D HNC-TOCSY-CH), 2D HCN, 3D HCCH-COSY and 3D HCCH-TOCSY (13–17). Spectra were processed and analyzed using Varian VNMRJ 2.1B and Accelrys Felix 2002 software packages.

NOE distance restraints for exchangeable protons were classified as strong (1.8–3.6 Å), medium (2.6–5.0 Å) and weak (3.5–6.5 Å). Inter-proton distance restraints of non-exchangeable protons were calculated from the volumes of NOE cross-peaks in relation to the reference distance of pyrimidine H5–H6 set to 2.45 Å. The upper and lower bounds were ± 20% (τ_m of 75 ms), ± 30% (τ_m of 150 ms) and ± 40% (τ_m of 300 ms). Torsion angles (α, β, γ, δ, ε, ζ and χ) were restrained as described earlier (10).

One-bond ^{1}H-^{15}N RDCs were measured from the splitting of the resonances along the ^{15}N dimension of 2D IPAP ^{15}N-HSQC spectra (18) recorded in the presence and the absence of Pf1 phages, by simulating the F1 doublets (error ± 2.0 Hz). One-bond ^{1}H-^{13}C RDC values were measured for H2–C2, H6–C6 and H8–C8 bonds in 2D CE-CT ^{13}C-HSQC spectra (19). Spectra showed no apparent signs of aggregation.

In the process of structure calculation we used 810 NOE distance restraints (on average 24 NOE restraints per residue), 138 torsion angle restraints, 32 hydrogen bond and 24 bp planarity restraints. Torsion angle, hydrogen bond and planarity restraints were applied only for residues in the upper and lower stems, except restraints for torsion angle χ of A17, which was set to anti, and torsion angles ν_0-ν_4 of U24 and A30, which were set to a S-type sugar pucker. Structures were calculated using simulated annealing with AMBER 9 (20) and Wang *et al.* (21) force field and consequently subjected to conjugate gradient minimization using protocols described earlier (10). A total of 52 RDCs were incorporated into restrained simulated annealing and conjugate gradient minimization. We refined the structure against a single alignment tensor. Initially, we compared differences in the axial (A_a) and rhombic (A_r) components of the alignment tensor obtained for LCS1co and LCS2co, considering the entire molecule as a single rigid unit on one hand, and each of the lower and upper stems as separate units, on the other hand. The values of A_a and A_r for LCS1co were 13.4 Hz and 0.13 (entire molecule), 13.4 Hz and 0.12 (lower stem) and 11.3 Hz and 0.19 (upper stem), respectively. A_a and A_r values for LCS2co were 12.3 Hz and 0.08 (entire molecule), 12.3 Hz and 0.08 (lower stem) and 12.4 Hz and 0.10 (upper stem), respectively. Pairs of values for the upper stems deviate only slightly from the values of the entire molecules and for the lower stems. List of residual dipolar restraints was prepared by makeDIP_RST. It contained also the limiting values for the observed dipolar splittings (in Hz). We increased the error limits for RDC values of residues in the asymmetric internal loop (U9 C6–H6, A10 C2–H2 and C8–H8, G23 C8–H8, U24 C6–H6) to avoid over fitting the RDC data. The plot of RDC values per base pairs of LCS1co construct reveals that the spread in their magnitude is similar for both helical stems and asymmetric internal loop (Supplementary Figure S1). A set of 10 minimized structures with the lowest energy and smallest NMR violations were analyzed using suppose and 3DNA software (22).

Protein Data Bank coordinates deposition

The coordinates were deposited in the Protein Data Bank with accession code 2KPV.

RESULTS

Base pairing

The base pairing network within LCS1co was initially analyzed through the imino–imino (Figure 2A) and imino–amino (data not shown) sequential walks in the 2D NOESY spectra as well as from the ^{15}N chemical shifts in the ^{15}N-HSQC spectra. Assignment was subsequently confirmed by the HNN-COSY spectrum

