Published in final edited form as:

Biochemistry. 2017 July 25; 56(29): 3733–3744. doi:10.1021/acs.biochem.7b00201.

NMR Structure of an 8x8 Nucleotide RNA Internal Loop Flanked on Each Side by Three Watson-Crick Pairs and Comparison to 3D Predictions

Andrew D. Kauffmann^{†,§}, Scott D. Kennedy^{‡,§}, Jianbo Zhao^{†,§}, and Douglas H. Turner^{†,§,*}

[†]Department of Chemistry, University of Rochester, Rochester, New York 14627, United States

[‡]Department of Biochemistry and Biophysics, University of Rochester School of Medicine and Dentistry, Rochester, New York 14642, United States

[§]Center for RNA Biology, University of Rochester, Rochester, New York 14627, United States

Abstract

The prediction of RNA 3D structure from sequence alone is a long-standing goal. High resolution, experimentally determined structures of simple noncanonical pairings and motifs are critical to the development of prediction programs. Here, we present the NMR structure of the duplex,

*Corresponding author. turner@chem.rochester.edu.

ASSOCIATED CONTENT

Supporting Information

- I. Proton and phosphorus chemical shift assignments for the duplex, (5'CCAGAAACGGAUGGA)₂.
- II. Experimentally assigned chemical shifts compared to the average and range of chemical shifts predicted by NUCHEMICS for the ensemble of structures generated by simulated annealing using the NMR data alone.
- III. Experimentally assigned chemical shifts compared to the average and range of chemical shifts predicted by NUCHEMICS on the ensemble of structures generated by simulated annealing using NMR data with the additional A7/A7* N1 to C8*/C8 H41, used to force the AN1-CH41 single hydrogen bond pair.
- IV. Carbon chemical shift assignments for the duplex, (5'CCAGAAACGGAUGGA)₂.
- V. Comparison of observed vs. average NUCHEMICS predicted chemical shifts for the ensemble of structures having the additional A7/A7* N1 to C8*/C8 H41 restraint.
- VI. Dinucleotide stacking of 5'AC/3'CA and 5'AA/3'GC from the ensemble of structures having the additional A7/A7* N1 to C8*/C8 H41 restraint.
- VII. Thermodynamics calculations for the sequence, 5'CCAGAAACGGAUGGA, using the nearest neighbor model.
- VIII. One dimensional ³¹P spectrum of (5'CCAGAAACGGAUGGA)₂ at 10 °C.
- IX. Ensemble of 33 accepted structures generated from restrained simulated annealing.
- X. Adenosine C2-H2 region of the ¹³C-¹H HSQC spectrum.
- XI. Distance restraints for AMBER modeling of the duplex.
- XII. Dihedral restraints for AMBER modeling of the duplex.
- XIII. Additional restraints for AMBER modeling of the duplex, forcing the AN1-CH41 single hydrogen bond pair.

ACCESSION CODES

NMR chemical shift assignments were deposited in the Biological Magnetic Resonance Data Bank as entry 30268. Structural coordinates were deposited in the RCSB Protein Data Bank as entry 5V2R.

The authors declare no competing financial interest.

(5'CCAGAAACGGAUGGA)₂, which contains an 8x8 nucleotide internal loop flanked by three Watson-Crick pairs on each side. The loop is comprised of a central 5'AC/3'CA nearest neighbor flanked by two 3RRs motifs, a known stable motif consisting of three consecutive sheared GA pairs. Hydrogen bonding patterns between base pairs in the loop, all-atom RMSD for the loop, and deformation index were used to compare the structure to automated predictions by MC-sym, RNA FARFAR, and RNAComposer.

INTRODUCTION

Advances in high throughput RNA-seq technologies and whole transcriptome sequencing have rapidly expanded the RNA sequence library.^{1, 2} Likewise, methods such as free energy minimization with restraints from sequence comparison,³⁻⁷ chemical mapping,⁸⁻¹⁰ NMR,^{11, 12} and combinations of these,^{13, 14} have provided rapid and accurate secondary structure prediction. Solving 3D structures from sequence and secondary structure, however, remains difficult.¹⁵⁻¹⁹ Development of programs for predicting 3D structure can be facilitated by availability of simple structural benchmarks with noncanonical RNA pairs and motifs.

Noncanonical pairs,²⁰ especially GA²¹⁻²⁶ and AC^{27, 28} (Figure 1), have conformations that are very sensitive to sequence and structural environment. GA pairs can be involved in a two hydrogen bond imino pair (cis Watson-Crick/Watson-Crick), two hydrogen bond sheared pair (trans Hoogsteen/sugar edge), and base triples.^{29, 30} They can create larger, multi-pair conserved motifs such as the Loop E motif,³¹ the sarcin-ricin loop,^{32, 33} kink-turn,³⁴ and the 3RRs motif.²⁷ The 3RRs motif consists of three consecutive sheared purine-purine pairs, including three GA pairs or a combination of GA, GG, and AA pairs and is known to be thermodynamically stabilizing.^{27, 35}

AC pairs also exhibit a range of conformations.^{20, 27, 28} In addition, the pK_a of A residues is highly sensitive to sequence and structural environment and thus A can be protonated at or near physiological pH's,^{36, 37} further increasing the number of available conformations.

Due to the range of possible interactions available to GA, AC, and other noncanonical pairs, experimentally determined three-dimensional structures can provide difficult and important benchmarks for the development of prediction programs. Herein, we report the solution structure of a 30 nucleotide self-complementary duplex, (5'CCAGAAACGGAUGGA)₂, with an 8x8 nucleotide internal loop consisting of six GA and two CA pairs flanked on each side by three Watson-Crick pairs and a 3' dangling end A.

MATERIALS AND METHODS

NMR Sample Preparation

A 5 μmol synthesis of the RNA oligonucleotide, 5'CCAGAAACGGAUGGA, was purchased from Integrated DNA Technologies. For NMR spectra of exchangeable protons, the RNA was suspended in 330 μL of 95:5 (v/v) H₂O/D₂O buffer (80 mM KCl, 0.05 mM Na₂EDTA, 20 mM potassium phosphate, pH 6.2) and placed into a Shigemi tube. The RNA was annealed in the NMR probe by raising the temperature to 80 °C and slowly cooling at a

rate of 10 °C/5 min to promote formation of the self-complementary duplex. The total RNA concentration of the sample was 3.1 mM.

For spectra of non-exchangeable protons, the sample was lyophilized in a vacuum centrifuge and resuspended in 99.9% D₂O (Cambridge Isotope Laboratories) twice. The sample was lyophilized again and resuspended in 330 µL of 99.96% D₂O (Cambridge Isotope Laboratories) from a fresh ampule to minimize protons in the sample. The sample was then placed in a Shigemi tube for collection of spectra.

NMR Spectroscopy and Assignment

Spectra were collected with a Varian Inova 600 MHz spectrometer except the natural abundance ¹³C-¹H sensitivity enhanced HSQC which was collected with a Bruker 800 MHz spectrometer. One dimensional imino proton spectra were collected using a 1-1-spin-echo water suppression pulse³⁸ at 0 and 10 °C. Two dimensional NOESY spectra in H₂O were collected using a WATERGATE pulse^{39, 40} with flipback for water suppression at 0 °C and mixing times of 125 and 400 msec.

