Supplementary information

NMR:

Table S1: Structural statistics for the 10 refined structures of the 100 structures calculated by torsion angle dynamics

Averaged total energy (kcal.mol^{-1})	-69557 +/- 1098
Distances	**217**
Intra-residual NOE	108
Inter-residual NOE	95
Hydrogen Bonds	14
Dihedral angles	**123**
Backbone	63
χ	15
Ring pucker	45
Restraints violations	
NOE (>5.0 Å)	0
Angles (> 5°)	0 - 1
RMS Deviation	
For NOE restraints (Å)	0.056 +/- 0.002
For dihedral restraints (°)	0.748 +/- 0.154
RMSD from mean structure	
Backbone	0.564+/- 0.132

Table S2: Average backbone torsion angles and standard deviations (between parentheses) were calculated from an ensemble of 10 structures. Values that deviate to accommodate the internal loop are indicated in bold. Backbone angles measured in the model calculated for the complex are indicated in italic.

Resid.	Alpha		beta		gamma		delta		epsilon		zeta	
1					61.7	(1.8)	89.1	(1.3)	207.6	(7.3)	286.6	(5.8)
					71.7		*76.8*		*188.9*		*290.4*	
2	289.2	(4.9)	175.9	(10.3)	60.9	(3.2)	87.9	(1.6)	200.6	(5.0)	285.3	(5.2)
	278.0		*175.0*		*62.2*		*82.2*		*195.5*		*283.5*	
3	291.9	(5.5)	171.8	(7.8)	63.3	(2.0)	88.0	(1.7)	200.9	(4.1)	286.6	(4.1)
	273.5		*179.8*		*69.0*		*78.8*		***171.9***		***78.0***	
4	299.5	(2.9)	**164.1**	(5.2)	67.7	(2.2)	87.3	(1.4)	203.5	(3.7)	298.0	(5.4)
	170.5		***89.8***		***186.7***		*82.2*		*195.6*		*286.8*	
5	288.6	(3.4)	193.2	(4.8)	57.8	(2.2)	87.8	(1.4)	220.1	(5.6)	285.7	(5.6)
	258.6		*188.3*		*61.1*		*80.7*		*225.8*		281.8	
6	286.7	(3.8)	177.5	(7.3)	58.4	(2.6)	89.4	(2.5)	214.6	(8.4)	277.0	(12.4)
	299.7		*167.0*		*44.8*		*74.9*		*200.3*		*303.1*	
7	131.4	(6.3)	213.7	(16.0)	178.7	(3.0)	92.1	(1.4)	217.3	(10.7)	280.2	(6.0)
	262.3		*194.8*		*52.3*		*82.8*		*235.0*		*292.6*	
8	159.0	(5.1)	155.1	(12.1)	60.0	(2.0)	93.7	(1.7)	191.6	(3.4)	284.4	(7.7)
	174.7		*153.2*		*51.3*		*90.1*		*224.7*		*277.7*	
9	142.7	(2.9)	215.9	(8.1)	173.2	(2.6)	98.8	(2.9)	221.8	(5.0)	306.9	(4.6)
	293.1		*164.5*		*73.1*		*93.4*		*199.6*		*291.1*	
10	280.0	(9.5)	170.7	(7.4)	55.2	(2.1)	89.5	(2.7)	228.6	(2.9)	262.8	(4.7)
	289.4		*168.1*		*57.3*		*89.4*		*236.3*		*305.0*	
11	315.3	(12.9)	127.8	(1.6)	70.7	(4.5)	85.5	(1.9)	193.5	(2.4)	291.7	(1.8)
	299.0		*158.6*		*61.6*		*81.1*		*204.3*		*298.5*	
12	286.4	(2.3)	167.6	(3.1)	64.7	(2.2)	84.3	(1.6)	189.1	(2.4)	288.1	(3.0)
	287.8		*171.2*		*70.6*		*75.6*		*203.6*		*295.0*	
13	290.7	(3.5)	180.5	(4.3)	59.9	(1.2)	85.8	(1.0)	216.0	(6.8)	**298.6**	(5.0)
	295.4		*169.8*		*61.3*		*84.6*		*205*.3		*281.7*	
14	299.0	(5.0)	**162.6**	(3.4)	64.3	(1.8)	89.5	(1.6)	**169.2**	(3.0)	**263.2**	(2.6)
	295.6		***158.9***		*61.6*		***69.6***		*176.7*		***268.4***	
15	288.5	(4.7)	174.1	(3.6)	60.5	(2.5)	86.7	(1.9)	196.0	(3.3)	284.2	(6.9)
	297.0		157.9		*63.6*		*77.1*		*187.6*		*286.3*	
16	292.5	(4.4)	173.1	(6.5)	63.4	(2.6)	89.4	(1.4)	198.8	(4.2)	286.6	(3.8)
	155.9		184.3		***189,0***		*104.2*		*217.2*		*299.6*	
17	285.0	(4.6)	173.8	(6.4)	60.7	(1.1)	88.2	(1.1)				
	275.5		176.8		*56.1*		*86.4*					