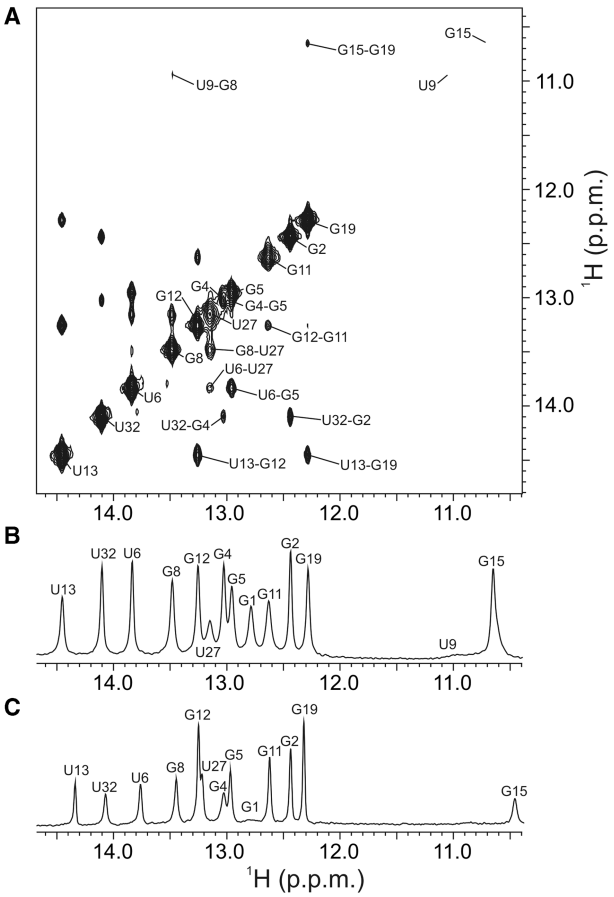

Figure 2. Imino–imino region of 2D NOESY spectrum of LCS1co at 278 K (**A**) and imino proton region of the 1D ^{1}H NMR spectrum at 278 K (**B**) and 303 K (**C**).

(Supplementary Figure S2). The cross-peaks in the HNN-COSY spectrum correlating each imino proton to two nitrogen atoms offered unequivocal identification of regular GC and AU Watson–Crick base pairs. With the assignment of Watson–Crick base pairs we were able to confirm the expected secondary structure of LCS1co presented in Figure 1B. The secondary structure displays the lower stem with eight Watson–Crick base pairs (residues G1–G8, C26–C29 and C31–C34) that includes the adenine bulge (A30), the asymmetric internal loop (residues U9–A10 and G23–U25), the upper stem with four Watson–Crick base pairs (residues G11–C14 and G19–C22) and the GAAA (residues G15–A18) tetraloop (schematically shown in Figure 1B).

Twelve narrow and resolved signals were found in the 1D ^{1}H NMR spectrum (Figure 2B) in the region typical for Watson–Crick imino proton resonances (from 12.3 to 14.5 p.p.m.). An additional narrow signal at δ 10.5 p.p.m. and a broad peak at δ 11.0 p.p.m. are outside the typical range for exchangeable protons involved in Watson–Crick base pairs. The chemical shift of G15 imino proton (δ 10.5 p.p.m. at 278 K) is indicative of a non-canonical base pair which is part of the GAAA loop structure. A broad U9 imino proton signal (δ 11.0 p.p.m.) was observed at 278 K, while G23, U24 and U25 imino proton signals could not be observed at any temperature (Figure 2B and C) presumably due to their exposure to fast exchange with bulk solvent. In full agreement, the HNCCCH spectrum (optimized for uridine residues; data not shown) does not show any cross-peaks between U9, U24 and U25 imino and the corresponding H6 protons. The chemical shifts of U6, U13, G15, U27 and U32 imino protons in the 1D ^{1}H NMR spectrum exhibit the highest changes as a function of temperature. An upfield shift of up to 0.1 p.p.m. upon increase of temperature from 278 to 303 K suggests some degree of flexibility of the above five residues within a 3D structure. The G1 imino signal broadens almost to base-line at 303 K. We also note that imino signals of G11 and U27 (Figure 2B and Supplementary Figure S2) are broader with respect to those involved in strong hydrogen bonds with canonical Watson–Crick base pairs.

Assignment and NMR spectral characteristics

A complete NOE sequential walk between aromatic and anomeric proton resonances could be established in the 2D NOESY spectrum (Figure 3) after the identification of almost all ribose protons using 3D HCCH-COSY, 3D HCCH-RELAY and 3D HCCH-TOCSY spectra. 3D NOESY-^{13}C-HSQC spectra were used to reduce an overlap problem in the ribose region of the 2D NOESY spectra. Weak sequential cross-peaks are observed between C14 H1′ and G15 H8 due to the long distances imposed by the GAAA tetraloop structure. Weak sequential NOE cross-peaks are also observed between A10 and G11 as well as among residues U24, U25, C26 and U27. NOE cross-peaks between intra-nucleotide H8/H6 and H1′ proton resonances display characteristic intensities of nucleobases in the anti-conformation for all residues, except for A16, A18 and U25, which adopt the *syn* conformation.

