
Mighty tiny

KATHLEEN B. HALL

Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, Missouri 63110, USA

In 1983, Woese, Gutell, Gupta, and Noller noted that 16S rRNA contained a disproportionate number of hairpins with only four nucleotides in the loop. The loop sequences of these “tetraloops” were not random, but consisted of two classes: GNRA and UNCG, where N is anything and R is purine. This observation was intriguing but also puzzling, for it begged the question of their functions.

The UUCG tetraloop

Five years later, the UUCG tetraloop was found in bacteriophage T4 mRNA. Taking advantage of the reductionist approach to RNA structure/function, the tetraloop was removed from its mRNA, and examined as an isolated hairpin. The dramatic finding was that this tiny RNA secondary structure, with its UUCG loop and C:G loop-closing base pair, was “extraordinarily stable.” The significance of this finding was not lost on the investigators in this first study: The authors include Stormo, Uhlenbeck, Tinoco, and Gold, all recognized leaders in RNA biology.

The realization that the 5'-CUUCGG hairpin was thermodynamically far more stable than it should have been was galvanizing to the RNA community. Since it is so small and stable, it would fold immediately upon transcription where it could potentially direct secondary structure formation and reduce alternative structures that were off-pathway. Its function was independent of its RNA context, as shown by its presence in 16S rRNA. The authors suggested a function: It acted as a nucleation site for RNA folding.

That first study established the CUUCGG as a uniquely stable tetraloop, so the next question was obvious: Why was it stable? Sequence variants were tested for their thermodynamic stabilities, RNA hairpins were compared to DNA hairpins, and different salts conditions were tested for their contribution to stability. The answer came from solution NMR structures of the CU1U2C3G4G tetraloop: There is a hydrogen bond between C3 amino hydrogen and the phosphate oxygen between U1 and U2, G4 was *syn* and makes a bifurcated hydrogen bond to U1, the U1 ribose 2' OH makes a hydrogen bond to the G4 carbonyl oxygen, the riboses of

U2 and C3 were *C2'-endo*, and U5 stacked over the C and G4 stacked over the G of the C:G loop closing base pair. The combinations of hydrogen binding partners were astounding, and confirmed the realization that predicting such an intricate network of interactions for other RNAs was going to be a challenge.

Because the CUUCGG tetraloop is so small and stable, it became a favorite subject of many investigators as they sought to discover the detailed contributions of each nucleotide to its unusual stability. There have been many solution NMR studies of U1U2C3G4 tetraloops and their variants. More recent experiments look at the dynamics of each nucleotide in the tetraloop using ¹⁵N and ¹³C-isotopically labeled nucleotides. Since U2 is the one nucleotide that can be replaced by others (the UNCG motif), it is not surprising that the U2 nucleobase samples conformational space, but so does its phosphodiester backbone and its ribose. The ribose of C3 is also dynamic, sampling different puckers. The phosphodiester backbone extending 3' to U1 through G4 undergoes rapid (ps-ns) motion. The UUCG tetraloop is the most stable RNA loop structure known, but the NMR studies illustrate that despite its thermodynamic stability, the CUUCGG loop is not rigid.

There have also been many thermodynamic studies of UUCG with substituted nucleotides. We could not resist the temptation to play with it, and examined the contribution of each ribose 2' hydroxyl to structural integrity (using NMR) and thermodynamic stability. Since the substituent at the 2' position affects the sugar pucker, replacing the 2' OH with H, F, or NH₃ does more than simply alter hydrogen bonding patterns. Several labs have studied the importance of the C:G loop-closing base pair. Some investigators have reduced the hairpin to its minimal form (a two base pair stem including the C:G loop closing base pair) to measure its folding rates. It has been subjected to T-jump, Raman, IR, and vibrational spectroscopy interrogations. A first attempt to use X-ray crystallography to solve the CUUCGG tetraloop structure ended in failure, since the RNA oligonucleotide formed a dimer. That experiment did yield some new data, however, since the dimer contained both a U:U and a U:C base pair.

Computational experiments with the UUCG tetraloop began with the work of Kollman's group in 1996, and are

Corresponding author: kathleenhal@gmail.com

Article and publication date are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.050567.115>. Freely available online through the RNA Open Access option.

