Published in final edited form as:

J Phys Chem B. 2015 October 22; 119(42): 13252–13261. doi:10.1021/acs.jpcc.5b06970.

Structures and Energetics of Four Adjacent G·U Pairs That Stabilize an RNA Helix

Xiaobo Gu^{1,2}, Blaine H.M. Mooers^{3,4,*}, Leonard M. Thomas¹, Joshua Malone¹, Steven Harris¹, and Susan J. Schroeder^{1,2,*}

¹Department of Chemistry and Biochemistry, University of Oklahoma, Norman, OK, 73019, USA

²Department of Microbiology and Plant Biology, University of Oklahoma, Norman, OK, 73019, USA

³Department of Biochemistry and Molecular Biology, University of Oklahoma Health Sciences Center, Oklahoma City, OK, 73104, USA

⁴Stephenson Cancer Center, University of Oklahoma Health Sciences Center, Oklahoma City, OK, 73104, USA

Abstract

Consecutive G·U base pairs inside RNA helices can be destabilizing while those at the ends of helices are thermodynamically stabilizing. To determine if this paradox could be explained by differences in base stacking, we determined the high-resolution (1.32 Å) crystal structure of (5'-GGUGGCUGUU-3')₂ and studied three sequences with four consecutive terminal G·U pairs by NMR spectroscopy. In the crystal structure of (5'-GGUGGCUGUU-3')₂, the helix is overwound but retains the overall features of A-form RNA. The penultimate base steps at each end of the helix have high base overlap and contribute to the unexpectedly favorable energetic contribution for the 5'-GU-3'/3'-UG-5' motif in this helix position. The balance of base stacking and helical twist contributes to the positional dependence of G·U pair stabilities. The energetic stabilities and similarity to A-form RNA helices suggest that consecutive G·U pairs would be recognized by RNA helix binding proteins, such as Dicer and Ago. Thus, these results will aid future searches for target sites of small RNAs in gene regulation.

INTRODUCTION

Donohue was the first to predict a structure for guanine pairing with thymidine or uracil¹, and Crick was the first to propose that G·U pairs have a function². Since then many important biological roles for G·U pairs have been discovered including roles in the triplet

*To whom correspondence should be addressed. Tel: (405) 271-8300 (B.H.M.M.), (405) 325-3092 (S.J.S.); Fax: (405) 271-3910 (B.H.M.M.), (405, 325-6111 (S.J.S); blaine-mooers@ouhsc.edu; susan.schroeder@ou.edu.
Present Address: Steven Harris, Department of Molecular Biology and Chemistry, Christopher Newport University, Newport News, VA, 23606, USA

SUPPORTING INFORMATION DESCRIPTION

The Supporting Information includes 3 tables and 4 figures containing the details of helical parameters in the crystal structure, crystallography models and metal ion binding, proton resonance assignments, NMR experimental parameters, and NMR spectra. This information is available free of charge via the Internet at <http://pubs.acs.org>

code during protein translation and tRNA synthetase activity^{3, 4}, in HIV genomic riboswitches⁵, and at the active site of RNA enzymes, such as group I⁶ and II introns⁷ and hepatitis delta virus⁸. Consecutive terminal G·U pairs occur in miRNA-mRNA interactions in cancer gene regulation⁹ and models of encapsidated Satellite Tobacco Mosaic Virus RNA¹⁰. Thus reliably predicting G·U pairing would be advantageous for deciphering RNA function from new transcriptome data. However, while Watson-Crick nucleotide pairing can be predicted reasonably accurately at the sequence level, predictions of pairing between guanines and uracils (G·U pairs) is much more difficult. Tandem G·U pairs inside a helix can be thermodynamically destabilizing, with the notable exception of the sequence 5'-GGUC-3'/3'-CUGG-5'¹¹⁻¹⁴. In contrast, all consecutive terminal G·U pairs are stabilizing (Table 1), and the amount of stability depends more on sequence and context than other consecutive terminal mismatches in RNA duplexes¹⁵. Thus the same idiosyncratic physical properties and non-nearest neighbor effects of G·U pairs that facilitate molecular recognition and specificity in biological function frustrate RNA structure prediction.

The previous unfavorable energy terms for consecutive G·U pairs in RNA structure prediction^{16, 17} may have underestimated the occurrence of this motif in predictions of natural RNA sequences. For example, some debate exists about whether to include more than one consecutive G·U pair in predictions of miRNA-mRNA interactions^{18, 19}. Recently, however, a validated virus-encoded small RNA-host mRNA interaction contains consecutive terminal G·U pairs and explains the observed phenotype of yellow spots on leaves²⁰. The results presented here show that up to four consecutive terminal G·U pairs can adopt a conformation similar to A-form that would be recognized by RNA-binding proteins, such as Dicer, Drosha, and Ago, and further support the case that consecutive terminal G·U pairs should be included in predictions of small RNA-mRNA interactions.

Terminal G·U wobble base pairs have been shown to make favorable energetic contributions to helix stability. Thus, we proposed a hypothesis that an RNA decamer with four adjacent G·U wobble base pairs on both ends will form stable base stacking interactions in an RNA double helix despite 8 of the 10 base pairs being non-Watson Crick base pairs. To test this hypothesis we did crystallographic and NMR studies. The crystallographic, NMR, and thermodynamic results show that four consecutive G·U pairs can form energetically stable, A-form-like RNA helices that could be recognized by RNA binding proteins. The base stacking and helical twist contribute to the sequence-dependent thermodynamic stabilities of terminal consecutive G·U pairs. The main goal of the crystallographic and NMR studies is to establish a structural basis for the observed thermodynamic stabilities of consecutive terminal G·U pairs. In addition, our results increase the diversity of known RNA structural motifs, improve RNA structure predictions, help identify functional sites in new noncoding RNAs, and help improve predictions for target sites of small RNAs.

EXPERIMENTAL METHODS

RNA Preparation

Oligoribonucleotides were purchased from Dharmacon (Thermo Fisher Scientific Inc.) and deblocked with acetic acid to remove the protecting group on its 2'-hydroxyl as per the manufacturer's instructions. The RNA oligonucleotides were lyophilized and resuspended in

Milli-Q water. 9.7 mM RNA samples for crystallography were used for screening and crystallization without further purification. For NMR studies, the RNA samples were further purified on a PD-10 sephadex G25 gel exclusion column to remove residual salts. Sep-pak C-18 reverse-phase columns were also used to remove salts from some RNA NMR samples. The RNA was lyophilized, resuspended in Milli-Q water again, and dialyzed in 0.02 mM Na₂ EDTA and then dilute NMR buffer using a 1000 molecular weight cutoff membrane. The final NMR samples contained 2 mM RNA, 10 mM NaCl, 10 mM Na₃PO₄, 0.2 mM EDTA, pH 6 with or without Co(NH)₆Cl₃. RNA concentrations were based on high-temperature UV absorbance measurements. Optical melting experiments were conducted as previously described¹⁵. All RNA samples for optical melting, NMR, and crystallography were greater than 90 % pure as assessed by ³²P labeling and a single band on a 20% polyacrylamide gel.

