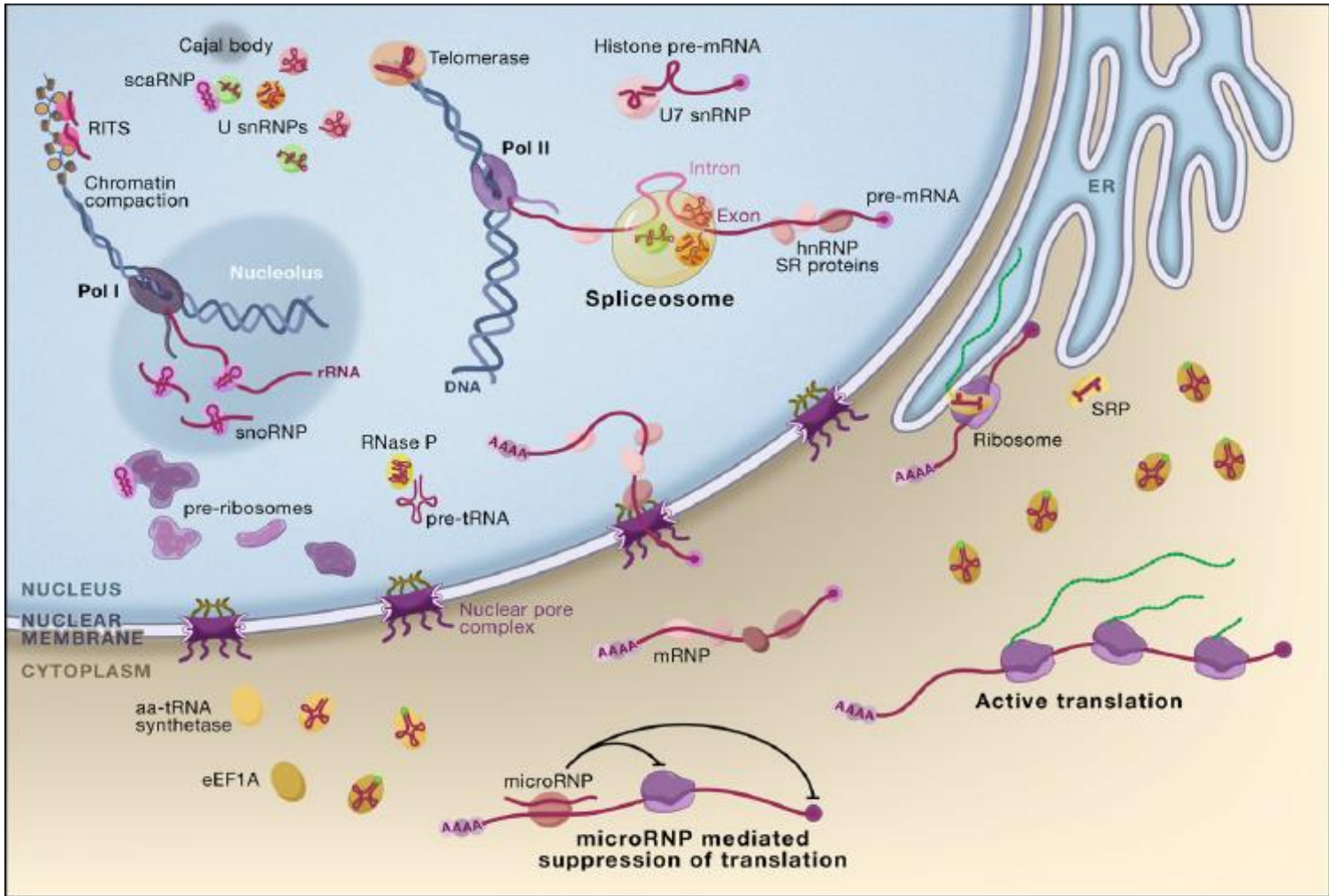
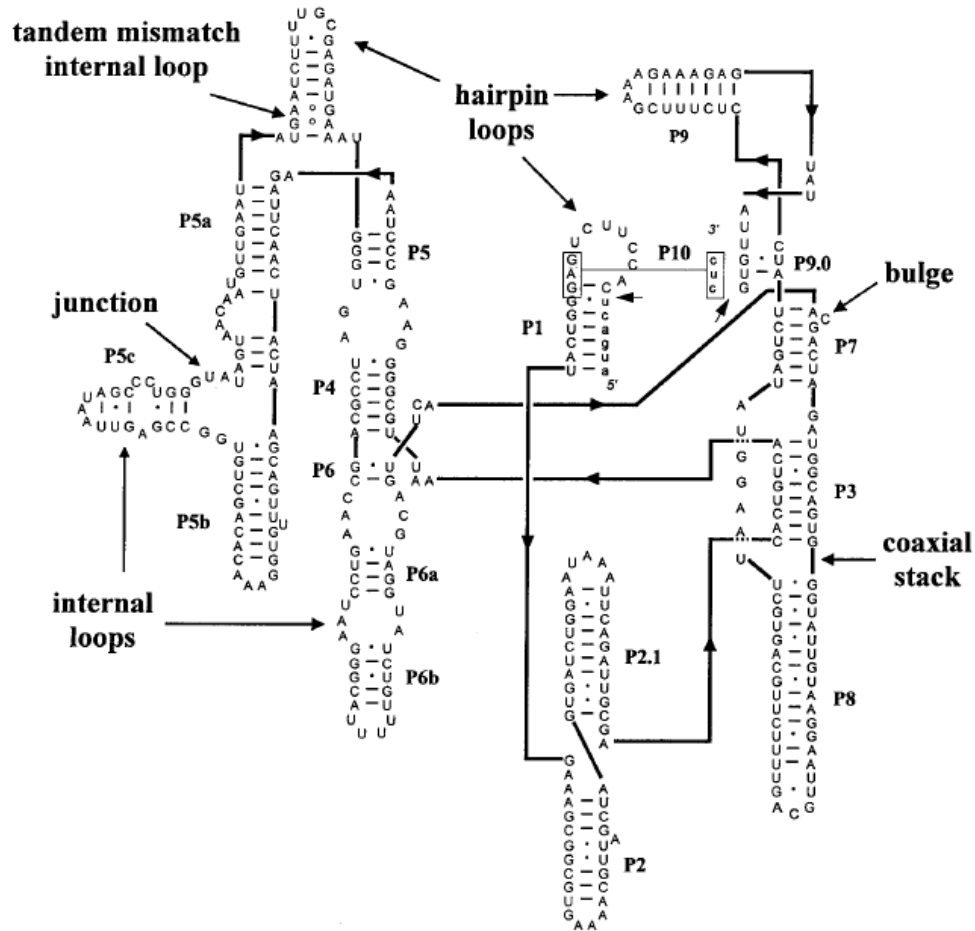


An introduction to RNA biology
and
why biophysical chemistry matters



RNAs are not noodles in the cell.