Assignment of non-exchangeable protons was achieved with 2D NOESY (75, 150, and 400 msec mixing times), TOCSY (32 msec mixing time), natural abundance ¹³C-¹H sensitivity enhanced HSQC, and ³¹P-¹H HETCOR spectra.

NMRpipe⁴¹ was used to process two dimensional data. Spectra were internally referenced to the chemical shift of H₂O or HDO, and chemical shifts are reported relative to 2,2-dimethylsilapentane-5-sulfonic acid (DSS). NMRFAM-SPARKY⁴² was used to assign protons.

Restraint Generation

Distances between nonexchangeable protons were obtained from spectra at 10 °C and 75 msec mixing time. NOE cross-peak volumes were (1/r)⁶ scaled and referenced to H1'-H2' cross-peak volumes of C2 and G13 (2.75 Å). For non-overlapped resonances, cross-peak volumes were measured using the sum over box integration method in NMRFAM-SPARKY.⁴² For partially overlapped resonances, volumes were measured using Gaussian fitting. Upper and lower bounds for distance restraints were calculated allowing a 4-fold error in NOE volume measurement.

For exchangeable protons, lower bounds were set to 2.0 Å and upper bounds were set to 4.0, 5.0 or 6.0 Å, respectively, for large, medium or small NOE volumes. Because of the ambiguity of the guanine amino protons in the loop, distances were restrained to N2 assuming a 1.0 Å N-H bond length.

Dihedral angles for the Watson-Crick stem were restrained to typical A-form values⁴³: α (-65 ± 90°), β (165 ± 75°), γ (60 ± 60°), δ (80 ± 35°), ε (-115 ± 125°), ζ (-70 ± 90°), and χ (-255 ± 85°). The relative sizes of intraresidue H8/H6-H1' NOE cross-peaks for residues not in Watson-Crick pairs were consistent with all residues having χ in the anti conformation, so χ was restrained to -255 ± 85°. Only G10/G10* had a TOCSY cross-peak

between H1' and H2'; thus G10/G10* were restrained to a C2'-endo sugar pucker, δ ($130 \pm 20^\circ$). All other riboses were restrained to a C3'-endo sugar pucker, δ ($80 \pm 35^\circ$).

Structure Generation

Three dimensional structures were generated with restrained simulated annealing starting with an A-form like structure generated from NUCGEN.⁴⁴ A generalized Born implicit solvent was used with a salt concentration of 0.1 M NaCl.⁴⁵ One hundred structures were generated with AMBER 14⁴⁶ using the following protocol: (1) The structure was held at 3000 K for 5 ps (5,000 steps); (2) cooled to 100 K over 93 ps (93,000 steps); (3) held at 0 K for 2 ps (2,000 steps). Force constants for NOE and dihedral angle restraints were 30 kcal mol⁻¹ Å⁻² and 30 kcal mol⁻¹ rad⁻², respectively, with a weighting factor of 0.1 for the first 5,000 steps and 1.0 for the remainder of the simulation. The resulting structures were sorted by restraint energy and structures with restraint energies of lower than 0.1 kcal mol⁻¹ with no distance violations greater than 0.1 Å were selected as the final ensemble. Structure analysis and figures were generated with PyMOL.⁴⁷

Structure Prediction

The RNA FARFAR (RNA de novo) protocol,^{48–50} part of the ROSETTA software package, was run on the ROSIE server.⁵¹ The program was given the sequence and secondary structure of the duplex. Two runs were submitted, one with no chemical shifts and one with all assigned chemical shifts except for H5' and H5''. Both submissions were run with the preset options with the addition of the *Allow Bulge* and *Use Updated (2012) Force Field* options as per recommendations in the user manual. The best scoring structure for each of the two runs was selected as the representative structure.

A new sequence was designed for submission to MC-sym^{52–54} and RNAComposer^{55, 56} because they do not allow submission of multi-strand motifs. The 3' dangling A residues were removed from the 30 nt duplex, and an A₄ hairpin loop was added connecting G14 with C16 to make a new 32 nt hairpin. All references to nucleotide numbers in this new hairpin are with respect to the equivalent nucleotide in the 30 nt duplex.

The 32 nt hairpin was submitted to the MC-sym server.^{52, 53} The input script was generated by MC-fold⁵⁴ with no restraints and contained the correct Watson-Crick base pairs. MC-sym was run with the preset protocol, generating 1000 structures. The resulting structures were then polished using the *Relieve* function. After refinement, structures were scored with the *Score* function and the best scoring structure was chosen for comparison to the solution structure determined by NMR.

Due to computational time requirements, a second run was submitted to MC-sym using the same sequence and input script, but generating only 10 structures. This allowed for both the *Relieve* and *Refine* function to be run as advised by the user manual. These structures were then scored and the best scoring structure was selected as the representative structure.

The sequence and secondary structure of the 32 nt hairpin were also submitted to the RNAComposer server.^{55, 56} No experimentally generated restraints were provided and no additional minimization was performed after the structure was generated. To test if the A₄

hairpin loop had any influence on the internal loop, A₆ and GAAA loops were also submitted and analyzed.

The *rms* function in PyMOL was used to calculate RMSDs between each predicted structure and the lowest restraint energy model from the NMR ensemble. Only residues involved in the internal loop (G4-A11 and G4*-A11*) were used for the calculation. The deformation index (DI) comparing each prediction with the lowest restraint energy model from the NMR ensemble was calculated with the RNA interaction network fidelity module on the MC-sym server using the default options.⁵⁷

Quantum Mechanics Energy Calculations

Quantum Mechanics (QM) calculations were performed with the Gaussian 09 package.⁵⁸ Due to the large size of the RNA, a model system containing A6-G9*, A7-C8*, C8-A7*, and G9-A6* nucleobase pairs was used. The sugar-phosphate backbones were replaced by methyl groups to reduce the total number of electrons and eliminate the negative charges. Electronic structure energies were calculated with B3LYP-D3⁵⁹ and M06-2X⁶⁰ methods at the 6-31+g(d,p) level of theory.⁶¹

RESULTS

Conformation of the Duplex

Because this sequence forms a self-complementary duplex, it has the possibility of forming a hairpin. Nearest neighbor thermodynamic parameters^{8, 62} for 1 M NaCl predict the absence of single strand and an equilibrium constant of 1.5 at 10 °C (Figure S3) for the hairpin to duplex reaction:

This predicts only 0.9% of the strands will be in duplex at the 3.1 mM concentration used for NMR (Figure S3). Furthermore, the 100 mM K⁺ concentration used in the NMR buffer should favor hairpin conformation.

One set of imino proton peaks were observed in a 1D NMR spectrum of the 3.1 mM RNA sample. To determine which conformation these peaks represented, the 3.1 mM sample was diluted to 150 μM and 15 μM. As the RNA concentration decreased, a second set of imino proton peaks grew in intensity relative to the set of peaks present at 3.1 mM RNA (Figure 2). This is consistent with the expectation that hairpin conformations are thermodynamically favored at lower RNA concentrations. Evidently, this sequence adopts only a self-complementary duplex conformation at 3.1 mM. Thus, the 1D NMR spectra reveal that the current nearest neighbor parameters^{8, 62} do not accurately predict the thermodynamics for this sequence.

Two-dimensional imino proton spectra confirmed the six expected Watson-Crick pairs in the duplex (Figure 3). Assignments were made using NOE cross-peak connections to cytosine

H5 and H6 for GC pairs and to adenosine H2 for AU pairs. Resonance assignments are listed in Supporting Information (Table S1).

