



# NMR Characterization of an Oligonucleotide Model of the MiR-21 Pre-Element

Sara Chirayil<sup>1</sup>, Qiong Wu<sup>2</sup>, Carlos Amezcua<sup>2,3</sup>, Kevin J. Luebke<sup>1\*</sup>✉

**1** Division of Cardiology, Department of Internal Medicine, University of Texas Southwestern Medical Center, Dallas, Texas, United States of America, **2** Department of Biophysics, University of Texas Southwestern Medical Center, Dallas, Texas, United States of America, **3** Structural Elucidation Group, Medical Products Division, Baxter Healthcare, Round Lake, Illinois, United States of America

## Abstract

We have used NMR spectroscopy to characterize an oligonucleotide stem loop structure based on the pre-element of an oncogenic microRNA, miR-21. This predicted stem-loop structure is cleaved from the precursor of miR-21 (pre-miR-21) by the nuclease Dicer. It is also a critical feature recognized by the protein complex that converts the primary transcript (pri-miR-21) into the pre-miRNA. The secondary structure of the native sequence is poorly defined by NMR due to rapid exchange of imino protons with solvent; however, replacement of two adjacent putative G•U base pairs with G•C base pairs retains the conformation of the hairpin observed by chemical probing and stabilizes it sufficiently to observe most of the imino proton resonances of the molecule. The observed resonances are consistent with the predicted secondary structure. In addition, a peak due to a loop uridine suggests an interaction between it and a bulged uridine in the stem. Assignment of non-exchangeable proton resonances and characterization of NOEs and coupling constants allows inference of the following features of the structure: extrahelicity of a bulged adenosine, deviation from A-form geometry in a base-paired stem, and consecutive stacking of the adenosines in the 5' side of the loop, the guanosine of the closing base pair, and a cross-strand adenosine. Modeling of the structure by restrained molecular dynamics suggests a basis for the interaction between the loop uridine, the bulged uridine in the stem, and an A•U base pair in the stem.

**Citation:** Chirayil S, Wu Q, Amezcua C, Luebke KJ (2014) NMR Characterization of an Oligonucleotide Model of the MiR-21 Pre-Element. PLoS ONE 9(9): e108231. doi:10.1371/journal.pone.0108231

**Editor:** Patrick van der Wel, University of Pittsburgh School of Medicine, United States of America

**Received:** April 3, 2014; **Accepted:** August 25, 2014; **Published:** September 24, 2014

**Copyright:** © 2014 Chirayil et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability:** The authors confirm that all data underlying the findings are fully available without restriction. Coordinates for the 14 lowest energy calculated structures are deposited in the RCSB Protein Data Bank, (<http://www.rcsb.org/pdb/home/home.do>) RCSB ID code rcsb103821 and PDB ID code 2 mnc. All other relevant data are within the paper.

**Funding:** This work was supported by funding from the National Institute of General Medical Science, R01GM088842, to KJL. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** Dr. Amezcua is employed by a commercial enterprise, Baxter Healthcare and ask that the authors declare the affiliation to this company, along with any other relevant declarations relating to employment, consultancy, patents, products in development or marketed products. All of Dr. Amezcua's contributions to this paper were made prior to his employment with Baxter Healthcare, while he was employed by the University of Texas Southwestern Medical Center. The contents of this paper are unrelated to his work at Baxter Healthcare. Dr. Amezcua is currently an employee of Baxter Healthcare. This does not alter the authors' adherence to PLOS ONE policies on sharing data and materials.

\* Email: kevin.luebke@gmail.com

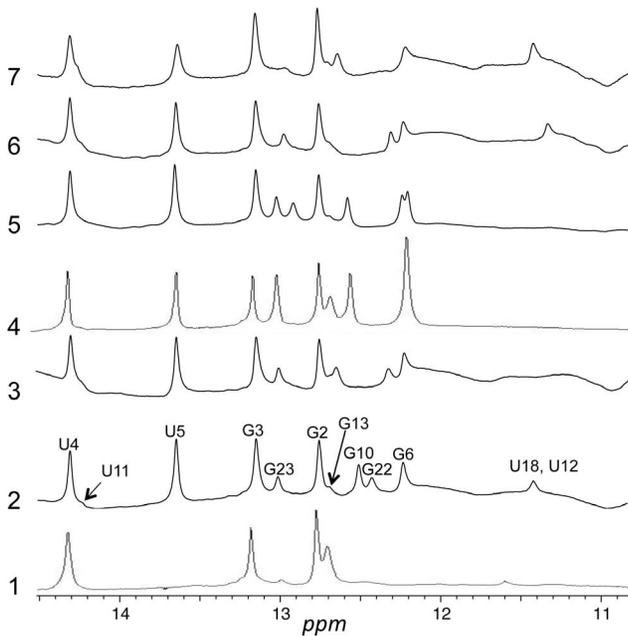
✉ Current address: Biosciences Division, SRI International, Harrisonburg, Virginia, United States of America

## Introduction

MicroRNAs (miRNAs) are short, non-coding RNAs that regulate gene expression by diminishing translation of their target messenger RNAs [1,2]. Whereas their normal function is regulation of development and cellular responses to stress [3], the aberrant expression of specific miRNAs is associated with a wide range of diseases, including cancer [4] and heart disease [5,6]. For example, miR-21 is a miRNA that is elevated in both cancer and heart disease. It is highly expressed in a variety of tumors [7], contributing to the cancer phenotype by diminishing translation of tumor suppressor genes [8–12]. It is also expressed in hypertrophic heart tissue, where it contributes to the fibrotic response to cardiac stress or injury [13]. An understanding of the factors that regulate miRNA expression is essential to efforts to therapeutically target specific disease-related miRNAs [14,15] and to gaining a basic understanding of the roles of miRNAs in biology.

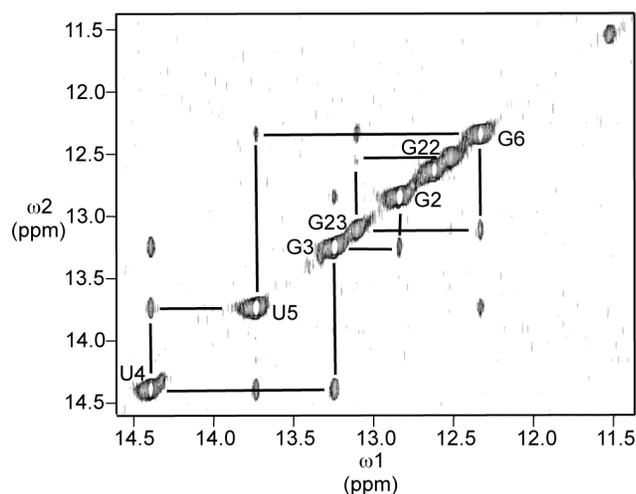
Expression of miRNAs is regulated post-transcriptionally by modulation of their maturation [16,17]. MiRNAs are initially transcribed within much longer RNAs (pri-miRNAs), which are subsequently processed in a series of steps to produce the mature miRNA [1] (Figure 1). The first step of this process is cleavage of a long hairpin structure, which contains the mature miRNA sequence, from the primary transcript. A multiprotein complex called the Microprocessor effects this processing step, distinguishing hairpins that contain miRNAs from the multitude of other hairpin structures in the transcriptome [18–22]. The excised hairpin is exported to the cytoplasm [23,24], where the nuclease Dicer cleaves the mature miRNA from the precursor hairpin (pre-miRNA) [25] by removing a structure called the pre-element, a short stem-loop comprising the terminal loop and a short region of predicted base pairing. This structure is also known as the terminal loop region or apical region. The remaining duplex associates with an Argonaute protein, leading to the retention of the single-stranded miRNA in the active miRNA-Argonaute complex [26].





