Supplemental Information for

2′-Fluoro-Modified Pyrimidines Enhance Affinity of RNA Oligonucleotides to HIV-1 Reverse Transcriptase

Paige R. Gruenke[1,2], Khalid K. Alam[1,2,†], Kamal Singh[2,3], and Donald H. Burke[1,2,4,5]

Author Affiliations:

[1]Department of Biochemistry, University of Missouri, Columbia, MO 65211

[2]Bond Life Sciences Center, University of Missouri, Columbia, MO 65211

[3]Department of Veterinary Pathobiology, University of Missouri, Columbia, MO 65211

[4]Department of Molecular Microbiology and Immunology, University of Missouri, Columbia, MO 65211

[5]Department of Biological Engineering, University of Missouri, Columbia, MO 65211

[†]Current affiliation: Stemloop, Inc., Evanston, IL 60204

Supplemental Information

Supplemental Text

Impact of 2′ Modifications on Previously Identified Anti-HIV-1 RT Aptamers

Previously identified RNA aptamers that bind and inhibit HIV-1 RT were transcribed using 2′-OH, 2′-F, 2′-NH_2, or 2′-OMe pyrimidines and were evaluated for RT inhibition. The aptamers tested belonged to the family 1 pseudoknot (F1Pk) motif (Tuerk et al. 1992; Burke et al. 1996; Ditzler et al. 2013), UCAA motif (Whatley et al. 2013), or (6/5) asymmetric loop(AL) motif (Ditzler et al. 2013) aptamer families (Figure S1A). Each of the three 2′ modifications had a different impact on the abilities of these aptamers to inhibit RT (Figure S1B), similar to the trends observed in Figure 1. When modified with 2′-F pyrimidines, aptamers from all three structural classes inhibited RT at least as well as the unmodified versions, and many seemed to inhibit to a greater extent compared to the unmodified aptamers. When 2′-OMe pyrimidines were incorporated, two F1Pk aptamers remained moderately inhibitory, but aptamers from the UCAA and (6/5)AL family aptamers were no longer able to inhibit RT. When these aptamers were modified with 2′-NH_2 pyrimidines, RT inhibition was no longer observed regardless of structural motif. These results were in agreement with the trends observed in Figure 1. As expected, the 2′ modifications increase the half-life of the aptamer 70N 1.1 in the presence of 10% fetal bovine serum from <<15sec (not shown) to 2-3 hours (Figure S1C).

Additional Analyses of Results from 2′-OMeY and 2′-NH_2Y Trajectories

As shown in Figure 1E, F1Pk aptamers were still able to inhibit RT when modified with 2′-OMe pyrimidines, while (6/5)AL aptamers lost ability to inhibit RT. The moderate reduction in inhibition by 2′-OMeY-modified F1Pk aptamers is consistent with a previous study that found that 2′-OMe substitution of C13, which is the first nucleotide of the 5′-CGGG-3′ element within stem 1, reduced affinity to HIV-1 RT by approximately 5-fold. For that same aptamer, an approximate 10-fold reduction in affinity was observed when all positions were modified with 2′-OMe pyrimidines (Green et al. 1995). Therefore, it is likely that the incorporation of 2′-OMe pyrimidines into F1Pk aptamers studied here decreased, but did not abolish, their affinity to HIV-1 RT. However, any decrease in affinity did not greatly affect HIV RT

inhibition by these aptamers under the conditions of this assay. For (6/5)AL aptamers, at least one cytosine and at least one uracil are required to be 2′-OH, as singly-substituted 2′-OMeY transcripts were non-inhibitory (Figure S2).

Four aptamer variants from cluster 11 (11.1, 11.184, 11.230, and 11.468) were highly enriched in the 2′-OMeY and 2′-NH_2Y trajectories. Their sequences do not contain recognizable motifs for F1Pk, UCAA, or (6/5)AL families, and their functional structure is not known. Each of these aptamers was transcribed with various NTP combinations and tested for RT inhibition. All four aptamers moderately inhibited HIV-1 RT as unmodified transcripts (Figure 1E and 1F). When modified with 2′-OMe pyrimidines, RT inhibition by aptamer 11.1 was slightly weaker (not significant) while inhibition by aptamer 11.230 was about the same as for the 2′-OH version. In contrast, aptamer 11.468 lost all ability to inhibit HIV-1 RT after 2′-OMe modification. Neither aptamer 11.1 nor 11.184 inhibited RT when modified with 2′-NH_2 pyrimidines. While sequences from cluster 11 were highly enriched, it seems that these sequences were artefacts from the selection. Therefore, studies further exploring the structural class of cluster 11 aptamers were not pursued.