The Sets of Consecutive GA Pairs Form Two 3RRs Motifs

Imino peaks at 10.03, 10.40, and 9.40 ppm, were assigned to G4, G9, and G10, respectively. These upfield shifts relative to shifts of imino protons in Watson-Crick pairs suggest these protons are not directly involved in a hydrogen bond to nitrogen. The sharpness of the resonances, however, suggests these protons are protected from exchange and the residues are stably paired.

Sheared pairs (Figure 1) are formed by G4-A11*, A5-G10*, and A6-G9* (along with their symmetric equivalents). These sheared pairs are confirmed by cross peaks of G4H1-A11*H8, G9H1-A6*H8, and G10H1-A5*H8. Also, the upfield chemical shifts of A7H1' and U12H1' to 5.02 and 4.02 ppm (Table S1), respectively, are consistent with positioning of A6 and A11, respectively, directly over these protons.^{27, 35} In addition, the G4 amino protons were separately assigned to 5.30 and 7.38 ppm. The markedly different chemical shifts suggest one of these protons is involved in a stable hydrogen bond.

Cross-strand stacking between the G4-A11* and A5-G10* sheared pairs is confirmed by: (1) cross peaks between G4H1-G10*H1', G4H1-U12*H6, G10*H1-G4H1', G10*H1-G4H2'; (2) cross peaks between A5H2-A11*H1' and A11*H2-A5H1'; and (3) a cross peak between A5H2-A11*H8. This cross-strand stacking pattern is consistent with the upfield chemical shifts of G9H2' and A5H2. Structures generated from restrained simulated annealing are consistent (Table 1) and show the ring of G10* positioned directly over G9*H2' explaining the upfield chemical shift of 3.43 ppm. Similarly, the rings of A6 and A11* are positioned directly over and under A5H2, resulting in the slight upfield chemical shift of this proton to 7.23 ppm.

Other data consistent with the 3RRS motif include the C2'-endo sugar pucker of G10, confirmed by a TOCSY cross peak between G10H1' and H2', and the downfield chemical shift of the phosphorus between G10 and A11 to 0.208 ppm, ~3.5 ppm separated from all other phosphorus signals (Figure S4). These features have previously been reported for the equivalent nucleotides in other structures with 3RRs motifs.^{27, 35}

AC Pairs May Adopt Multiple Conformations

The pairing of A7 with C8* and C8 with A7* are likely dynamic. Resonances for A7/A7* H1', H2, and H8 and for C8/C8* H5 and H6 are significantly broader than other resonances for protons in the loop (Figure 4). The resonance for C8/C8* H1' is too broad to assign.

Despite the sparse restraints available for the AC region, simulated annealing structures always contained AC pairs with a single hydrogen bond between the AH61 amino proton and the C carbonyl oxygen (Figure 5). Because a previous structure of an AC pair next to a 3RRs motif²⁷ had a single hydrogen bond between the C amino proton and A N1, it was explored further. A new restraint was added to force A7/A7* N1 to be between 1.8 and 2.4 Å from C8/C8* H41, consistent with the previously reported CA structure. After simulated annealing, this new pairing did not introduce new restraint violations or raise the restraint

energy significantly. It is likely that both pairings, or more, could be exchanging, thus explaining the peak broadening in this region. Due to the lack of upfield shifted adenosine C2 chemical shifts, indicative of a protonated adenosine residue,³⁷ the A⁺-C wobble pair was not considered (Figure S6 and Table S4).

Quantum Mechanics Calculations

QM was used to explore the energy difference expected between base-base interactions of the tandem AC pairs in the structures generated by simulated annealing with NMR data alone, i.e. having AH61-CO2 hydrogen bond pairs, and the structure generated with the additional restraint to force the AN1-CH41 single hydrogen bond pairs. Single point QM energies were calculated for the A6-G9*, A7-C8*, C8-A7*, and G9-A6* base pairs held in the positions generated for the two different structures (see Materials and Methods). Using the B3LYP-D3 and M06-2X methods, the AH61-CO2 pairing was more stable by 27.01 and 24.05 kcal/mol, respectively.

Chemical shift predictions of NMR structures using NUCHEMICS

To check that the solution structure of the duplex is well represented in the ensemble of structures produced by restrained simulated annealing, NUCHEMICS⁶³ was used to predict the chemical shifts for each non-exchangeable proton. Chemical shifts were considered to agree with prediction if they fell within the range of shifts calculated by NUCHEMICS for either of its symmetric equivalents, allowing for a 0.16 ppm margin of error, which is the RMS deviation reported for NUCHEMICS.⁶³ Of the 89 experimentally assigned chemical shifts, 84 agreed with the predictions. Of the 5 protons that did not fall within the predicted range, only 3 were within the 8x8 internal loop, G4H3', G4H8, and G10H3' (Figure 6, Table S2). These resonances were outside of the predicted range by 0.234, 0.425, and 0.113 ppm, respectively.

In an attempt to rule out one of the AC pairings via chemical shift agreement, NUCHEMICS was also used to predict the proton chemical shifts for the structures generated from the NMR data with the additional restraint, A7/A7* N1 to C8*/C8 H41, used to force the AN1 – CH41 single hydrogen bond pair (Figures 1 and S1, Table S3). Using these structures, 82 of the 89 chemical shifts agreed with prediction, with 5 of those that did not fall within the range predicted contained in the 8x8 internal loop. Of the 7 proton chemical shifts that were outside the predicted range, 5 matched protons that were also outside the predicted range in the models that used the NMR data alone. The 2 protons outside the prediction range unique to these models were G9H8 and G10H8, which were outside of the range by 0.107 and 0.046 ppm, respectively. Because both A7 and C8 had no proton chemical shifts outside the NUCHEMICS predicted range, neither conformation can be ruled out.

Interestingly, the chemical shift of C8H1', which could not be assigned, was predicted to be 4.27 ppm from the ensemble generated with NMR data alone. These structures have AC pairs with an AH61-CO2 hydrogen bond. For the ensemble using the additional A7/A7* N1 - C8*/C8 H41 restraint, C8H1' was predicted to resonate at 5.46 ppm, where it should be observable and assignable. The difficulty in assigning the peak is understandable if it is broad and at 4.27 ppm where it would overlap many other signals. Thus, the absence of a

resonance at 5.46 ppm is consistent with the AH61-CO2 hydrogen bonded AC pairs of the NMR model.

Comparison of the NMR Model to Automated 3D Structure Predictions

RNA FARFAR predicted a duplex structure with a DI of 6.74 and an RMSD for the loop of 3.71 Å compared to the experimental structure (Table 2). All six of the GA pairs in the loop were predicted to be imino GA pairs with two hydrogen bonds, GH1 to AN1 and AH61 to GO6 (Figure 1). This is not supported by the NMR data. Both AC pairs were predicted to be in the same conformation, a single hydrogen bond between AN1 and CH41.

RNA FARFAR has been expanded to allow restraints from assigned non-exchangeable proton chemical shifts. When these chemical shifts were provided, RNA FARFAR predicted a structure with a DI of 5.44 and an RMSD for the loop of 3.49 Å (Table 2). Of the six GA pairs, three were in the imino conformation (G4-A11*, A5-G10*, and A6-G9*), two pairs had no hydrogen bonds (G9-A6* and G10-A5*), and the remaining A11-G4* pair was sheared. NMR data only agrees with the prediction for the A11-G4* pair. As with the prediction with no chemical shift data provided, both AC pairs were predicted to be in the CH41-AN1 single hydrogen bond conformation.