Crystallization and cryoprotection

Crystals were grown by vapor diffusion using the hanging drop method. Optimized crystallization conditions for the sequence 5'-GGUGGCUGUU-3' were 4.6 mM RNA, 26.2% (v/v) 2-methyl-2,4-pentanediol (MPD), 40 mM lithium chloride, 30 mM cobalt hexammine chloride and 40 mM sodium cacodylate pH 5.5. Crystals were not obtained in the absence of cobalt hexammine. Crystallization drops were assembled by mixing 1 μ L of RNA solution and 1 μ L of reservoir solution. Plate-shaped crystals with dimensions of 100 $\times$ 100 $\times$ 15 μ m appeared within two weeks. To cryoprotect the crystals, they were transferred with Rayon cryoloops through synthetic reservoir solutions containing 25, 27, 29, 31, and 33% (v/v) MPD and then submerged in liquid nitrogen.

Diffraction data collection

The diffraction data were collected at 100 K with synchrotron radiation and an ADSC Quantum 315r detector at the Stanford Synchrotron Radiation Lightsource (SSRL) beam line 7-1. One crystal diffracted to 1.32 Å; its diffraction data were used for refinement. The diffraction data were collected at three distances starting with the longest distance and attenuated beam intensity to avoid saturation of the detector with the strongest reflections and to measure the lowest resolution reflections before radiation damage became significant. Another crystal diffracted X-rays to 1.55 Å and was used to collect diffraction data for a multi-wavelength anomalous diffraction (MAD) phasing experiment at the cobalt absorption edge. Diffraction data were collected in the inverse beam mode at three wavelengths (inflection, peak, and high energy remote). The diffraction data were indexed and integrated using XDS²¹. The data were then scaled with SCALA²² and negative intensities were treated with TRUNCATE²³.

Crystal structure determination

The molecular structure was determined with a MAD phasing experiment using cobalt as the anomalous scatter. The program SHELXD was used to find 8 cobalt hexammine sites²⁴. The first seven sites were used with SHELXE to refine the positions of the heavy atoms before phase extension by density modification to 1.32 Å with the data that were collected with 0.97 Å radiation and that were subsequently used in structure refinement²⁴. Four additional sites were found and added to the heavy atom substructure. The "free lunch"

algorithm was used with extrapolated data to 0.9 Å during density modification¹⁸. The estimated mean figure of merit was 0.758 and the pseudo-free correlation coefficient was 77.16% with data to 1.32 Å. Initial models of the RNA were built automatically with the CCP4 program NAUTILUS²⁰ and were then completed and corrected manually using the molecular graphics package COOT²⁵ and the CRANE plugin²⁶. The structure was refined with PHENIX using the Bijvoet pairs (F^+ and F^-) and the maximum likelihood target²⁷. The backbone torsion angles were corrected with ERRASER²⁸. The sites of bound cobalt hexammines were validated with an electron density map made by using the F_A values from the MAD experiment and the program ANODE²⁹. This map and an independent map made with the anomalous differences in the 1.32 Å dataset confirmed the presence of twelve cobalt hexammines. The structure was eventually refined with anisotropic atomic displacement parameters for the nonhydrogen RNA atoms, the cobalt hexamine ions, and chloride anions while isotropic temperature factors were retained for the water oxygen atoms. The occupancies of the chlorides and cobalt hexamine ions were refined while the occupancies of the waters were not refined.

Crystal structure validation and analysis

The distribution of the anisotropic temperature factors were checked the PARVATI webserver³⁰. Close contacts, bad geometry, and backbone torsion angles were checked with MOLPROBITY³¹ and RNABC³². The helical parameters were measured with 3DNA³³ and Curves+³⁴. The local helical parameters are defined with regard to base steps and without regard to a global axis. PDB2PQR was used to add hydrogens and assign partial charges and atomic radii³⁵ using the AMBER99 forcefield³². Electrostatic surfaces were calculated with APBS version 1.4.1³⁶ using the nonlinear Poisson-Boltzmann equation, 150 mM NaCl, a solute dielectric constant of 2.0 and a solvent dielectric of 78.5. The molecular graphics figures were made with PyMOL (Schrödinger, LLC). The statistics for data collection and structure refinement are included in Tables S1 and S2.

NMR Spectroscopy

For the (5'-GGUGGCUGUU-3')₂ NMR sample, the H₂O 2D NOESY and ¹⁵N HMQC spectra were collected on a Varian 800 MHz spectrometer with a cryoprobe and a 600 MHz spectrometer, respectively, at the National Magnetic Resonance Facility in Madison, Wisconsin (NMRFAM). Data processing was performed with NMRPipe³⁷ and SPARKY³⁸. For 5'-GUGUCGGUGU-3', SNOESY, COSY, ³¹P HETCOR, ¹³C and ¹⁵N HSQC spectra were collected on a Varian 500 MHz spectrometer using Biopack software³⁹. Data processing and analysis was performed with Varian software and SPARKY³⁸. Detailed experimental parameters are shown in Supporting Information Table 2. No imino proton resonances in the ultimate and penultimate G·U pairs for all three sequences were observed at any temperature. The lack of cross-strand NOEs for the ultimate and penultimate G·U pairs precludes rigorous structure determination by NMR constraints. Thus modelling with ROSIE, which has the ability to incorporate constraints from NMR data, was pursued rather than simulated annealing methods for NMR structure determination.

Computational predictions of RNA duplexes with terminal G-U pairs

Three RNA sequences, 5'-GUGUCGGUGU-3', 5'-UGUGGCUGUG-3', and 5'-GGUGGCUGUU-3' form self-complementary duplexes with four consecutive terminal G-U pairs and were submitted for RNA structure prediction to ROSIE^{40, 41}, MCSYM⁴², and RNA123⁴³. The predictions used no additional experimental data and followed the standard protocols and parameters suggested by each website. The ROSIE program uses the most updated thermodynamic prediction rules for consecutive terminal GU pairs¹⁵ and gave the best results. Because the structures predicted by ROSIE were consistent with all the assigned NOEs and NMR data for the three duplexes without directly including the NMR data as constraints, no further modelling with constraints was necessary. The results of the computational predictions were analyzed prior to knowledge of the crystal structure for the duplex (5'-GGUGGCUGUU-3')₂.

The crystal structure coordinates and diffraction data have been deposited in the Protein Data Bank and have accession code 4PCO. All chemical shift information and structures have been submitted to the Biological Magnetic Resonance Data Bank (Supporting Information Table 2).