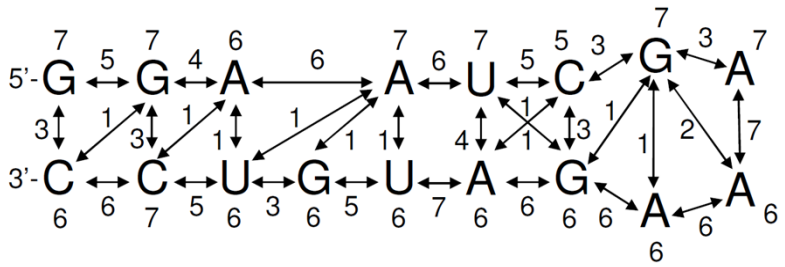


Figure S1: Distribution of distance restraints extracted from the NOE.

Figure S2: Section of the 2D NOESY spectrum in 90%H_2O and 10%D_2O. Assignment of sharp imino signals is added for the 1D spectrum at the top. The box in dashed lines indicates interaction between G11:H1 and U5:H3.

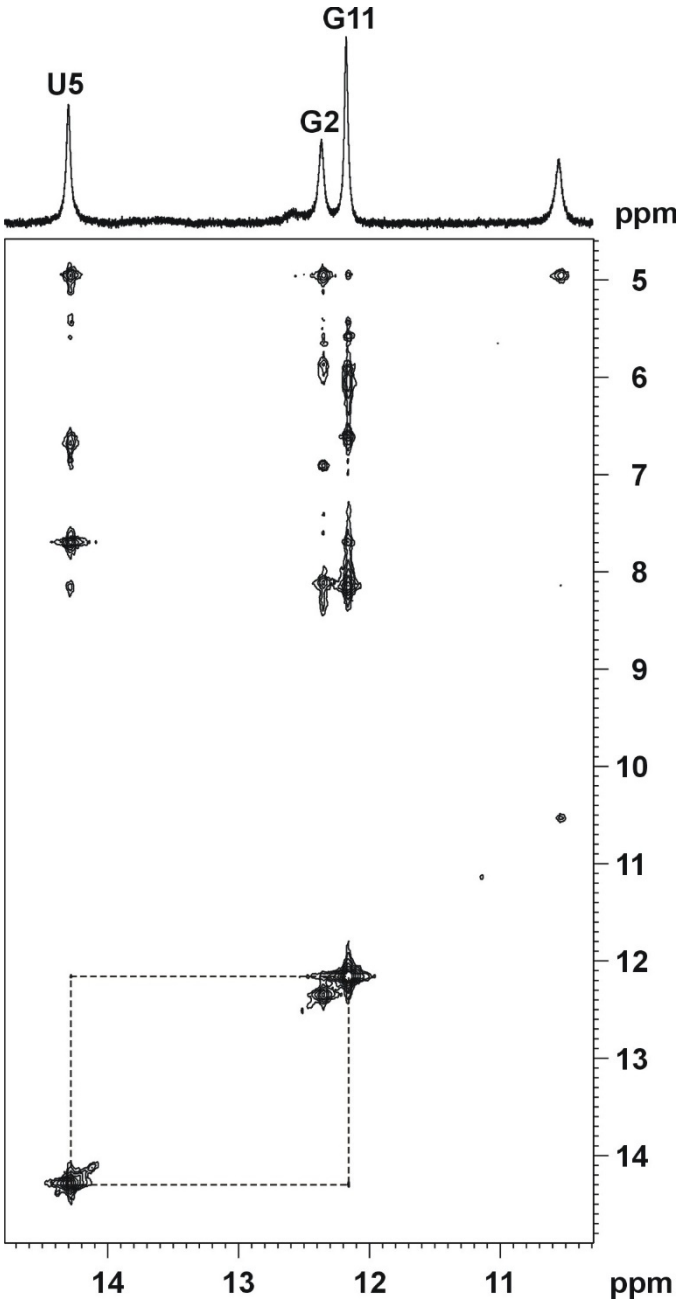


Figure S3: Overlay of corresponding sections from 2D HMBC (black contours) and HSQC (red contours) spectra of the RNA 17mer. HSQC signals from purine H8 and pyrimidine H6 are labeled in red. Dashed lines in red correlate for each adenine the H8 and H2 signals through their interaction with C4 in HMBC. Dashed lines in green correlate uridine H6/C6 signals in HSQC to their C4 in HMBC.