A chemical shift of A18 H3′ (δ 5.24 p.p.m.) is characteristic for the terminal adenine in the known structures of GAAA tetraloops (10,23). In addition, G19 H1′ (δ 3.80 p.p.m.) which follows the GAAA tetraloop at the 3′-end exhibits an expected chemical shift. Internucleotide NOE contacts in the asymmetric internal loop and around the adenine bulge are displayed schematically in Supplementary Figure S3. We note that the asymmetric internal loop (Supplementary Figure S3A) exhibits a much

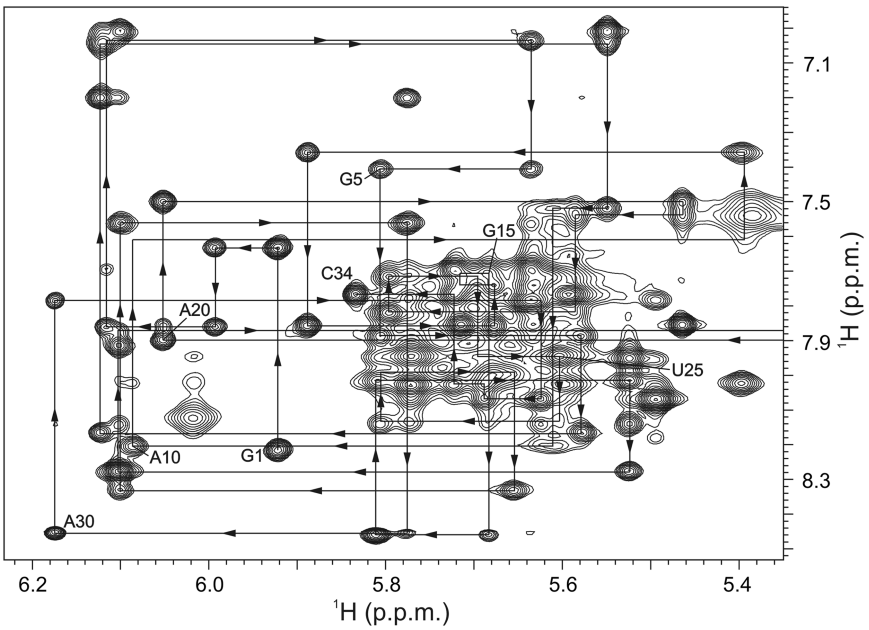

Figure 3. Aromatic-anomeric region of 2D NOESY spectrum (τ_m = 300 ms) at 303 K. Lines indicate aromatic–anomeric sequential walk. Every fifth residue is labeled.

smaller number of NOEs per residue than the rest of the molecule (18 NOEs for U9, 18 for A10, 16 for G23, 20 for U24 and nine for U25). No NOE signals between exchangeable and non-exchangeable protons and a small number of sequential NOEs between ribose and aromatic protons are observed for residues in the asymmetric internal loop. Proton–carbon cross-peaks of ribose moieties of U9, G11, A16, A17, U24, U25 and C26 are weak in the ^{13}C-HSQC spectrum at 303 K and suggest their conformational flexibility.

Most of the potential 34 H1′-H2′ cross-peaks are absent in the 2D DQF-COSY spectrum which suggests that sugar residues adopt predominantly an N-type conformation. However, H1′-H2′ cross-peaks of G1, U24, C29, A30 and C34 exhibit high $^{3}J_{1'2'}$ coupling constants (~8 Hz) implying the predominance of an S-type conformation for the respective sugar moieties.

A separate sample was prepared in the presence of filamentous Pf1 phages in order to measure RDCs. Quality of the spectra of the weakly aligned sample did not deteriorate. The upper and lower stems of LCS1co exhibit small differences in ^{1}H-^{15}N and ^{1}H-^{13}C RDC values (Supplementary Figure S1). However, the lower stem exhibits more negative one-bond ^{1}H-^{15}N RDC values (from −14 to −18 Hz) than the upper stem (from −8 to −12 Hz). Residues G5 and A7 from the lower stem exhibit the highest ^{1}H-^{13}C RDC values of 36 and 40 Hz, respectively, while residues G15, A16, A17 and A18 constituting the GAAA tetraloop and the bulged residue A30 exhibit the lowest RDC values amongst H2–C2, H6–C6 and H8–C8 couplings. A simple comparison of these and other RDC values has enabled us to predict that the two stems are not collinear and that the asymmetric internal loop induces a small bend in the overall structure.

Solution structure of *let-7*-binding site

Simulated annealing and energy minimization of over 100 structures converged to a set of 10 lowest energy structures (Figure 4). Most NOE violations were below 0.2 Å. However, few NOE distance deviations exceeded 0.3 Å with ~70% of them originating from the internucleotide distance restraints between the bulged A30 and residues C29 and C31. Overall pairwise heavy atom RMSD to the mean was 1.21 Å for the 10 superimposed structures (Table 1). Comparison of these structures without considering residues of the asymmetric internal loop gives a lower RMSD value of 0.99 Å, which demonstrates that the loop is not well-defined.