Secondary Structure Insights for Cluster 31 Aptamers from the 2′-NH_2Y Trajectory

As shown in Figure 1F, only the two aptamers from cluster 31 were able to inhibit RT when modified with 2′-NH_2 pyrimidines. The cluster 31 aptamers do not belong to any of the previously defined structural families of HIV-1 RT RNA aptamers, and the required functional structure is not known. A series of variants of aptamer 31.56 were evaluated for RT inhibition to gain insight into the functional requirements of this novel aptamer, using 2′-OH transcripts to simplify the analysis. Removing ten nucleotides at a time from either the 5′ or 3′ end abolished RT inhibition (Figure S3A), indicating that nucleotides near each end are required for the aptamer to remain fully functional. Among the 1596 unique sequences present in this cluster, 1334 of these contained two or more nucleotide differences relative to the cluster seed sequence, aptamer 70N 31.1 (Figure S3B), and were therefore capable of covariation. The 300 most abundant of these sequences were used to generate two covariation models using the

RNAalifold web server (Bernhart et al. 2008) by applying two different parameter settings (Figure S3C). Model 1 predicted that portions of the 5′ and 3′ constant regions base paired with each other to form a 4 base pair stem. Model 2, in contrast, predicted that three nucleotides from the 5′ constant region base paired with three nucleotides from the random region and five nucleotides from the 3′ constant region base paired with five nucleotides from the random region. Both models are consistent with initial 5′ and 3′ truncation data, but invariance within the constant regions precludes using covariation data to discriminate between these models. Therefore, sequence variants were generated that preferentially disrupted, rescued, or stabilized one or the other of the two structures, and these variants were tested for RT inhibition. Specifically, each predicted stem was mutated in one strand to disrupt its formation, and the original predicted pairing pattern was restored by introducing compensatory mutations in the other strand. For Model 1, the stem disruption remained fully inhibitory, while introducing compensatory mutations to restore pairing, or extending the stem to reinforce base pairing between the 5′ and 3′ constant regions, eliminated RT inhibition (Figure S3D). These data rule out Model 1. Sequence variants that disrupted or rescued the predicted stems in Model 2 all inhibited RT to the same degree as the full-length unmodified or 2′-NH_2 versions of aptamer 31.56, while an internal deletion of 19 nucleotides abolished RT inhibition. Model 2 predicted structures at the 5′ and 3′ ends, while the data do not provide evidence that these structures are functionally required. Further, non-inhibition by the internal deletion could mean that a portion of the deleted sequence may be involved in making the active structure or induces a misfold of the aptamer. Finally, when aptamer 70N 31.56 was transcribed with 2′-FY in place of 2′-NH_2Y, the 2′-FY version strongly inhibited RT (Figure S3E). These data were in agreement with the other primer extension assays using 2′-FY-modified aptamers in Figure 1 and suggested that the 2′-FY modification was either well tolerated by the RT aptamers tested or enhanced non-specific binding of the RNA to RT.

Supplemental Methods

Serum Nuclease Assays. 5′ radiolabeled RNA was incubated in DMEM medium containing 10% fetal bovine serum at 37°C. At specified time points, an aliquot of the reaction was removed, added to an equal

volume of denaturing gel loading buffer, and placed in an ethanol and dry ice bath to halt the reaction. Samples were run on a denaturing polyacrylamide gel (6% TBE-PAGE, 8 M urea). The gel was imaged using a Typhoon FLA 9000 phosphoimager (GE Healthcare Life Sciences). The fraction of full-length RNA present in the sample was quantified by measuring the intensity of the band corresponding to the full-length RNA and dividing it by the intensity of all bands present in the lane of the gel using Multiguage software (Fujifilm).

Supplemental References

Bernhart SH, Hofacker IL, Will S, Gruber AR, Stadler PF. 2008. RNAalifold: improved consensus structure prediction for RNA alignments. *BMC Bioinformatics* **9**: 474.

Burke DH, Scates L, Andrews K, Gold L. 1996. Bent pseudoknots and novel RNA inhibitors of type 1 human immunodeficiency virus (HIV-1) reverse transcriptase. *J. Mol. Biol.* **264**: 650-666.