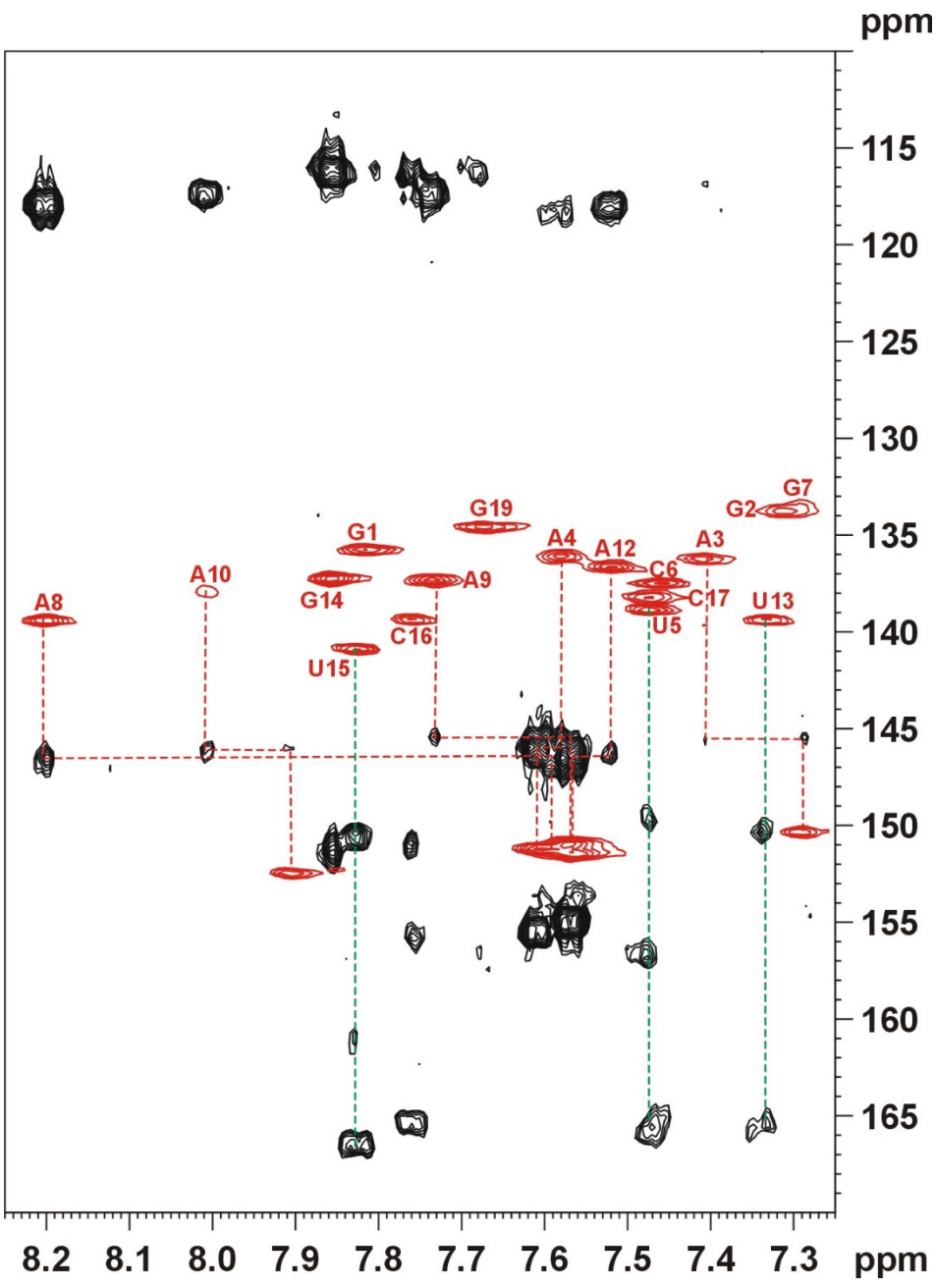
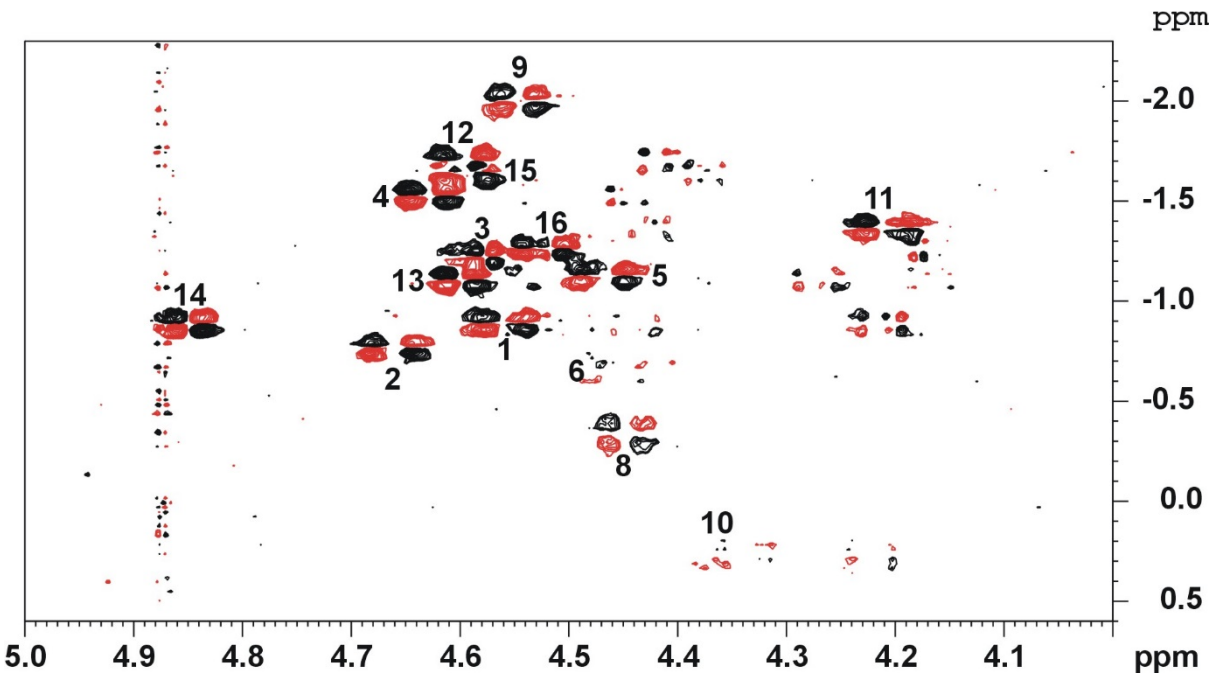