**Figure 3. NMR spectra of imino protons for RNAs 1–7 in H<sub>2</sub>O.** Peaks for **2** are labeled according to assignments described in the text. Conditions were as described in Materials and Methods. doi:10.1371/journal.pone.0108231.g003

by the appearance of peaks for at least 11 of the 14 imino protons in the molecule. The correspondence in chemical shifts between the imino protons that are visible in both **1** and **2** suggests that **1** has a similar conformation to **2**. Therefore, **1** is likely dynamic rather than disordered. A series of sequential NOEs in the NOESY spectrum of **2** in H<sub>2</sub>O (Figure 4) from G2 through G22 indicates formation of the predicted base pairs by each of these nucleotides. An NOE between the imino protons of G6 and G23 indicates continuous stacking of the base pairs formed by these nucleotides.



**Figure 4. Imino proton region of NOESY spectrum of **2** in H<sub>2</sub>O.** Conditions were as described in Materials and Methods. Sequential NOEs are indicated by lines to peaks due to proximal protons. Peaks are labeled according to assignments based on sequential NOEs. doi:10.1371/journal.pone.0108231.g004

In addition to the sequentially proximal imino protons, four other imino protons are evident in the spectrum. A shoulder at 14.2 ppm is consistent with formation of the A•U base pair between A20 and U11. A very broad peak between 11 and 12 ppm is consistent with an unpaired uridine. Peaks at 12.7 and 11.4 ppm could indicate the presence of an additional G•C base pair (12.7 ppm) such as the predicted pair between G13 and C19 and a uridine (11.4 ppm) engaged in a non-Watson-Crick interaction. Alternatively, it could signal the presence of an unexpected G•U base pair, with the guanosine imino resonance at 11.4 ppm and the uridine imino resonance at 12.7 ppm.

The absence of an NOE between the peaks at 11.4 ppm and 12.7 ppm argues against the involvement of these two protons in a G•U base pair. Furthermore, though replacement of U18, the most likely participant in a G•U pair with G13, by a cytosine (RNA **3**) results in disappearance of the peak at 11.4 ppm, it does not result in the appearance of a new peak corresponding to a new G•C base pair. Similarly, replacement of U12 with a cytosine (RNA **4**) also eliminates the peak at 11.4 ppm as well as the broad peak from 11 ppm to 12 ppm. This replacement does not result in appearance of a new peak due to a G•C base pair either but does result in intensification of the peak at 12.7 ppm.

To confirm the assignment of the shoulder at 14.2 ppm to the base pair between A20 and U11, we replaced these nucleotides with a guanosine and cytosine, respectively (RNA **5**). As anticipated, the shoulder at 14.2 ppm is absent in the spectrum of this RNA and a new peak at 12.9 ppm, corresponding to a G•C base pair, appears. Somewhat unexpectedly, both the broad peak at 11 ppm–12 ppm and the small peak at 11.4 ppm are absent in the spectrum of **5**.

To confirm the assignments of G22 and G10 and assess the overall effect of the replacement of a G•C base pair for each individual G•U pair in the upper stem, spectra were acquired for RNAs **6** and **7**, in which C21 and C9 of **2** were replaced with uridines, respectively. As expected, peaks at 12.5 ppm and 12.4 ppm, respectively, were absent. Furthermore, these alterations had little effect on the spectra beyond those peaks.

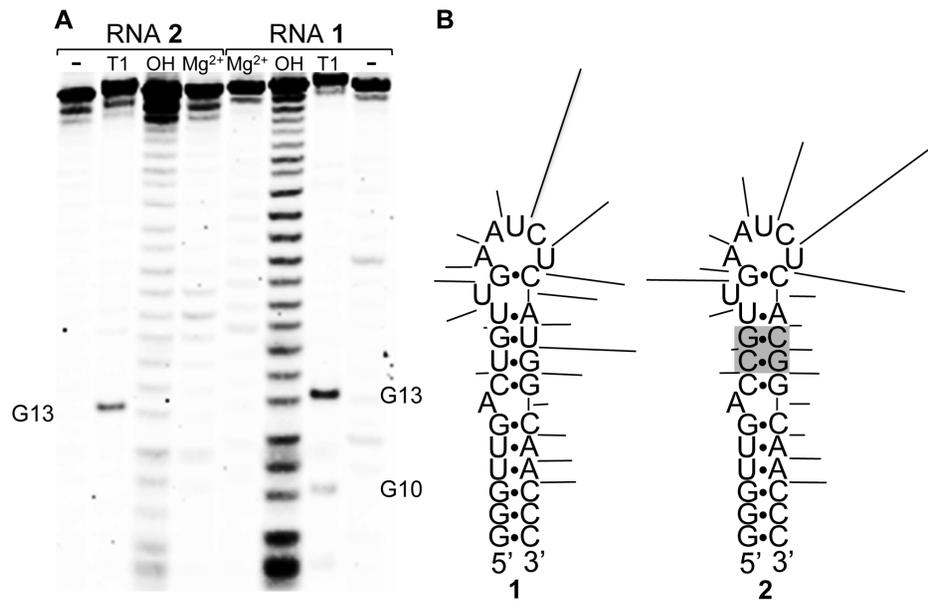
### In-Line Probing of RNAs 1 and 2

Mg<sup>2+</sup>-induced hydrolytic cleavage was used to characterize and compare the conformations of RNAs **1** and **2**. Hydrolytic cleavage of the RNA backbone occurs principally through nucleophilic attack of a 2'-hydroxyl on the adjacent phosphodiester, displacing the 5'-hydroxyl of the following nucleotide. For this displacement to occur, the attacking 2'-hydroxyl must be in-line with the scissile phosphorus–oxygen bond [45]. Thus, in-line cleavage, which is stimulated by divalent metal ions such as Mg<sup>2+</sup>, is a useful probe of conformation and flexibility.

Electrophoretic analysis of Mg<sup>2+</sup>-stimulated cleavage of RNAs **1** and **2** is shown in Figure 5. The cleavage patterns are similar, with most intense cleavage occurring after the 3 nucleotides at the 3' side of the predicted loop. Both also show cleavage along the 3' strand of the stem. At a quantitative level, some differences in the cleavage patterns are apparent. Specifically, cleavage after U12 is stronger in **2**; whereas, cleavage after nucleotide 21 (C21 in **2**, U21 in **1**) is stronger in **1**. Also, cleavage in the loop is strongest after U16 in **1** but strongest after C17 in **2**.