RNAs have secondary and tertiary structures.

Determination of an RNA secondary structure

Thermodynamics can be used when there are no other options. If, for example, an RNA has been identified as a noncoding RNA, but lacks functionality, biochemical structure probing, or phylogenetic data, then a first pass at a secondary structure can be obtained from thermodynamic calculations.

Phylogenetic comparisons are the most powerful methods for secondary structure determination, especially when the RNA is long (rRNA is the obvious example). There needs to be a deep sequence data base for this method to be robust, especially if a function is preserved but sequence is not (RNase P RNA or a Group I intron).

Biochemical mapping should accompany any secondary structure prediction. There are enzymes and chemical methods for identifying single stranded regions and flexible regions, and humans to map the data onto the sequence.

Determination of an RNA tertiary structure

X-ray crystallography has been used for solving RNA structures, but continues to be challenging and idiosyncratic (depending on the RNA and the human).

NMR is useful for small RNAs or cleverly constructed larger RNAs, and is most useful for looking at dynamics.

Phylogenetic comparisons of RNA sequences can reveal secondary structure conservation.

Secondary structure can be important even when the sequence is not preserved, since an A-form duplex is a duplex is a duplex. Compensatory mutations that preserve a duplex can be parsed by phylogenetics.

What class of molecules is the most highly conserved in all living organisms?

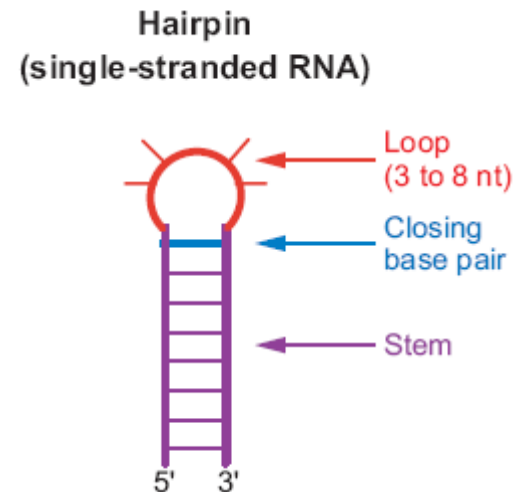
The **ribosome**, which synthesizes protein in all living systems, is one of life's most ancient molecular machines. The ribosome is our most direct macromolecular connection to the distant evolutionary past and to early life. "Translation is not just another molecular structure to be solved. It represents, it is, the evolutionary transition from some kind of nucleic acid-based world to the protein-based world of modern cells" (Woese 2001). It is believed that the ribosome in its present form was well established before the last universal common ancestor of life (LUCA), that is, beyond the root of the phylogenetic tree (Fox and Ashinikumar 2004). Much of the diversity of conformation and sequence between bacterial and archaeal ribosomes is believed to predate the LUCA.

Hsiao C...Williams LD 2009. Peeling the Onion: Ribosomes Are Ancient Molecular Fossils. *Molec Biol Evol* 26 (11): 2415-2425.

Comparison of linear rRNA sequences is a well established method for determination of phylogenetic relationships. Woese and Fox (1977) used sequence data in their discovery of Archaea, the third kingdom of life (Magrum et al. 1978; Woese et al. 1978). Their results produced the phylogenetic tree that includes prokaryotes, protozoa, fungi, plants, and animals (Woese 1987).

Over 10,000 16S and 16S-like rRNA and over 1,000 23S and 23S-like rRNA genes have been sequenced (Cannone et al. 2002).

In 1983, Woese et al., reported that 16S rRNAs contained a disproportionate number of 'tetraloops': RNA stemloops with specific 4-nucleotide loops. Woese, C. R., Gutell, R., Gupta, R. & Noller, H. (1983) *Microbiol. Rev.* 47, 621-669.



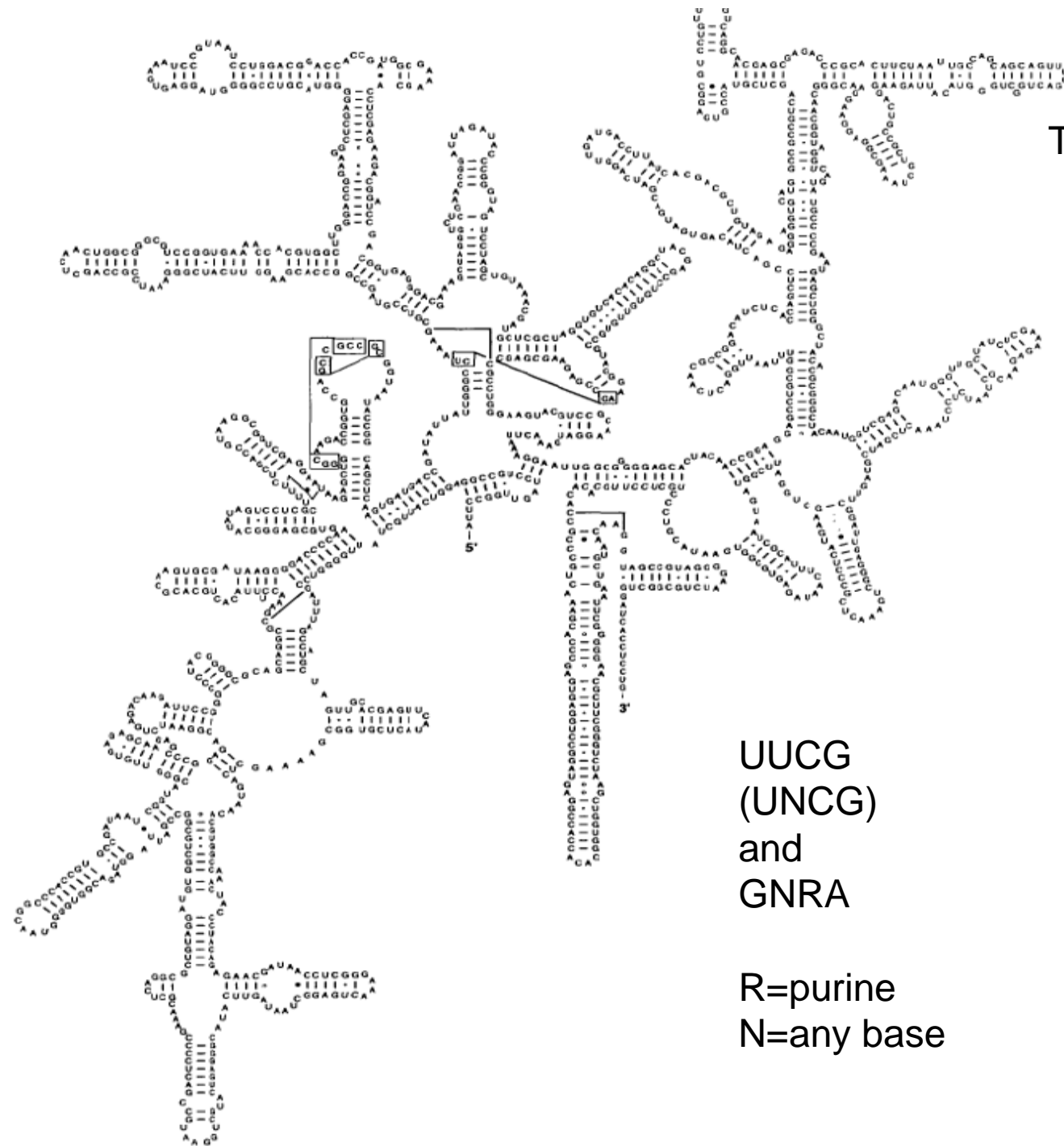
Tetraloops in 16S rRNA.