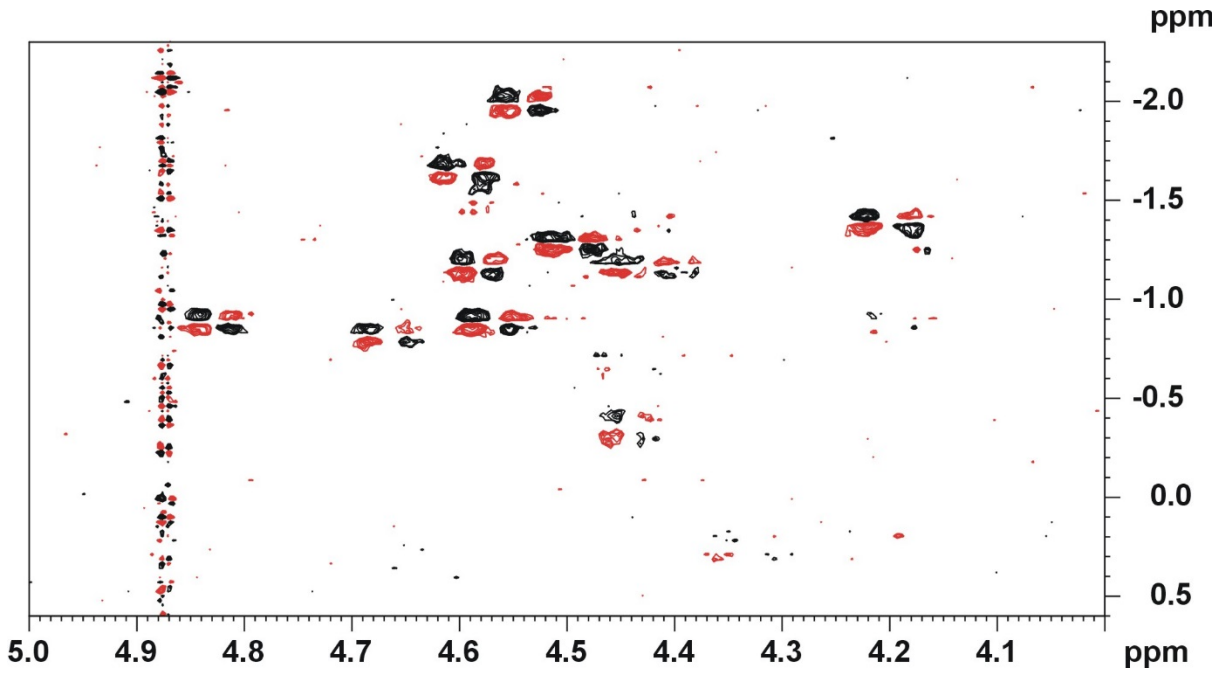


