Supporting Information

Elucidation of Catalytic Strategies of Small Nucleolytic Ribozymes from Comparative Analysis of Active Sites

Daniel D. Seith,[1,‖] Jamie L. Bingaman,[1,‖] Andrew J. Veenis,[1] Aileen C. Button,[1,2] Philip C. Bevilacqua[1,3,*]

[1]Department of Chemistry and Center for RNA Molecular Biology, The Pennsylvania State University, University Park, Pennsylvania 16802

[2]Department of Biochemistry, The University of Vermont, Burlington, Vermont 05405

[3]Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, Pennsylvania 16802

[‖]These two authors contributed equally to this work.

*Corresponding author: pcb5@psu.edu

Contents

Supporting Methods

PyMOL Plugin and Python Script Details

Python code was written for use with Python 2.7.

A.) γ, β, and δ Scissile Phosphate Plugin

1. Protons are added to all relevant atoms using PyMOL's cmd.h_add function.
2. All atoms within 5 Å of the AOI are searched for. Their distance to the AOI, atom name, residue name, residue number, contact angle, b-factor, and chain letter are appended to an array.
3. Angles are gathered for nitrogen CAs that are not rotatable. For exocyclic CAs, angles are measured using the Cartesian coordinates of the CA's associated protons (Figure S1). For endocyclic CAs, the method detailed in Step 4 is used.
4. If the CA is an endocyclic nitrogen, the coordinates of where its proton would be are calculated. For example, if the CA is the N1 of an adenine, the following will occur (See Figure S1). The atoms neighboring CA (NA1 and NA2) are searched for, and their coordinates are recorded. If they cannot be found, an angle of zero is reported. If they are found, a parametric equation for a line connecting the points NA1 and NA2 is created. The coordinates for the midpoint of this line are then determined. An equation for the line connecting the midpoint of this line and CA is then created. This line is then used to calculate the coordinates of a point 1 Å from CA, representing the hypothetical proton. The angle is then determined using the Cartesian coordinates of CA, the calculated coordinates just described, and the coordinates of the AOI. Figure S1 serves as a reference for this method.
5. If neither of the above two methods detailed in Steps 3 and 4 are possible, an angle of "N/A" is reported.
6. The CA's index, residue number, residue name, distance to the AOI, hydrogen bonding angle to the AOI, b factor, and chain letter are written to a text file with the PDBID and AOI in the filename.

B.) α Plugin

1. This plugin generates data for both the scissile phosphate and the all phosphate analysis and removes all atoms belonging to conformation B (if present) when run. When initiated, the plugin prompts the user to enter the residue number of the –1 nucleotide of the scissile phosphate and the name of the organic cofactor (if present).
2. The plugin determines whether or not the residue numbers repeat.
3. If there are no repeating residue numbers, the method unique_residue_numbers runs and the following actions are performed:
    a. If present and defined by the user, the organic cofactor is removed from the structure.
    b. Arrays are created that collect information on all O2′, phosphorus, and O5′ atoms.
    c. Phosphorus and O5′ atoms residing upstream of the first O2′ are removed.
    d. O2′s downstream from the last phosphorus are removed.

e. The appropriate O2′, phosphorus, and O5′ atoms are removed from the arrays when there is a break in the phosphorus backbone so that an angle isn't determined over the break.
f. The O2′–P–O5′ angle at each non-scissile phosphate is determined and written to a text file while the angle at the scissile phosphate is determined and written to a separate text file.

4. If the residue numbers repeat, the plugin prompts the user to enter the chains of interest, starting with the chain containing the scissile phosphate and then runs the method repetitive_residue_numbers and performs the following actions:
    a. If present and defined by the user, the organic cofactor is removed from the structure.
    b. For each chain defined by the user, the plugin performs steps b through f as detailed above.