RESULTS

The duplex sequences in Table 1 form the basis for the present structural studies. The 5'-GGUGGCUGUU-3' forms a self-complementary, antiparallel helix with four consecutive G-U pairs on each end. The sequence motif (5'-GGUGG-3'/3'-UUGUC-5') has not been previously reported in the Protein Databank. The duplex (5'-GGUGGCUGUU-3')₂ forms one of each possible G-U pair stacking conformations, 5'-UG-3'/3'-GU-5', 5'-GU-3'/3'-UG-5', and 5'-UU-3'/3'-GG-5'. The duplex (5'-GGUGGCUGUU-3')₂ has only the final base pair orientation changed. The duplexes (5'-UGUGGCUGUG-3')₂ and (5'-GUGUCGGUGU-3')₂ have the same nucleotides in reverse 5' to 3' order. Thus, comparisons of these three duplexes provide a structural context for evaluating the sequence variation in the thermodynamic stabilities of terminal G-U pairs.

An RNA duplex with four consecutive terminal G-U pairs on each end can adopt an A-form conformation

A crystal structure was obtained for the self-complementary RNA duplex (5'-GGUGGCUGUU-3')₂ at a resolution of 1.32 Å resolution with R_{free} and R_{work} values of 0.1857 and 0.1569, respectively. The crystallographic and refinement statistics are summarized in Tables 2 and 3. The crystals in question presented an interesting crystallography challenge because there are five RNA strands in the asymmetric unit, rather than the even number that have been anticipated. The slight differences in conformation observed between the helices in the asymmetric unit of the crystal are consistent with NMR data obtained from this molecule which show that its ultimate and penultimate base pairs are structurally dynamic. The G-U pairs in these helices are all in the cis Watson-Crick-Watson-Crick edge conformation (Figure 1). The overall geometry of these helices is that of a slightly overwound A-form helix, which aligns well with a model of an all AU A-form RNA helix (Figure 2, Supporting Information Figure S1 and Table S1).

Base stacking and helical twist contribute to thermodynamic stabilities

The hydrogen bonding in wobble pairs, base stacking between adjacent base pairs, and variations in helical twist contribute to the thermodynamic stability of the duplex (5'-GGUGGCUGUU-3')₂. The cross-strand guanine stacking contributes to the sequence dependence of the thermodynamic stabilities for terminal G·U pairs. Base stacking and helical twist angles are inherently correlated. Overwinding (large helical twist) leads to cross-strand overlap, while underwinding leads to intrastrand overlap. The base stacking contributes to favorable free energies, while underwound or overwound helical twist angles that cause large deviations from the mean value of 32.2° may destabilize a base step. The base stacking and helical twist between consecutive G·U pairs is first analyzed in the crystal structure of (5'-GGUGGCUGUU-3')₂ and compared with the NMR and thermodynamic data. Then the NMR-consistent models for all three duplexes in Figure 1 are analyzed. These structural analyses form the basis for understanding the positional dependence of G·U pair thermodynamic stabilities.

In the crystal structure of (5'-GGUGGCUGUU-3')₂, the first base step 5'-GG-3'/3'-UU-5' has very little base overlap (Figure 3A). The addition of a fourth G·U pair in either orientation, [ie the comparison of the octamer duplex (5'-GUGGCUGU-3')₂ to the decamer duplexes (5'-GGUGGCUGUU-3')₂ and (5'-UGUGGCUGUG-3')₂], does not affect the thermodynamic stability of the duplexes (Table 1)¹⁵. In the NMR spectra, no imino proton resonances are observed for the ultimate G·U pairs, suggesting fast exchange of the imino protons with water and only weak, dynamic hydrogen bonding. Thus, the crystallography and NMR data support the observed thermodynamic data.

The second base step 5'-GU-3'/3'-UG-5' in the (5'-GGUGGCUGUU-3')₂ duplex is underwound with high intrastrand overlap (Figures 3B, 4). This base step sequence has an extra bonus of -0.4 kcal/mol in addition to the -0.5 kcal/mol for G·U pairs in this position of the helix (Table 1)¹⁵. The third base step 5'-UG-3'/3'-GU-5' is highly overwound and has high inter-strand stacking between the guanines that stabilizes the large helical twist while the uracils are unstacked (Figures 3C, 4). The stacking is consistent with -0.7 kcal/mol for G·U pairs in this helix position. The fourth base step 5'-GG-3'/3'-UC-5' is only slightly underwound and has good intrastrand base overlap (Figures 3D, 4). The central base step of two Watson-Crick base pairs 5'-GC-3'/3'-CG-5' has an A-form twist angle with high intrastrand base overlap and almost no interstrand overlap (Figures 3E, 4). The thermodynamic parameters for the nearest neighbor terms for the fourth and central base stacks are -1.9 and -3.4 kcal/mol, respectively. Thus, analysis of the duplex structure provides a clear explanation for the thermodynamic stability.

NMR spectroscopy data show A-form helices and G·U pair formation for three duplexes with different sequences containing four consecutive terminal G·U pairs

The solution NMR data for (5'-GGUGGCUGUU-3')₂ validates both the crystal structure and the ROSIE (Rosetta Online Server that Includes Everyone) models⁴⁰. All chemical shift information and structures have been submitted to the Biological Magnetic Resonance Data Bank (Supporting Information Table 2). Eight internucleotide cross-strand NOEs and 15 internucleotide sequential NOEs are all correctly predicted by the ROSIE model.

Specifically, strong NOEs between the imino protons of the guanine and uracil (i.e., G2H1-U9H3) confirm the wobble pairs. NOEs between imino protons of sequential bases (i.e., G2H1-U3H3), intrastrand NOEs (i.e., G2H1-U3H1'), and cross-strand NOEs to ribose protons (i.e., G5H1-U7H1') define the base stacking between the consecutive pairs (Figure 5). The changes in the NMR spectrum when adding cobalt hexammine and multiple NOES to the amino protons of the cobalt hexammine ions are consistent with the multiple metal ion binding sites observed crystallographically (Supporting Information Figure S2). The conformational dynamics observed in the NMR spectroscopy are consistent with the structural heterogeneity in the three slightly different helices in the asymmetric unit of the crystal structure.

The sequences (5'-GUGUCGGUGU-3')₂ and (5'-UGUGGCUGUG-3')₂ were studied by NMR spectroscopy and predicted well by ROSIE. All the NOES observed in H₂O SNOESY experiments for these duplexes were consistent with the models predicted by ROSIE and the formation of A-form helices and thus validate the structure prediction. For the (5'-GUGUCGGUGU-3')₂, A-form-like geometry is confirmed by the imino proton NOEs, the pattern of ribose proton-base aromatic proton NOES, phosphorous chemical shifts, and P-H3' coupling constants (Supporting Information Figures S3 and S4). In the absence of cobalt hexammine ions, dynamic flexibility in the ultimate and penultimate pairs is observed as broad imino proton resonances and H1'-H2' COSY crosspeaks revealing the conformational dynamics between C2'-endo and C3'-endo sugar conformations. The absence of observable imino protons and cross strand NOEs in all three duplexes with or without cobalt hexammine ions limits the ability to experimentally determine by NMR the base stacking in both the ultimate and penultimate GU pairs. While the imino protons in the ultimate base pair are often not observed, the fast exchange that leads to the absence of imino protons in the penultimate pair is unusual. However, all NMR data and ROSIE models confirm the formation of A-form like helices.