Figure S4: 2D ^{1}H-detected [^{1}H,^{31}P]HETCOR of the RNA oligomer prior (top) and after (bottom) adding the ligand. H3' to P cross peaks are labeled with the residue number of H3'.





Molecular modeling & simulation:

Molecular dynamics simulations with the AMBER14 software[1] were performed on the unliganded RNA hairpin (Scheme S1), the liganded hairpin with compound 1 and a double helical structure with G bulge (Scheme S2) with ligand 2.[2] The parameters for RNA structure was taken from the Amber ff14SB force field.[3] Small ligand molecules were parametrized using the general Amber force field (gaff).

The unliganded hairpin structure.

The average experimental NMR structure (from this study) was used as starting model. It was solvated in an octahedral TIP3P[4] water box followed by neutralization with Na^+ ions, resulting in periodic simulation systems with 10112 atoms for the system. The particle mesh Ewald method[5] was used to handle electrostatic interactions using a 10Å cutoff with default parameters. Lennard-Jones interactions were also treated with a 10Å cutoff. A short energy minimization was done to remove possible clashes with solvent molecules. A heating MD simulation was then started (50ps, to 300K). The loop structure atom positions were restrained using a force constant of 2.0 kcal/(mole*Å). In all MD simulations a time-step of 2 fs was used while keeping all covalent bonds to hydrogen atoms fixed (SHAKE).[6] The restraints were then removed and the system was allowed to equilibrate first at constant volume (NVT, 50ps) and then at constant pressure (NPT) during 500ps. Then the system was simulated for an additionally 20 ns using the PMEMD program on a linux workstation equipped with 2 NVIDIA Titan Black GPU cards. Average linkage clustering in 10 clusters based on root mean square deviation (rmsd) of the atomic positions was performed in cpptraj (Table S3).[7, 8]