Structural calculations have shown that LCS1co adopts a family of structures with two stem regions. The upper and lower stems are interrupted by the asymmetric internal loop which adopts two families of structures that are energetically comparable and both satisfy NMR data (Figure 5). In both families of structures of the asymmetric internal loop, A10 is stacked on U9 which is in turn stacked on G8 (Figure 5A and B). The NOE correlation peak between A10 H2 and U24 H1′ proves that U24 is oriented into the interior of the asymmetric internal loop. In the first family, U24 adopts a coplanar orientation with U9 and A10, while G23 points into the minor groove and is perpendicular to G11–C22 and G12–C21 base pairs (Figure 5A). U25 is flipped out into the solution. G23 and U24 exhibit an S-type sugar pucker while U25 exhibits an N-type sugar conformation. In the second family, residue U24 is placed in coplanar arrangements with A10, while G23 is looped-out and U25 points inside the major groove (Figure 5B). Sugar rings of G23, U24 and U25 are in an S-type conformation. Torsion angles of residues in the asymmetric internal loop do not deviate from values found in an A-form RNA, except ζ in C22, β in U9 and ε in G23, U24 and U25.

Preliminary relaxation measurements of ^{13}C R_2 values for C8, C6 and C2 atoms of individual residues have been performed to assess the flexibility of the asymmetric internal loop with respect to the stem regions. Residues U6, A10, A18, U24, U27, C29, C31, C33 and C34 exhibit

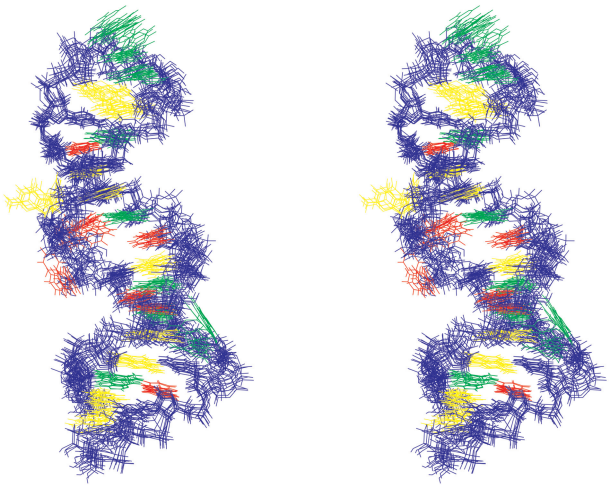

Figure 4. Stereo view of the family of 10 lowest energy structures. Sugar–phosphate backbone and cytosine bases are colored in blue, adenine bases in green, uracil bases in red and guanine bases in yellow.

Table 1. NMR restraints and structural statistics

Distance restraints	
Intranucleotide NOEs	483
Internucleotide NOEs (n, $n+1$)	268
Long range NOEs (n, $n+m$, $m>1$)	59
Torsion angle restraints	138
Hydrogen bond restraints	32
Residual dipolar couplings	52
Base pair planarity restraints	24
Total number of restraints	1056 (~31/nt)
NOE violations >0.3 Å	1.8 ± 1.0
Maximum NOE violations (Å)	0.39 ± 0.08
Deviations from idealized covalent geometry	
Bonds (Å)	0.009 ± 0.000
Angles (°)	1.85 ± 0.12
Pairwise all heavy atom RMSD from average structure (Å)	
Overall	1.21 ± 0.19
Without the asymmetric internal loop	0.99 ± 0.24
Without the GAAA tetraloop	1.20 ± 0.23
Without the adenine bulge	1.18 ± 0.19
Without the asymmetric internal loop and the adenine bulge	0.94 ± 0.26