C.) γ, β, and δ All-Phosphates Plugin

1. When initiated, the plugin prompts the user to enter the residue number of the –1 and +1 nucleotides, the letters of the chains to be considered and whether an amino is substituted for the 2′OH on the –1 nucleotide.
2. The user has the option to click one of four buttons labeled with O2′, NBO, O5′, or All. The plugin takes one of the following routes depending on the button pressed:
    a. **O2′ Button:** The plugin identifies all O2′ atoms present in the chains specified. O2′ atoms on the last nucleotide of the chains are removed, as they are not a part of a nucleotide step. If the user specified that an amino was substituted for the 2′OH, the plugin searches for a nitrogen within 1.8 Å of the C2′ and catalogs it if found. All atoms within 5 Å of each O2′ (and N2′, if applicable) are then identified and their distances to the O2′ are cataloged. For the O2′ at the –1 nucleotide, the plugin writes identifying information for the O2′ and its respective contact atoms along with the distances to a file. The plugin writes this same information for the rest of the O2′s to an additional text file.
    b. **NBO Button:** The plugin identifies all OP1 atoms present in the chains specified. OP1 atoms on the first nucleotide of the chains are removed as they are not a part of a nucleotide step. This same process is carried out for OP2 atoms. All atoms within 5 Å of each OP1 are then identified and their distances to the OP1 are cataloged. For the OP1 at the +1 nucleotide, the plugin writes identifying information for the OP1 and its respective contact atoms along with the distances to a file. The plugin writes this same information for the rest of the OP1s to an additional text file. Data collection and handling for OP2 is performed in the same manner and included with the previously mentioned two text files.
    c. **O5′ Button:** The plugin collects and handles data on the O5′ atoms just as it does for the OP1 atoms, as described above.
    d. **All Button:** The plugin performs all three of the actions described above in a, b and c.

D.) Scissile Phosphate Downstream Processing Script

1. First, text files are opened. If a filename is present in a list of non-catalytic PDB's, that file is skipped.
2. The residue numbers of the –1/+1 bases are extracted so that contacts to the –1/+1 site (scissile phosphate) can be separated from the rest of the ribozyme.
3. Residue names are modified to account for numbering discrepancies.
4. CA distances are averaged and contacts to the –1/+1 bases are kept separate.
5. The contacts made to the –1/+1 nucleobases are sorted to give the top 15 contacts as follows: all sugar-phosphate, carbon, and hydrogen atoms are excluded. The occurrence of CAs are first checked against the occurrence threshold (25%) and then ranked based on distance. If a CA has an average angle greater than 140° it is added to the list of regular contacts.
6. The list of regular contacts is then sorted to find the top 15 contacts as follows: Hydrogen, carbon and phosphorus CAs are automatically excluded. The occurrence of CAs are first checked against the occurrence threshold (25%) and then ranked based on distance.
7. The script then searches the text files for the top 15 contacts to find all observed distances for graphing purposes. Observed distances and angles are then appended to an array.
8. The top 5 contacts are plotted. CAs from mutant ribozymes are plotted in the same category as the corresponding CA from the WT ribozyme when the WT ribozyme CA is present in the top 5. Angles less than 140° are depicted as red numbers in the plots or "N/A" according to the criteria above.

E.) All-Phosphate Downstream Processing Script

1. Data from each ribozyme class and each AOI type (O2′, NBO or O5′) are handled separately. The script iterates through the pertinent data text files in a given directory, performing steps 2 through 4 for each data file.
2. The script stores the data to an array, keeping scissile and non-scissile data separate.
3. If the AOIs are NBOs, the script performs the following steps:
    a. A maximum of 2 CAs to each OP1 within 5 Å of the OP1 are counted given that the CAs are not hydrogen, carbon, phosphorus, oxygen (except for O2′s and nucleobase oxygens), or any atoms of the -1 and +1 nucleotides.
    b. A maximum of 2 CAs to each OP2 are counted following the same criteria as above.
    c. The number of CAs to each NBO within each linkage are combined and stored.
4. If the AOIs are O2′s or O5′s, the script counts and stores the number of CAs to each AOI within 5 Å of the AOI given that the CAs are not hydrogen, carbon, phosphorus, oxygen (except for O2′s and nucleobase oxygens), or any atoms of the -1 and +1 nucleotides.
5. The number of CAs to each AOI from the data text files are binned into five categories: 0, 1, 2, 3, and 4+ contacts. Data from the different ribozyme classes and the different AOI types (O2′, NBO or O5′) are binned separately. Scissile and non-scissile data is also binned separately.