The sequences (5'-GUGUCGGUGU-3')₂ and (5'-UGUGGCUGUG-3')₂ are the same nucleotides in reverse 5' → 3' order, but differ by 2 kcal/mol in thermodynamic stabilities (Table 1, Figure 6). Differences in base stacking, especially at the penultimate base step, provide a structural basis for the difference in thermodynamic stabilities. The two most stable helices (5'-UGUGGCUGUG-3')₂ and (5'-GGUGGCUGUU-3')₂ have the extra bonus for the 5'-GU-3'/3'-UG-5' base step in the penultimate position, while the other helix does not have this base step sequence at this helix position. The 5'-GU-3'/3'-UG-5' base step in the penultimate position has good base stacking while the 5'-UG-3'/3'-GU-5' base step has a cross-strand guanine stack that leaves the uracil bases unstacked. Thus, the comparison of the two duplexes provides a structural basis for the thermodynamic bonus parameter for the 5'-UG-3'/3'-GU-5' base step.

In order to test the effects of electrostatics in base stacking interactions, the thermodynamic stability of (5'-GIUGGCUIUU-3')₂ was measured. This duplex has a similar sequence as the crystal structure shown in figures 1–4, with inosine substitutions for G2 and G8. The inosine substitutions at positions G2 and G8 would be predicted to show the largest effect in thermodynamic stability because the amino groups in nucleotides G2 and G8 show the most overlap in base stacking among the 8 G·U pairs in the duplex (Figures 3C, D). Some guanine

amino groups in the G-U pairs show no overlap with the adjacent bases. The free energy (37 °C), enthalpy, entropy, and melting temperature at 1×10^{-4} M are -5.0 kcal/mol, -39.2 kcal/mol, -110.1 eu, and 31.7 °C, respectively. The net change in thermodynamic stability for the inosine substitutions in 4 G-U pairs is 1.0 kcal/mol. This small change in thermodynamic stability for each inosine substitution is consistent with the small amount of overlap between the guanine amino group and the adjacent nucleotide base.

DISCUSSION

G-U rich RNA helix adopts a slightly overwound A-form conformation

Although G-U wobble base pairs are not isosteric with Watson-Crick base pairs, the duplex $(5'-\text{GGUGGCUGUU}-3')_2$ retains an overall A-form-like geometry. Given the structural diversity of G-U pairs in RNA tertiary contacts and active sites of RNA enzymes, the high similarity between a 10-mer helix with 8 G-U pairs and A-form RNA is surprising. The similarity to A-form RNA suggests that consecutive terminal G-U mismatches would still be bound by proteins such as DICER that recognize the overall shape of A-form RNA rather than a particular sequence in miRNA-mRNA interactions. Thus, this motif should be included in bioinformatics searches for the target mRNA sequences of microRNA identified by high throughput sequencing.

Base stacking and helical twist provide an explanation for the thermodynamics stabilities of consecutive terminal GU pairs

Although consecutive internal G-U pairs may destabilize a helix, consecutive terminal G-U pairs are thermodynamically favorable; and the free energy of a G-U pair depends on both sequence and helix position (Table 1)¹⁵. The crystal structure and ROSIE models provide insight into the sequence and position dependence for G-U pair stabilities, why G-U pairs are more stable than other terminal mismatches, and why the additional stability does not continue past three terminal G-U pairs. $(5'-\text{GGUGGCUGUU}-3')_2$ contains an example of each possible orientation of two G-U pairs: $5'-\text{GU}-3'/3'-\text{UG}-5'$, $5'-\text{GG}-3'/3'-\text{UU}-5'$, and $5'-\text{UG}-3'/3'-\text{GU}-5'$. Internal tandem G-U wobble base pairs have been classified into three categories⁴⁴. In $(5'-\text{GGUGGCUGUU}-3')_2$, the base step $5'-\text{UG}-3'/3'-\text{GU}-5'$ has a cross-strand G stack similar to a type I motif (Figure 3C). The penultimate base step $5'-\text{GU}-3'/3'-\text{UG}-5'$ has high base overlap in a type II motif (Figure 3B), and the ultimate base step $5'-\text{UU}-3'/3'-\text{GG}-5'$ shows little stacking in the type III motif (Figure 3A). Thus, the high-resolution crystal structure provides a complete profile for analyzing structure-energetics relationships in tandem G-U pairs.

The sequence-dependent base stacking is one reason why consecutive terminal G-U pairs show more sequence variation than other consecutive terminal mismatches

The stabilities of consecutive terminal mismatches depend only on purine or pyrimidine stacking in the 5' to 3' direction⁴⁵. The stability of a G-U pair depends on its position in the helix, which is a non-nearest neighbor effect. The additional thermodynamic stability for this first stacked terminal G-U pair ranges from -2.3 kcal/mol for $5'-\text{GU}-3'/3'-\text{CG}-5'$ to -0.8 kcal/mol for $5'-\text{AG}-3'/3'-\text{UU}-5'$. (As a point of reference, 1.4 kcal/mol is approximately one order of magnitude in a binding constant at 37 °C.) The next terminal G-U pair adds -0.7

kcal/mol. The third stacked terminal G-U pair adds -0.5 kcal/mol, and an extra bonus of -0.4 kcal/mol is added when the third terminal G-U pair creates a 5'-GU-3'/3'-UG-5' motif. The crystal structure of (5'-GGUGGCUGUU-3')₂ contains the 5'-GU-3'/3'-UG-5' motif in the third nearest neighbor position, which is the penultimate base pair step. The base stacking in the penultimate 5'-GU-3'/3'-UG-5' base step observed in the crystal structure supports the thermodynamic bonus for this motif (Figure 3B). NOEs such as the NOE from the G2H1 proton to both the U3H3 and U3H1' protons are consistent with a highly stacked conformation for 5'-GU-3'/3'-UG-5' (Figure 5) and would not be observed in *A*-form Watson-Crick pair steps. Thus, the NMR data confirms the thermodynamic bonus for this motif.