The Nuchemics software was used to calculate proton NMR chemical shifts for the centroid structures of the 10 clusters (van der Werf 2013).[9] A simple linear regression y = ax+b was used to compare the shifts from residues Ade3, ade4, Uri13, Gua14, Uri15 (in the neighborhood if the G bulge) with the experimentally obtained chemical shifts (Table S4). For comparison, also the regression of the 10 NMR structures is given in table S5. The Nuchemics results for the G bulge region in Table S4 and Table S5 show that the centroid structure from cluster 0, the largest one, is most similar to the NMR structures. However, the fit of the Nuchemics results for the average NMR structure (NMR structure 3 is closest to the NMR average structure) are still better than the centroid structures from the simulation. A Nuchemics calculation describes the structure locally. Those local differences for the NMR structure and MD structures may be explained by: 1) Table S6 shows the stacking interactions for G14 with surrounding nucleotides. Small differences are seen in the stacking pattern of G14 between the NMR average and cluster centroid structures which may contribute to the different proton environment resulting in different chemical shifts calculated. 2) The MD simulation was only 20ns long. A longer simulation may be needed to sample fully the NMR structure conformation. 3) Although the ff14SB force field was used in the simulations, this force field is probably not perfect yet to describe the RNA accurately. Force field improvement is continuously in progress.[10] Another observation is that in the average NMR structure G14 flips out of the helical stem while in the MD simulation this G is more turned in the helical stem, hydrogen bonding with A3.

From the superposition of the centroid structure C0 from the MD simulation on the experimental NMR structure by the R3D Align server (Figure S6)[11] it is seen that the stem region of the MD structure and NMR structure do not fit very well, although in both structures a correct Watson-Crick base pairing pattern is observed. An analysis of the stem (GGA:CCU) by curves+[12] points to a higher inclination for the MD structure (10.5) than for the NMR structure (2.5).

The stabilizing elements of the GAAA loop are given in Table S6. The A10-A11-A12 stacking is maintained in the MD structures. However, the expected stabilizing hydrogen bonds across the GAAA tetraloop are poorly sampled during the 20 ns MD simulation. The loop spanning stabilizing hydrogen bonds that are present in the NMR structure (and also in a tetraloop structure 1ZIF)[13] are missing in many clusters (table S6). Clusters C0, C5, C6 and C7 show some cross loop hydrogen bonds but different from the one from the NMR structures. Cluster C1, C2, C3 and C8 show the G7 nucleotide turned away from the loop towards A4 making there a hydrogen bond with the phosphate group. These simulations prove again that force fields are continuously improving over the years but are not perfect yet in reproducing the experimental comformations.[10]

*Hairpin +ligand **1** MD simulation*.

Additionally, two more MD simulations were performed with the hairpin structure and ligand **1**. One simulation starting from the average NMR structure in which molecule **1** was docked using Autodock4.2 (figure S5)[14]. As seen in this figure, the optimal docked structure has hydrogen bonding from the N2 of G14 to compound1 in stead of a Watson-Crick like basepairing as shown in figure 5 in the main manuscript. To make this possible G14 has to be flipped out slightly.

Another simulation starting from one where the double ring system of compound1 is 180 degrees flipped around the bond between the aromatic ring system and the amide nitrogen was also started. The second simulation became

unstable immediately. The first simulation however was stable for at least 20 ns. The largest cluster of 10 after clustering had 3100 structures of 6000. The hydrogen bonding of figure S5 was present for 48, 33, 25% of the simulation time for Gua N2 to the amide N, Gua N2 to amide CO and Gua N2 to the ring Nitrogen, corresponding to the largest cluster, spread out over all the simulation time. (Prod20.pdb is part of that firat cluster)

*Double helix with G bulge + ligand **2** MD simulation.*

Finally a simulation of a double helical structure with compound **2** having Watson-Crick base pairing similar as in figure 5 of the main manuscript was performed. This simulation keeps compound **2** in the cavity for about 3 ns but then the G14 started turning out of the helix making the WC basepairing impossible and leaving the compound unrestrained. The compound however keeps stacked between dG6:dC15 and dC5:dG17.

The G in the bulge in the double helix is even flipped out more than in the loop: trying to model compound **2** having the hydrogen bonding as in figure S6 puts the compound outside the helix, without stacking interactions.

Table S3. Average linkage clustering of the 20 ns MD trajectory generating 10 clusters.