Ditzler MA, Lange MJ, Bose D, Bottoms CA, Virkler KF, Sawyer AW, Whatley AS, Spollen W, Givan SA, Burke DH. 2013. High-throughput sequence analysis reveals structural diversity and improved potency among RNA inhibitors of HIV reverse transcriptase. *Nucleic Acids Res.* **41**: 1873-1884.

Green L, Waugh S, Binkley JP, Hostomska Z, Hostomsky Z, Tuerk C. 1995. Comprehensive chemical modification interference and nucleotide substitution analysis of an RNA pseudoknot inhibitor to HIV-1 reverse transcriptase. *J. Mol. Biol.* **247**: 60-68.

Katoh K, Misawa K, Kuma Ki, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**: 3059-3066.

Tuerk C, MacDougal S, Gold L. 1992. RNA pseudoknots that inhibit human immunodeficiency virus type 1 reverse transcriptase. *Proc. Natl. Acad. Sci. U. S. A.* **89**: 6988-6992.

Whatley AS, Ditzler MA, Lange MJ, Biondi E, Sawyer AW, Chang JL, Franken JD, Burke DH. 2013. Potent inhibition of HIV-1 reverse transcriptase and replication by nonpseudoknot, "UCAA-motif" RNA aptamers. *Mol. Ther. Nucleic Acids* **2**: e71.

Supplemental Tables

Table S1: Read counts (total, unique, reads per sequence) and number of unique sequence clusters from high throughput sequences of aptamer populations.

Library ID	Total Reads	Unique Sequences	Average Reads per Sequence[a]	Sequence Clusters[b]
70N Round 14	2,160,216	72,921	30	1,250
70N 2'-OH (17A)	2,028,611	44,811	45	604
70N 2'-OH (17B)	2,760,126	42,998	64	
70N 2'-FY (17A)	3,542,268	129,540	27	2,360
70N 2'-FY (17B)	2,774,057	115,242	24	
70N 2'-OMeY (17A)	2,564,164	73,009	35	188
70N 2'-OMeY (17B)	1,862,002	62,405	30	
70N 2'-NH_2Y (17A)	977,733	25,481	38	818
70N 2'-NH_2Y (17B)	2,913,710	73,550	40	

[a]Average reads per sequences was rounded to the nearest whole number.

[b]The number of unique sequence clusters was combined for the two replicates of the same trajectory.

Table S2: RNA sequences used in this study[a]

RNA	Sequence
70N 1.1	ACGUUGUCGAAAGCCUAUGCAAAUUAAGGACUGUCGACGAAACCUUGCGUAGACUCGCCACGCUUGGUGU
70N 2.1	CAUAGCGACUGUCCACGAAUCCGAAGCCUAACGGGACAAAAGGCAAGAGCGCGAUACCAAUGCUGGACUG
70N 3.1	AACCGCAAGCAACACCCAGCAAGAAACAUCCGACGCACGACGGGAGAAAGUGCAUUACCACGAUGUCGAU
70N 7.1	GCGAACCAAACCCAGAUUACUAACCGUGGGCCUGAAACACGGGACAAAACAGGCAUCAAUGGAGUGGUAC
70N 8.1	ACGUUGUGCACGGAUGCCCACGGUCGCACGAAACCUUGUGUGGGAUAGCGCGAAUACUGACGAGUGUGCC
70N 9.1	ACCAAAUCCCGAACUACAAAUCCGAACGCUAACGGGACAAUUGCGAAAUGGAACAUACGGGCCUGGUUGA
70N 10.1	GUGCGCUACCACAUGAUCCGAGGCAAAACGGGAAAAGAUAGCAUCGAUUACGGAACCGGCCACGCACA
70N 11.1	ACCAAGAUAAACCGAGGUGUAAACGGGAAAACACAUGGAAUGACGUGCUGUAGUAGGUGUGAUCACUGC
70N 11.184	ACCAAGAUAAACCGAGGUGUAAAACGGGAAAACACAUGGAAUGACGUGCUGUAGUAGGUGUGAUCACUGC
70N 11.230	ACCAAGAUAAACCGAGGUGUAAACGGGAUAACACAUGGAAUGACGUGCUGUAGUAGGUGUGAUCACUGC
70N 11.468	ACCAAGAUAAACCGAGGUGUAAACGGGAAAUCACAUGGAAUGACGUGCUGUAGUAGGUGUGAUCACUGC
70N 28.1	GGCGCACACUACCAAUUGAAUCCGAAAGACACGGGACAAGUCUAAAUGCGAUGAAAUACCGAAUCGAGA
70N 28.46	GGCACACUACCAAUUGAAUCCGAAAGACACGGGACAAGUCUAAAUGCGAUGAAAUACCGAAUCGAGA
70N 29.1	GCGCGAUUGCCAUAGAAUCCGAGGGUCAACGGGAAAAGGACCAACCAACGCGUCGAGCAGCCACAUGCAG
70N 31.8	UAACCGCACAUCCGAAAACGAUCCCACGGCAAGGGGAUAAUGCCAGCGAUUAAUGGAGCCUACGUCAACA
70N 31.56	UAACCGCACAUCCGAGAACGAUCCCAUGGCAAGGGGAUAAUGCCAGCGAUUAAUGGAGCCUACGUCAACA
70N 51.1	ACAUUGGCAAACCACAGAAUCCGAGGCGACACGGGAUAACGCAACCUUAAGGUCGACAUGUUUCGCACAA
70N 67.1	AAGACUAACACAUGACCCAAUACACGAAUCCGAAGGCAAACGGGACAACUGCCGAUACCUGUGUACCGAGCUAC

