Supplementary Material

for

High-resolution structure of stem-loop 4 from the 5'-UTR of SARS-CoV-2 solved by solution state NMR

Jennifer Vögele[1,2], Daniel Hymon[2,3], Jason Martins[2,3], Jan Ferner[2,3], Hendrik R. A. Jonker[2,3], Amanda E. Hargrove[4], Julia E. Weigand[5], Anna Wacker[2,3], Harald Schwalbe[2,3], Jens Wöhnert[1,2], Elke Duchardt-Ferner[1,2]*

[1]Institute for Molecular Biosciences, Goethe-University Frankfurt, Max-von-Laue-Str. 9, 60438 Frankfurt/M., Germany.

[2]Center for Biomolecular Magnetic Resonance (BMRZ), Goethe-University Frankfurt, Max-von-Laue-Str. 7, 60438 Frankfurt/M., Germany.

[3]Institute for Organic Chemistry and Chemical Biology, Goethe-University Frankfurt, Max-von-Laue-Str. 7, 60438 Frankfurt/M., Germany.

[4]Department of Chemistry, Duke University, Durham, North Carolina 27708, USA.

[5]Philipps-University Marburg, Department of Pharmacy, Institute of Pharmaceutical Chemistry, Marbacher Weg 6, 35037 Marburg, Germany.

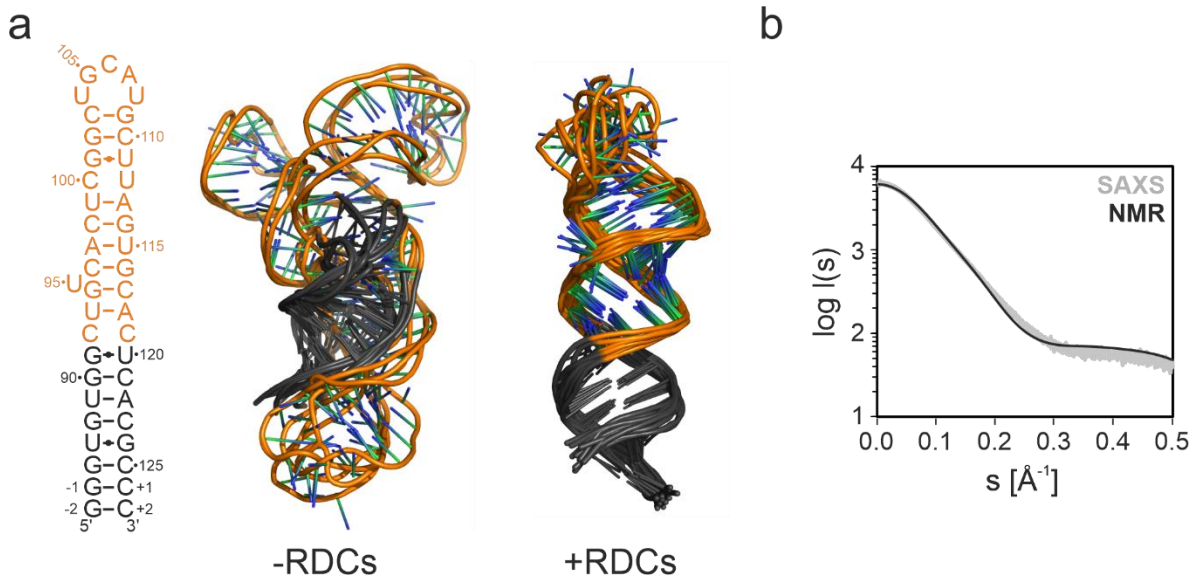
Supplementary Tables:

Supplementary Table 1.