### Assignment of Non-exchangeable protons of RNA 2

Assignment of non-exchangeable protons followed standard procedures based on sequential NOE connectivities and through-bond correlations [46]. The chemical shifts of assigned non-exchangeable and exchangeable protons of RNA **2** are listed in Table 1. Pyrimidine H5 and H6 resonances were identified by



**Figure 5. Mg<sup>2+</sup>-stimulated (in-line) cleavage of RNAs 1 and 2.** A. Electrophoretic analysis of cleavage. Mg<sup>2+</sup>: RNA treated with 5 mM MgCl<sub>2</sub>. OH: hydrolysis ladder. T1: Ribonuclease T1 digestion. Untreated RNA in lanes labeled “-”. Bands due to guanosines identified from RNase T1 digestion are indicated. B. Mapping of cleavage onto predicted secondary structures of **1** and **2**. Cleavage after each nucleotide is indicated by a line with length proportional to band intensity.  
doi:10.1371/journal.pone.0108231.g005

their strong crosspeaks in the double-quantum filtered COSY spectrum of the molecule. Cytosines were further distinguished from uridines by the chemical shifts of their C5 carbons, determined in a natural abundance <sup>1</sup>H-<sup>13</sup>C HSQC spectrum. NOE connectivities were observed in a 400 ms mixing-time NOESY experiment with identifiable purine-pyrimidine patterns, leading to the sequential assignment of aromatic and H1' protons. Sequential H1' to aromatic connectivities were nearly continuous through the molecule, broken only between C21 and G22 and broken or obscured by overlap between C8 and C9 and between U12 and G13 (Figure 6). Assignments were confirmed by sequential aromatic to aromatic and H1' to H1' crosspeaks.

The 2' protons were assigned by their crosspeaks with 1' protons in a short mixing time (60 ms) NOESY spectrum and, for nucleotides with significant C2'-endo character, in the DQF-COSY spectrum. The 2' assignments were confirmed by sequential H2'-H6/H8 NOE connectivities in the short mixing-time NOESY spectrum. H4' assignments were made from the H1'-H2'/H3'/H4'/H5'/H5'' region of a NOESY spectrum with a mixing time of 150 ms, and H3' assignments were made from the H1'-H2'/H3'/H4'/H5'/H5'' region of a NOESY spectrum with a mixing time of 400 ms.

Adenosine H2 protons were identified by the chemical shifts of bound carbons, determined in the natural abundance <sup>1</sup>H-<sup>13</sup>C HSQC spectrum, and for adenosines 7, 20, 25, and 26, NOEs to cross-strand H1' protons. The H2 protons of adenosines A14 and A15 had nearly identical chemical shifts, 7.66 ppm, but a small offset allowed specific assignment, based on a strong NOE between H2 of A14 and the H1' of A15 (Figure 6B). This NOE is comparable in intensity to the intranucleotide H5-H6 NOEs and much stronger than the intranucleotide H2-H1' NOEs for adenosines in the *anti*- conformation about the glycosidic bond, ruling out its assignment as the A15 H2-H1' intranucleotide NOE. Further supporting this assignment, the assigned A14 H2 to A15 H1' NOE overlaps with an NOE between the same H2 proton and C19 H1'. This overlap was resolved in a NOESY spectrum

taken at 15°C (data not shown) and is most consistent with an NOE between A14H2 and C19H1'.

### Conformational Features of RNA 2

The NMR data for the lower stem of **2** (nucleotides 1–6 and 24–29) are consistent with a Watson-Crick base paired A-form double helix as anticipated. The observation of imino proton peaks with a continuous sequence of NOEs between them for these residues in addition to standard internucleotide NOEs between the non-exchangeable protons indicate a typical RNA duplex. The NOE between the imino protons of G6 and G23, without interruption by A7, indicates that the bulged adenosine is not stacked into the helix. The internucleotide NOEs between G6 and A7 and between A7 and C8 are very weak, supporting this view. However, a very weak cross-strand NOE between A7 H2 and C24 H1' suggests that the purine heterocycle of A7 is partially associated with a groove of the duplex.

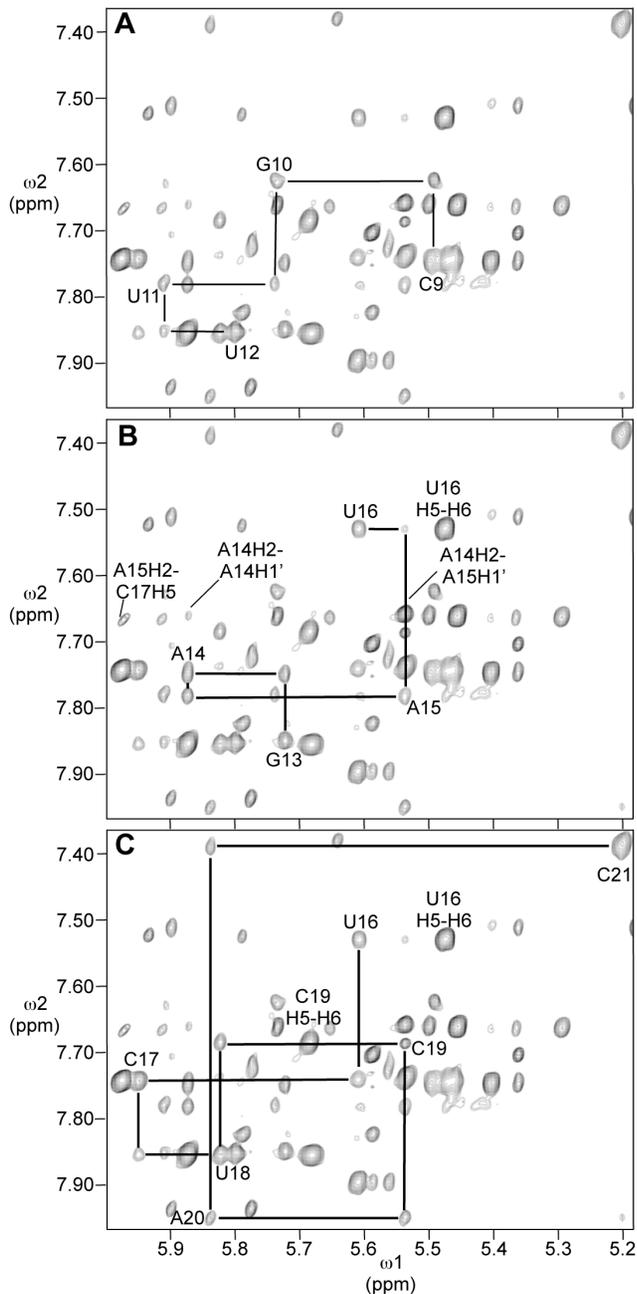
Crosspeaks between H1' and H2' in the DQF-COSY spectrum indicate a significant C2'-endo character for 13 of the sugars in the molecule. The 1'-2' scalar couplings for these sugars and an estimate of the equilibrium percentage of each in the C3'-endo conformation [47] are listed in Table 2. These sugars indicate points of backbone flexibility or regions where the backbone spans a greater distance than in A-form double-helical structure. In addition to the terminal nucleotides (G1 and C29) and the bulged adenosine (A7) and preceding nucleotide (G6), many of the loop and adjacent nucleotides display C2'-endo character. Of the five nucleotides formally included in the loop, only A14 does not show significant C2'-endo character, consistent with the requirement that it span a bulged uridine. Though G10, G13, and U11 or U12 are also apparently base paired, as indicated by the observation of a peak due to an imino proton for each, their adoption of partial C2'-endo character indicates that they are distorted from a canonical duplex.