#cluster	Frames	Frac	AvgDist	StDev	Centroid	AvgCDist
0	2048	0.205	1.439	0.322	1147	2.426
1	1968	0.197	1.317	0.284	5.980	2.572
2	1436	0.144	1.401	0.313	7551	2.215
3	1210	0.121	1.281	0.302	8754	2.309
4	818	0.082	1.266	0.303	9607	2.268
5	767	0.077	1.405	0.344	2240	3.095
6	737	0.074	1.481	0.330	2697	2.362
7	474	0.047	1.413	0.330	3562	2.233
8	440	0.044	1.301	0.306	7381	2.507
9	102	0.010	1.082	0.266	9299	2.719

Table S4. Linear regression fit y=ax+b of the Nuchemics chemical shift calculations for residues A3, A4, U13, G14 and U15 for different centroid structures of the clusters compared to the experimental NMR chemical shifts.

#cluster	slope	intercept	R2
0	0.955559	0.0940987	0.939353
1	0.997454	-0.0354726	0.929674
2	0.988471	-0.0309243	0.856705
3	1.02188	-0.146312	0.86629
4	1.06054	-0.313755	0.891289
5	1.06132	-0.508222	0.894202
6	1.02237	-0.149627	0.919078
7	1.06028	-0.379927	0.919877
8	1.06024	-0.382907	0.856888
9	1.03259	-0.185898	0.907587

Table S5. Linear regression fit y=ax+b of the Nuchemics chemical shift calculations for residues A3, A4, U13, G14 and U15 for different NMR model structures compared to the experimental NMR chemical shifts.

#	Slope	Intercept	R2
1	1.05022	-0.306469	0.933680
2	1.06216	-0.320136	0.954285
3	1.04215	-0.236304	0.962027
4	1.05096	-0.270173	0.956723
5	1.02534	-0.168624	0.954644
6	1.06893	-0.326383	0.967120
7	1.03631	-0.211635	0.964518
8	1.03690	-0.213320	0.966656
9	1.05461	-0.270224	0.955711
10	1.04704	-0.230811	0.964724

Table S6. Some hydrogen bonds and stacking interactions in different structures (1ZIF, average NMR styructure, centriod structures of 10 clusters from 20 ns MD simulation) *Numbering of this loop in 1ZIF: G5-A6-A7-A8; hb = hydrogen bond; st = stacking.

Structure	1ZIF*	NMR average	C0	C1	C2	C3	C4	C5	C6	C7	C8	C9
hb G7.O2‘-A9.N7	2.4	2.8										
hb G7.O2‘-A9.N6	2.6											
hb G7.O2‘-A9.OP2			2.6					2.6	2.8	3.3		
hb G7.O2‘-A10.N6			3.3							3.1		
hb G7.O4‘-A10.N6	3.2											
hb G7.N2-A9.OP1/2		3.0							2.8	2.9		
hb G7.N2-A10.OP2	2.9											
hb G7.N1-A4.OP2				3.0	2.8	3.3					2.8	
hb G7.N2-A4.OP2					3.0						3.3	
st A8-A9	+	+	+	+	+	+	+	+	+	+	+	+
st A9-A10		+	-	+	+	+	+	+	-	-	+	+
st A10-G11		+	+	+	+	+	+	+	+	+	+	-
st A4-G14		-	+	+	+	+	+	+	+	+	+	+
st G14-U13		+	-	+	+	+	+	-	+	+	+	+
st G14-U15		-	+	+	+	+	+	+	+	+	+	+

Scheme S1: Hairpin structure:

5’ - G - G - A- - - - -A - U - C - G - A

||| ||| || || || ||| >

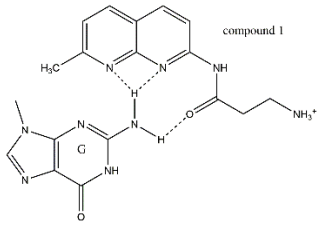
3’ - C - C - U - G - U - A - G - A – A

Scheme S2: Double helix with G bulge:

5’-dT- dC - dC- dA – dG[5] - - - dG[6] - dC - dA - dA - dC - 3’

|| ||| ||| || ||| ||| ||| || || |||

3’-dA - dG - dG -dT- dC - dG - dC - dG - dT - dT - dG - 5’