6. The percent to which each bin contributes to the total for each given data set (e.g. O2′ hammerhead non-scissile data) is plotted.

Supporting Results

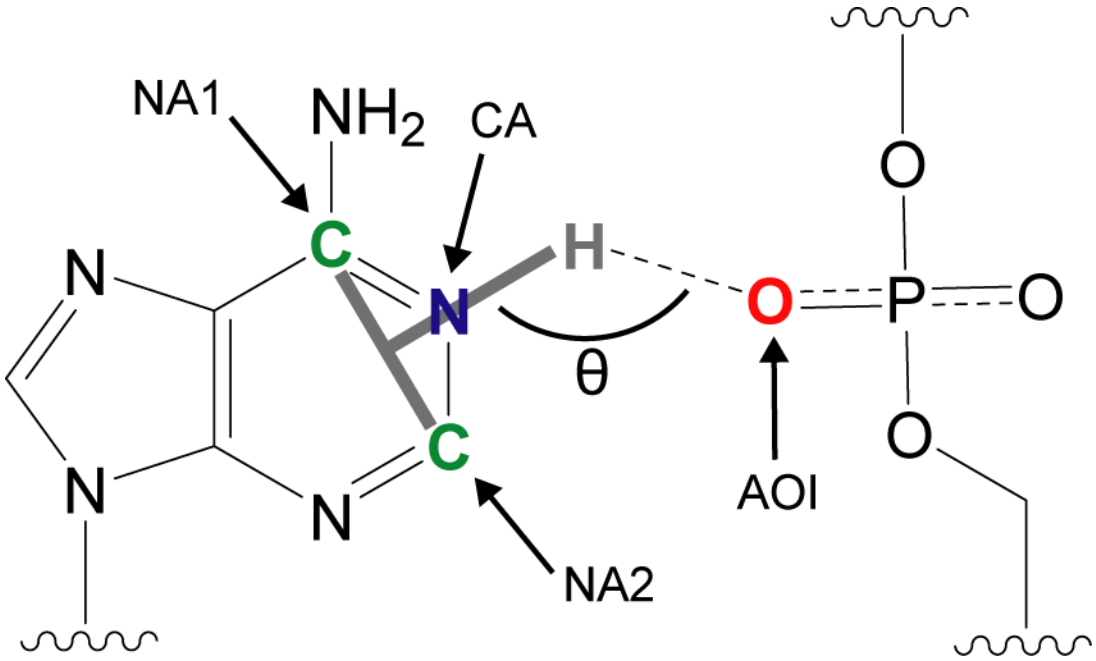

Figure S1. Process for calculating the angle of a contact. The two neighboring atoms (labeled NA1 and NA2 and colored green) of the contact atom (labeled CA and colored blue) are found and their coordinates recorded. A parametric equation for a line connecting the atoms NA1 and NA2 is calculated (gray solid line from NA1 to NA2), its midpoint determined, and an equation for the line connecting this midpoint and the coordinates of CA is determined (gray solid line through CA). This line then serves to provide the coordinates of a theoretical proton 1 Å from CA (colored gray). The contact angle, θ, is then calculated using the Cartesian coordinates of CA, the calculated proton, and the AOI, which is colored red.