In 1988, Tuerk* et al. described the presence of an unusual tetraloop in bacteriophage T4 mRNA and then went on to show that it was thermodynamically unusually stable.

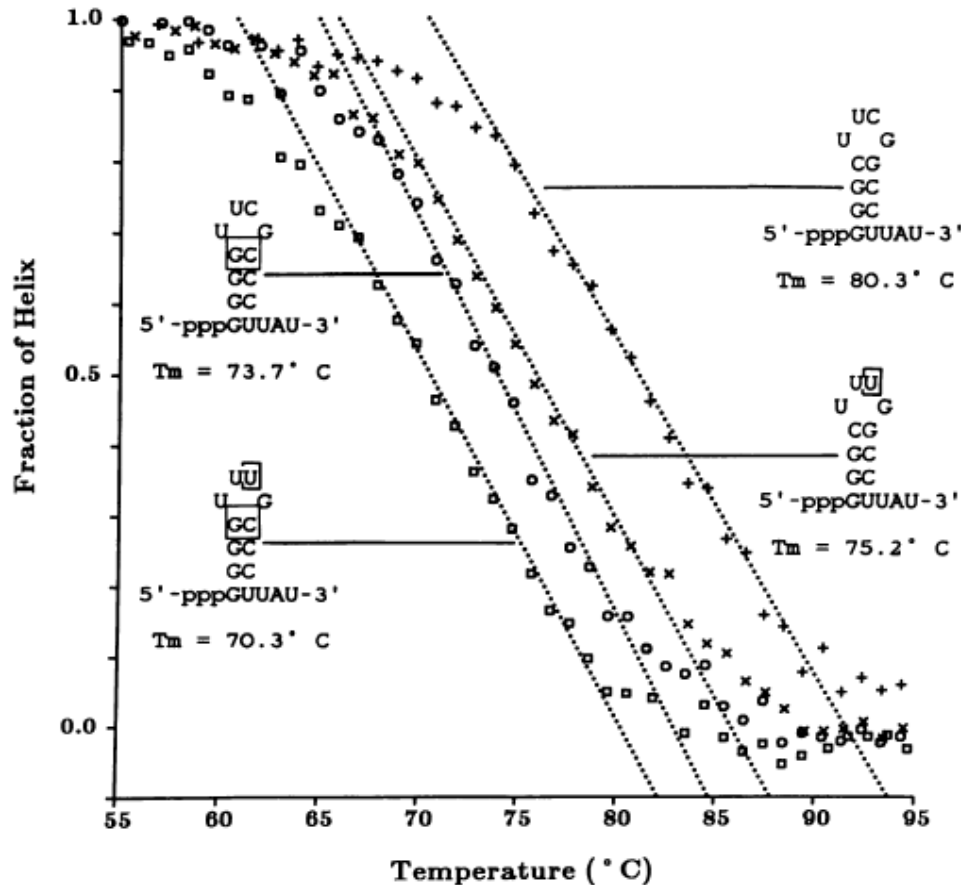
*Tuerk, ,, Stormo, Uhlenbeck, Tinoco, ..Gold..1988. [Proc Natl Acad Sci U S A](#). 85(5):1364-8.

UUCG
(UNCG)
and
GNRA

R=purine
N=any base



Cut out the hairpins and study them*.



*can you do the same with proteins?

Denaturation experiments carried out in 0.1 M NaCl/10 mM sodium phosphate/0.1 mM EDTA, pH 7. Each hairpin concentration varied from 1-20 μM ; three independent runs were averaged to give T_m . Linear regression is shown. (Tuerk et al, 1988)

For an RNA, hairpin loop formation is critical, since the RNA strand folds locally as it is being transcribed.

This is an intramolecular interaction, and thus unimolecular (like protein folding!). Consequently, it is concentration Independent.

$$u \rightleftharpoons N \quad K = \frac{[N]}{[u]} = \frac{\alpha}{1-\alpha}$$

*u is unfolded
N is native (folded)*

$$\alpha = \frac{[N]}{[N] + [u]}$$

$$C_T = [u] + [N]$$

Hairpin properties that are biologically important:

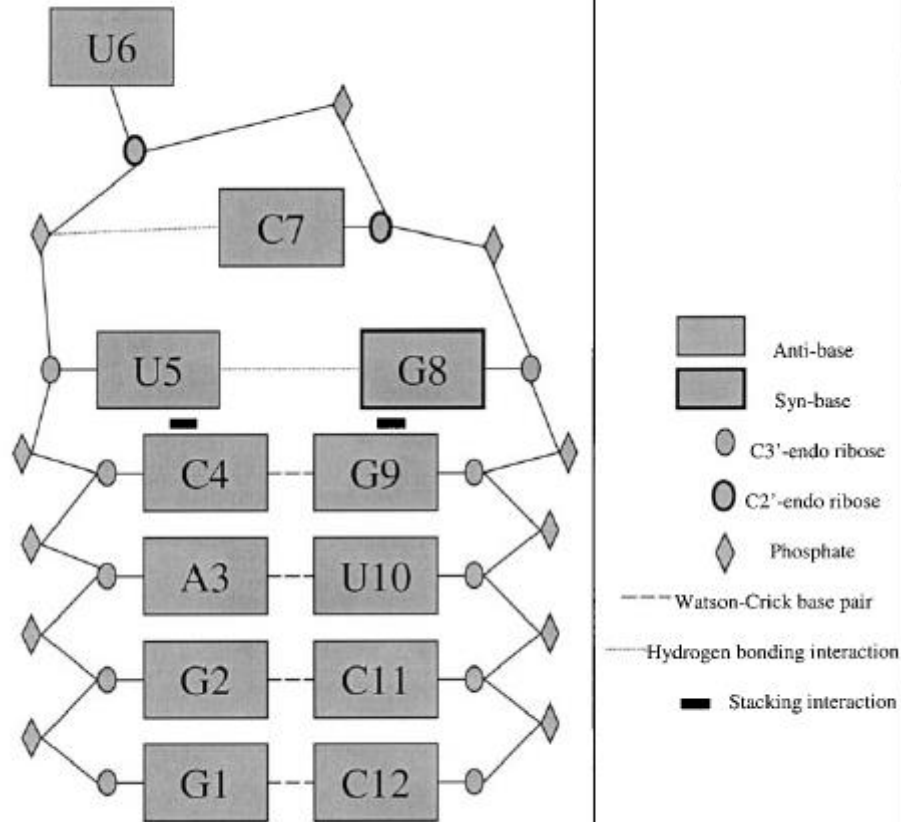
1. Thermodynamics. What is the folding free energy? ΔG°
2. Kinetics. How fast does it fold, compared to rates of transcription, protein/ligand binding.
3. Alternative structures that compete for the 'correct' fold