**Table 1.** Chemical Shifts (ppm) of Assigned Protons<sup>a</sup>.

	H8/H6	H5/H2	H1'	H2'	H3'	H4'	imino <sup>b</sup>	amino <sup>b</sup>
G1	8.06	NA <sup>c</sup>	5.79	4.92	4.30	4.38		
G2	7.52	NA	5.93	4.74	4.61	4.51	12.84	8.45/6.85
G3	7.17	NA	5.77	4.54	4.39	4.49	13.25	8.36/6.97
U4	7.72	5.08	5.56	4.44	4.61	4.46	14.39	NA
U5	7.90	5.61	5.59	4.34	4.51	4.38	13.74	NA
G6	7.82	NA	5.79	4.30	4.04	4.38	12.34	8.43/6.56
A7	8.40	8.22	6.17	4.87	4.21	4.56	NA	
C8	7.73	5.56	5.45	4.24	3.98	4.16	NA	
C9	7.74	5.46	5.49	4.48	4.56	4.40	NA	
G10	7.62	NA	5.74	4.45	4.60	4.50	12.64	8.25/6.71
U11	7.78	5.41	5.91	4.41	4.71	4.46	14.34	NA
U12	7.85	5.68	5.80	4.41	4.64	4.48		NA
G13	7.85	NA	5.72	4.73	4.48	4.48	12.76	
A14	7.75	7.66	5.87	4.58	4.53	4.50	NA	
A15	7.78	7.66	5.54	4.46	4.41	4.50	NA	
U16	7.53	5.47	5.61	4.12	4.29	4.61		NA
C17	7.74	5.98	5.95	4.36	4.66	4.53	NA	
U18	7.85	5.88	5.82	4.23	4.50	4.58	11.53	NA
C19	7.69	5.68	5.54	4.62	4.30		NA	
A20	7.95	7.22	5.84	4.55	4.45	4.48	NA	
C21	7.39	5.20	5.20	4.39	4.31	4.47	NA	
G22	7.38	NA	5.64	4.42	4.62		12.52	8.13/6.67
G23	7.03	NA	5.65	4.60	4.34	4.65	13.12	8.03/6.48
C24	7.66	5.29	5.37	4.46	4.38	4.55	NA	
A25	8.05	6.78	5.78	4.46			NA	7.87/6.76
A26	7.94	7.71	5.90	4.40	4.35	4.42	NA	7.70/
C27	7.51	5.17	5.37	4.35	4.00	4.16	NA	
C28	7.75	5.41	5.51	4.24	4.00	4.34	NA	
C29	7.66	5.46	5.73	3.98	4.19	4.16	NA	8.30/6.99

<sup>a</sup>Assignments of nonexchangeable protons are at 25°C. All chemical shifts are reported relative to TSP.<sup>b</sup>Assignments of imino and amino protons are at 10°C.<sup>c</sup>NA: not applicable.

doi:10.1371/journal.pone.0108231.t001



**Figure 6. Portion of the NOESY spectrum showing NOEs between H8/H6/H2 (7.4–8.0 ppm) and H1'/H5 (5.2–6.0 ppm) protons.** Mixing time was 400 ms and temperature was 25°C. Cross-peaks due to NOEs from a nucleotide aromatic proton to the H1' proton of its own sugar are labeled. A. Sequential NOEs from C9 to U12. B. Sequential NOEs from G13 to U16. Crosspeaks due to H2 of A14 and A15 and the U16 H5 to U16 H6 are labeled. C. Sequential NOEs from U16 to C21.  
doi:10.1371/journal.pone.0108231.g006

Structure predictions indicate that A20 base pairs with U11, leaving U12 unpaired, but the NMR data do not clearly distinguish between that possibility and the alternative of A20 base pairing with U12. Overlap between U12 H6 and G13 H8 resonances obscures the presence or absence of an NOE between those protons. Typical sequential internucleotide NOEs, (H8/H6-H8/H6 and H1'-H1') are weak or not seen from G10-G13,

consistent with either possibility. The in-line cleavage data, however, suggest the greatest propensity for an unpaired conformation around U12. The most effective hydrolytic cleavage flanks that nucleotide, especially following it between U12 and G13. This pattern is observed for RNA 1 and RNA 2.

Continuous sequential NOEs from G13 through A15 suggest continuous stacking of those bases. The relatively weak  $Mg^{2+}$ -induced cleavage between those nucleotides further supports that view. An NOE between A15 H1' and U16 H6 is observed, but relatively weak (Figure 6B), and no NOE is observed between U16 H6 and C17 H6. These facts taken together suggest that the backbone trajectory turns at U16. The NOE between A15H2 and C17H5 (Figure 6B) suggests that C17 is oriented toward the inside of the loop, facing across it toward the stacked adenines.

The dependence of the appearance of the imino proton resonance at 11.4 ppm on U18 and U12, as well as the A20•U11 base pair indicates an interaction between these nucleotides. The interaction could be direct, such as through hydrogen bonding between the interacting nucleotides, or indirect, such as a structural perturbation that affects spatially remote nucleotides.

### Structure modeling by restrained molecular dynamics

Nucleotides 8–23 were modeled using 80 torsion angle restraints, 39 distance restraints to constrain the experimentally determined base pairs (C8•G23, C9•G22, G10•C21, U11•A20, and G13•C19) to appropriate hydrogen bonding distances and base pair planarity, 29 NOE-derived intranucleotide distance restraints, 87 NOE-derived internucleotide restraints, and 100 distance restraints to model the two terminal base pairs as A-form duplex. Eighty structures were calculated and the fourteen lowest energy output structures were analyzed.

An alignment of the fourteen structures is shown in Figure 7A (BMRB accession code: 19887; PDB ID: 2MNC); however, the orientation of the base paired stem (lower three base pairs) with respect to the loop is poorly defined, and separate alignment of loop nucleotides 11–20 (Figure 7B) provides a clearer view of the common structural features. This alignment illustrates that U11 and U12 are the least well-defined residues in the model. An average structure was calculated, and the output structure with the smallest RMSD (1.31) from the average is shown in stereoview in Figure 8. This structure is taken to be the best representative of the ensemble. The average RMSD of the ensemble from the average structure is 2.53.

The structure is shown schematically in Figure 9. In addition to stacking of sequential purines G13–A15, a sharp twist between the A20•U11 and C19•G13 base pairs places A20 under G13 in a cross-strand stacking interaction (Figure 10A). The loop turns at U16, and the pyrimidine ring is oriented out of the loop at this position. On the other hand, the pyrimidine ring of C17 is oriented toward the interior of the loop. U18 and U12 both protrude into the major groove, proximal to the edge of the A20•U11 base pair. Their distance and orientation with respect to each other are not well defined (Figure 10B); however, their Watson-Crick faces are generally oriented toward each other, and in several of the output structures they approach hydrogen bonding distance of each other. In a separate set of calculations, simulated annealing was carried out with these nucleotides restrained to within 2 Å (U18 H3–U11 O4) of each other. The lowest energy output structures had NOE and total energies equal to the lowest energies obtained without that constraint and shared the major conformational features of those structures.