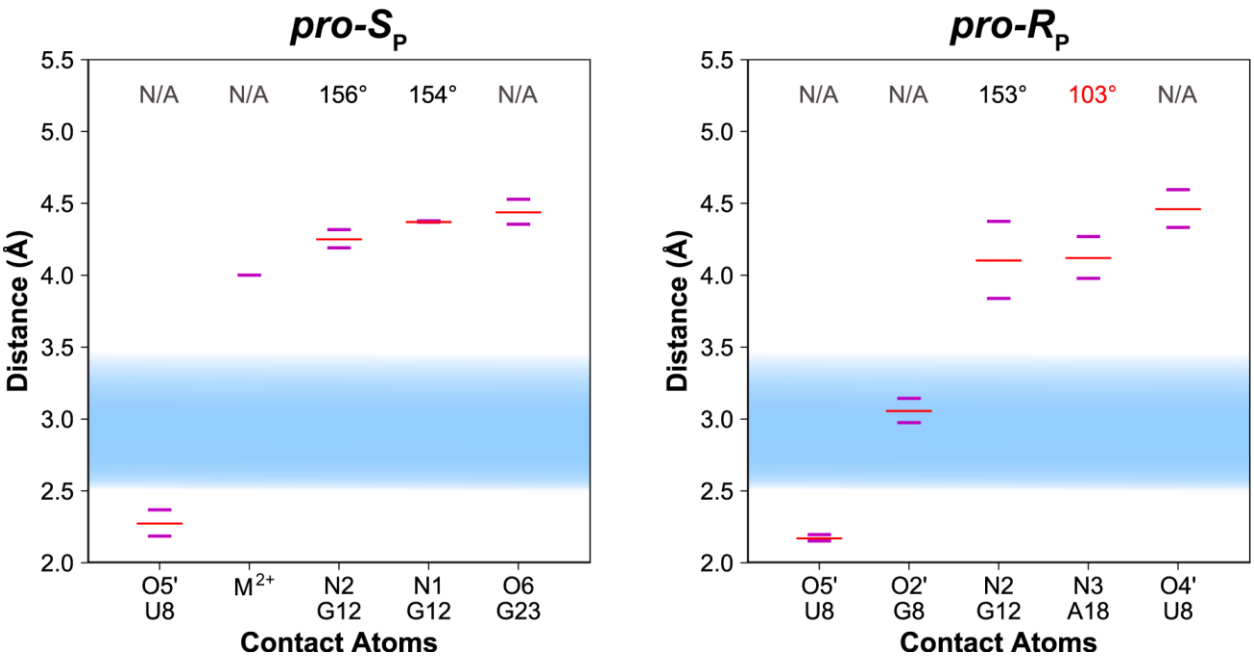

Figure S2. Contacts for β in Hammerhead Ribozyme PDBs 5EAQ and 5EAO: Neutralization of the negative charge on the NBO atoms. Distances between the *pro*-R_P and *pro*-S_P NBO atoms and nearby atoms are shown for the hammerhead ribozyme in PDBs 5EAQ and 5EAO.[1] Distances observed in each crystal structure are shown as horizontal purple (vanadate or vanadate-like) lines; average distances are shown as horizontal red lines; angles for each contact are shown at the top of the plot; and the typical hydrogen bonding distance in RNA (2.5 Å – 3.5 Å) is shown as a blue shaded region.