The cUUCGg hairpin was unusually stable. Why?

Table 1. RNA and DNA Hairpins of the (UNCG)/(TNCG) and the (CUUG)/(CTTG) Families in 1 M sodium chloride, 0.01 M sodium phosphate, 0.1 mM EDTA, at pH 7.

RNA GGAX(NNNN)X'UCC Hairpin Parameters					DNA GGAX(NNNN)X'TCC Hairpin Parameters				
RNA Loop Sequence	T _m (°C)	ΔH° (kcal/mol)	ΔS° (e.u.)	ΔG°(37) (kcal/mol)	DNA Loop Sequence	T _m (°C)	ΔH° (kcal/mol)	ΔS° (e.u.)	ΔG°(37) (kcal/mol)
C(UUCG)G	76.2	-55.9	-159.9	-6.3	C(TTCG)G	60.4	-31.3	-93.9	-2.2
C(UUUG)G	70.3	-44.0	-128.0	-4.2	C(TTTG)G	59.8	-31.0	-93.0	-2.1
C(UUUU)G	69.6	-44.3	-129.3	-4.2	C(TTTT)G	61.4	-33.5	-100.1	-2.4
G(UUCG)C	67.7	-44.8	-131.4	-4.0	G(TTCG)C	53.8	-30.2	-92.5	-1.5
G(UUUG)C	*	*	*	*	G(TTTG)C	54.6	-31.7	-96.8	-1.7
G(UUUU)C	*	*	*	*	G(GCTT)C	51.2	-27.2	-83.7	-1.2
G(CUUG)C	*	*	*	*	G(CTTG)C	70.7	-42.0	-122.2	-4.1
C(GCUU)G	70.9	-45.0	-130.8	-4.4	C(GCTT)G	58.5	-29.3	-88.3	-1.9
C(UACG)G	73.8	-53.6	-154.5	-5.7	C(dUdUCG)G	60.8	-31.8	95.3	-2.3

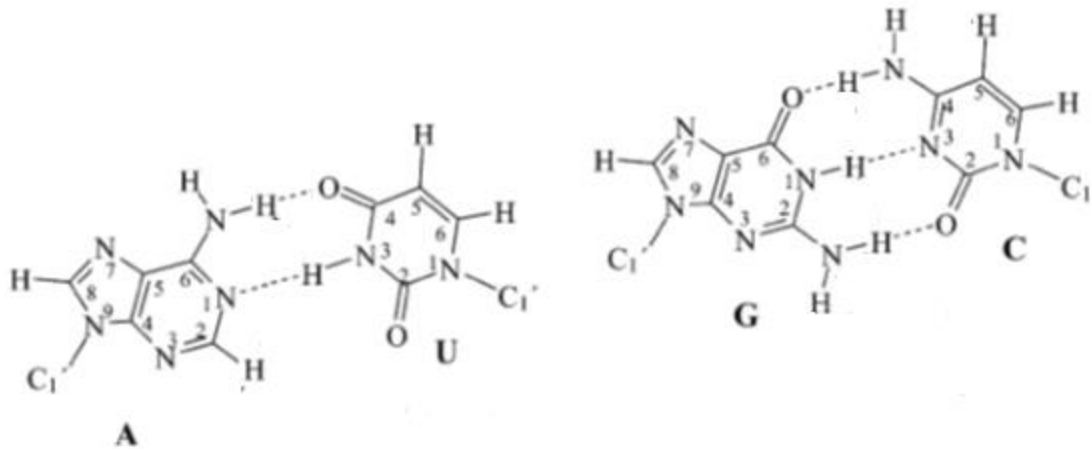
*Evidence for duplex formation was seen in the melting curves



The UUCG structure was first solved in solution by NMR

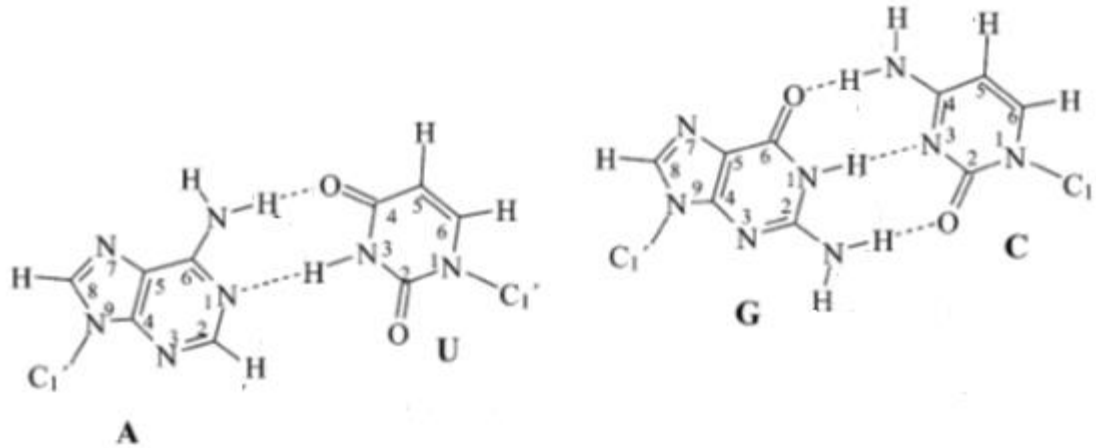
A brief excursion into NMR of RNA before more UUCG tetraloop stories.