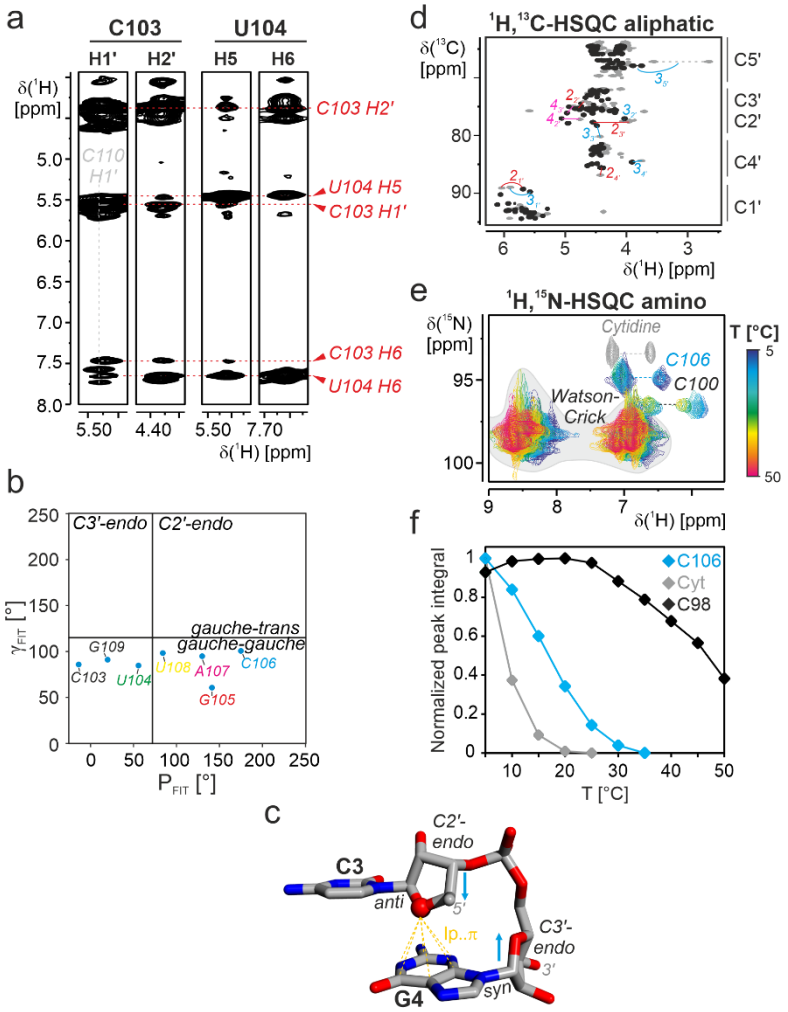
$^{3}J_{(H1',H2')}$ and $^{3}J_{(H3',H4')}$ coupling constants for residues 103 to 109 of the apical loop and the tandem mismatch TM2 extracted from a forward-directed HCCH-TOCSY-CCH-E.COSY (1,2) spectrum on 5_SL4sh. $^{3}J_{(H1',H2')}$ coupling constants < 2 Hz and $^{3}J_{(H3',H4')}$ coupling constants > 8 Hz indicate a *C3'-endo* conformation. $^{3}J_{(H1',H2')}$ coupling constants > 8 Hz and $^{3}J_{(H3',H4')}$ coupling constants < 2 Hz indicate a *C2'-endo* conformation.

Residue	$^{3}J_{(H1',H2')}$ [Hz]	$^{3}J_{(H3',H4')}$ [Hz]	**Ribose Pucker**
C103	0.3	n.d.	*C3'-endo*
U104	3.3	n.d.	mixed
G105	8.2	0.6	*C2'-endo*
C106	9.0	1.4	*C2'-endo*
A107	4.5	7.2	mixed
U108	4.7	3.8	mixed
G109	1.0	n.d.	*C3'-endo*
C100	0.7	n.d.	*C3'-endo*
G101	0.1	n.d.	*C3'-endo*
U111	0.1	n.d.	*C3'-endo*
U112	0.5	n.d.	*C3'-endo*