A hairpin sequence (see Materials and Methods) was submitted to MC-sym because it does not accept duplexes. The model generated using only the *Relieve* function had a DI of 5.28 and an RMSD for the loop of 3.99 Å (Table 2). The G9-A6*, G10-A5*, and A11-G4* pairs were predicted to be sheared and to form the 3RRs motif, including the cross-strand stacking pattern between the G10-A5* and A11-G4* pairs. The other set of GA pairs were not as accurately predicted as only the pair between A6-G9* was sheared while G4-A11* and A5-G10* showed no hydrogen bonding. The pair between A7-C8* was predicted to have the same hydrogen bonding pattern as seen from the NMR structure, A7H61-C8*O2, but the C8-A7* pair showed no hydrogen bonding.

The model predicted by MC-sym when both the *Relieve* and *Refine* steps were performed showed marked improvement, giving a DI of 3.26 and a loop RMSD of 2.46 Å. All six GA pairs in the internal loop were predicted correctly as sheared pairs, generating two 3RRs motifs in agreement with the NMR structure. The A7-C8* and C8-A7* pairs, respectively, had A7N1-C8*CH41 and C8O2-A7*H61 hydrogen bonding. That is, both CA pairings largely consistent with NMR spectra were predicted to be present simultaneously. NMR spectra do not definitively exclude this possibility.

The models generated by RNAComposer using an A₄, A₆, or GAAA connecting hairpin loop (see Materials and Methods) had DI's of 4.12, 5.48, and 5.60, and RMSDs for the internal loop of 2.69, 3.37, and 3.38 Å, respectively (Table 2). In all three models, G4-A11* is predicted as a sheared pair; A5 and G10* do not hydrogen bond, but are stacked on top of each other. The A6-G9* pair is predicted to be sheared using the GAAA connecting loop but no pair is predicted to form using the A₄ or A₆ loop. The A7-C8* and C8-A7* pairs show almost no consistencies between the three models. Interestingly, the G9-A6*, G10-A5*, and A11-G4* pairs are predicted to form a 3RRs motif in the model with the A₄ connecting hairpin loop, but no pairing is predicted for these pairs using the A₆ or GAAA hairpin loops.

DISCUSSION

Accurate prediction of 3D structure from sequence is a major goal of RNA research.^{17–19} Hydrogen bonding between GA^{21–24} and AC^{27, 28} pairs is very sensitive to sequence context and difficult to predict.^{64–66} To provide an indication of the current state of the field, the NMR structure was determined for the duplex, (5'CCAGAAACGGAUGGA)₂, and compared to automated predictions with three methods, RNA FARFAR,^{48–51} MC-sym,^{52, 53} and RNAComposer^{55, 56} (Table 2).

The duplex, (5'CCAGAAACGGAUGGA)₂, has six Watson-Crick pairs, two AC pairs, and two 3RRs motifs with each comprised of three sheared GA pairs. The 3RRs motif is a structural element present in a broad range of sequence environments.²⁷ The isolated 3x3 loop, 5'NGAAN/3'NAGGN, adopts a 3RRs structure while being tolerant of canonical closing base pairs.³⁵ The 3RRs motif can also form in the context of a larger loop when adjacent to a noncanonical pair, such as the CA pair in the loop, 5'GAAGCC/3'CAGGAG, from the *B. mori* R2 retrotransposon (PDB ID: 2L8F).²⁷ Many examples²⁷ have also been seen in crystal structures of ribosomes. The structure of (5'CCAGAAACGGAUGGA)₂ shows that the two 3RRs motifs are structurally stable even in the context of an 8x8 nucleotide internal loop.

While the two 3RRs motifs each have three sheared purine-purine pairs as observed in other contexts, the two CA pairs are less well defined by NMR. Although very similar in sequence to the internal loop structure for 5'GAAGCC/3'CAGGAG, the A7C8* pair reported here behaves very differently from the CA pair of 5'GAAGCC/3'CAGGAG. The latter CA had a very large NOE between the AH2 and C amino protons, which only allows the AN1-CH41 single hydrogen bond pair (Figure 1) with no possibility of dynamics. The A7C8* pair described herein does not show this distinctive NOE and the NMR data is consistent with the possibility that the A7H61-C8*O2 and A7N1-C8*H41 pair (Figure 1) could be rapidly exchanging with each other. Both the A7C8* pair and the AC pair in the loop, 5'GAAGCC/3'CAGGAG, from Lerman et. al.²⁷ are on the side opposite of the cross-strand stacked pair in the 3RRs motif, suggesting the difference is not influenced by the directionality of the 3RRs motif. Although the structure reported here has the A7C8* pair stacked on top an AG rather than a GG pair as previously reported, inspection suggests there is no stacking interaction of AC on GG that would favor one AC conformation over the other.

While no NMR spectra discriminate between the two possible AC pairs, modeling suggests the AN1-CH41 tandem AC pairs are less favorable because all four relatively positive amino groups in the 5'AC/3'CA tandem pair are in a very small volume (Figure S2). In contrast, the NMR model with the AH61-CO2 single hydrogen bond 5'AC/3'CA tandem has the positive 5'A amino group positioned directly over the relatively negative 3'C carbonyl group, giving a favorable electrostatic stacking interaction (Figure 7). Furthermore, when these AC pairs stack on the adjacent sheared GA pair, the same interaction occurs between the C amino group and the G carbonyl oxygen (Figure 7). The quantum calculations are consistent with the expectation that the differences in overlap of functional groups could favor the structures shown in Figures 5 and 7. Similar overlap between amino and carbonyl

groups is also seen in the 5'AU/3'GA nearest neighbor at the stem/loop interface (Figure 7D). A UV melting study of (5'GCAUGCU)₂ and (5'GCAUG²PU)₂, where ²P is 2-pyrimidinone, which replaces the C amino group with a hydrogen, indicated such amino-carbonyl stacking can be favorable by at least 0.5 kcal/mol for each occurrence.⁶⁷

The x-ray structure of a similar loop from SRP RNA (PDB ID: 1LNT),²⁸ 5'AGCAG/3'GGACU, with 5'CA/3'AC rather than 5'AC/3'CA nearest neighbors shows a *trans* WC/H CA pair with CH41-AN7 and CN3-AH62 hydrogen bonds stacked on a sheared GG pair followed by a *cis* H/S AC pair with a AH62-CO2 hydrogen bond. While the SRP pairing yields a total of three hydrogen bonds for the 5'CA/3'AC tandem pairs, compared to the two formed by the 5'AC/3'CA tandem pairs in (5'CCAGAAACGGAUGGA)₂, it also results in a large backbone distortion. The structures suggest that the hydrogen bonding in AC pairs results from a subtle balance between the strength of hydrogen bonds and sequence dependent stacking interactions. It follows that the 5'AC/3'CA orientation and sheared GA closing pairs may be required for exploration of the AH61-CO2 single hydrogen bond conformation in 5'AC/3'CA.