NMR can define the base-pairing pattern. This includes standard and nonstandard Watson - Crick-type base pairs and allows verification and prediction of the secondary structure elements of RNA and determination of the basepair stability.



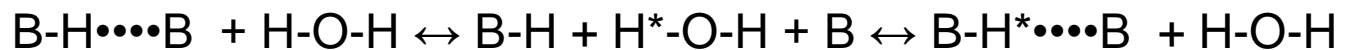
Imino protons are indicators of hydrogen bonding and therefore base-pairing.

Exchangeable vs nonexchangeable protons.



Imino and Amino protons are exchangeable. They are in equilibrium with bulk water, and their exchange is base-catalyzed. That means that at lower pH, exchange is slower.

The model for imino proton exchange:



In open-limited exchange, every time the base pair breaks, its imino proton exchanges.

U⁷ C⁸
 U⁶.....G⁹
 C⁵—G¹⁰
 A⁴—U¹¹
 C³—G¹²
 G²—C¹³
 G¹—C¹⁴

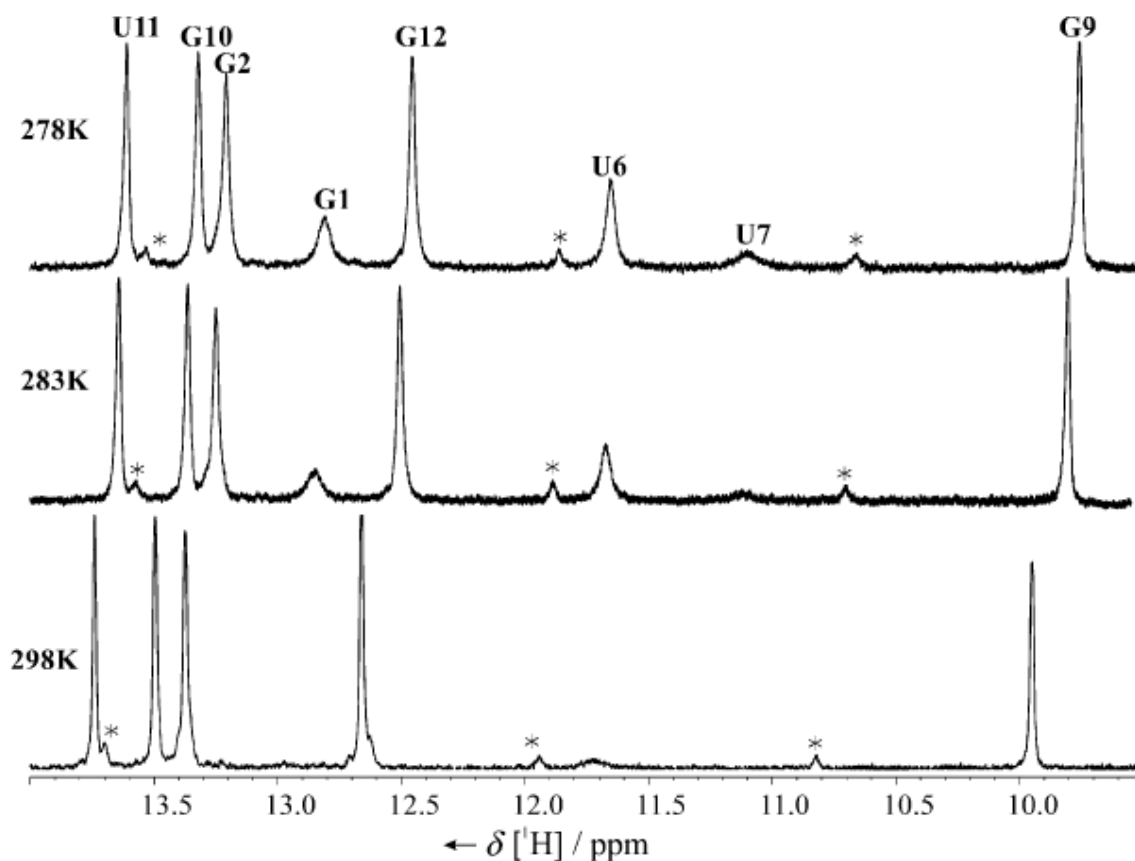
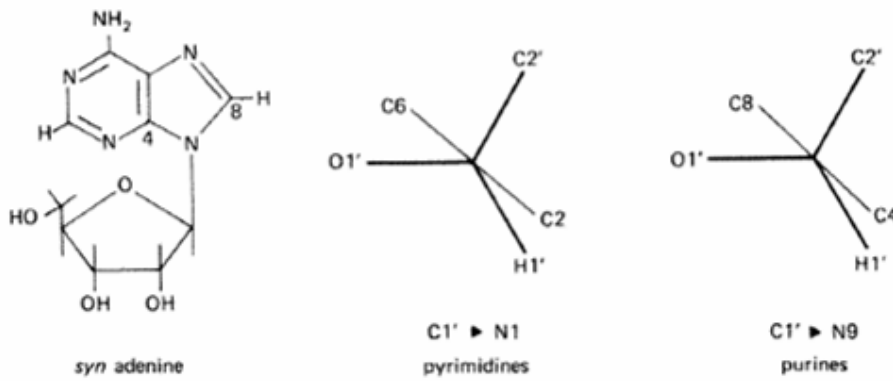
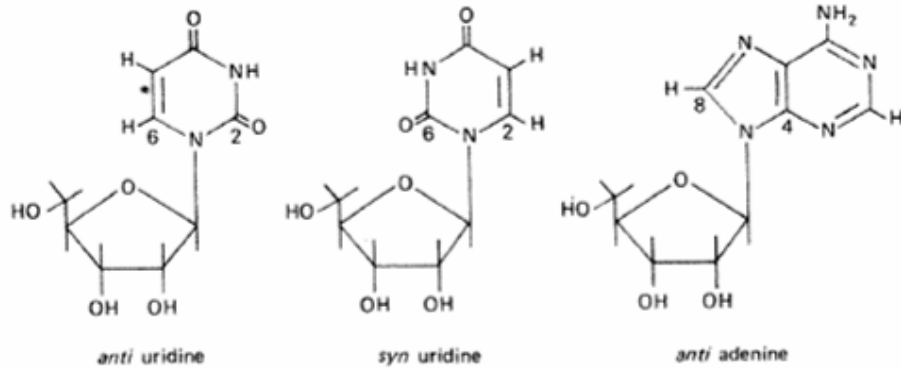
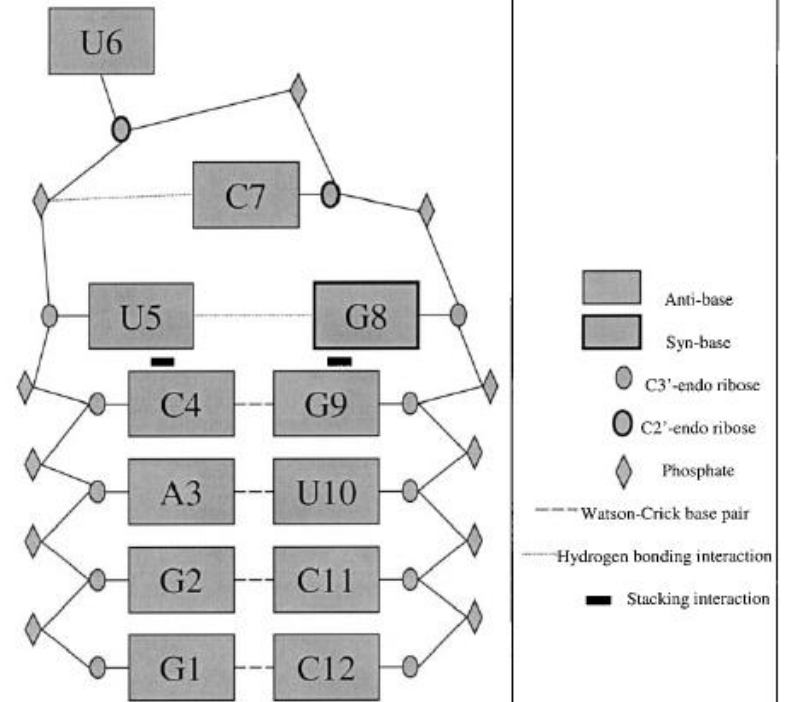