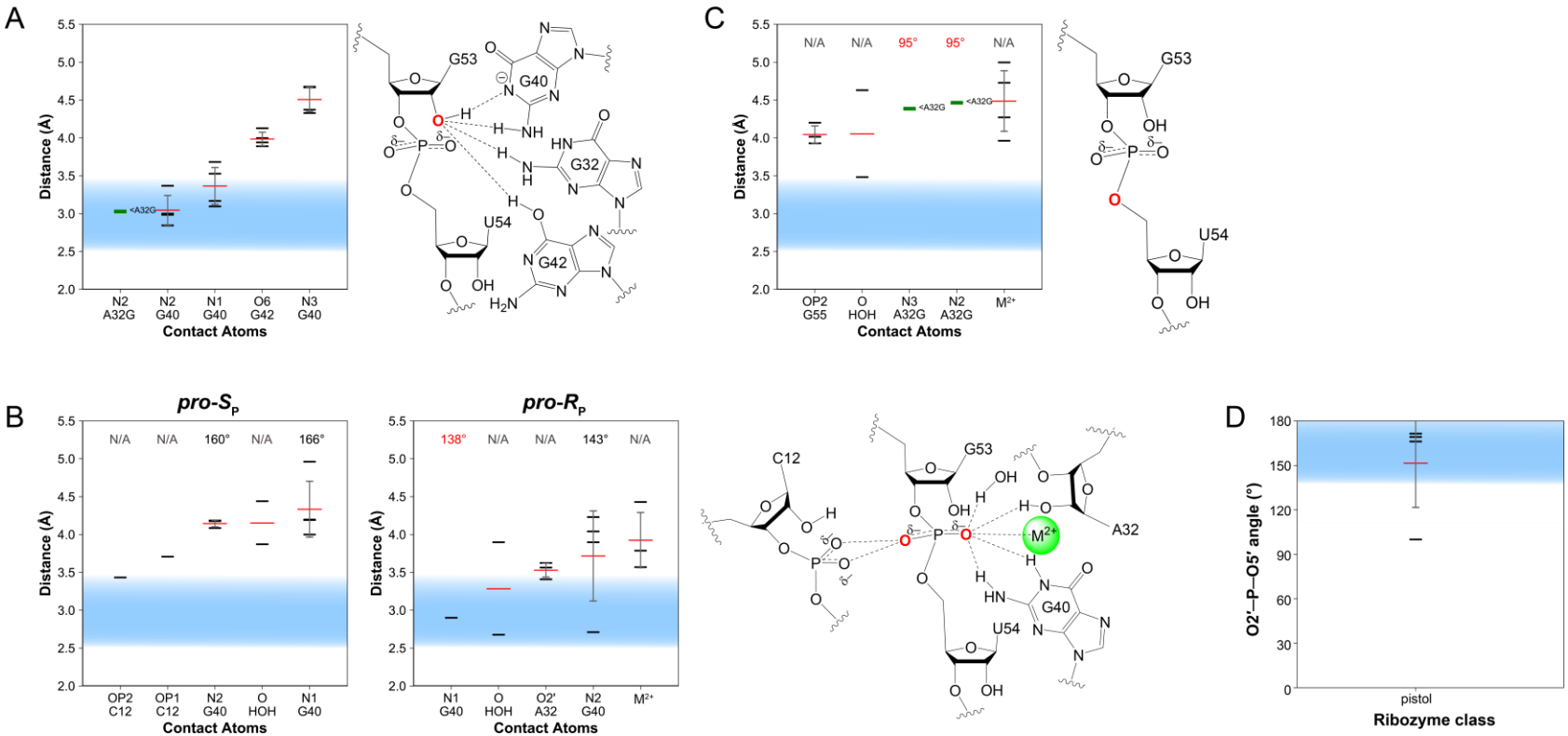

Figure S3. Contacts and angles for γ, γ′, β, γ′′, δ, and α for pistol. Distances between the (A) O2′, (B) NBO atoms, and (C) O5′ and nearby atoms are shown for the pistol ribozyme. Distances observed in each crystal structure are shown as horizontal black (WT) lines; average distances are shown as horizontal red lines; standard deviation is shown as a gray vertical line; angles for each contact are shown at the top of the plot; and the typical hydrogen bonding distance in RNA (2.5 Å – 3.5 Å) is shown as a blue shaded region. Illustrations to the right of each plot visualize contacts within 4 Å of the AOI. Angles are not provided for the scissile phosphate gamma plot (A) because the H2′ was substituted with an OH2′ and the pucker for the sugar is uncertain. Contacts between the scissile phosphate NBO atoms and other NBO atoms are not depicted. Contacts between the scissile phosphate *pro*-S_P NBO atom and the ligands of a $Co(NH_3)_6^{3+}$ are not shown since the Co^{3+} itself is not a top contact. (D) Values corresponding to the O2′–P–O5′ angle of the scissile phosphate for the pistol ribozyme. Angles observed in each crystal structure are shown as horizontal black (WT) lines; the average angle is shown as a horizontal red line; standard deviation is shown as a gray vertical line; and the optimal angles for in-line nucleophilic attack (140-180°) are shown as a blue shaded region.

Table S1. Crystal structures of small ribozymes in this study.