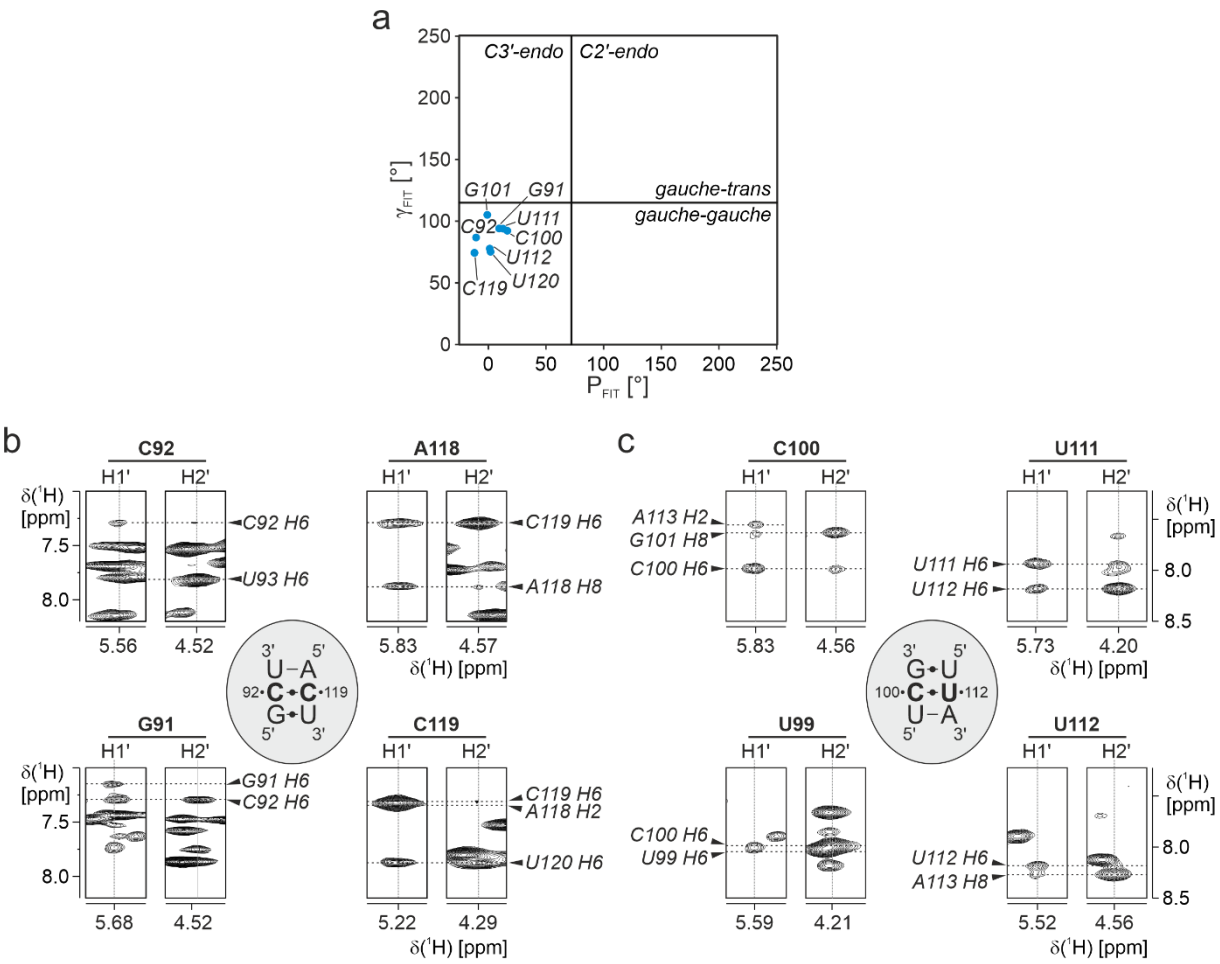
Supplementary Figures:



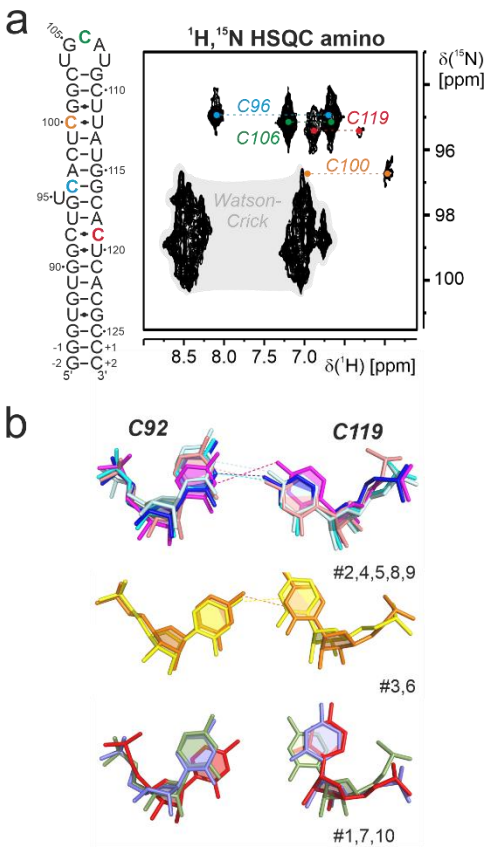
Supplementary Figure 1: a Secondary structure of 5_SL4 (left) and overlay of the 10 NMR structures with the lowest target function prior (middle) and after (right) the incorporation of NH and CH residual dipolar couplings. The lower stem (residues G_{-1} to G91; U120 to C_{+1}) is overlayed. The lower part of 5_SL4 comprising the base pairs G_{-2}-C_{+2} to G91•U120 is shown in black, the upper part is colored orange. **b** SAXS data of 5_SL4 (grey) and CRYSOL (ATSAS 3.1.3) evaluation of the lowest energy NMR structure (black).



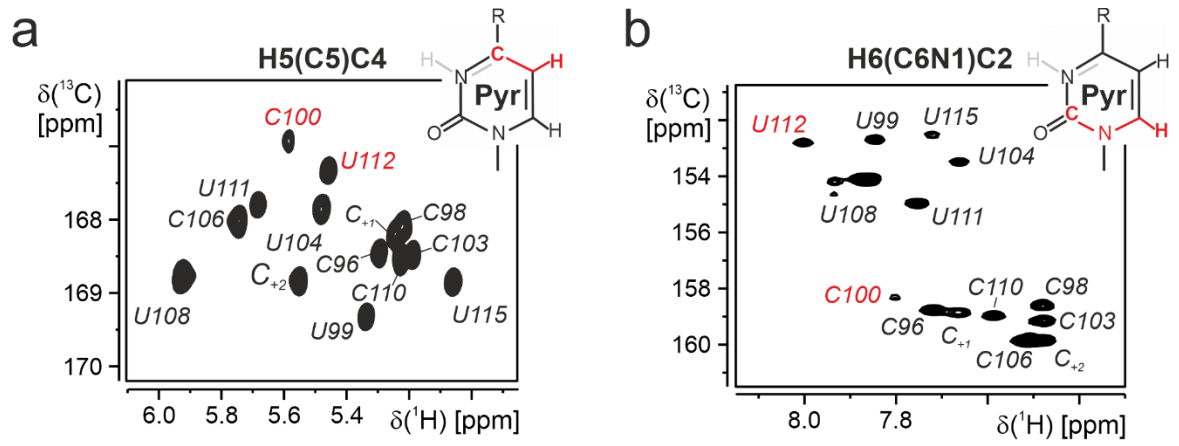
Supplementary Figure 2: a ^{1}H,^{1}H slices from 3D-^{1}H,^{13}C-NOESY-HSQC spectra at the resonance positions of C1' and C2' of C103 and C5 and C6 of U104 showing a sequential NOE cross peak pattern indicative of stacking of the two residues. **b** Canonical coordinates for the residues 103 to 109 of the apical loop and its closing base pair. **c** Z-step formed by the 3rd and 4th residue of a UUCG tetraloop (PDB-entry 2koc (3)). The characteristic features of this structural motif are indicated. Cyan arrows indicate the head-to-head orientation of the ribose moieties. Ribose puckers and glycosidic bond conformations are indicated. The nucleotides are shown in stick representation colored according to atom type: carbons are shown in grey, oxygens in red, nitrogens in blue, phosphates in orange. O4'-atoms are shown as spheres. The lone-pair...π interaction between the 5'-ribose and the 3'-nucleobase is indicated as yellow dashes. **d** Aliphatic region of a ^{1}H,^{13}C-HSQC spectrum recorded for 5_SL4sh (black) and a 14 nt stem-loop containing a UUCG tetraloop (grey) (3). Assignments for the second, the third and the 2'- and 3'-CH group of the 4th loop nucleotides of both RNAs are indicated in red, blue and magenta, respectively, and connected by solid lines. **e** Cytosine amino group region of ^{1}H,^{15}N-HSQC spectra recorded for 5_SL4sh at temperatures between 5 and 50 °C in colors from blue to red. Assignments for C106 and C100 are given. The amino region of cytosines in Watson-Crick base pairs is highlighted in grey. The amino group ^{1}H,^{15}N-HSQC spectrum of free cytidine at 10 °C is shown in grey. **f** Temperature-dependence of the amino resonance integrals of C106 (blue), free cytidine (grey) and C98 (black), which is part of a canonical Watson-Crick base pair.