Figure 6. Imino region of the 1D ¹H spectra of the 14-mer cUUCGg tetraloop RNA recorded at 600 MHz and three different temperatures (278, 283, and 298 K). Assignments are annotated for the imino proton resonances. Resonances stemming from the duplex form are indicated with *.

Now for the LOOP.



anti ($\chi = 45^\circ$, $\phi_{CN} = -45^\circ$)



U⁷ C⁸
 U⁶.....G⁹
 C⁵—G¹⁰
 A⁴—U¹¹
 C³—G¹²
 G²—C¹³
 G¹—C¹⁴

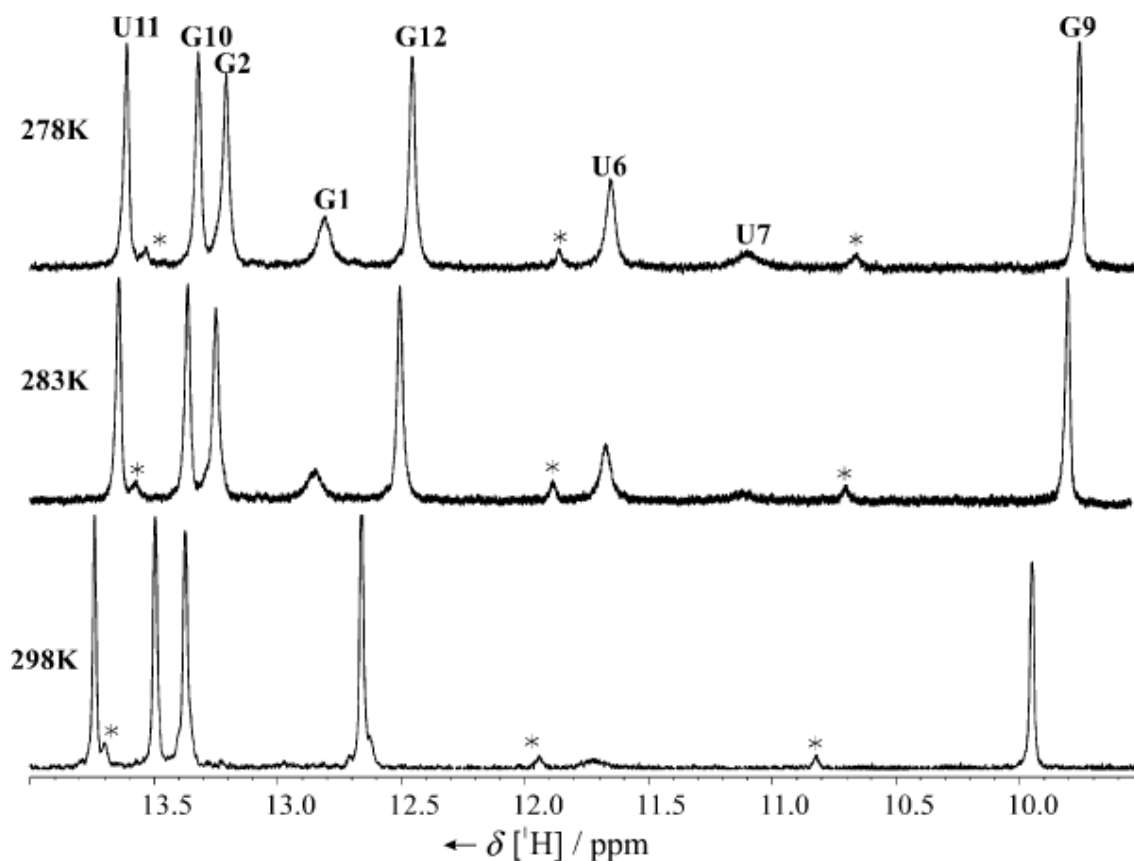
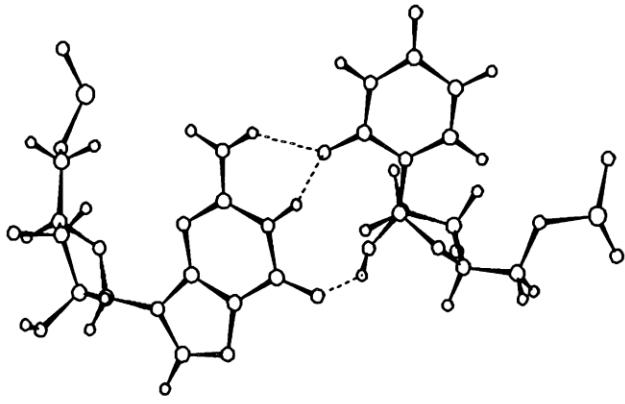
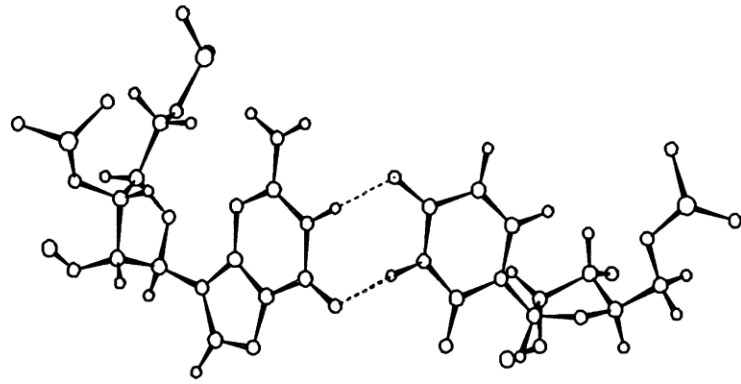