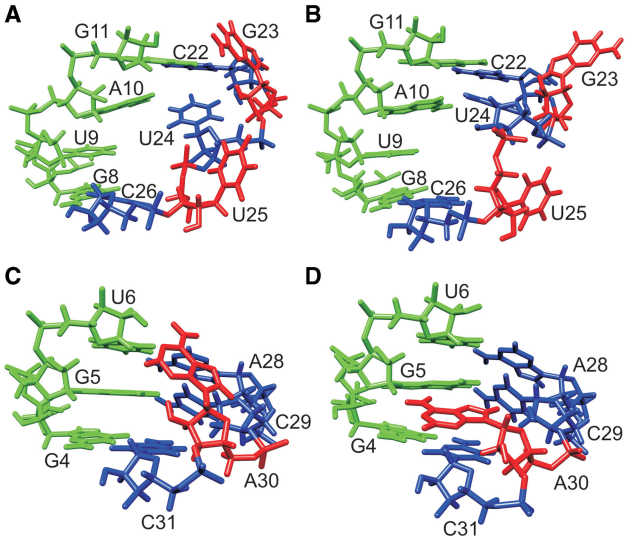

Figure 5. Structures of the asymmetric internal loop and the adenine bulge. The asymmetric internal loop has been found to adopt two families of structures shown in (**A**) and (**B**) that are in agreement with the NMR data. Structure in (A) is energetically slightly more favorable. Residues from the *let-7* sequence are colored in green, while residues from the *lin-41* sequence are colored in blue except G23 and U25 which are shown in red. The adenine bulge adopts two families of structures (**C** and **D**). Of 10 structures, 8 adopt structure shown in (C), which is energetically slightly favored. Residues are colored similar to (A) and (B) except for A30 which is shown in red color.

higher R_2 values (Supplementary Figure S4). Unfortunately, in the asymmetric internal loop we could only establish ^{13}C R_2 values of residues A10, G23 and U24. ^{13}C R_2 values for C8 of A10 and G23 do not deviate from stem residues. Cross-peak for A10 C2 is weak and as a result the error limit of otherwise high R_2 value, which suggests fast motion on ps–ns time scale, is large.

The lower stem is composed of eight Watson–Crick base pairs and the bulged residue A30. ^{1}H-^{15}N or ^{1}H-^{13}C RDC values suggest that the bulged residue A30 has a different orientation than the other residues in the lower stem. The members of the final set of structures presented in Figure 5C and D show that A30 adopts two positions. In the first family of structures the nucleobase of residue A30 points into the minor groove and is positioned perpendicular to G5–C29 and U6–A28 base pairs (Figure 5C). Flanking helices are stacked among themselves with the average twist of 32.7° and rise of 2.8 Å, which are perfectly comparable to the parameters characteristic for an A-form RNA. A30 exhibits an S-type sugar pucker with the phase angle of pseudorotation (P) equal to 142° and maximum puckering amplitude (Ψ_{max}) of 32°. In the second family of structures, A30 is turned into the minor groove, is not perfectly coplanar with G5–C29 base pair and is partially stacked on C31 (Figure 5D). The average rise in the lower stem for this family of structures is around 3.0 Å which is slightly higher than the rise in an A-form RNA, while the average twist of 32.7° is similar to A-form RNA. The sugar ring of A30 exhibits an N-type conformation with $P = 33°$ and $\Psi_{max} = 12°$. The virtual angle between the glycosidic bonds and base pair C1′-C1′ line (λ angle) is ~54° for all the residues in the lower stem for both families of structures and is similar to what is observed in an A-form RNA helix. Torsion angles ζ in C29, β and ε in A30 and β in C31 are outside the usual range found in A-form RNA and they allow positioning of the bulged residue A30 in the way described earlier. Internucleotide NOE contacts (Supplementary Figure S3B) around the bulged residue A30 are similar to the NOE contacts around the adenine bulged RNA duplex published recently (24). Gdaniec and Adamiak's NMR structure reveals formation of the dinucleotide platform in the C:G-A base triple (24). The supporting experimental evidence for exclusion of A30 of LCS1co from the helix comes from the internucleotide NOE contacts between C29 and C31, which imply that these two nucleotides are spatially close as expected for a normal sequential step.

The upper stem of LCS1co is stabilized by four Watson–Crick base pairs and shows conformational features of an A-form RNA with the average twist of 31.0° and rise parameters of 2.7 Å. The GAAA tetraloop closes the upper stem and adopts a structure that resembles the structure of the loop in a GAAA RNA hairpin (25). RDC values imply that residues in the GAAA tetraloop have different positions in comparison to the residues of the upper stem. The tetraloop consists of a sheared G15·A18 mismatch base pair and stacked adenines A16, A17 and A18. Residue A18 is positioned directly above G19 H1′, which is experimentally evidenced by its upfield chemical shift induced by ring current effects. A network of several hydrogen bonds stabilizes the GAAA tetraloop. For example, G15 imino proton with chemical shift of 10.5 p.p.m. at 278 K (Figure 2B) is involved in hydrogen bond with A18 phosphate oxygen.