**Table 2.** Summary of Ribose Sugar Conformations.

nucleotide <sup>a</sup>	$J_{1'-2'}$ (Hz)	%C3'-endo ( $\pm 10\%$ ) <sup>b</sup>
G1	4.8	45
G6	6.4	21
A7	6.5	21
G10	5.1	41
U11	5.9	29
U12	6.2	26
G13	4.7	47
A15	5.6	33
U16	6.7	18
C17	7.2	11
U18	2.8	75
A20	6.4	21
C29	4.6	49

<sup>a</sup>Nucleotides not included in the table are >90% in the C3'-endo conformation.

<sup>b</sup>Values were calculated from  $J_{1'-2'}$  using the empirical equation of van den Hoogen: %C3'-endo =  $114.9 - 14.5(J_{1'-2'})$  [47].

doi:10.1371/journal.pone.0108231.t002

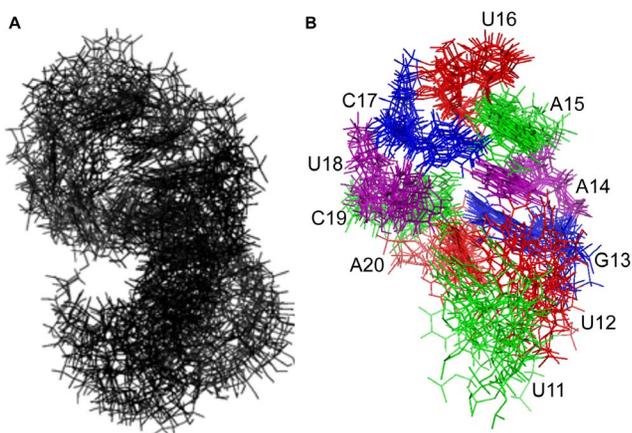
## Discussion

The pre-element of miR-21 is an important subject of structural study because of its significance as a site of molecular recognition by the miRNA processing apparatus and because of its interesting predicted secondary structure. In addition to being important for recognition of pri-miR-21 by the microprocessor and pre-miR-21 by Dicer, it is a site of recognition by other endogenous factors that regulate interaction of the microprocessor with the primary transcript of miR-21 [31]. The strong sequence conservation of the apical loop, beyond what is required for recognition by factors that are known to bind this structure, suggests that there are still-undiscovered agents that form critical associations with the loop [27].

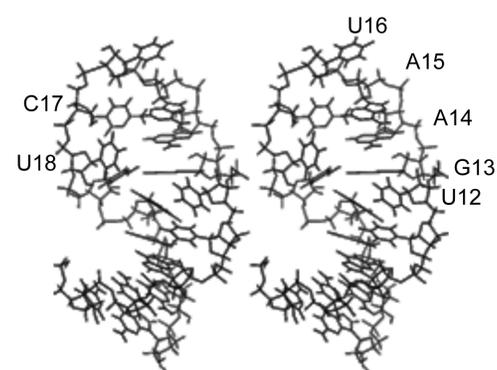
The predicted secondary structure of the loop, formally a five-nucleotide loop with a bulged nucleotide adjacent to the closing base pair, is unlike hairpin loops that have been studied previously

by NMR [48–50]. Much of the structural analysis of hairpin loops has been directed to four-nucleotide loops such as the stable UNCG tetraloops [51] and the GNRA loop motif [52]. Several six-nucleotide hairpin loops have also been analyzed by solution methods [53–55]. However, there is relatively little structural information for five-nucleotide RNA loops, particularly in the context of a proximal bulge nucleotide as in the loop studied here.

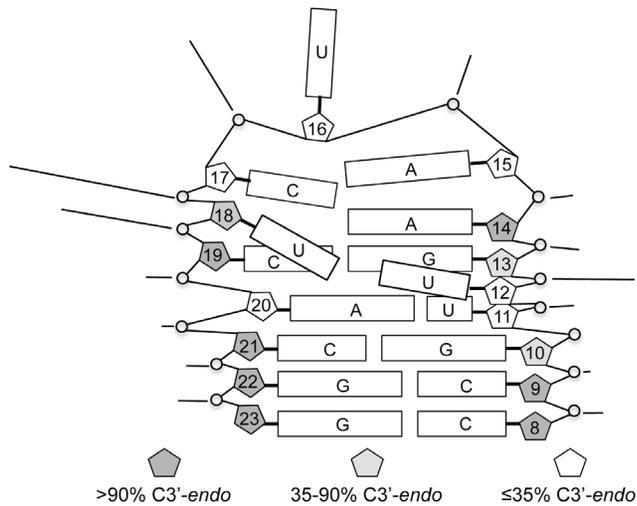
The significance of the bulged uridine, U12, adjacent to the loop is underscored by its interesting and unexpected interaction with U18. Our NMR-derived model, in conjunction with the imino proton spectra of **2** and related hairpins, suggest a direct hydrogen bonding interaction between these nucleotides across the major groove. The orientation of U18 is similar to a feature seen in the solution structure of a five-nucleotide hairpin loop modeled after a loop in the 18S ribosomal RNA [48]. In that case, a cytosine at the 3' end of the loop sequence is oriented toward the major groove of the adjacent double helix. That RNA, however, does not have an unpaired nucleotide available for interaction with the fifth loop nucleotide in the major groove. In contrast, the bulged uridine of **2**, also protruding into the major groove, provides an interaction partner for this nucleotide.



**Figure 7. Alignment of the 14 lowest energy output structures for **2** (PDB ID 2MNC).** A. Alignment of 16 nucleotides (C8–G23) included in restrained molecular dynamics calculation. B. Alignment of loop nucleotides U11 to A20. Only the aligned nucleotides are shown. doi:10.1371/journal.pone.0108231.g007



**Figure 8. Stereoview of a representative structure for nucleotides 8–23 of RNA **2** (PDB ID 2MNC).** doi:10.1371/journal.pone.0108231.g008



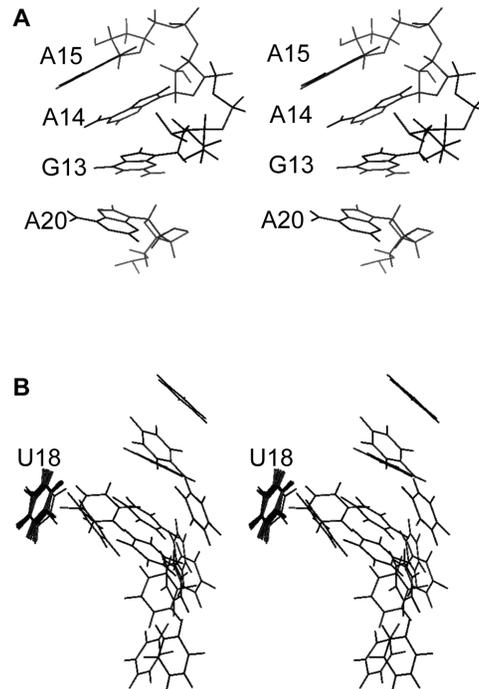
**Figure 9. Schematic illustration of RNA 2 structure.** Rectangles represent the bases of the indicated nucleotides. Base stacking is indicated by adjacent, parallel rectangles. Sugar pucker, as estimated from  $J_{1'-2'}$ , is indicated by shading of pentagons representing the ribose ring for each nucleotide. Lines pointing to internucleotide linkages indicate relative intensity of  $Mg^{2+}$ -induced cleavage at each position. doi:10.1371/journal.pone.0108231.g009

Molecular modeling suggests that the most likely hydrogen bonding partners for the interaction between U18 and U12 are U18 H3 and U12 O4. Such an interaction directs H3 of U12 inward toward the G13•C19 base pair, consistent with the partial protection of this proton from exchange with  $H_2O$ . This arrangement also places the interacting uridines proximal to the major groove edge of the A20•U11 base pair, consistent with the loss or destabilization of the U12–U18 interaction upon alteration of this base pair. Though there is no direct evidence for a hydrogen bonding interaction between either of the interacting uridines and this A•U base pair, a hydrogen bond between U18 O2 and the exocyclic amine of A20 is a possibility.