70N 112.2	CUAAUCCUGGAGAAGGAACAGGGACAAACACCUUCAACAUCGAACCAAAUCCCGAACUGCACCGCGCCAU
70N 134.1	ACAAUCCGAGGCAUAACGGGAGAAGAUGCAAAAAUUGAGUGGAAAACACAAAAGUGCCGUCGUAACGUAC
70N 181.1	ACUGAAGGGGAAAGAGUGTGUAUCCCUUCCACACAACAUGGCGCGCAAUAUGAUUGGUGAUGCUUGGC
Arb1	AAUUUGGACUUUCCGCCCUUCUUGGCCUUUAUGAGGAUCUCUCUGAUUUUUCUUGCGUCGAGUUUUCCGG

[a]The full RNA sequence consists of the 5′ constant region (gggaaaagcgaaucauacacaaga), the ~70 nucleotide sequence displayed above, and the 3′ constant region (gggcauaagguauuuaauuccaua).

Table S3: Enrichment Values for Candidate Aptamers[a]

RNA	2'OH		2'F		2'OMe		2'NH_2	
	A	B	A	B	A	B	A	B
70N 1.1	0.042	0.0036	0.053	0.088	0.0027	0.00018	0.000026	0.027
70N 2.1	1.11	1.06	-	-	-	-	1.28	0.78
70N 3.1	0.63	0.27	-	-	0.99	0.81	-	-
70N 7.1	0.0017	0.00024	0.011	0.050	-	-	-	-
70N 8.1	0.032	0.0018	0.14	0.48	0.00041	0.00032	0.00015	0.064
70N 9.1	0.70	0.51	48.13	47.09	8.64	5.53	-	-
70N 10.1	0.72	0.10	-	-	5.23	2.32	-	-
70N 11.1	390.18	486.84	-	-	364.20	384.43	47.2	374.01
70N 11.184	246.48	422.15	-	-	-	-	733.74	3039.61
70N 11.230	1996.43	1961.97	-	-	2534.93	3151.13	-	-
70N 11.468	38.59	41.74	-	-	4268.70	5927.48	-	-
70N 28.1	16.47	25.2	673.67	447.92	-	-	-	-
70N 28.46	NP	NP	-	-	22.89	656.15	-	-
70N 29.1	1.74	0.34	-	-	-	-	19.15	4.24
70N 31.8	44.33	6.26	-	-	-	-	408.37	35.31
70N 31.56	52.50	NP	-	-	-	-	647.02	56.70
70N 51.1	9.02	0.75	2671.60	673.52	-	-	-	-

70N 67.1	0.37	0.11	1084.92	67.45	-	-	-	-
70N 112.2	NP	NP	-	-	-	-	NP	218.76
70N 134.1	244.80	256.63	1288.5	401.54	-	-	-	-
70N 181.1	12.26	NP	225.08	72.68	-	-	-	-

[a]Enrichment values are listed for the modification trajectories that the RNA was modified with and tested for RT inhibition. NP = No enrichment value was present for the RNA in a particular trajectory.