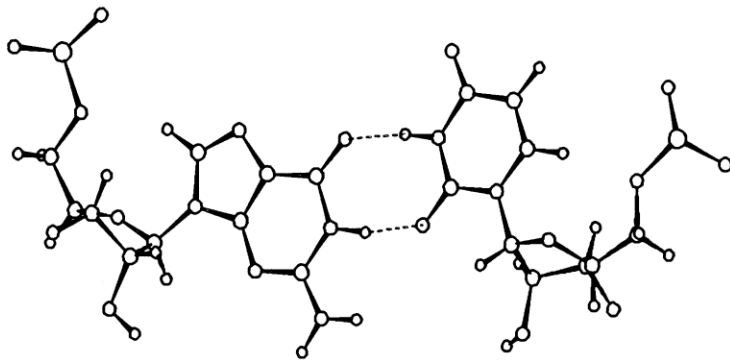
Figure 6. Imino region of the 1D ¹H spectra of the 14-mer cUUCGg tetraloop RNA recorded at 600 MHz and three different temperatures (278, 283, and 298 K). Assignments are annotated for the imino proton resonances. Resonances stemming from the duplex form are indicated with *.



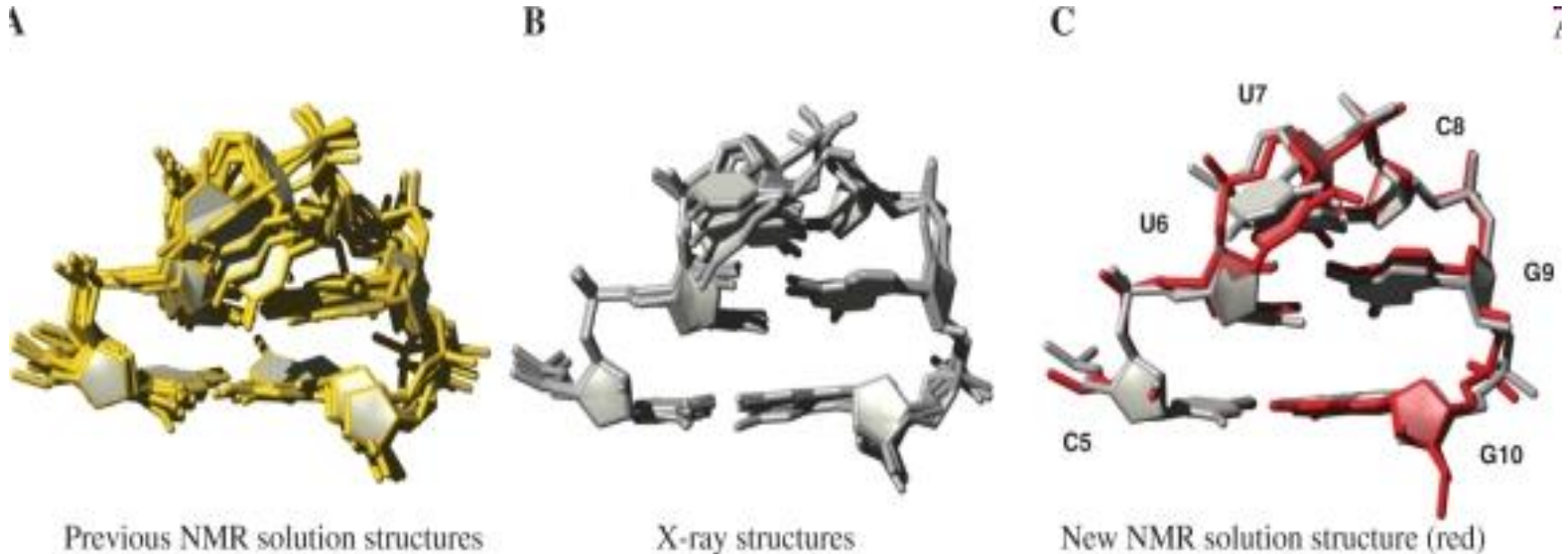
UUCG LOOP PAIR



REVERSE WOBBLE PAIR

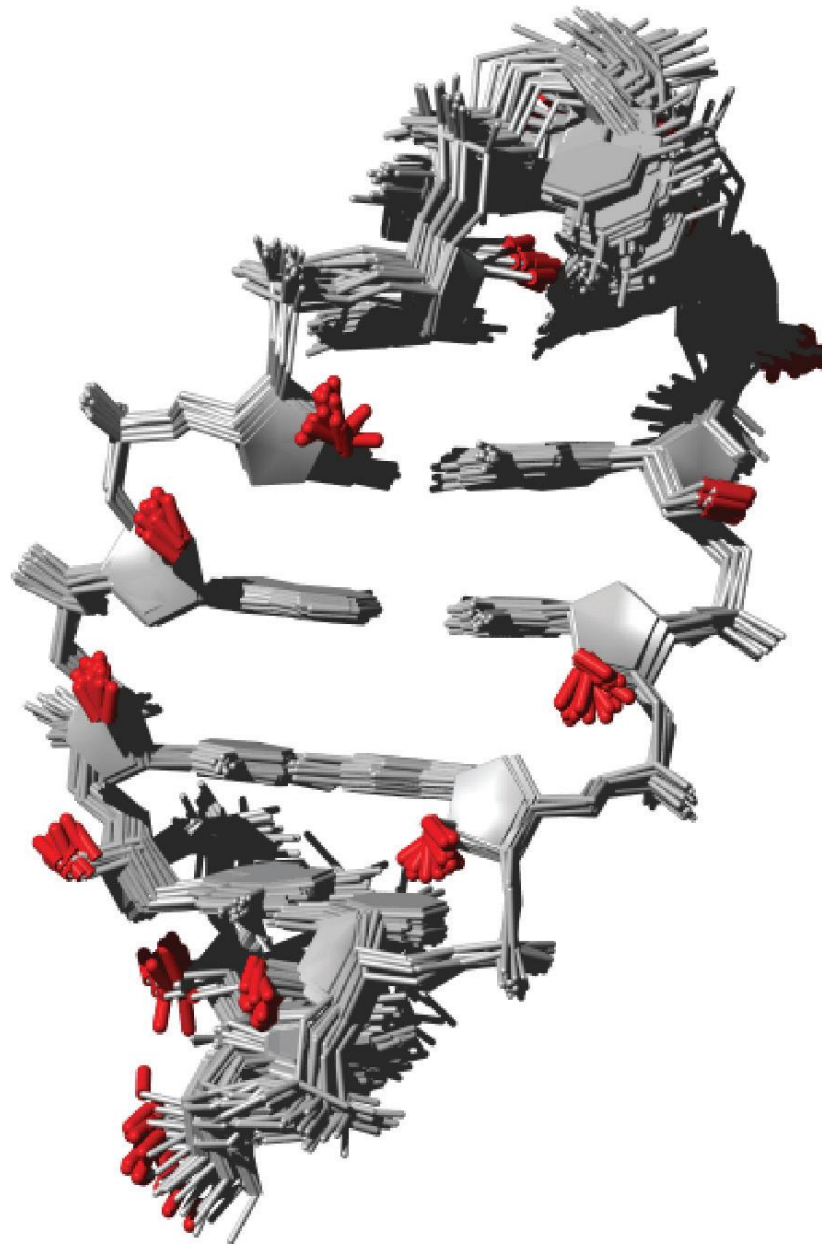
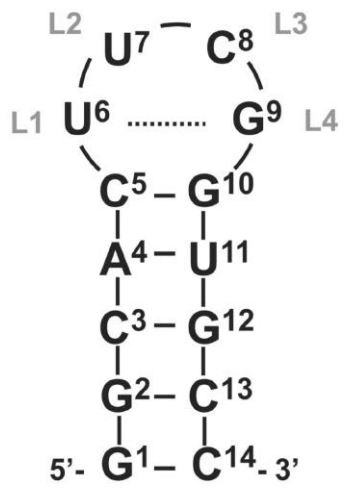