© 2015 Hall This article, published in *RNA*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

continuing. With high resolution starting structures (there are now many X-ray crystal structures of it) the UUCG tetraloop has the potential to be a benchmark for accurate molecular dynamics descriptions of its structure and dynamics. It does serve this function admirably, especially since it requires no divalent ions to adopt its structure or to increase its stability, but for now it is still a challenge. With only a two base pair stem, attempts have been made to fold it *in silico*, without much success.

Today the CUUCGG tetraloop has found its niche as an experimental tool to stabilize the secondary structure of a larger RNA. No proteins bind to it, and it makes no RNA:RNA interactions. It is a tiny autonomous RNA structural element that has had a mighty contribution to RNA biology.

The GNRA tetraloop

The GRNA sequence was also over-represented in rRNA, and it became the focus of thermodynamic and structural investigations. Here the question was the same: What is the function(s) of these overabundant RNA hairpins?

The GNRA tetraloops are thermodynamically less stable than the UUCG tetraloops, but they are anomalously stable when compared to other four-nucleotide loops. Curiously, experiments with RNA and DNA tetraloop sequences showed that both d(CGNAAG) and r(CGNAAG) hairpins were both unexpectedly stable, providing a first clue that these tetraloops have unique properties.

The first GNRA structures to be described were the G1A2A3A4 and GCAA tetraloops with C:G loop closing base pairs. Solution NMR structures from the Pardi group revealed that G1 and A4 formed a G:A base pair, with the amino group of each purine making hydrogen bonds to an aromatic nitrogen (GN7 or AN3). A3 was stacked over A4, and the G:A pair was stacked over the C:G loop closing base pair, but in the CGCAAG tetraloop, the C was unconstrained. There are additional hydrogen bonds from the G4 amino hydrogen to the A4 phosphate oxygen, and a bifurcated hydrogen bond from the G4 ribose 2' OH to both A3 N7 and A4 amino hydrogen. Pardi's group later posited that the GAAA structure is an example of a U-turn. Since many RNA U-turn hairpin loops can make RNA:RNA interactions, this

was an inspired conclusion given what we now know about GNRA tetraloops.

As originally suggested by Westhof and Michel, tiny tetraloops can and do participate in tertiary interactions with other RNA elements. Most recently, single molecule fluorescence experiments are being used to probe these interactions. A particular focus is the docking of the G1A2A3A4 tetraloop with its 11 nucleotide receptor, to investigate both the thermodynamics and kinetics of the interactions. A2 stacks with an A-A platform within the receptor, and also makes hydrogen bonds with receptor bases. Meanwhile, A3 and A4 make hydrogen bonds with the base pairs in the receptor helical region (base triples), to anchor the two elements together. Some ingenious constructs that allow the tetraloop to transiently dock with its receptor have produced finely resolved kinetic traces of its attachment/detachment and its residence time in the docked state. Since in a natural RNA, these interactions will staple the two sites together, the lifetime of the interaction could be mighty important for tertiary stability.

In the Tetrahymena Group I intron, a GAAA:receptor interaction anchors the P456 domain together. Formation of the docked state is part of the folding pathway of the intron and also of its stability. P456 is one of the model systems for the study of GAAA kinetics and thermodynamics. Ribosomal rRNA contains many GNRA tetraloops, so are there many examples of GNRA:receptor interactions? More pointedly, can the receptors be identified by sequence/phylogenetic conservation, and then can they be paired with their tetraloop without relying on the crystal structure of the ribosome? In general, can tetraloop/receptor pairs be predicted from the secondary structure of an RNA? This is a challenge.

Summary

Where the UUCG tetraloop is a loner, the GNRA tetraloop is gregarious. Where UUCG accommodate few substitutions, GNRA can accept many. GNRA often makes long-range hydrogen-bonding interactions to other RNA elements, the most famous being the tetraloop receptor. The GNRA tetraloop illustrates how RNA:RNA interactions work, while the UUCG tetraloop illustrates how RNA can be self-contained. These tiny RNAs have mighty roles in the biology of RNA!