DISCUSSION

Knowledge on miRNA:mRNA interactions is important for understanding the function of miRNA, its target prediction and recognition. It has been demonstrated that there are additional structural elements besides base pairing with (partial) complementarity within a miRNA:mRNA complex which play a critical role in regulation of protein expression. Two molecules of *let-7* miRNA bind to two separate *let-7* complementary sites (i.e. LCS) on *lin-41* mRNA. Both binding sites are required for the regulation of development in *C. elegans* (5). Structural features of LCS1 and LCS2 prevent other *let-7* family members such as *miR-48*, *miR-84* and *miR-241* (26) to target these binding sites. Although such type of binding sites appear less frequently and have been harder to maintain during evolution, it is important to study their structure due to the interesting structural features, which are essential for the downregulation of *lin-41*.

We have designed a 34-nt RNA construct, herein named LCS1co, which mimics the interaction between *let-7* miRNA and *lin-41* mRNA (LCS1 site) from *C. elegans*.

We have shown in our previous study that the unimolecular nature of the LCS2co construct is not detrimental for understanding the structural details within a bimolecular complex of *let-7* miRNA to the LCS2 site of *lin-41* mRNA (10). NMR experiments were performed to evaluate the effect of the adenine bulge and of the asymmetric internal loop on the overall topology and mutual orientation of the two double helical regions. We have established a 3D structure of the construct composed of upper and lower stems. The two stems are separated by the asymmetric internal loop, which has been found to adopt two energetically comparable families of structures that exhibit the lowest energies and restraint violations. In both families, U9 and A10 are stacked. U24 in the first family is orientated into the helix, G23 points into the minor groove and U25 is flipped out. On the other hand, in the second family, U24 is in a coplanar arrangement with A10, G23 is looped-out and U25 points into the major groove. Our structural calculations show that there are also other structures that satisfy NMR observables. For example, in one of the structures G23, U24 and U25 are all looped-out, and in another structure G23 is looped-out, while U24 and U25 are stacked-in. They are however energetically unfavorable and are thus not included in the final family of structures.

Residue A30 within the lower stem produces only very small deviations in the geometry of the helical parameters. A30 itself has been found to adopt two positions. In the first, the nucleobase of residue A30 points into the minor groove and is positioned perpendicular to G5–C29 and U6–A28 base pairs. A30 exhibits an S-type sugar pucker. In the second family of structures, A30 is in the stacked-in conformation and adopts an N-type sugar conformation. It is not perfectly coplanar with the G5-C29 base pair and is partially stacked on C31.

It is interesting to compare the structures of LCS1co and LCS2co constructs. Both RNA molecules adopt stem-loop structures that feature a lower and an upper stem separated by an asymmetric internal loop (Figure 6). Stem regions of both structures show conformational features of an A-form RNA. The overall topology and structural details of the GAAA tetraloop are similar. The main differences between the two structures lie in the details of the asymmetric internal loops. The asymmetric internal loop of LCS1co studied here (Figure 6A) is not defined to a level that has been found for the corresponding loop of LCS2co which adopts a well-defined structure (Figure 6B). In LCS2co, U9, U24 and U25 form a base triple, while adenines form an A10·A23 base pair (Figure 6C). Structure of the internal loop of LCS2co is corroborated by the numerous internucleotide NOE contacts in the asymmetric internal loop with an average of 22.4 NOEs per residue in the internal loop (10). In the case of LCS1co a lot of internucleotide NOE contacts, which could define the structure of the asymmetric internal loop, were not observed. The average number of NOEs per residue in the internal loop of LCS1co is only 16.2. Higher flexibility of the asymmetric internal loop in LCS1co is verified through weak (if any) hydrogen bonding which is experimentally evidenced by the absence