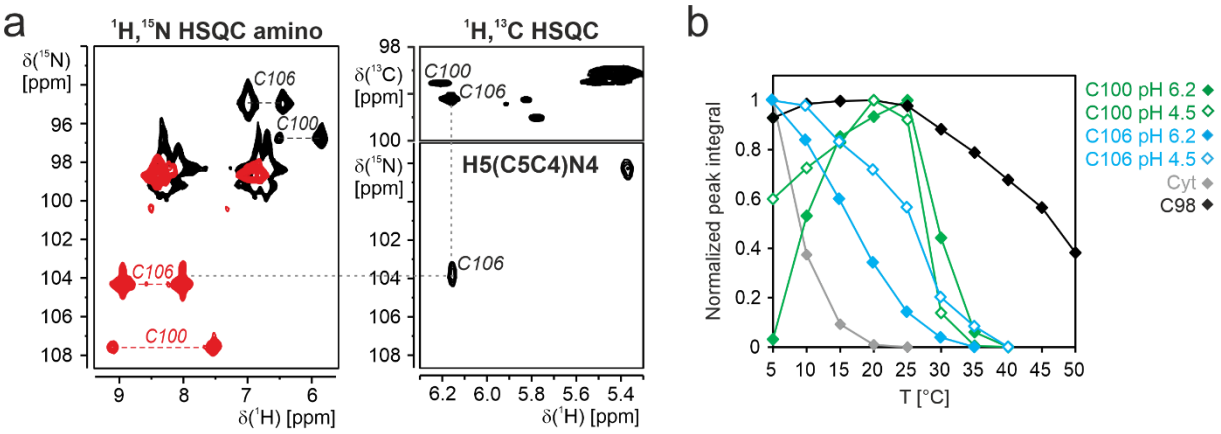
Supplementary Figure 3: a Canonical coordinates for the residues G91, C92, C119 and U120 of TM1 and C100, G101, U111 and U112 of TM2. Assignments are given. **b** ^{1}H,^{1}H slices from 3D-^{1}H,^{13}C-NOESY-HSQC spectra at the resonance positions of C1' and C2' for the residues G91, C92, A118 and C119 of TM1 showing sequential aromatic-aliphatic NOE patterns indicative of an A-form helical arrangement. Assignments are given. **c** ^{1}H,^{1}H slices from 3D-^{1}H,^{13}C-NOESY-HSQC spectra at the resonance positions of C1' and C2' for the residues U99, C100, U111 and U112 of TM2 showing sequential aromatic-aliphatic NOE pattern indicative of an A-form helical arrangement. Assignments are given.