### Relevance of the structure of RNA 2 to RNA 1 and the miR-21 pre-element

Whereas RNA 1 most closely models the miR-21 pre-element, RNA 2 provides a better system for structural study by NMR, because its secondary structure is more clearly defined by the observation of resonances due to imino protons. This added structural definition extended beyond the two predicted G•U base pairs that were changed in 2 to include observation of the imino protons of U5 and G6. It is important, however, to evaluate the relationship between the structures of 1 and 2.

All of the peaks that are visible in the imino proton spectrum of 1 correspond to peaks in the spectrum of 2. Those corresponding to G13 and G23, which are proximal to the altered nucleotides, have the same chemical shifts in both. Thus, the chemical environments of G13 and G23 are not dramatically altered between the two molecules. There is also a peak in the spectrum of 1 that corresponds to that assigned to U18 in 2, though it is shifted downfield 0.2 ppm. The structural features that give rise to protection of this imino proton from exchange with  $H_2O$  are apparently present in both RNAs. Reversion of only one of each of the altered base pairs to a G•U pair (RNAs 6 and 7) results only in a loss of that G's imino proton from the spectrum and small changes in the chemical shift or intensity of peaks due to neighboring imino protons. These observations support the



**Figure 10. Stereoviews of structural features apparent in the NMR-based model of 2.** A. Stacking of consecutive purines G13–A15 and cross-strand stacking of A20. B. Orientation of U12 with respect to U18. The fourteen lowest energy structures were aligned on U18 and the pyrimidine rings of U18 and U12 are shown. doi:10.1371/journal.pone.0108231.g010

conclusion that 1 and 2 share most of their significant structural features.

The similarity of the in-line cleavage patterns for the two molecules further supports that conclusion. One of the primary differences in the patterns, stronger cleavage after U21 in 1 than after C21 in 2 can be attributed simply to lower stability of the base pairing in that stem. The other differences surround U12 and U18, which interact with each other. A perturbation in this interaction likely accounts for the alteration in backbone conformation or dynamics that alters the cleavage efficiency. The downfield shift of the imino proton peak for U18 between 2 and 1 is consistent with a perturbation of this interaction, but its presence in both suggests that the change does not entirely disrupt (or direct) the interaction but modifies its conformational details or dynamics. The lower stability of the duplex in 1 is sufficient to account for this modification. Thus, our data indicate that the conformation of RNA 2 resembles that of 1 in its most pronounced features.

### Relevance to maturation of miR-21

Zeng and co-workers have proposed a model in which the microprocessor and Dicer preferentially recognize conformations of the pre-element in which the base pairs are disrupted. They conclude that the predicted base paired regions as well as the loops of miRNA pre-elements are flexible [28,29]. Our data for RNA 1 confirm the flexibility of the stem region of the miR-21 pre-element. Backbone conformations that allow in-line cleavage and sufficient breathing of the base pairs to allow exchange of imino protons with water are evidence of its dynamic nature.

The model proposed by Zeng and co-workers is supported by data for processing of pri-miR-21 mutants corresponding to RNAs 6 and 7. Both of these mutations somewhat diminish cleavage by

Drosha in vitro and maturation of pri-miRNA to the active miRNA in cultured cells [28], as predicted from the model in which disruption of base pairs in the pre-element aids recognition by the processing machinery. It is noteworthy, however, that both of these mutant pri-miRNAs are nonetheless processed appreciably [28], confirming that the structures of these mutant pri-miRNAs are relevant in a biological context. Furthermore, the imino proton spectra of **6** and **7** are essentially identical to that of **2**, except for the peaks due to G10 and G22, which are directly involved in the altered base pairs. This similarity establishes the structural similarity of **6** and **7** to **2** (as well as to **1**, as noted above). Thus, the structural features we have identified for **2** are consistent with processing of a pri-miRNA to functional maturity. Similarly, even a pri-miR-21 mutant with both G•U base pairs and the A•U base pair predicted in the pre-element converted to G•C base pairs (i.e., a mutant corresponding to **2** with the additional stabilization of U11•A20 converted to a G•C pair) is processed poorly but measurably [28].

The structural characteristics of **2** that we have determined are those of the isolated element of RNA secondary structure. Interactions with protein factors can refashion conformationally labile secondary structures. However, the high degree of conservation of the miR-21 pre-element suggests conservation of its distinctive conformation as well as its sequence. Thus, auxiliary factors may recognize this structure in the regulation of miR-21 processing. In the case of factors that do alter the conformation (i.e., bind and stabilize an altered conformation), the conformation of the free RNA provides a basis for probing and understanding the changes induced. It also provides a basis for characterization of the interaction of artificial ligands with pri-miR-21.

## Materials and Methods

### Sample preparation

RNA was obtained from ThermoFisher with 2'OH protective groups in place, deprotected according to vendor protocol, and precipitated with sodium acetate and ethanol.

### NMR Spectroscopy

Spectra of exchangeable protons were measured with the sample dissolved in a 10:1 mixture of H<sub>2</sub>O and D<sub>2</sub>O containing 10 mM sodium phosphate, pH 6.7, and 50 μM EDTA. To measure spectra of non-exchangeable protons, the sample was dissolved in buffer and lyophilized to dryness before reconstituting in 600 μL of 99.96% D<sub>2</sub>O (Aldrich) to an RNA concentration of 1 mM. After reconstitution, the solution was heated briefly to 95°C and allowed to cool slowly to room temperature.

NMR experiments were recorded on Varian INOVA NMR spectrometers operating at proton frequencies of 600 and 800 MHz and equipped with cryogenic triple resonance probes. All experiments were acquired using standard pulse sequences from the library provided by the vendor's software. Spectra were processed using NMRPipe [56] and visualized and analyzed using NMRViewJ.

One-dimensional exchangeable proton spectra in H<sub>2</sub>O were collected at 5°C with a 1-1 water suppression sequence and a 1.2 s delay between pulses. A NOESY spectrum of the sample in H<sub>2</sub>O was acquired at 10°C (800 MHz spectrometer). Water suppression was achieved with a 1-1 pulse sequence having a 1.5 s delay between pulses. 160 scans were taken for each of 256 FIDs. The spectral width in both dimensions was 13587.0 Hz.