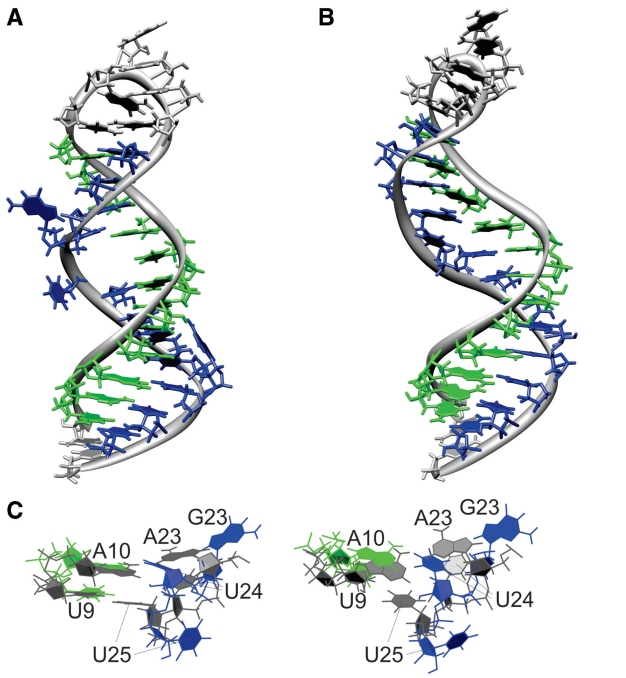

Figure 6. Comparison of structures of LCS1co (**A**) and LCS2co (PDB code 2JXV) (**B**). Residues originating from sequences of *let-7* miRNA and *lin-41* mRNA are colored in green and blue, respectively. (**C**) Overlay of the asymmetric internal loops of LCS1co and LCS2co. Residues from LCS1co are colored in green and blue, while residues from LCS2co are colored in gray.

of exchangeable imino protons of U9, G23, U24 and U25 in NMR spectra. Non-existence of intra- and cross-strand NOEs is corroborated by broad and weak aromatic and sugar proton signals of U9, U24 and U25. It could be hypothesized that freedom of the asymmetric internal loop to sample a larger conformational space by adopting two energetically favorable structures is important for interaction with Argonaute proteins.

Additionally, we have compared thermodynamic stabilities of the two constructs. UV melting experiments suggest a single melting transition for both RNA oligonucleotides (Supplementary Figure S5). The melting temperature of LCS2co (341 K) is 9 K higher than that of LCS1co, which implies that stacking of the nucleobases and hydrogen bonds in the asymmetric internal loop of the former contribute to a well-defined and more stable structure.

Structural analysis of the helical parameters of LCS1co showed that helical axes of the lower and upper stems exhibit an angular bend of ~10°, which is not as pronounced as the angular bend between the upper and lower stems in LCS2co (~20°). Smaller angular bend induced by the asymmetric internal loop between the stems in LCS1co was expected based on the small differences in RDC values between the upper and lower stems. It seems that the structure of the asymmetric internal loop in LCS1co, with residues G23 and U25 looped-out, does not induce any significant bend and does not widen the major groove. In addition we did not notice an angular bend within the lower stem at the bulged

A30 residue. In contrast, an earlier study has shown that a conserved non-paired U bulge in the apical stem–loop of the human hepatitis B virus encapsidation signal bends the upper and lower stems to an average value of 21 ± 9° (27). The main reason that the adenine bulge in LCS1co does not create a bend in the lower stem lies in the fact that it does not stack onto residues within the helix and thus change the overall structure of the backbone. In fact, the calculated bending angle in the lower stem is similar to the calculated bending angle in an A-form RNA (28). The high-resolution structure presented here demonstrates that the bulged adenine, which is essential for downregulation of *let-41* (8), could represent a protein-binding site. On the other hand, it could also confirm the results of Wang *et al.* (6) that the target strand makes no contacts with the Argonaute scaffold and can accommodate a single bulge when sufficient pairing between miRNA and mRNA is retained.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank Gregor Ilc for the relaxation measurements of ^{13}C R_2 values.

FUNDING

Slovenian Research Agency, the Ministry of Higher Education, Science and Technology of the Republic of Slovenia (grant numbers Z1-2138 and P1-0242); EN-FIST Center of Excellence. Funding for open access charge: Slovenian Research Agency (ARRS).

Conflict of interest statement. None declared.