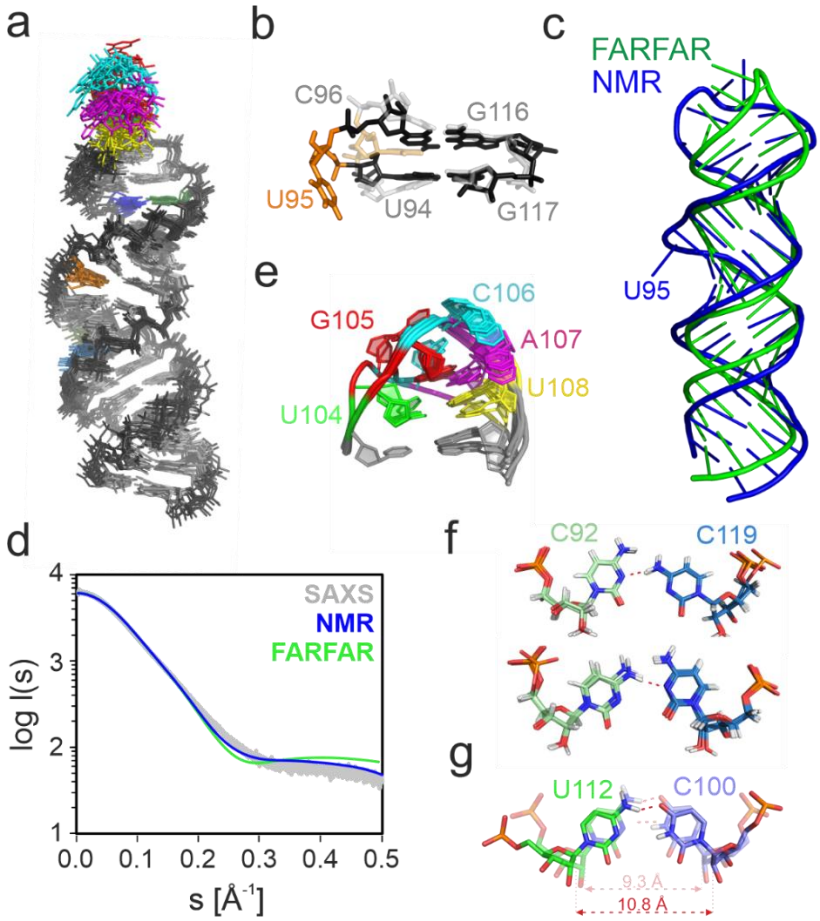
Supplementary Figure 4: a Cytosine amino group region of a ^{1}H,^{15}N-HSQC recorded for 5_SL4. Assignments for C96, C100, C106 and C119 are given. The position of these residues in the sequence is highlighted in the respective colors in the secondary structure (left). The amino region of cytidines of Watson-Crick base pairs is highlighted in grey. **b** Different conformations of the C92•C119 mismatch observed in different structures of the NMR bundle. The numbers indicate the number of the structure in the structural bundle sorted by increasing target function. Putative hydrogen bonds are indicated by dashed lines (heavy atom distance ≤ 3.0 Å).



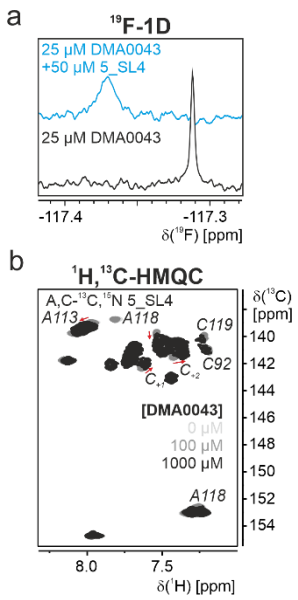
Supplementary Figure 5: a 2D-H5(C5)C4 spectrum recorded for 5_SL4sh at pH 6.2. Assignments are given. C100 and U112 are highlighted in red. The chemical structure of a pyrimidine base is displayed. The H5 and C4 atoms are shown in red. R is NH_2 for cytosine and O for uracil. **b** 2D-H6(C6N1)C2 spectrum recorded for 5_SL4sh at pH 6.2. Assignments are given. C100 and U112 are highlighted in red. The chemical structure of a pyrimidine base is displayed. The H6, N1 and C2 atoms are shown in red. R is NH_2 for cytosine and O for uracil.