WOBBLE PAIR

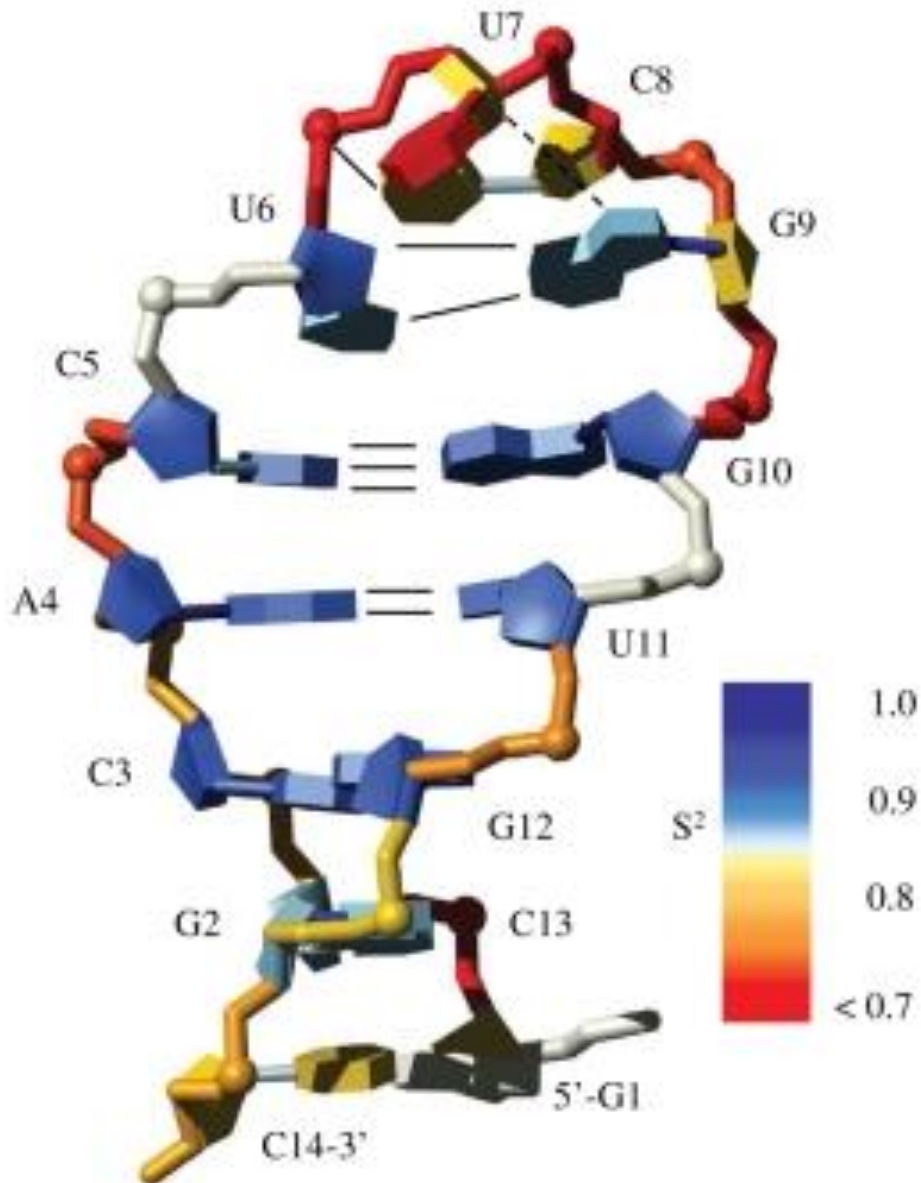


Nozinovic S, Fürtig B, Jonker HR, Richter C, Schwalbe H. High-resolution NMR structure of an RNA model system: the 14-mer cUUCGg tetraloop hairpin RNA. *Nucleic Acids Res.* 2010 Jan;38(2):683-94.

Fürtig B, Richter C, Wöhnert J, Schwalbe H. NMR spectroscopy of RNA. *Chembiochem.* 2003 4(10):936-62.



Hydroxyl groups

B

(B) Dynamics. Structure is coloured by the order parameter (S^2) determined for the sugar (C1'H1') and base (C6H6, C8H8) at 317 K and backbone (^{31}P) at 310 K as published earlier ([6](#),[10](#),[13](#)), no data for the base and backbone of G1 (grey). Hydrogen bonds are indicated by solid lines. Dashed line indicates a potential hydrogen bond between the 2'-OH of U7 and N7 of G9 that is found in a third of calculated structures.

A word about base stacking.

Stacking occurs through electronic interactions between bases.

Stacking is enthalpically driven ($\Delta H^\circ < 0$)

Entropy of stacking is unfavorable.

If stacking were hydrophobic, then it would be accompanied by a release of water and an increase in entropy, which is not observed. Stacking is not hydrophobic.

A schematic structural comparison of the two most thermodynamically stable tetraloops.