REFERENCES

1. Reinhart,B.J., Slack,F.J., Basson,M., Pasquinelli,A.E., Bettinger,J.C., Rougvie,A.E., Horvitz,H.R. and Ruvkun,G. (2000) The 21-nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature*, **403**, 901–906.
2. Büssing,I., Slack,F.J. and Großhans,H. (2008) *let-7* microRNAs in development, stem cells and cancer. *Trends Mol. Med.*, **14**, 400–409.
3. Bernstein,E., Caudy,A.A., Hammond,S.M. and Hannon,G.J. (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature*, **409**, 363–366.
4. Jannot,G., Boisvert,M.E.L., Banville,I.H. and Simard,M.J. (2008) Two molecular features contribute to the Argonaute specificity for the microRNA and RNAi pathways in *C. elegans*. *RNA*, **14**, 829–835.
5. Bartel,D.P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, **136**, 215–233.
6. Wang,Y.L., Juranek,S., Li,H., Sheng,G., Tuschl,T. and Patel,D.J. (2008) Structure of the argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. *Nature*, **456**, 921–926.
7. Vella,M.C., Choi,E.Y., Lin,S.Y., Reinert,K. and Slack,F.J. (2004) The *C. elegans* microRNA *let-7* binds to imperfect *let-7* complementary sites from the *lin-41* 3′ UTR. *Genes Dev.*, **18**, 132–137.
8. Vella,M.C., Reinert,K. and Slack,F.J. (2004) Architecture of a validated microRNA::target interaction. *Chem. Biol.*, **11**, 1619–1623.
9. Didiano,D. and Hobert,O. (2008) Molecular architecture of a miRNA-regulated 3′ UTR. *RNA*, **14**, 1297–1317.
10. Cevec,M., Thibaudeau,C. and Plavec,J. (2008) Solution structure of a *let-7* miRNA:*lin-41* mRNA complex from *C. elegans*. *Nucleic Acids Res.*, **36**, 2330–2337.
11. Milligan,J.F. and Uhlenbeck,O.C. (1989) Synthesis of small RNAs using T7 RNA polymerase. *Method Enzymol.*, **180**, 51–62.
12. Kao,C., Zheng,M. and Rüdisser,S. (1999) A simple and efficient method to reduce nontemplated nucleotide addition at the 3′ terminus of RNAs transcribed by T7 RNA polymerase. *RNA*, **5**, 1268–1272.
13. Fürtig,B., Richter,C., Wöhnert,J. and Schwalbe,H. (2003) NMR spectroscopy of RNA. *ChemBioChem*, **4**, 936–962.
14. Dingley,A.J. and Grzesiek,S. (1998) Direct observation of hydrogen bonds in nucleic acid base pairs by internucleotide $^{2}J_{NN}$ couplings. *J. Am. Chem. Soc.*, **120**, 8293–8297.
15. Wijmenga,S.S. and van Buuren,B.N.M. (1998) The use of NMR methods for conformational studies of nucleic acids. *Prog. Nucl. Magn. Reson. Spectrosc.*, **32**, 287–387.
16. Nikonowicz,E.P. and Pardi,A. (1993) An efficient procedure for assignment of the proton, carbon and nitrogen resonances in $^{13}C/^{15}N$ labeled nucleic acids. *J. Mol. Biol.*, **232**, 1141–1156.
17. Hansen,M.R., Mueller,L. and Pardi,A. (1998) Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nat. Struct. Biol.*, **5**, 1065–1074.
18. Ottiger,M., Delaglio,F. and Bax,A. (1998) Measurement of *J* and dipolar couplings from simplified two-dimensional NMR spectra. *J. Magn. Reson.*, **131**, 373–378.
19. Tian,F., Al-Hashimi,H.M., Craighead,J.L. and Prestegard,J.H. (2001) Conformational analysis of a flexible oligosaccharide using residual dipolar couplings. *J. Am. Chem. Soc.*, **123**, 485–492.
20. Case,D.A., Darden,T., Cheatham,T.E. III, Simmerling,C., Wang,J., Duke,R.E., Luo,R., Merz,K.M., Pearlman,D.A., Crowley,M. *et al.* (2006) AMBER 9. University of California, San Francisco, CA.
21. Wang,J.M., Cieplak,P. and Kollman,P.A. (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.*, **21**, 1049–1074.
22. Lu,X.J. and Olson,W.K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.*, **31**, 5108–5121.
23. Rüdisser,S. and Tinoco,I. (2000) Solution structure of cobalt(III)hexammine complexed to the GAAA tetraloop, and metal-ion binding to GA mismatches. *J. Mol. Biol.*, **295**, 1211–1223.
24. Popenda,L., Bielecki,L., Gdaniec,Z. and Adamiak,R.W. (2009) Structure and dynamics of adenosine bulged RNA duplex reveals formation of the dinucleotide platform in the C:G-A triple. *Arkivoc*, **III**, 130–144.
25. Jucker,F.M., Heus,H.A., Yip,P.F., Moors,E.H.M. and Pardi,A. (1996) A network of heterogeneous hydrogen bonds in GNRA tetraloops. *J. Mol. Biol.*, **264**, 968–980.
26. Lim,L.P., Lau,N.C., Weinstein,E.G., Abdelhakim,A., Yekta,S., Rhoades,M.W., Burge,C.B. and Bartel,D.P. (2003) The microRNAs of *Caenorhabditis elegans*. *Genes Dev.*, **17**, 991–1008.
27. Flodell,S., Petersen,M., Girard,F., Zdunek,J., Kidd-Ljunggren,K., Schleucher,J. and Wijmenga,S. (2006) Solution structure of the apical stem-loop of the human hepatitis B virus encapsidation signal. *Nucleic Acids Res.*, **34**, 4449–4457.
28. Popenda,L., Adamiak,R.W. and Gdaniec,Z. (2008) Bulged adenosine influence on the RNA duplex conformation in solution. *Biochemistry*, **47**, 5059–5067.