Supplementary Figure 6: a Overlay of the amino region of $^{1}H,^{15}N$-HSQC spectra of 5_SL4sh at pH 6.2 (black) and pH 4.5 (red) (left). C5H5 region of a $^{1}H,^{13}C$-HSQC (right, top) and a 2D-H5(C5C4)N4 spectrum (right, bottom) recorded for 5_SL4sh at pH 4.5. Assignments for C100 and C106 are given. The pH dependence of the amino group as well as the C4 shifts (main text, Figure 3) of C106 indicates that this residue becomes protonated in the pH range between 6.2 and 4.5. A fit of the pH dependence of the C4 chemical shifts yields a pKa of 5.1±0.1, which is elevated compared to a pKa of 4.3 of free cytidine. In the main text, we suggest that the amino group of C106 forms a hydrogen bond to the phosphate group of G105 at least transiently in analogy to the YNMG(N) tetraloop fold of the 5_SL4 apical loop. In this conformation, the C106 N3 would be in close proximity to the hydrogen bond acceptor phosphate group, rendering the formation of a bifurcated hydrogen bond of both the amino and the protonated imino group of C106 to the G105 phosphate favorable. **b** Temperature-dependence of the amino resonance integrals of C100 at pH 6.2 (filled green diamonds) and pH 4.5 (open green diamonds), C106 at pH 6.2 (filled cyan diamonds) and pH 4.5 (open cyan diamonds), free cytidine at pH 6.2 (grey diamonds) and C98 at pH 6.2 (black diamonds). In contrast to the amino group of C100, the amino group of C106 is stabilized at lower pH. It follows that in this case the additional hydrogen bond provided by the protonated N3 group results in a stabilization.



Supplementary Figure 7: a Overlay of the 10 lowest energy structures of 5_SL4 predicted by FARFAR2. The RNA is shown in stick representation. Non-canonical residues are highlighted according to Figure 2b of the main text. **b** Overlay of U95 and its two adjacent base pairs of the NMR structure with the lowest target function (intense colors) and the lowest energy structure predicted by FARFAR2 (pale colors). Residues are assigned. **c** Overlay of the NMR structure with the lowest target function (blue) and the lowest energy FARFAR2 structure (green). **d** SAXS data of 5_SL4 (grey) and CRYSOL (ATSAS 3.1.3) evaluation of the NMR structure with the lowest target function (blue) and the lowest energy FARFAR2 structure (green). **e** Overlay of the apical loop region of the 10 lowest energy structures predicted by FARFAR2 in cartoon representation. Residues are assigned. **f** Different conformations of the C92•C119 mismatch predicted by FARFAR2. Residues are assigned. **g** Overlay of the C100•U112 mismatch of the NMR structure with the lowest target function (intense colors) and the lowest energy structure predicted by FARFAR2 (pale colors). The C1'-C1' distances are given.



Supplementary Figure 8: a Binding of 5_SL4 to DMA0043 followed by ^{19}F 1D spectra of the compound DMA0043. To 25 µM compound (black spectrum) 50 µM 5_SL4 RNA have been titrated (blue spectrum). The experiments were carried out on 600 MHz Bruker Neo NMR-spectrometer equipped with a 5-mm cryogenic quadruple resonance QCI probe. Measurements were performed at 25 °C. **b** Titration of DMA0043 to 5_SL4 followed by sf-^{1}H,^{13}C-HMQC spectra of the aromatic pyrimidine C6-H6 and purine C8-H8 and C2-H2 moieties using a selectively A,C-^{13}C,^{15}N-labeled sample. The spectra of the different titrations steps are shown in different shades of grey as indicated. For residues displaying significant chemical shift perturbations in the course of the titration, assignments are given and the peak shifts with increasing ligand concentration are indicated by arrows.

References

1. Glaser, S.J., Schwalbe, H., Marino, J.P., Griesinger, C. (1996). Directed TOCSY, a method for selection of directed correlations by optimal combinations of isotropic and longitudinal mixing. J Magn Reson B, 112 (2), 160–180.
2. Schwalbe, H., Marino, J.P., Glaser, S.J., Griesinger, C. (1995). Measurement of H,H-Coupling Constants Associated with.nu.1,.nu. 2, and.nu.3 in Uniformly 13C-Labeled RNA by HCC-TOCSY-CCH-E.COSY. J Am Chem Soc, 117 (27), 7251–7252.
3. Nozinovic, S., Fürtig, B., Jonker, H.R. A., Richter, C., Schwalbe, H. (2010). High-resolution NMR structure of an RNA model system: the 14-mer cUUCGg tetraloop hairpin RNA. Nucleic Acids Res, 38 (2), 683–694.