The balance of base stacking and helix twist contributes to the difference in stabilities of internal and terminal G-U pairs

The ROSIE models for (5'-GGUGGCUGUU-3')₂ and (5'-UGUGGCUGUG-3')₂ both show strong base overlap for the 5'-GU-3'/3'-UG-5' step (Figure 5). In contrast, (5'-GUGUCGGUGU-3')₂ differs from (5'-UGUGGCUGUG-3')₂ in the 5' → 3' order of nucleotides and contains a 5'-UG-3'/3'-UG-5' penultimate base step. The 5'-UG-3'/3'-GU-5' base step shows less base overlap than the 5'-GU-3'/3'-UG-5' motif (Figure 5). The cross-strand G base stacking in the 5'-UG-3'/3'-UG-5' leaves the uridine bases with no base overlap. The difference in thermodynamic stabilities for (5'-GUGUCGGUGU-3')₂ and (5'-UGUGGCUGUG-3')₂ is 2.0 kcal/mol. This key difference in thermodynamic stabilities in the penultimate stacks can account for most of the difference between the stabilities of the two helices. At the end of a helix, the base stacking determines the conformation and energetics of the terminal consecutive G-U pairs without any constraints or strain to resume Watson-Crick *A*-form helices and thus no further penalty for high twist base steps (Figure 4).

For consecutive G-U pairs internal to a Watson-Crick helix, the strain on the backbone for high-twist G-U base pair steps may counteract favorable base stacking. For example, in the duplexes (5'-GGAGUUCC-3')₂ and (5'-GGAUGUCC-3')₂, both internal consecutive G-U pairs and neighbors (underlined) have similar base stacking overlap, but the duplex (5'-GGAGUUCC-3')₂ has two high twist steps (40.0°) between the G-U pair and the adjacent AU pair and a low helical twist between the G-U pairs (24.0°)¹⁴. As a result of both the base stacking and the helical twist, the free energies for the 5'-GU-3'/3'-UG-5' and 5'-UG-3'/3'-GU-5' steps are $+0.7$ and -0.6 kcal/mol, respectively¹³. Similar overtwisting of the helix occurs in the duplex (5'-GAGGUCC-3')₂ compared to the duplex (5'-GAGUCC-3')₂^{46, 47}. Thus the overall energetics for internal tandem G-U pairs is often unfavorable or near zero, while terminal G-U pairs with fewer, less extreme helical twist have favorable energetics.

Inosine substitutions destabilize internal G-U pairs more than terminal G-U pairs. The free energy changes for two inosine substitutions in internal tandem G-U pairs in 5'-CGUG-3'/3'-GUGC-5', 5'-CUGG-3'/3'-GGUC-5', and 5'-CGGG-3'/3'-GUUC-5' are 3.0, 4.8, and 5.2 kcal/mol, respectively⁴⁸ (Note underlined guanine nucleotides are inosine substitution sites). In contrast, the four inosine substitutions in terminal G-U pairs in (5'-

GGUGGCUGUU-3')₂ have a net change of only 1.0 kcal/mol. Thus, internal tandem G·U pairs are more sensitive to subtle electrostatic changes in stacking than terminal G·U pairs. The amount of overlap between the amino group and the adjacent nucleotide base is similar in both internal and terminal G·U pairs. In internal G·U pairs, however, the high twist steps and Watson-Crick pairs on both sides may restrict the base stacking and dynamic motions. In contrast, terminal G·U pairs are structurally dynamic as observed in the NMR data and the heterogeneity of the 5 strands in the asymmetric unit of the crystal structure (Table S1). The fewer high-twist base steps and nucleotide dynamics in terminal G·U pairs may contribute to the thermodynamic stability of this motif.

The crystal structure, NMR data, and ROSIE models all show poor stacking in the ultimate G·U pair

These results are consistent with the observation that a fourth G·U pair added onto the end of a helix does not increase thermodynamic stability, in contrast to the stability of adding an A·U pair. The crystal structure of (5'-GGUGGCUGUU-3')₂ shows little stacking in the ultimate base step (Figure 3A) and some conformational heterogeneity. The NMR data shows few NOES and no imino protons for the ultimate G·U pair (Figure 5), which is likely due to rapid base pair opening and imino protons exchanging with water. (5'-GGUGGCUGUU-3')₂ differs from (5'-UGUGGCUGUG-3')₂ in only the orientation of the ultimate G·U pair. The ROSIE models for these sequences have nearly identical ultimate G·U pairs with no stacking. The thermodynamic stabilities for (5'-GGUGGCUGUU-3')₂, (5'-UGUGGCUGUG-3')₂, and (5'-GUGGCUGU-3')₂ are -6.1, -6.0, and -6.0 kcal/mol, respectively, which are all very similar¹⁵. If stacking of the fourth G·U pair added onto the end of a helix was significant, then one would expect the thermodynamic stabilities to vary. In contrast, adding a fourth A·U pair at the end of a helix increases the thermodynamic stability; for example, the thermodynamic stabilities of (5'-UAUCGAUA-3')₂ and (5'-UUAUCGAUAA-3')₂ are -7.1 and -9.2 kcal/mol, respectively¹⁵. Base stacking in isosteric Watson-Crick A·U pairs has intermediate base overlap and intermediate twist at each step whether the orientation is 5'-AU-3'/3'-UA-5', 5'-AA-3'/3'-UU-5', or 5'-UA-3'/3'-AU-5'. The thermodynamic stability of these base steps does not depend strongly on its position in a helix. The non-isostericity of G·U pairs, the highly overwound nature of 5'-UG-3'/3'-GU-5' base steps, and the poor base overlap in 5'-UU-3'/3'-GG-5' base steps may prevent the indefinite continuation of favorable stacking for consecutive G·U pairs. The stabilities of consecutive terminal G·U pairs and helical base stacking extend to only three G·U pairs, which is consistent with other consecutive terminal mismatches and the approximate persistence length of RNA⁴⁵. These results may partly explain why attempts to form a helix solely of G·U pairs were unsuccessful¹⁵.

Accurate thermodynamic parameters for RNA secondary structure motifs can improve tertiary structure prediction

The program ROSIE is the only RNA three-dimensional structure prediction program that currently uses the experimentally measured thermodynamic parameters for terminal G·U pairs and also the only program to accurately predict the structures of consecutive terminal G·U pairs. Some other RNA structure prediction programs predicted an overall unfavorable free energy for the duplex and are not designed to predict structures with unfavorable

energies. The recent blind competition RNA Puzzles tests the accuracy of several different RNA structure prediction programs^{49, 50}. The most successful programs rely on a thermodynamic database for RNA motifs⁵¹, to which thermodynamic parameters for terminal G-U pairs will be added. Thus, continued efforts to improve RNA thermodynamic parameters will improve not only RNA secondary structure but also RNA tertiary structure prediction and the next challenge of predicting function.