In all spectra of nonexchangeable protons, the residual HDO resonance was suppressed by presaturation during a 1.2 s relaxation delay. A double-quantum filtered COSY spectrum

was acquired with a spectral width of 5421 Hz (600 MHz spectrometer), 16 scans, 1322 points in the directly detected dimension and 800 complex points in the indirect dimension. A series of NOESY experiments with 60 ms, 100 ms, 150 ms, and 400 ms mixing times was acquired on a 600 MHz spectrometer with 5421 Hz spectral width, 16 scans, 1322 FIDs, and 800 complex points. A NOESY spectrum with 400 ms mixing time was acquired on a 800 MHz spectrometer with 7227 Hz spectral width, 8 scans, 764 FIDs and 800 complex points. A natural abundance heteronuclear <sup>1</sup>H-<sup>13</sup>C single quantum coherence (HSQC) spectrum was acquired (600 MHz spectrometer) using a standard pulse sequence and wurt140 for carbon decoupling during acquisition. The spectral width was 8012.8 Hz (13.4 ppm) in the <sup>1</sup>H dimension and 16000 Hz (106 ppm) in the <sup>13</sup>C dimension. 128 FIDs were collected with 2048 scans of 2728 complex points.

### In-line cleavage analysis

In-line cleavage, alkaline hydrolysis, and ribonuclease T1 digestion reactions were carried out and analyzed as previously described [42]. 5 mM MgCl<sub>2</sub> was included in each in-line cleavage reaction.

### Structure Modeling

The three-dimensional structure of the pre-element (nucleotides 8–23 of RNA **2**) was modeled using a restrained molecular dynamics protocol incorporating NMR-derived distance and torsion angle restraints. Where a Watson-Crick base pair was indicated by the observation of an imino proton resonance, distance restraints were applied to maintain the appropriate hydrogen bonding distances and coplanarity of the bases. The distance between imino protons for which an NOE was observed was restrained to the range 2.0–4.5 Å.

NOE cross-peak intensities were used semiquantitatively to assign distance ranges to nonexchangeable protons. Crosspeak intensities were characterized as strong, medium, weak, and very weak according to the NOESY mixing times at which they were observable. Strong cross-peaks, observable with a mixing time of 60 ms, were assigned the range 1.8–3.0 Å. Medium-intensity cross-peaks, observable with a NOESY mixing time of 100 ms or longer, were assigned the range 2.0–4.0 Å. Weak cross-peaks, observable with mixing times of 150 ms or longer, were assigned the range 2.5–5.0 Å. Very weak cross-peaks were only seen in NOESY spectra with a 400 ms mixing time and were assigned the range 2.5–6.0 Å. Internucleotide distance constraints between C8 and C9 and between G22 and G23 were applied to model an A-form geometry for these residues, clearly indicated by presence of base pairing and internucleotide NOE intensities.

The sugar conformations were characterized by the H1'-H2' scalar couplings evident in the DQF-COSY spectrum. Residues for which no H1'-H2' coupling was observed were constrained to the C3'-endo conformation with the endocyclic torsion angles  $\nu_0$ ,  $\nu_1$ ,  $\nu_2$ , and  $\nu_3$ . Sugars with coupling  $\geq 5.5$  Hz were constrained to the C2'-endo conformation. Sugars with coupling  $< 5.5$  Hz were constrained to the range of conformations including C2'-endo, O4'-endo, and C3'-endo. Because all intranucleotide H1' to aromatic NOESY crosspeaks were similar in intensity to those in the base paired lower stem (nucleotides 1–6 and 24–29), the glycosidic torsion angle,  $\chi$ , was constrained to the *anti* conformation for all nucleotides.

The molecular dynamics program CNS 1.3 was used to generate three-dimensional structures consistent with the NMR data. The covalent structure of the RNA was created in an extended conformation and subjected to a simulated annealing

protocol, varying the initial velocities for multiple structure calculations. Torsion dynamics and the CNS default parameters for nucleic acids were used in each step. High temperature annealing was simulated with 4000 steps, 15 ps, at 36,000 K. Subsequently, 1000 steps of slow cooling were followed by 10 cycles of 200 final minimization steps. Structures were displayed using PyMol.

## Acknowledgments

The authors wish to thank the UTSW Departments of Biophysics and Biochemistry for kindly allowing our use of spectrometers. We also wish to