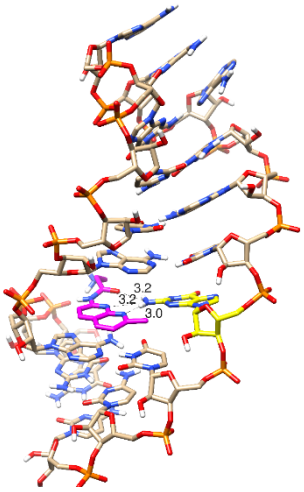
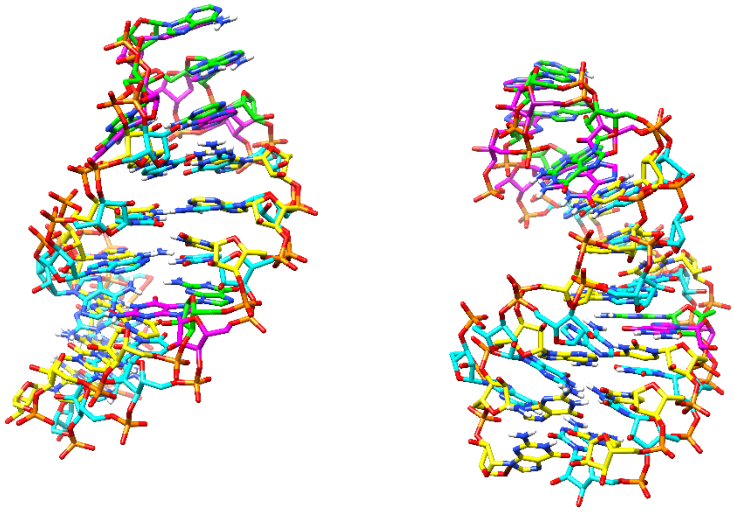


Figure S5. Interaction of compound **1** with G14.

Figure S6. Superposition of MD structure from cluster0 onto the NMR structure by R3Dalign (front and side view). The NMR structure has yellow carbons and ribbon (G14 and the GAAA motif starting at residue 7 have green carbon atoms) while the MD structure has cyan carbon atoms and ribbon (G14 and the GAAA hairpin carbons are magenta). Image made by Chimera.[15]



References.

1. A. Pérez, I. Marchán, D. Svozil, J. Sponer, T. E. Cheatham, 3rd, C. A. Laughton and M. Orozco, *Biophys J*, 2007, **92**, 3817-3829.
2. K. Nakatani, S. Sando and I. Saito, *J. Am. Chem. Soc.*, 2000, **122**, 2172-2177.
3. D. A. Case, J. T. Berryman , R. M. Betz, D. S. Cerutti, I. Cheatham, T.E., T. A. Darden, R. E. Duke, T. J. Giese, H. Gohlke, A. W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T. S. Lee, S. LeGrand, P. Li, T. Luchko, R. Luo, B. Madej, K. M. Merz, G. Monard, P. Needham, H. Nguyen, H. T. Nguyen, I. Omelyan, A. Onufriev, D. R. Roe, A. Roitberg, R. Salomon-Ferrer, C. L. Simmerling, W. Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu, D. M. York and P. A. Kollman, *Amber 2014*, 2014, University of California, San Francisco. .
4. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926-935.
5. U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577-8593.
6. J.-P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327-341.
7. J. Shao, S. W. Tanner, N. Thompson and T. E. Cheatham, *J. Chem. Theory Comput.*, 2007, **3**, 2312-2334.
8. D. R. Roe and T. E. Cheatham, *J. Chem. Theory Comput.*, 2013, **9**, 3084-3095.
9. R. M. van der Werf, M. Tessari and S. S. Wijmenga, *J Biomol NMR*, 2013, **56**, 95-112.
10. A. A. Chen and A. E. García, *Proc. Natl. Acad. Sci. USA*, 2013, **110**, 16820-16825.
11. R. R. Rahrig, A. I. Petrov, N. B. Leontis and C. L. Zirbel, *Nucleic Acids Res.*, 2013, **41**, W15-21.
12. C. Blanchet, M. Pasi, K. Zakrzewska and R. Lavery, *Nucleic Acids Res.*, 2011, **39**, W68-73.
13. F. M. Jucker, H. A. Heus, P. F. Yip, E. H. Moors and A. Pardi, *J. Mol. Biol.*, 1996, **264**, 968-980.
14. G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell and A. J. Olson, *J. Comput. Chem.*, 2009, **30**, 2785-2791.
15. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *J. Comput. Chem.*, 2004, **25**, 1605-1612.