Comparisons between the NMR model and those automatically generated from sequence alone provide some insight into challenges for prediction of RNA 3D structure. While the *refined* MC-sym model had the best RMSD (2.46 Å) and the same noncanonical hydrogen bonding compared to the NMR model, the MC-sym model generated with only the *relieve* function had the worst RMSD (3.99 Å) (Table 2). This suggests that knowledge based approaches can generate structures in the appropriate energy basin, but that all-atom force fields are required to optimize local 3D structures.

By providing RNA FARFAR with chemical shift data, the structure predicted improved from an RMSD for the loop of 3.71 Å to 3.49 Å (Table 2). Additionally, with no chemical shift data, RNA FARFAR did not predict any of the GA pairs to be sheared as shown by the NMR data, but with chemical shift data, A11-G4* is predicted to be sheared. Allowing more data to be used with the RNA FARFAR program should improve predictions. For example, assigned chemical shifts of exchangeable protons and of ¹³C nuclei can now be predicted through programs such as LARMOR^D.⁶⁸ Adding distance restraints from NOE volumes should also improve modeling.

The various approaches for generating 3D structures produced three different predictions for structures of G opposite A (Table 2). This reflects the range of determined structures that contain a G opposite an A in loops. For example, two GA pairs in tandem can be imino or sheared depending on the orientation of adjacent GC base pairs.^{25, 26} Moreover, two GA pairs in tandem can be in slow exchange between imino and sheared conformations when flanked by iGiC pairs.²¹ Clearly, the energetics of stacking and hydrogen bonding are closely balanced as also suggested for CA pairs. The different types of potential GA pairs (Figure 1) provide different availability of functional groups and dramatically different backbone conformations.^{25, 26, 69, 70} Evidently, accurate prediction of 3D structures of RNA loops will require quite accurate approximations of the energetics of stacking and hydrogen bonding.

Surprisingly, the 3D predictions of RNAComposer depended on the hairpin loop sequence added (Table 2). This suggests that more than one hairpin loop sequence should be tested when RNAComposer is used to model a duplex.

Commonly used thermodynamic parameters^{8, 62} predict that 5'CCAGAAACGGAUGGA would fold into a hairpin at the conditions used for NMR (Figure S3) rather than the duplex observed. Presumably, the observation of duplex reflects the favorable G° for the 3RRs motif,^{22, 27, 35} which has yet to be fully explored. The 5'CCAGAAACGGAUGGA sequence was originally designed to mimic a short hairpin loop from the pseudoknotted structure of the 3' splice site (3PSS) of Influenza A segment 7 mRNA (nt 720-733).^{71, 72} Rearrangement of this hairpin is responsible for a transition between a pseudoknot and a different hairpin conformation that unveils the 3PSS. We speculate that because segment 7 mRNA is expressed late in the infection cycle,⁷³ a high concentration of it could build up in the nucleus of the host cell, allowing interactions between two mRNAs to form the duplex reported here. Such duplex formation could have biological significance.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Minimizations by simulated annealing and QM energy calculations were performed with computational resources provided by the University of Rochester Center for Integrated Research Computing (CIRC).

We thank Prof. Ryszard Adamiak for suggesting testing RNAComposer with several hairpin loop sequences.

We thank David Kiemle for his help and advice collecting spectra from the 800 MHz spectrometer.

Funding

This work was supported by National Institutes of Health grant GM22939 and National Institutes of Health shared instrumentation grant 1S10 OD012254.

References

1. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 2009; 10:57–63. [PubMed: 19015660]
2. Ding Y, Tang Y, Kwok CK, Zhang Y, Bevilacqua PC, Assmann SM. In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature.* 2014; 505:696–700. [PubMed: 24270811]
3. Luck R, Steger G, Riesner D. Thermodynamic prediction of conserved secondary structure: application to the RRE element of HIV, the tRNA-like element of CMV and the mRNA of prion protein. *J. Mol. Biol.* 1996; 258:813–826. [PubMed: 8637012]
4. Hofacker IL, Fekete M, Stadler PF. Secondary structure prediction for aligned RNA sequences. *J. Mol. Biol.* 2002; 319:1059–1066. [PubMed: 12079347]
5. Mathews DH, Turner DH. Dynalign: an algorithm for finding the secondary structure common to two RNA sequences. *J. Mol. Biol.* 2002; 317:191–203. [PubMed: 11902836]
6. Mathews DH. Using the RNAstructure software package to predict conserved RNA structures. *Curr. Protoc. Bioinformatics.* 2014; 46:12.14.11–22.
7. Fu Y, Sharma G, Mathews DH. Dynalign II: common secondary structure prediction for RNA homologs with domain insertions. *Nucleic Acids Res.* 2014; 42:13939–13948. [PubMed: 25416799]