## References

- Bartel DP (2004) MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell* 116: 281–297.
- Chekulaeva M, Filipowicz W (2009) Mechanisms of miRNA-mediated post-transcriptional regulation in animal cells. *Curr Opin Cell Biol* 21: 452–460.
- Ambros V (2004) The function of animal microRNAs. *Nature* 431: 350–355.
- Garzon R, Calin G, Croce C (2009) MicroRNAs in Cancer. *Annu Rev Med* 60: 167–179.
- van Rooij E, Sutherland LB, Liu N, Williams AH, McAnally J, et al. (2006) A signature pattern of stress-reponsive microRNAs that can evoke cardiac hypertrophy and heart failure. *Proc Natl Acad Sci USA* 103: 18255–18260.
- Divakaran V, Mann DL (2008) The Emerging Role of MicroRNAs in Cardiac Remodeling and Heart Failure. *Circ Res* 103: 1072–1083.
- Volinia S, Calin G, Liu C-G, Ambs S, Cimmino A, et al. (2006) A microRNA expression signature of human solid tumors defines cancer gene targets. *Proc Nat Acad Sci USA* 103: 2257–2261.
- Asangani IA, Rasheed SAK, Nikolova DA, Leupold JH, Colburn NH, et al. (2008) MicroRNA-21 (miR-21) post-transcriptionally downregulates tumor suppressor Pdc4 and stimulates invasion, intravasation and metastasis in colorectal cancer. *Oncogene* 27: 2128–2136.
- Zhu S, Si M-L, Mo Y-Y (2007) MicroRNA-21 targets the tumor suppressor gene tropomyosin 1 (TPM1). *J Biol Chem* 282: 14328–14336.
- Zhu SH, Wu F, Wu D, Nie S, Sheng, et al. (2008) MicroRNA-21 targets tumor suppressor genes in invasion and metastasis. *Cell Res* 18: 350–359.
- Meng F, Henson R, Wehbe-Janeck H, Ghoshal K, Jacob ST, et al. (2006) MicroRNA-21 regulates expression of the PTEN tumor suppressor gene in human hepatocellular cancer. *Gastroenterology* 133: 647–658.
- Frankel LB, Christoffersen NR, Jacobsen A, Lindow M, Krogh A, et al. (2008) Programmed cell death 4 (PDCD4) is an important functional target of the microRNA miR-21 in breast cancer cells. *J Biol Chem* 283: 1026–1033.
- Thum T, Gross C, Fiedler J, Fischer T, Kissler S, et al. (2008) MicroRNA-21 contributes to myocardial disease by stimulating MAP kinase signaling in fibroblasts. *Nature* 456: 980–986.
- Esquela-Kerscher A, Slack FJ (2006) Oncomirs-microRNAs with a role in cancer. *Nature Rev Cancer* 6: 259–269.
- van Rooij E, Marshall WS, Olson EN (2008) Toward microRNA-based therapeutics for heart disease. *Circ Res* 103: 919–928.
- Obernosterer G, Leuschner PJ, Alenius M, Martinez J (2006) Posttranscriptional regulation of microRNA expression. *RNA* 12: 1161–1167.
- Thomson JM, Newman M, Parker JS, Morin-Kensicki EM, Wright T, et al. (2006) Extensive post-transcriptional regulation of microRNAs and its implications for cancer. *Genes Dev* 20: 2202–2207.
- Gregory RI, Yan KP, Amuthan G, Chendrimada T, Doratotaj B, et al. (2004) The Microprocessor complex mediates the genesis of microRNAs. *Nature* 432: 235–240.
- Han J, Lee Y, Yeom KH, Kim YK, Jin H, et al. (2004) The Drosha-DGCR8 complex in primary microRNA processing. *Genes and Development* 18: 3016–3027.
- Han J, Lee Y, Yeom KH, Nam JW, Heo I, et al. (2006) Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell* 125: 887–901.
- Lee Y, Ahn C, Han J, Choi H, Kim J, et al. (2003) The nuclear RNase III Drosha initiates microRNA processing. *Nature* 425: 415–419.
- Landthaler M, Yalcin A, Tuschl T (2004) The human DiGeorge syndrome critical region gene 8 and Its D. melanogaster homolog are required for miRNA biogenesis. *Curr Biol* 14: 2162–2167.
- Lund E, Guttinger S, Calado A, Dahlberg JE, Kutay U (2004) Nuclear export of microRNA precursors. *Science* 303: 95–98.
- Yi R, Qin Y, Macara IG, Cullen BR (2003) Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev* 17: 3011–3016.
- Hutvagner G, McLachlan J, Pasquinelli AE, Balint E, Tuschl T, et al. (2001) A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science* 293: 834–838.
- Khvorova A, Reynolds A, Jayasena SD (2003) Functional siRNAs and miRNAs exhibit strand bias. *Cell* 115: 209–216.
- Michlewski G, Guil S, Semple CA, Caceres JF (2008) Posttranscriptional regulation of miRNAs harboring conserved terminal loops. *Mol Cell* 32: 383–393.
- Zeng Y, Yi R, Cullen BR (2005) Recognition and cleavage of primary microRNA precursors by the nuclear processing enzyme DROSHA. *EMBO J* 24: 138–148.
- Zhang X, Zeng Y (2010) The terminal loop region controls microRNA processing by Drosha and Dicer. *Nucleic Acids Res* 38: 7689–7697.
- Newman MA, Thomson JM, Hammond SM (2008) Lin-28 interaction with the Let-7 precursor loop mediates regulated microRNA processing. *RNA* 14: 1539–1549.
- Trabucchi M, Briata P, Garcia-Mayoral M, Haase AD, Filipowicz W, et al. (2009) The RNA-binding protein KSRP promotes the biogenesis of a subset of microRNAs. *Nature* 459: 1010–1014.
- Viswanathan SR, Daley GQ, Gregory RI (2008) Selective blockade of microRNA processing by Lin28. *Science* 320: 97–100.
- Chan JA, Krichevsky AM, Kosik KS (2005) MicroRNA-21 is an antiapoptotic factor in human glioblastoma cells. *Cancer Res* 65: 6029–6033.
- Moriyama T, Ohuchida K, Mizumoto K, Yu J, Sato N, et al. (2009) MicroRNA-21 modulates biological functions of pancreatic cancer cells including their proliferation, invasion, and chemoresistance. *Mol Cancer Ther* 8: 1067–1074.
- Hatley ME, Patrick DM, Garcia MR, Richardson JA, Bassel-Duby R, et al. (2010) Modulation of K-Ras-dependent lung tumorigenesis by microRNA-21. *Cancer Cell* 18: 282–293.
- Tatsuguchi M, Seok HY, Callis TE, Thomson JM, Chen J-F, et al. (2007) Expression of microRNAs is dynamically regulated during cardiomyocyte hypertrophy. *J Mol Cell Cardiology* 42: 1137–1141.
- Bose D, Jayaraj G, Suryawanshi H, Agarwala P, Pore SK, et al. (2012) The tuberculosis drug streptomycin as a potential cancer therapeutic: inhibition of miR-21 function by directly targeting its precursor. *Angew Chem Int Ed* 51: 1019–1023.
- Chirayil S, Chirayil R, Luebke KJ (2009) Discovering Ligands for a microRNA Precursor with Peptoid Microarrays. *Nucleic Acids Res* 37: 5486–5497.
- Schöniger C, Arenz C (2013) Perspectives in targeting miRNA function. *Bioorg Med Chem* 21: 615–618.
- Murata A, Fukuzumi T, Umamoto S, Nakatani K (2013) Xanthone derivatives as potential inhibitors of miRNA processing by human Dicer: Targeting secondary structures of pre-miRNA by small molecules. *Bioorg Med Chem Lett* 23: 252–255.
- Maiti M, Nauwelaerts K, Herdewijn P (2012) Pre-microRNA binding aminoglycosides and antitumor drugs as inhibitors of Dicer catalyzed microRNA processing. *Bioorg Med Chem Lett* 22: 1708–1711.
- Diaz JP, Chirayil R, Chirayil S, Tom M, Head KJ, et al. (2014) Association of a peptoid ligand with the apical loop of pri-miR-21 inhibits cleavage by Drosha. *RNA* 20: 528–539.
- Lunse CE, Michlewski G, Hopp CS, Rentmeister A, Caceres JF, et al. (2010) An Aptamer Targeting the Apical-Loop Domain Modulates pri-miRNA Processing. *Angew Chem Int Ed* 49: 4674–4677.
- Griffiths-Jones S (2004) The microRNA registry. *Nucleic Acids Res* 32: D109–D111.
- Soukup GA, Breaker RR (1999) Relationship between internucleotide linkage geometry and the stability of RNA. *RNA* 5: 1308–1325.
- Varani G, Tinoco I (1991) RNA structure and NMR spectroscopy. *Quart Rev Biophys* 24: 479–532.
- van den Hoogen YT, Erkelens C, de Vroom E, van der Marel GA, van Boom JH, et al. (1988) Influence of uracil on the conformational behaviour of RNA oligonucleotides in solution. *Eur J Biochem* 173: 295–303.
- Sich C, Ohlenschläger O, Ramachandran R, Gorchach M, Brown LR (1997) Structure of an RNA hairpin loop with a 5'-CGUUUCG-3' loop motif by heteronuclear NMR spectroscopy and distance geometry. *Biochem* 36: 13989–14002.
- Schwalbe M, Ohlenschläger O, Marchanka A, Ramachandran R, Hafner S, et al. (2008) Solution structure of stem-loop  $\alpha$  of the hepatitis B virus post-transcriptional regulatory element. *Nucleic Acids Res* 36: 1681–1689.

50. Yoo J-S, Cheong H-K, Lee BJ, Kim Y-B, Cheong C (2001) Solution structure of the SL1 RNA of the M1 double-stranded RNA Vvirus of *Saccharomyces cerevisiae*. *Biophys J* 80: 1957–1966.
51. Allain FH, Varani G (1995) Structure of the P1 helix from group I self-splicing introns. *J Mol Biol* 250: 333–353.
52. Butcher SE, Dieckmann T, Feigon J (1997) Solution structure of a GAAA tetraloop receptor RNA. *EMBO J* 16: 7490–7499.
53. Laing LG, Hall KB (1996) A model of the iron responsive element RNA hairpin loop structure determined from NMR and thermodynamic data. *Biochem* 35: 13586–13596.
54. Fountain MA, Serra MJ, Krugh TR, Turner DH (1996) Structural features of a six-nucleotide RNA hairpin loop found in ribosomal RNA. *Biochem* 35: 6539–6548.
55. Jaeger JA, Tinoco I (1993) An NMR study of the HI V- 1 TAR element hairpin. *Biochem* 32: 12522–12530.
56. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, et al. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6: 277–293.