Ribozyme	Relevant details[a]	PDB ID	Reference	Resolution (Å)
glmS				
	G33A	3B4A	[2]	2.7
	2′,5′ linkage	3B4B	[2]	2.7
	2′,5′ linkage	3B4C	[2]	3
	2′-deoxy at A–1	2Z75	[3]	1.7
	2′-deoxy at A–1, bound to Glc6P	2Z74	[3]	2.2
	Unliganded	3G8S	[4]	3.1
	G33A	3G8T	[4]	3
	Bound to Glc6P	3L3C	[4]	2.85
	Bound to MaN6P	3G96	[4]	3.01
	Unliganded	2GCS	[5]	2.1
	2′-deoxy at A–1, unliganded	2H0S	[5]	2.35
	2′-deoxy at A–1, bound to Glc6P	2H0Z	[5]	2.7
	Bound to Glc6P	2HO7	[5]	2.9
	2′-amino at A–1, unliganded	2H0X	[5]	2.3
	2′-methoxy at A–1	2NZ4	[6]	2.5
Hammerhead				
	Tethered	1NYI	[7]	2.85
	Minimal Tethered	1Q29	[7]	3
	2′-methoxy at C17	3ZP8	[8]	1.55
	No divalent metal ions	299D	[9]	3
	2′-methoxy at C17	300D	[9]	3
	2′-methoxy at C17	301D	[9]	3
	G12A	2QUS	[10]	2.4
	Bound to inhibitor Tb(lll)	359D	[11]	2.9
	2′-methoxy at C17	2OEU	[12]	2
	G12A	3ZD4	[13]	2.2
	2′-methoxy at C17	3ZD5	[13]	2.2
	2′-methoxy at C17	1MME	[14]	3.1
	DNA substrate strand	1HMH	[15]	2.6
	2′-deoxy at C17, pH 8.0	5DI2	[16]	2.99
	2′-deoxy at C17, pH 5.0	5DI4	[16]	2.95
	2′-deoxy at C17	5DQK	[16]	2.71
	G12A, 2′-deoxy at C17	5DH6	[16]	2.78
	G12A, 2′-deoxy at C17	5DH7	[16]	3.06
	G12A, 2′-deoxy at C17	5DH8	[16]	3.3
	Vanadate	5EAO	[16]	2.99
	Vanadate	5EAQ	[16]	3.2

Twister				
	2′-methoxy at U5	5DUN	[17]	2.64
	2′-deoxy at U6	4OJI	[18]	2.3
	2′-deoxy at U5	4RGE	[19]	2.89
	2′-deoxy at U5	4RGF	[19]	3.2
	Simulated Structure	Triple_RS_ Active	[20]	n/a
	Simulated Structure	Triple_RS_ Neutral	[20]	n/a
	Simulated Structure	Triple_TS	[20]	n/a
	Disordered active site, 2′-deoxy at U5	4QJD	[21]	3.1
	2′-deoxy at U9	4QJH	[21]	3.88
Hairpin				
	2′,5′ linkage	3CQS	[22]	2.8
	Vanadate	2P7E	[23]	2.05
	2′,5′ linkage	2P7F	[23]	2.35
	2′-amino U at A–1, S9 linker at position 14	2NPY	[24]	2.65
	C3 linker at position 14, 2′-deoxy at A–1	2NPZ	[24]	3.35
	Purine substitution at A38, 2′-methoxy at A–1	4G6P	[25]	2.64
	G8I, 2′,5′ linkage	4G6R	[25]	2.83
	Purine substitution at A38, 2′,5′ linkage	4G6S	[25]	2.84
	2′-methoxy at A–1	2D2K	[26]	2.65
	U39C	1X9K	[26]	3.17
	U39C, 2′-methoxy at A–1	1X9C	[26]	2.19
	Propyl linker (C3) at U39, 2′-methoxy at A–1	2D2L	[26]	2.5
	5-iodo-uracil substitution at U5, 2′-methoxy at A–1	1M5K	[27]	2.4
	N1-deazaadenosine at A38	3GS1	[28]	2.85
	Transition-state mimic, N1-deazaadenosine at A38	3GS8	[28]	2.85
	N1-deazaadenosine at A9, 2′-methoxy at A–1	3I2Q	[29]	2.9
	2′,5′ linkage, N1-deazaadenosine at A9	3I2R	[29]	2.8
	N1-deazaadenosine at A10, 2′-methoxy at A–1	3I2S	[29]	2.75
	2′,5′ linkage, N1-deazaadenosine at A10	3I2U	[29]	2.8
	A38G, 2′-methoxy at A–1	3B5A	[30]	2.35
	2′,5′ linkage, A38G	3B58	[30]	2.65
	2′,5′ linkage, A38DAP	3B5F	[30]	2.7
	A38DAP, 2′-methoxy at A–1	3B5S	[30]	2.25
	2′,5′ linkage, A38AP	3B91	[30]	2.75
	2′,5′ linkage, A38C	3BBK	[30]	2.75
	A38C, 2′-methoxy at A–1	3BBM	[30]	2.65
	A38C, 2′-methoxy at A–1	3CR1	[30]	2.25