8. Mathews DH, Disney MD, Childs JL, Schroeder SJ, Zuker M, Turner DH. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl. Acad. Sci. U.S.A.* 2004; 101:7287–7292. [PubMed: 15123812]
9. Deigan KE, Li TW, Mathews DH, Weeks KM. Accurate SHAPE-directed RNA structure determination. *Proc. Natl. Acad. Sci. U.S.A.* 2009; 106:97–102. [PubMed: 19109441]
10. Hajdin CE, Bellaousov S, Huggins W, Leonard CW, Mathews DH, Weeks KM. Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots. *Proc. Natl. Acad. Sci. U.S.A.* 2013; 110:5498–5503. [PubMed: 23503844]
11. Hart JM, Kennedy SD, Mathews DH, Turner DH. NMR-assisted prediction of RNA secondary structure: identification of a probable pseudoknot in the coding region of an R2 retrotransposon. *J. Am. Chem. Soc.* 2008; 130:10233–10239. [PubMed: 18613678]
12. Chen JL, Bellaousov S, Tubbs JD, Kennedy SD, Lopez MJ, Mathews DH, Turner DH. Nuclear magnetic resonance-assisted prediction of secondary structure for RNA: incorporation of direction-dependent chemical shift constraints. *Biochemistry.* 2015; 54:6769–6782. [PubMed: 26451676]
13. Kierzek E, Christensen SM, Eickbush TH, Kierzek R, Turner DH, Moss WN. Secondary structures for 5' regions of R2 retrotransposon RNAs reveal a novel conserved pseudoknot and regions that evolve under different constraints. *J. Mol. Biol.* 2009; 390:428–442. [PubMed: 19397915]
14. Mathews DH, Moss WN, Turner DH. Folding and finding RNA secondary structure. *Cold Spring Harbor Perspect. Biol.* 2010; 2:a003665.
15. Moore PB. Structural motifs in RNA. *Annu. Rev. Biochem.* 1999; 68:287–300. [PubMed: 10872451]
16. Doherty EA, Doudna JA. Ribozyme structures and mechanisms. *Annu. Rev. Biophys. Biomol. Struct.* 2001; 30:457–475. [PubMed: 11441810]
17. Cruz JA, Blanchet MF, Boniecki M, Bujnicki JM, Chen SJ, Cao S, Das R, Ding F, Dokholyan NV, Flores SC, Huang L, Lavender CA, Lisi V, Major F, Mikolajczak K, Patel DJ, Philips A, Puton T, Santalucia J, Sijenyi F, Hermann T, Rother K, Rother M, Serganov A, Skorupski M, Soltysinski T, Sripakdeevong P, Tuszynska I, Weeks KM, Waldsich C, Wildauer M, Leontis NB, Westhof E. RNA-Puzzles: a CASP-like evaluation of RNA three-dimensional structure prediction. *RNA.* 2012; 18:610–625. [PubMed: 22361291]
18. Miao Z, Adamiak RW, Blanchet MF, Boniecki M, Bujnicki JM, Chen SJ, Cheng C, Chojnowski G, Chou FC, Cordero P, Cruz JA, Ferre-D'Amare AR, Das R, Ding F, Dokholyan NV, Dunin-Horkawicz S, Kladwang W, Krokhotin A, Lach G, Magnus M, Major F, Mann TH, Masquida B, Matelska D, Meyer M, Peselis A, Popena M, Purzycka KJ, Serganov A, Stasiewicz J, Szachniuk M, Tandon A, Tian S, Wang J, Xiao Y, Xu X, Zhang J, Zhao P, Zok T, Westhof E. RNA-Puzzles Round II: assessment of RNA structure prediction programs applied to three large RNA structures. *RNA.* 2015; 21:1066–1084. [PubMed: 25883046]
19. Miao Z, Adamiak RW, Antczak M, Batey RT, Becka AJ, Biesiada M, Boniecki MJ, Bujnicki JM, Chen SJ, Cheng CY, Chou FC, Ferre-D'Amare AR, Das R, Dawson WK, Ding F, Dokholyan NV, Dunin-Horkawicz S, Geniesse C, Kappel K, Kladwang W, Krokhotin A, Lach GE, Major F, Mann TH, Magnus M, Pachulska-Wieczorek K, Patel DJ, Piccirilli JA, Popena M, Purzycka KJ, Ren A, Rice GM, Santalucia J Jr, Sarzynska J, Szachniuk M, Tandon A, Trausch JJ, Tian S, Wang J, Weeks KM, Williams B 2nd, Xiao Y, Xu X, Zhang D, Zok T, Westhof E. RNA-Puzzles Round III: 3D RNA structure prediction of five riboswitches and one ribozyme. *RNA.* 2017; 23:655–672. [PubMed: 28138060]
20. Leontis NB, Stombaugh J, Westhof E. The non-Watson-Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res.* 2002; 30:3497–3531. [PubMed: 12177293]
21. Chen G, Kierzek R, Yildirim I, Krugh TR, Turner DH, Kennedy SD. Stacking effects on local structure in RNA: changes in the structure of tandem GA pairs when flanking GC pairs are replaced by isoG-isoC pairs. *J. Phys. Chem. B.* 2007; 111:6718–6727. [PubMed: 17411085]
22. Chen G, Kennedy SD, Turner DH. A CA(+) pair adjacent to a sheared GA or AA pair stabilizes size-symmetric RNA internal loops. *Biochemistry.* 2009; 48:5738–5752. [PubMed: 19485416]
23. Hammond NB, Tolbert BS, Kierzek R, Turner DH, Kennedy SD. RNA internal loops with tandem AG pairs: the structure of the 5'GAGU/3'UGAG loop can be dramatically different from others, including 5'AAGU/3'UGAA. *Biochemistry.* 2010; 49:5817–5827. [PubMed: 20481618]

24. Tolbert BS, Kennedy SD, Schroeder SJ, Krugh TR, Turner DH. NMR structures of (rGCUGAGGCU)₂ and (rCGGAUGCU)₂: probing the structural features that shape the thermodynamic stability of GA pairs. *Biochemistry*. 2007; 46:1511–1522. [PubMed: 17279616]
25. Santa Lucia J, Turner DH. Structure of (rGGCGAGCC)₂ in solution from NMR and restrained molecular dynamics. *Biochemistry*. 1993; 32:12612–12623. [PubMed: 8251479]
26. Wu M, Turner DH. Solution structure of (rCGGGACGC)₂ by two-dimensional NMR and the iterative relaxation matrix approach. *Biochemistry*. 1996; 35:9677–9689. [PubMed: 8703939]
27. Lerman YV, Kennedy SD, Shankar N, Parisien M, Major F, Turner DH. NMR structure of a 4 × 4 nucleotide RNA internal loop from an R2 retrotransposon: identification of a three purine-purine sheared pair motif and comparison to MC-SYM predictions. *RNA*. 2011; 17:1664–1677. [PubMed: 21778280]
28. Deng J, Xiong Y, Pan B, Sundaralingam M. Structure of an RNA dodecamer containing a fragment from SRP domain IV of Escherichia coli. *Acta Crystallogr, Sect. D: Struct. Biol.* 2003; 59:1004–1011.
29. Cash DD, Feigon J. Structure and folding of the Tetrahymena telomerase RNA pseudoknot. *Nucleic Acids Res.* 2017; 45:482–495. [PubMed: 27899638]
30. Abu Almakarem AS, Petrov AI, Stombaugh J, Zirbel CL, Leontis NB. Comprehensive survey and geometric classification of base triples in RNA structures. *Nucleic Acids Res.* 2012; 40:1407–1423. [PubMed: 22053086]
31. Wimberly B, Varani G, Tinoco I Jr. The conformation of loop E of eukaryotic 5S ribosomal RNA. *Biochemistry*. 1993; 32:1078–1087. [PubMed: 8424938]
32. Szewczak AA, Moore PB, Chang YL, Wool IG. The conformation of the sarcin/ricin loop from 28S ribosomal RNA. *Proc. Natl. Acad. Sci. U.S.A.* 1993; 90:9581–9585. [PubMed: 8415744]
33. Correll CC, Beneken J, Plantinga MJ, Lubbers M, Chan Y-L. The common and the distinctive features of the bulged-G motif based on a 1.04 Å resolution RNA structure. *Nucleic Acids Res.* 2003; 31:6806–6818. [PubMed: 14627814]
34. Klein DJ, Schmeing TM, Moore PB, Steitz TA. The kink-turn: a new RNA secondary structure motif. *EMBO J.* 2001; 20:4214–4221. [PubMed: 11483524]
35. Chen G, Znosko BM, Kennedy SD, Krugh TR, Turner DH. Solution structure of an RNA internal loop with three consecutive sheared GA pairs. *Biochemistry*. 2005; 44:2845–2856. [PubMed: 15723528]
36. Tang CL, Alexov E, Pyle AM, Honig B. Calculation of pK_as in RNA: on the structural origins and functional roles of protonated nucleotides. *J. Mol. Biol.* 2007; 366:1475–1496. [PubMed: 17223134]
37. Legault P, Pardi A. Unusual dynamics and pK_a shift at the active site of a lead-dependent ribozyme. *J. Am. Chem. Soc.* 1997; 119:6621–6628.
38. Sklenář V, Bax A. Spin-echo water suppression for the generation of pure-phase two-dimensional NMR spectra. *J. Magn. Reson.* (1969–1992). 1987; 74:469–479.
39. Grzesiek S, Bax A. The importance of not saturating water in protein NMR. Application to sensitivity enhancement and NOE measurements. *J. Am. Chem. Soc.* 1993; 115:12593–12594.
40. Piotto M, Saudek V, Sklenar V. Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR.* 1992; 2:661–665. [PubMed: 1490109]
41. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR.* 1995; 6:277–293. [PubMed: 8520220]
42. Lee W, Tonelli M, Markley JL. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics.* 2015; 31:1325–1327. [PubMed: 25505092]
43. Richardson JS, Schneider B, Murray LW, Kapral GJ, Immormino RM, Headd JJ, Richardson DC, Ham D, HersHKovits E, Williams LD, Keating KS, Pyle AM, Micaleff D, Westbrook J, Berman HM. RNA backbone: Consensus all-angle conformers and modular string nomenclature (an RNA Ontology Consortium contribution). *RNA*. 2008; 14:465–481. [PubMed: 18192612]
44. Bansal M, Bhattacharyya D, Ravi B. NUPARM and NUCGEN: software for analysis and generation of sequence dependent nucleic acid structures. *CABIOS, Comput. Appl. Biosci.* 1995; 11:281–287. [PubMed: 7583696]