CONCLUSION

We present a high-resolution crystal structure for (5'-GGUGGCUGUU-3')₂ and NMR data for three RNA sequences with terminal G-U wobble base pairs. The observed high base overlap in the penultimate 5'-GU-3'/3'-UG-5' base step is consistent with the thermodynamic bonus for this motif. The insights in RNA structure-energetics relationships gained through these biophysical studies will facilitate future *de novo* predictions of structure from sequence and fundamental physical principles. The similarities of consecutive terminal GU pairs to A-form RNA suggest that this motif will be recognized by DICERs and other RNA-binding proteins.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Drs. Clyde Smith and Tzanko Doukov for assistance during X-ray diffraction data collection at SSRL BL 7-1. We thank Dr. Marco Tonelli for assistance with NMR data collection at the National Magnetic Resonance Facility at Madison (NMRFAM). This work was supported by the SSRL Structural Molecular Biology Program, which is supported by the DOE Office of Biological and Environmental Research and the NIH NCCR Biomedical Technology Program. This study made use of the National Magnetic Resonance Facility at Madison, which is supported by NIH grant P41GM66326 (NIGMS). Equipment was purchased with funds from the University of Wisconsin-Madison, the NIH (P41RR02301, P41GM66326, S10RR02781, S10RR08438, S10RR023438, S10RR025062, S10RR029220), the NSF (DMB-8415048, OIA-9977486, BIR-9214394), and the USDA. The Macromolecular Crystallization and X-ray Facilities at the University of Oklahoma and the Laboratory of Biomolecular Structure and Function at the University of Oklahoma Health Sciences Center are supported by the Oklahoma COBRE in Structural Biology (NIH P20 GM103640 PI Ann West) and the National Science Foundation grant 0922269. This work was supported by grants from the National Science Foundation (CAREER grant no. 0844913 [PI:SJS]), the Oklahoma Center for the Advancement of Science and Technology (HR13-206 [PI:SJS]), and the NIH (R01 AI088011 [PI: BHMM]).

REFERENCES

1. Donohue J. Hydrogen-Bonded Helical Configurations of Polynucleotides. *Proc. Natl. Acad. of Sci. USA*. 1956; 42:60–65. [PubMed: 16589818]
2. Crick FH. Codon--anticodon pairing: the wobble hypothesis. *Journal of molecular biology*. 1966; 19:548–555. [PubMed: 5969078]
3. Hou YM, Schimmel P. A simple structural feature is a major determinant of the identity of a transfer RNA. *Nature*. 1988; 333:140–145. [PubMed: 3285220]
4. Musier-Forsyth K, Usman N, Scaringe S, Doudna J, Green R, Schimmel P. Specificity for aminoacylation of an RNA helix: an unpaired, exocyclic amino group in the minor groove. *Science*. 1991; 253:784–786. [PubMed: 1876835]
5. Lu K, Heng X, Garyu L, Monti S, Garcia EL, Kharytonchyk S, Dorjsuren B, Kulandaivel G, Jones S, Hiremath A, et al. NMR detection of structures in the HIV-1 5'-leader RNA that regulate genome packaging. *Science*. 2011; 334:242–245. [PubMed: 21998393]

6. Pyle AM, Moran S, Strobel SA, Chapman T, Turner DH, Cech TR. Replacement of the conserved G.U with a G-C pair at the cleavage site of the Tetrahymena ribozyme decreases binding, reactivity, and fidelity. *Biochemistry*. 1994; 33:13856–13863. [PubMed: 7947794]
7. Peebles CL, Zhang M, Perlman PS, Franzen JS. Catalytically critical nucleotide in domain 5 of a group II intron. *Proc. Natl. Acad. of Sci. USA*. 1995; 92:4422–4426. [PubMed: 7538669]
8. Perrotta AT, Been MD. Core sequences and a cleavage site wobble pair required for HDV antigenomic ribozyme self-cleavage. *Nucl. Acids Res*. 1996; 24:1314–1321. [PubMed: 8614636]
9. Doench JG, Sharp PA. Specificity of microRNA target selection in translational repression. *Genes & Devel*. 2004; 18:504–511. [PubMed: 15014042]
10. Schroeder SJ, Stone JW, Bleckley S, Gibbons T, Mathews DM. Ensemble of secondary structures for encapsidated satellite tobacco mosaic virus RNA consistent with chemical probing and crystallography constraints. *Biophys. J*. 2011; 101:167–175. [PubMed: 21723827]
11. Turner, DH. Conformational Changes. In: Bloomfield, VA.; Crothers, DM.; Tinoco, I., Jr, editors. *Nucleic Acids: Structures, Properties, and Functions*. Sausalito, CA: University Science Books; 2000.
12. McDowell JA, Turner DH. Investigation of the structural basis for thermodynamic stabilities of tandem GU mismatches: solution structure of (rGAGGUCUC)2 by two-dimensional NMR and simulated annealing. *Biochemistry*. 1996; 35:14077–14089. [PubMed: 8916893]
13. Chen JL, Dishler AL, Kennedy SD, Yildirim I, Liu B, Turner DH, Serra MJ. Testing the nearest neighbor model for canonical RNA base pairs: revision of GU parameters. *Biochemistry*. 2012; 51:3508–3522. [PubMed: 22490167]
14. McDowell JA, He L, Chen X, Turner DH. Investigation of the structural basis for thermodynamic stabilities of tandem GU wobble pairs: NMR structures of (rGGAGUUGC)2 and (rGGAUGUCC)2. *Biochemistry*. 1997; 36:8030–8038. [PubMed: 9201950]
15. Nguyen MT, Schroeder SJ. Consecutive terminal GU pairs stabilize RNA helices. *Biochemistry*. 2010; 49:10574–10581. [PubMed: 21067139]
16. Mathews DH, Sabina J, Zuker M, Turner DH. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol*. 1999; 288:911–940. [PubMed: 10329189]
17. 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. of Sci. USA*. 2004; 101:7287–7292. [PubMed: 15123812]
18. Ragan C, Cloonan N, Grimmond SM, Zuker M, Ragan MA. Transcriptome-wide prediction of miRNA targets in human and mouse using FASTH. *PloS ONE*. 2009; 4:e5745. [PubMed: 19478946]
19. Martin HC, Wani S, Steptoe AL, Krishnan K, Nones K, Nourbakhsh E, Vlassov A, Grimmond SM, Cloonan N. Imperfect centered miRNA binding sites are common and can mediate repression of target mRNAs. *Genome Biol*. 2014; 15:R51. [PubMed: 24629056]
20. Shimura H, Pantaleo V, Ishihara T, Myojo N, Inaba J, Sueda K, Burgyan J, Masuta C. A viral satellite RNA induces yellow symptoms on tobacco by targeting a gene involved in chlorophyll biosynthesis using the RNA silencing machinery. *PLoS Path*. 2011; 7:e1002021.
21. Kabsch W. Xds. *Acta crystallographica. Section D, Biological crystallography*. 2010; 66:125–132. [PubMed: 20124692]
22. Evans P. Scaling and assessment of data quality. *Acta Crystallogr. Section D, Biological crystallography*. 2006; 62:72–82. [PubMed: 16369096]
23. French S, Wilson K. On the Treatment of Negative Intensity Observations. *Acta Crystallogr*. 1978; A34:517–525.
24. Sheldrick GM. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr.. Section D, Biological crystallography*. 2010; 66:479–485. [PubMed: 20383001]
25. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallogr. Section D, Biological crystallography*. 2010; 66:486–501. [PubMed: 20383002]
26. Keating KS, Pyle AM. RCrane: semi-automated RNA model building. *Acta Crystallogr. Section D, Biological crystallography*. 2012; 68:985–995. [PubMed: 22868764]

27. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallogr. Section D, Biological crystallography*. 2010; 66:213–221. [PubMed: 20124702]
28. Chou FC, Sripakdeevong P, Dibrov SM, Hermann T, Das R. Correcting pervasive errors in RNA crystallography through enumerative structure prediction. *Nature Meth*. 2013; 10:74–76.
29. Thorn A, Sheldrick GM. ANODE: anomalous and heavy-atom density calculation. *J. of Applied Cryst*. 2011; 44:1285–1287. [PubMed: 22477786]
30. Zucker F, Champ PC, Merritt EA. Validation of crystallographic models containing TLS or other descriptions of anisotropy. *Acta crystallogr. Section D, Biological crystallography*. 2010; 66:889–900. [PubMed: 20693688]
31. Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta crystallogr. Section D, Biological crystallography*. 2010; 66:12–21. [PubMed: 20057044]
32. Wang X, Kapral G, Murray L, Richardson D, Richardson J, Snoeyink J. RNABC: forward kinematics to reduce all-atom steric clashes in RNA backbone. *J. of Math. Biol*. 2008; 56:253–278. [PubMed: 17401565]
33. Lu XJ. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucl. Acids Res*. 2003; 31:5108–5121. [PubMed: 12930962]
34. Lavery R, Moakher M, Maddocks JH, Petkeviciute D, Zakrzewska K. Conformational analysis of nucleic acids revisited: Curves+ *Nucl. Acids Res*. 2009; 37:5917–5929. [PubMed: 19625494]
35. Dolinsky TJ, Czodrowski P, Li H, Nielsen JE, Jensen JH, Klebe G, Baker NA. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucl. Acids Res*. 2007; 35:W522–W525. [PubMed: 17488841]
36. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. of Sci. USA*. 2001; 98:10037–10041. [PubMed: 11517324]
37. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. of Biomolec. NMR*. 1995; 6:277–293.
38. Goddard, TD.; Kneller, DG. SPARKY 3. San Francisco: University of California; 2008.
39. Lukavsky PJ, Puglisi JD. RNAPack: an integrated NMR approach to RNA structure determination. *Methods*. 2001; 25:316–332. [PubMed: 11860286]
40. Das R, Karanicolas J, Baker D. Atomic accuracy in predicting and designing noncanonical RNA structure. *Nature Meth*. 2010; 7:291–294.
41. Lyskov S, Chou FC, Conchuir SO, Der BS, Drew K, Kuroda D, Xu J, Weitzner BD, Renfrew PD, Sripakdeevong P, et al. Serverification of molecular modeling applications: the Rosetta Online Server that Includes Everyone (ROSIE). *PloS ONE*. 2013; 8:e63906. [PubMed: 23717507]
42. Parisien M, Major F. The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature*. 2008; 452:51–55. [PubMed: 18322526]
43. Sijenyi, F.; Saro, P.; Ouyang, Z.; Damm-Ganamet, K.; Wood, M.; Jiang, J.; Santa Luica, J, Jr. The RNA Folding Problems: Different Levels of RNA Structure Prediction. In: Leontis, N.; Westof, E., editors. *RNA 3D Structure Analysis and Prediction*. Springer; 2011.
44. Masquida B, Westhof E. On the wobble GoU and related pairs. *RNA*. 2000; 6:9–15. [PubMed: 10668794]
45. Clanton-Arrowood K, McGurk J, Schroeder SJ. 3' Terminal Nucleotides Determine Thermodynamic Stabilities of Mismatches at the Ends of RNA Helices. *Biochemistry*. 2008; 47:13418–13427. [PubMed: 19053257]
46. Chen X, McDowell JA, Kierzek R, Krugh TR, Turner DH. Nuclear magnetic resonance spectroscopy and molecular modeling reveal that different hydrogen bonding patterns are possible for GU pairs: one hydrogen bond for each GU pair in RGGCGUGCC)2 and two for each GU pair in RGAGUGCUC)2. *Biochemistry*. 2000; 39:8970–8982. [PubMed: 10913310]

47. McDowell JA, Turner DH. Investigation of the structural basis for thermodynamic stabilities of tandem GU mismatches: solution structure of (rGAGGUCUC)₂ by two-dimensional NMR and simulated annealing. *Biochemistry*. 1996; 35:14077–14089. [PubMed: 8916893]
48. Serra MJ, Smolter PE, Westhof E. Pronounced instability of tandem IU base pairs in RNA. *Nucl. Acids Res*. 2004; 32:1824–1828. [PubMed: 15037659]
49. Cruz JA, Blanchet MF, Boniecki M, Bujnicki JM, Chen SJ, Cao S, Das R, Ding F, Dokholyan NV, Flores SC, et al. RNA-Puzzles: a CASP-like evaluation of RNA three-dimensional structure prediction. *RNA*. 2012; 18:610–625. [PubMed: 22361291]
50. Miao Z, Adamiak RW, Blanchet MF, Boniecki M, Bujnicki JM, Chen SJ, Cheng C, Chojnowski G, Chou FC, Cordero P, et al. RNA-Puzzles Round II: assessment of RNA structure prediction programs applied to three large RNA structures. *RNA*. 2015; 21:1066–1084. [PubMed: 25883046]
51. Turner DH, Mathews DH. NNDB: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. *Nucl. Acids Res*. 2009; 38:D280–D282. [PubMed: 19880381]

Figure 1. Electron density map of the G8A:U3B base pair in the duplex (5'GGUGGCUGUU)₂. The electron density is contoured at the 1 σ level. NCO is the residue code for cobalt hexamine in the Chemical Component Dictionary of the Protein Data Bank.

Figure 2.

Overlay of the crystal structure of the duplex (5'GGUGGCUGUU)₂ with a model of an all AU pair helix using standard A-form RNA helix parameters. The rmsd of the superposition is 1.27 Å for all atoms in common between the structures.

Figure 3.

Base stacking. The 5' base in the left strand is part of a base pair that is colored gray and is on the bottom. The 3' base in the left strand is part of a base pair that is colored black and is on the top. A. 5'-GG-3'/3'-UU-5'. B. 5'-GU-3'/3'-UG-5'. C. 5'-UG-3'/3'-GU-5'. D. 5'-GG-3'/3'-UC-5'. E. 5'-GC-3'/3'-CG-5'.