	A38AP, 2′-methoxy at A–1	3BBI	[30]	2.35
	Vanadate	1M5O	[31]	2.2
	2′-methoxy at A–1	2OUE	[32]	2.05
	G8I, 2′-methoxy at A–1	1ZFT	[32]	2.33
	G8DAP, U39C	2FGP	[32]	2.4
	G8A	1ZFV	[32]	2.4
	G8U	1ZFX	[32]	2.38
	G8AP, U39C	2BCY	[32]	2.7
	G8I, U39C, 2′deoxy at A–1	2BCZ	[32]	2.4
Pistol				
	2′-deoxy at G53, A32G	5KTJ	[33]	2.97
	2′-deoxy at G53	5K7C	[34]	2.73
	2′-deoxy at G53	5K7D	[34]	2.68
	2′-deoxy at G53	5K7E	[34]	3.27

[a]Shading of crystal structure corresponds to the type of structure it is classified as: white (WT), green (variant), purple (vanadate or vanadate-like), yellow (simulation-averaged), and orange (not catalytically relevant).

References

1. Mir, A.; Golden, B. L., *Biochemistry* **2016,** *55*, 633-636.
2. Klein, D. J.; Been, M. D.; Ferré-D'Amaré, A. R., *J. Am. Chem. Soc.* **2007,** *129*, 14858-14859.
3. Klein, D. J.; Wilkinson, S. R.; Been, M. D.; Ferré-D'Amaré, A. R., *J. Mol. Biol.* **2007,** *373*, 178-189.
4. Cochrane, J. C.; Lipchock, S. V.; Smith, K. D.; Strobel, S. A., *Biochemistry* **2009,** *48*, 3239-3246.
5. Klein, D. J.; Ferré-D'Amaré, A. R., *Science* **2006,** *313*, 1752-1756.
6. Cochrane, J. C.; Lipchock, S. V.; Strobel, S. A., *Chem. Biol.* **2007,** *14*, 97-105.
7. Dunham, C. M.; Murray, J. B.; Scott, W. G., *J. Mol. Biol.* **2003,** *332*, 327-336.
8. Anderson, M.; Schultz, E. P.; Martick, M.; Scott, W. G., *J. Mol. Biol.* **2013,** *425*, 3790-3798.
9. Scott, W. G.; Murray, J. B.; Arnold, J. R. P.; Stoddard, B. L.; Klug, A., *Science* **1996,** *274*, 2065-2069.
10. Chi, Y.-I.; Martick, M.; Lares, M.; Kim, R.; Scott, W. G.; Kim, S.-H., *PLOS Biol.* **2008,** *6*, e234.
11. Feig, A. L.; Scott, W. G.; Uhlenbeck, O. C., *Science* **1998,** *279*, 81-84.
12. Martick, M.; Lee, T.-S.; York, D. M.; Scott, W. G., *Chem. Biol.* **2008,** *15*, 332-342.
13. Martick, M.; Scott, W. G., *Cell* **2006,** *126*, 309-320.
14. Scott, W. G.; Finch, J. T.; Klug, A., *Cell* **1995,** *81*, 991-1002.
15. Pley, H. W.; Flaherty, K. M.; McKay, D. B., *Nature* **1994,** *372*, 68-74.
16. Mir, A.; Chen, J.; Robinson, K.; Lendy, E.; Goodman, J.; Neau, D.; Golden, B. L., *Biochemistry* **2015,** *54*, 6369-6381.