45. Still WC, Tempczyk A, Hawley RC, Hendrickson T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* 1990; 112:6127–6129.
46. Case, DA., Babin, V., Berryman, JT., Betz, RM., Cai, Q., Cerutti, DS., Cheatham, TEI., Darden, TA., Duke, RE., Gohlke, H., Goetz, AW., Gusarov, S., Homeyer, N., Janowski, P., Kaus, J., Kolossváry, I., Kovalenko, A., Lee, TS., LeGrand, S., Luchko, T., Luo, R., Madej, B., Merz, KM., Paesani, F., Roe, DR., Roitberg, A., Sagui, C., Salomon-Ferrer, R., Seabra, G., Simmerling, CL., Smith, W., Swails, J., Walker, RC., Wang, J., Wolf, RM., Wu, X., Kollman, PA. AMBER 14. University of California; San Francisco: 2014.
47. Schrödinger, LLC. The PyMOL Molecular Graphics System, Version 1.7.4.4. 2015.
48. Sripakdeevong P, Cevic M, Chang AT, Erat MC, Ziegeler M, Zhao Q, Fox GE, Gao X, Kennedy SD, Kierzek R, Nikonowicz EP, Schwalbe H, Sigel RKO, Turner DH, Das R. Structure determination of noncanonical RNA motifs guided by 1H NMR chemical shifts. *Nat. Methods.* 2014; 11:413–416. [PubMed: 24584194]
49. Das R, Karanicolas J, Baker D. Atomic accuracy in predicting and designing noncanonical RNA structure. *Nat. Methods.* 2010; 7:291–294. [PubMed: 20190761]
50. Yesselman JD, Das R. Modeling small noncanonical RNA motifs with the Rosetta FARFAR server. *Methods Mol. Biol.* 2016; 1490:187–198. [PubMed: 27665600]
51. Lyskov S, Chou FC, Conchúir S, Der BS, Drew K, Kuroda D, Xu J, Weitzner BD, Renfrew PD, Sripakdeevong P, Borgo B, Havranek JJ, Kuhlman B, Kortemme T, Bonneau R, Gray JJ, Das R. Serverification of molecular modeling applications: The Rosetta Online Server that Includes Everyone (ROSIE). *PLoS ONE.* 2013; 8
52. Major F, Turcotte M, Gautheret D, Lapalme G, Fillion E, Cedergren R. The combination of symbolic and numerical computation for three-dimensional modeling of RNA. *Science (New York, N.Y.).* 1991; 253:1255–1260.
53. Major F. Building three-dimensional ribonucleic acid structures. *Comput. Sci. Eng.* 2003; 5:44–53.
54. Parisien M, Major F. The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature.* 2008; 452:51–55. [PubMed: 18322526]
55. Popena M, Szachniuk M, Antczak M, Purzycka KJ, Lukasiak P, Bartol N, Blazewicz J, Adamiak RW. Automated 3D structure composition for large RNAs. *Nucleic Acids Res.* 2012; 40:e112–e112. [PubMed: 22539264]
56. Biesiada M, Purzycka KJ, Szachniuk M, Blazewicz J, Adamiak RW. Automated RNA 3D structure prediction with RNAComposer. *Methods Mol. Biol.* 2016; 1490:199–215. [PubMed: 27665601]
57. Parisien M, Cruz JA, Westhof E, Major F. New metrics for comparing and assessing discrepancies between RNA 3D structures and models. *RNA.* 2009; 15:1875–1885. [PubMed: 19710185]
58. Frisch, M., Trucks, G., Schlegel, H., Scuseria, G., Robb, M., Cheeseman, J., Scalmani, G., Barone, V., Mennucci, B., Petersson, G. Gaussian 09, revision D. 01. Gaussian, Inc; Wallingford CT: 2009.
59. Grimme S, Antony J, Ehrlich S, Krieg H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* 2010; 132:154104. [PubMed: 20423165]
60. Zhao Y, Truhlar DG. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* 2008; 120:215–241.
61. Kruse H, Mladek A, Gkionis K, Hansen A, Grimme S, Sponer J. Quantum chemical benchmark study on 46 RNA backbone families using a dinucleotide unit. *J. Chem. Theory Comput.* 2015; 11:4972–4991. [PubMed: 26574283]
62. Turner DH, Mathews DH. NNDB: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. *Nucleic Acids Res.* 2010; 38:D280–D282. [PubMed: 19880381]
63. Cromsig JA, Hilbers CW, Wijmenga SS. Prediction of proton chemical shifts in RNA. Their use in structure refinement and validation. *J. Biomol. NMR.* 2001; 21:11–29. [PubMed: 11693565]
64. Morgado CA, Svozil D, Turner DH, Sponer J. Understanding the role of base stacking in nucleic acids. MD and QM analysis of tandem GA base pairs in RNA duplexes. *Phys. Chem. Chem. Phys.* 2012; 14:12580–12591. [PubMed: 22722325]

65. Yildirim I, Stern HA, Sponer J, Spackova N, Turner DH. Effects of restrained sampling space and nonplanar amino groups on free-energy predictions for RNA with imino and sheared tandem GA base pairs flanked by GC, CG, iGiC or iCiG base pairs. *J. Chem. Theory Comput.* 2009; 5:2088–2100. [PubMed: 20090924]
66. Aytenfisu A, Spasic A, Seetin MG, Serafini J, Mathews DH. Modified amber force field correctly models the conformational preference for tandem GA pairs in RNA. *J. Chem. Theory Comput.* 2014; 10:1292–1301. [PubMed: 24803859]
67. Burkard ME, Kierzek R, Turner DH. Thermodynamics of unpaired terminal nucleotides on short RNA helices correlates with stacking at helix termini in larger RNAs. *J. Mol. Biol.* 1999; 290:967–982. [PubMed: 10438596]
68. Frank AT, Law SM, Brooks CL. A simple and fast approach for predicting ¹H and ¹³C chemical shifts: toward chemical shift-guided simulations of RNA. *J. Phys. Chem. B.* 2014; 118:12168–12175. [PubMed: 25255209]
69. Kennedy SD, Kierzek R, Turner DH. Novel conformation of an RNA structural switch. *Biochemistry.* 2012; 51:9257–9259. [PubMed: 23134175]
70. Turner DH. Fundamental interactions in RNA: Questions answered and remaining. *Biopolymers.* 2013; 99:1097–1104. [PubMed: 23733402]
71. Moss WN, Priore SF, Turner DH. Identification of potential conserved RNA secondary structure throughout influenza A coding regions. *RNA.* 2011; 17:991–1011. [PubMed: 21536710]
72. Moss WN, Dela-Moss LI, Kierzek E, Kierzek R, Priore SF, Turner DH. The 3' splice site of influenza A segment 7 mRNA can exist in two conformations: a pseudoknot and a hairpin. *PLoS ONE.* 2012; 7:e38323. [PubMed: 22685560]
73. Lamb, RA., Krug, RM. Orthomyxoviridae: The Viruses and Their Replication. In: Knipe, DM., Howley, PM., editors. *Fundamental Virology*. 4. Lippincott Williams & Wilkins; Philadelphia: 2001. p. 749-751.