Table S4: Structural Motif Enrichment Data for Family 1 Pseudoknot Aptamers

Library ID	Number Sequences Matched	Total Motif Reads	Motif RPM (reads per million)	Enrichment Value[a]
70N Round 14	40,066	1,166,477	539,981.65	
70N 2'-OH (17A)	19,157	712,192	351,073.71	0.65
70N 2'-OH (17B)	16,875	759,350	275,114.25	0.51
70N 2'-FY (17A)	103,009	3,235,054	913,271.95	1.69
70N 2'-FY (17B)	86,412	2,388,807	861,123.98	1.60
70N 2'-OMeY (17A)	30,908	969,288	378,013.26	0.70
70N 2'-OMeY (17B)	25,039	600,101	322,288.05	0.60
70N 2'-NH_2Y (17A)	17,139	753,248	770,402.55	1.43
70N 2'-NH_2Y (17B)	39,599	1,200,166	411,903.04	0.76

[a]Enrichment value is calculated by taking the motif RPM at round 17 divided by the motif RPM at round 14.

Table S5: Structural Motif Enrichment Data for (6/5) Asymmetric Loop Aptamers

Library ID	Number Sequences Matched	Total Motif Reads	Motif RPM (reads per million)	Enrichment Value
70N Round 14	16,086	818,995	379,126.44	
70N 2'-OH (17A)	2,516	41,014	20,217.77	0.053
70N 2'-OH (17B)	834	6,019	2,180.70	0.0058
70N 2'-FY (17A)	9,470	150,785	42,563.76	0.11
70N 2'-FY (17B)	9,952	186,152	67,104.61	0.18
70N 2'-OMeY (17A)	413	3,105	1,210.92	0.0032
70N 2'-OMeY (17B)	71	192	103.11	0.00027
70N 2'-NH_2Y (17A)	57	280	286.38	0.00076
70N 2'-NH_2Y (17B)	1,970	34,168	11726.63	0.031

Table S6: Primer Sequences to Append Illumina Adapters and Sequencing Indices

Library	PCR 1 Forward	PCR 1 Reverse	PCR 2 Forward	PCR 2 Reverse	PCR 3 Forward	PCR 3 Reverse[a]
70N Round 14	TAATACGACTCACTATAGGGAGAATCGAATCATACACAAGA	TATGGAATTAAATACCTTATGCCC	TTTCCCTACACGACGCTCTTCCGATCTGAATCGAATCATACACAAGA	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTATGGAATTAAATACCTTAT	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT	CAAGCAGAAGACGGCATACGAGAT**TCGTAAGCCGTC**GTGACTGGAGTTCAGACGTG
70N 2'OH Round 17 A/B	TAATACGACTCACTATAGGGAAAAGCGAATGATACACAAGA	TATGGAATTAAATACCTTATGCCC	TTTCCCTACACGACGCTCTTCCGATCTGCGAATGATACACAAGA	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTATGGAATTAAATACCTTAT	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT	CAAGCAGAAGACGGCATACGAGAT**GCTACTGGTATG**GTGACTGGAGTTCAGACGTG
70N 2'FY Round 17 A/B	TAATACGACTCACTATAGGGAAAAGCCAATCATACACAAGA	TATGGAATTAAATACCTTATGCCC	TTTCCCTACACGACGCTCTTCCGATCTAGCCAATCATACACAAGA	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTATGGAATTAAATACCTTAT	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT	CAAGCAGAAGACGGCATACGAGAT**CAGCGTTTAGCC**GTGACTGGAGTTCAGACGTG
70N 2'OMeY Round 17 A/B	TAATACGACTCACTATAGGGAGAATCGAATCATACACAAGA	TATGGAATTAAATACCTTATGCCC	TTTCCCTACACGACGCTCTTCCGATCTGAATCGAATCATACACAAGA	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTATGGAATTAAATACCTTAT	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT	CAAGCAGAAGACGGCATACGAGAT**GCCAAATCGCTC**GTGACTGGAGTTCAGACGTG
70N 2'NH_2Y Round 17 A/B	TAATACGACTCACTATAGGGAACAGCGAATCATACACAAGA	TATGGAATTAAATACCTTATGCCC	TTTCCCTACACGACGCTCTTCCGATCTCAGCGAATCATACACAAGA	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTATGGAATTAAATACCTTAT	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT	CAAGCAGAAGACGGCATACGAGAT**GCCATAGTGTGT**GTGACTGGAGTTCAGACGTG

[a]The index sequence is bolded in the PCR 3 reverse primer.