17. Košutić, M.; Neuner, S.; Ren, A.; Flür, S.; Wunderlich, C.; Mairhofer, E.; Vušurović, N.; Seikowski, J.; Breuker, K.; Höbartner, C.; Patel, D. J.; Kreutz, C.; Micura, R., *Angew. Chem. Int. Ed. Engl.* **2015,** *54*, 15128-15133.
18. Liu, Y.; Wilson, T. J.; McPhee, S. A.; Lilley, D. M. J., *Nat. Chem. Biol.* **2014,** *10*, 739-744.
19. Ren, A.; Košutić, M.; Rajashankar, K. R.; Frener, M.; Santner, T.; Westhof, E.; Micura, R.; Patel, D. J., *Nat. Commun.* **2014,** *5*, 5534.
20. Gaines, C. S.; York, D. M., *J. Am. Chem. Soc.* **2016,** *138*, 3058-3065.
21. Eiler, D.; Wang, J.; Steitz, T. A., *Proc. Natl. Acad. Sci. U.S.A.* **2014,** *111*, 13028-13033.
22. Torelli, A. T.; Spitale, R. C.; Krucinska, J.; Wedekind, J. E., *Biochem. Biophys. Res. Commun.* **2008,** *371*, 154-158.
23. Torelli, A. T.; Krucinska, J.; Wedekind, J. E., *RNA* **2007,** *13*, 1052-1070.
24. MacElrevey, C.; Spitale, R. C.; Krucinska, J.; Wedekind, J. E., *Acta Crystallogr. D* **2007,** *63*, 812-825.
25. Liberman, J. A.; Guo, M.; Jenkins, J. L.; Krucinska, J.; Chen, Y.; Carey, P. R.; Wedekind, J. E., *J. Am. Chem. Soc.* **2012,** *134*, 16933-16936.
26. Alam, S.; Grum-Tokars, V.; Krucinska, J.; Kundracik, M. L.; Wedekind, J. E., *Biochemistry* **2005,** *44*, 14396-14408.
27. Rupert, P. B.; Ferre-D'Amare, A. R., *Nature* **2001,** *410*, 780-786.
28. Spitale, R. C.; Volpini, R.; Heller, M. G.; Krucinska, J.; Cristalli, G.; Wedekind, J. E., *J. Am. Chem. Soc.* **2009,** *131*, 6093-6095.
29. Spitale, R. C.; Volpini, R.; Mungillo, M. V.; Krucinska, J.; Cristalli, G.; Wedekind, J. E., *Biochemistry* **2009,** *48*, 7777-7779.
30. MacElrevey, C.; Salter, J. D.; Krucinska, J.; Wedekind, J. E., *RNA* **2008,** *14*, 1600-1616.

31. Rupert, P. B.; Massey, A. P.; Sigurdsson, S. T.; Ferré-D'Amaré, A. R., *Science* **2002,** *298*, 1421.
32. Salter, J.; Krucinska, J.; Alam, S.; Grum-Tokars, V.; Wedekind, J. E., *Biochemistry* **2006,** *45*, 686-700.
33. Nguyen, L. A.; Wang, J.; Steitz, T. A., *Proc. Natl. Acad. Sci. U.S.A.* **2017,** *114*, 1021-1026.
34. Ren, A.; Vusurovic, N.; Gebetsberger, J.; Gao, P.; Juen, M.; Kreutz, C.; Micura, R.; Patel, D. J., *Nat. Chem. Biol.* **2016,** *12*, 702-708.