Figure 4.

Helical twist angle between consecutive C1'—C1' vectors by base steps. Helical parameters are defined by the C1'—C1' vectors of each base pair. The parameters from the experimental structures are colored black. The AB, CD, and EE' helices are the three duplexes formed by the 5 RNA strands (A, B, C, D, and E) in the asymmetric unit of the crystal structure. The EE' helix is formed by the E strands in two asymmetric units in the C2 unit cell. The parameters from the calculated AU A-form duplex and the ROSIE structure are colored gray.

Figure 5. NOESY Spectrum for (5'-GGUGGCUGUU-3')₂. Assignments in blue identify the diagonal peaks for imino protons. Assignments in green identify NOES between imino protons in the wobble G-U pairs. Assignments in red identify imino proton-ribose NOEs that define base stacking.

Figure 6.

Base stacking in the penultimate base step. The stacking interactions in the 5'-GU-3'/3'-UG-5' step have a bonus in the thermodynamic parameters for prediction rules for consecutive terminal G-U pairs. The free energy values (G_{37}°) are for the duplex measured in 1 M NaCl, 10 mM sodium cacodylate, 0.1 mM Na₂EDTA, pH 7¹⁵. Cross-strand G stacking occurs in the 5'-UG-3'/3'-GU-5' steps, and there is a clear contrast in the sequence-dependent base-stacking. The base stacking geometries are from structures that were predicted by ROSIE⁴¹ and are consistent with all crystallography and NMR data.

Table 1

Experimentally Derived Thermodynamic Parameters for G·U Nearest Neighbors, Terminal Pairs, and Helices

Internal G·U Nearest Neighbor Parameters		
sequence	2012 parameters ^a G° (kcal/mol)	1999 parameters ^b G° (kcal/mol)
5'GU3' 3'UG5'	+0.7 ± 0.2	+1.3 ± 0.6 ^c
5'GG3' 3'UU5'	−0.3 ± 0.2	+0.5 ± 1.0
5'UG3' 3'GU5'	−0.6 ± 0.2	+0.3 ± 0.5
Terminal G·U Parameters ^d		G° (kcal/mol)
First terminal G·U pair ^e		−0.8 to −2.3 ± 0.1
Second terminal G·U pair		−0.7 ± 0.1
Third terminal G·U pair		−0.5 ± 0.1
5'GU3'/3'UG5' bonus ^f		−0.4 ± 0.1
Duplexes ^g		G° (kcal/mol)
(5'GGUGGCUGUU3') ₂		−6.1 ± 0.1
(5'GUGGCUGU3') ₂		−6.0 ± 0.1
(5'UGUGGCUGUG3') ₂		−6.0 ± 0.1
(5'GUGUCGGUGU3') ₂		−4.0 ± 0.1

^aParameters were derived in reference ¹³, which contains new thermodynamic data for more sequences.^bParameters were derived in reference ¹⁶. Note that both the 1999 and 2012 parameters are currently used in RNA structure prediction software.^cThe exception for this parameter is the specific context 5'GGUC3'/3'CUGC5' which has a value of −4.1 ± 0.5 kcal/mol ¹⁶. All other closing pairs for the internal 5'GU3'/3'UG5' nearest neighbour have a value of +1.3 ± 0.6 kcal/mol.^dTerminal G·U parameters were derived in reference ¹⁵ and count how many G·U pairs are added onto the end of a Watson-Crick base paired helix, which is a slightly different nomenclature than the crystallography analysis.^eThe thermodynamic stability of the first terminal G·U pair depends on the identity and orientation of the last Watson-Crick pair in the helix ¹⁵.^fThe bonus only applies to the case when the sequence 5'GU3'/3'UG5' occurs in the third nearest neighbour position ¹⁵.^gData from reference ¹⁵.

Table 2
X-ray diffraction data statistics for the duplex (5' GGUGGUGUU)₂. The space group is C2.

Diffraction data	Native	Inflection	HREM ¹	Peak
Wavelength (Å)	0.9795	1.6053	1.3624	1.6068
Cell Dimensions				
a (Å)	36.62	36.59	--	--
b (Å)	43.22	43.17	--	--
c (Å)	83.26	83.25	--	--
$\alpha = \gamma$ (°)	90.0	90.0	--	--
β (°)	102.48	102.47	--	--
Resolution range	27.55-1.32	27.62-2.44	27.62-2.44	27.67-2.44
High resolution shell	(1.39-1.32)	(2.82-2.44)	--	--
R _{merge}	6.0 (66.7)	5.3 (17.3)	3.4 (9.5)	4.1 (13.4)
R _{pim} (all I+ & I-)	3.1 (40.5)	2.9 (8.5)	1.7 (4.5)	1.9 (6.2)
<I/σ(I)>	18.2 (2.7)	26.8 (9.6)	41.8 (17.0)	35.6 (11.8)
<I> half-set correlation CC(1/2)	99.9 (82.4)	99.9 (99.8)	99.9 (99.2)	99.9(98.6)
Total reflections	236794	29986	23116	23313
Unique Reflections	(29968)	(7662)	(552)	(564)
	29968	4344	3216	3239
	(4325)	(1189)	(86)	(88)
Completeness (%)	100.0 (99.8)	98.6 (95.3)	98.7 (83.3)	99.4(88.3)
Multiplicity	7.9 (7.2)	6.9 (6.4)	7.2 (6.4)	7.2 (6.4)
Anomalous multiplicity		3.6 (3.4)	3.8 (3.3)	3.8 (3.2)
Average mosaicity (°)	0.2	0.27	0.34	0.28
Mid-slope Anom Normal Prob. Plot	1.12	2.083	1.917	1.499
Wilson B	11.98			

¹HREM: High energy remote wavelength.

Table 3Phasing and refinement statistics for the duplex (5'GGUGGCUGUU)₂.

Phasing statistics:	
Correlation coefficient between datasets (3.2-3.0 Å)	
HREM/PEAK	43.8
HREM/INFL	62.4
PEAK/INFL	53.5
Refinement statistics:	
Number of reflections used in refinement (F^+ and F^-)	58561
Number of unique reflections	29961
Resolution Range (Å)	27.2-1.32
$R_{\text{work}}/R_{\text{free}}$	0.1569/0.1857
RMS deviation bond lengths (Å)	0.005
RNA deviation bond angles (°)	1.002
Molprobit clash score	
Maximum likelihood coordinate uncertainty (Å)	0.12
No. nucleotides	50
No. water molecules	196
No. cobalt hexammines	12
No. chlorides	3
Average RNA B-factor, (Å ²)	21.8
Average solvent B-factor, (Å ²)	28.8