GA cis WC/WC (imino pair)

AC cis WC/WC (AN1-CH41)

GA cis S/H (sheared pair)

AC cis WC/S (AH61-CO2)

Figure 1.

A selection of conformations of AC and GA pairs discussed in this paper. Abbreviations are: WC (Watson-Crick), H (Hoogsteen), S (Sugar edge).²⁰

Figure 2.

Imino proton spectra for $(5'CCAGAAACGGAUGGA)_2$ at concentrations of 3.1 mM, 150 μ M, and 15 μ M, at 0 °C. Water was suppressed with a 1-1-spin-echo pulse. The dotted lines show the chemical shifts of the imino protons in the hairpin conformation.

Figure 3.
Two dimensional NOESY spectrum of $(5' \text{CCAGAAACGGAUGGA})_2$ in 95:5 (v/v) $\text{H}_2\text{O}/\text{D}_2\text{O}$ buffer at 0°C with a 125 msec mixing time.

Figure 4.

Two dimensional NOESY spectrum in 100% D₂O at 10 °C and a mixing time of 400 msec showing the H8/H6/H2 to H5/H1' region. C and U H5-H6 cross peaks are labeled in red.

A)

B)

Figure 5.

(A) Representative model of a portion of the internal loop from the duplex, (5'CCAGAAACGGAUGGA)₂, showing the 3RRs motif flanked by an AC and a Watson-Crick AU pair. Cross-strand stacking can be seen between the G4-A11* and A5-G10* pairs. Additionally, the backbone kink between G10* and A11* is consistent with the downfield chemical shift of the A11* phosphorus. (B) The structures of the pairs involved in the 3RRs motif and the AC pair for the NMR model. Pairs are shown looking down the long axis of the duplex with no rotation with respect to the representative model. Potential hydrogen bond donors and acceptors connected, respectively, with black and red dashed lines are within 2.5 Å and > 2.5 Å in the average of the 33 accepted structures. Note that only G10/G10* nucleotides have a C2'-endo ribose, consistent with previously reported structures of 3RRs motifs.^{27, 35}

Chemical Shift (ppm)

Figure 6.

Observed (solid) and predicted (dashed) chemical shifts of each atom type plotted versus residue number. Chemical shifts were predicted with NUCHEMICS⁶³ using the ensemble of 33 accepted structures and averaged together.

Figure 7.

(A) Representative model of the tandem pairs, 5' AC/3' CA from the internal loop of (5'CCAGAAACGGAUGGA)₂. The AC pairs are described as cis WC/S (Figure 1) using the Leontis-Westhof notation.²⁰ (B) Rotation of the representative model by 90° about the short axis, looking down the long axis of the duplex. (C) Stacking of A7*-C8 pair on the A6*-G9 pair looking down the long axis of the duplex. (D) Stacking of A11-G4* pair on the U12-A3* closing base pair look down the long axis of the duplex.

Table 1

NMR statistics for the 33 accepted structures generated by restrained simulated annealing for the duplex, (5' CCAGAAACGGAUGGA)₂.

Number of restraints	
All distance restraints	186
All NOE restraints	170
Intranucleotide	82
Internucleotide	88
Long Range ($ i-j \geq 2$)	29
Hydrogen bond	16
Backbone dihedral	106
RMSD of experimental restraints	
Distances (Å)	6.9×10^{-4}
Dihedral angles (deg)	2.9×10^{-4}
Heavy atom RMSD for structures (Å)	
All residues	1.54 ± 0.53
8x8 internal loop (residues 4–11 and 19–26)	1.12 ± 0.40
3RRs motif 1 (residues 4–6 and 24–26)	0.53 ± 0.16
3RRs motif 2 (residues 9–11 and 19–21)	0.50 ± 0.15
Tandem AC pairs (residues 7, 8, 22, and 23)	0.83 ± 0.30
Number of violations	
$> 0.1 \text{ Å}$	0
$> 5^\circ$	0

Table 2

Comparison of base pair classifications between the NMR structure and representative structures from automated 3D prediction programs^a.

Sequence	Pairing Partner	NMR Structure	RNA FAREAR (no CS)	RNA FAREAR (with CS)	MC-sym (relieved)	MC-sym (refined)	RNAComposer A ₄ loop	RNAComposer A ₆ loop	RNAComposer GAAA loop
C1	G14*	WC/WC	WC/WC	WC/WC	WC/WC	WC/WC	WC/WC	WC/WC	WC/WC
C2	G13*	WC/WC	WC/WC	WC/WC	WC/WC	WC/WC	WC/WC	WC/WC	WC/WC
A3	U12*	WC/WC	WC/WC	WC/WC	WC/WC	WC/WC	WC/WC	WC/WC	WC/WC
G4	A11*	Sheared	Imino	Imino	None ^a	Sheared	Sheared	Sheared	Sheared
A5	G10*	Sheared	Imino	Imino	None ^a	Sheared	None ^{a,c}	None ^{a,c}	None ^{a,c}
A6	G9*	Sheared	Imino	Imino	Sheared	Sheared	None ^a	None ^a	Sheared
A7 ^d	C8* ^d	AH61-CO2	AN1-CH41	AN1-CH41	AH61-CO2	AN1-CH41	AH61-CO2	None ^a	None ^a
C8 ^d	A7* ^d	CO2-AH61	CH41-AN1	CH41-AN1	None ^a	CO2-AH61	CH41-AN1	None ^a	CO2-AH61
G9	A6*	Sheared	Imino ^b	None ^a	Sheared	Sheared	Sheared	None ^a	None ^a
G10	A5*	Sheared	Imino	None ^a	Sheared	Sheared	Sheared ^b	None ^a	None ^a
A11	G4*	Sheared	Imino	Sheared	Sheared	Sheared	Sheared	None ^a	None ^a
U12	A3*	WC/WC	WC/WC	WC/WC	WC/WC	WC/WC	WC/WC	WC/WC	WC/WC
G13	C2*	WC/WC	WC/WC	WC/WC	WC/WC	WC/WC	WC/WC	WC/WC	WC/WC
G14	C1*	WC/WC	WC/WC	WC/WC	WC/WC	WC/WC	WC/WC	WC/WC	WC/WC
All-Atom Loop ^e RMSD (Å)			3.71	3.49	3.99	2.46	2.69	3.37	3.38
Deformation Index			6.74	5.44	5.28	3.26	4.12	5.48	5.60

^aHydrogen bonds were considered if the distance between heavy atom was ≤ 4.0 Å. If no hydrogen bonds were present between a residue and its pairing partner, this pair is classified as None. Watson-Crick pairs are denoted as WC/WC and noncanonical pairs classified as shown in Figure 1.

^bOnly one of the two hydrogen bonds are within 4.0 Å but the orientation of the other donor and acceptor suggest it is likely to form.

^cPairing partners are stacked on each other rather than in the same plane.

^dThe model generated by simulated annealing using the AMBER 14 force field and NMR distance restraints had an AH61-CO2 hydrogen bond, but the NMR data do not rule out an AN1-CH41 hydrogen bonded structure.

^e“Loop” was defined as nucleotides G4 through A11 and G4* through A11*.