Supplemental Data

Isolation of MLL1 inhibitory RNA aptamers

Short title: MLL1 binding ssRNA aptamers

Asad Ul-Haq, Ming Li Jin, Kwang Won Jeong, Hwan-Mook Kim, and Kwang-Hoon Chun[†]

Gachon Institute of Pharmaceutical Sciences, College of Pharmacy, Gachon University, Incheon 21936, Republic of Korea

[†]**Correspondence to: Kwang-Hoon Chun, College of Pharmacy, Gachon University, Incheon 21936, Republic of Korea**

Tel.: +82-32-820-4951

E-mail: khchun@gachon.ac.kr

Supplementary Materials and Methods

Prediction of secondary and G-quadruplex structure

Secondary structures of five highly popular aptamers were predicted by Mfold algorithm (http://mfold.rna.albany.edu) [1]. Default parameters were used. G-quadruplex structure formation was predicted by QGRS algorithm (http://bioinformatics.ramapo.edu/QGRS/analyze.php) [2]. Putative G-quadruplex is evaluated from the motif $G_xN_{y1}G_xN_{y2}G_xN_{y3}G_x$ in this algorithm. "x" represents the number of guanine tetrads and "y1-y3" represents gaps on the loops.

In silico evaluation of MLL1-aptamer interaction by *cat*RAPID

RNA-binding proteins (RBPs) recognize their target RNAs through the RNA-binding domains (RDs). The classical RDs are domains well established such as RNA-recognition motif while non-classical RDs has no annotation yet. catRAPID signature predicts RNA-binding ability and RDs in the proteins using 80 different physico-chemical properties. The interaction probability among protein-nucleotide pairs was evaluated using *cat*RAPID algorithm (http://s.tartaglialab.com/page/catrapid_group) [3]. *cat*RAPID *signature* calculates overall RNA-binding ability and RNA-binding regions. This algorithm utilizes physic-chemical features for prediction instead of sequence similarity searches. *cat*RAPID *fragments* divides protein and oligonucleotides into fragments and predicts the interaction propensities. *cat*RAPID *strength* calculates the strength of a protein-RNA pair based on a reference set. Reference sequences have the same lengths of the protein-RNA pair of interest. The primary sequences of aptamers and MLL1 SET domain sequence (Protein data bank (PDB) - 2W5Y) were used.

Supplementary Tables

Table S1

	G-quadruplex structure	G-score
APT1	5′-GGCUCGAGGAACGUACAGA-***GG***GU***GG***AGAGU***GG***GU***GG***-*AAGCUUACGGUACCUAG*-3′	**18**
APT2	5′-GGCUCGAGGACGUAACAGA-***GG***GA***GG***CGAGU***GG***GU***GG***-AAGCUUACGGUACCUAGC-3′	**18**
APT3	5′-GGCUCGAGGACCGAAGUCGA-***GG***G***GG***ACGUGA***GG***G***GG***-AAGCUUACGGUACCUAGC-3′	**16**
APT4	5′-GGCUCGAGGACCUAAGU-***GG***GAA***GG***UGAGC***GG***GUG***GG***-AAGCUUACGGUACCUAGC-3′	**19**
APT5	5′-GGCUCGAGGAACGUACAGA-***GG***GC***GG***AGAGU***GG***GU***GG***-AAGCUUACGGUACCUAGC-3′	**18**

Table S1. QGRS sequences found in five aptamers

G-quadruplex structure formation was predicted by Quadruplex forming G-Rich Sequences (QGRS). Suggested G-quardruplex structures were shown in italic and guanine nucleotides were underlined in bold.

Table S2

	#	Protein region	Interaction propensity	Discriminative power	Normalized score
APT1	1	101–152	9.97	28	1.77
	2	67–118	8.44	26	1.29
	3	26–77	7.41	24	0.97
	4	76–127	5.68	22	0.43
	5	42–93	4.1	20	-0.07
	6	117–168	1.54	17	-0.87
	7	92–143	1.04	17	-1.02
	8	51–102	0.67	17	-1.14
APT2	1	101–152	12.16	35	1.77
	2	67–118	10.42	32	1.31
	3	26–77	9.19	28	0.98
	4	76–127	7.16	24	0.44
	5	42–93	5.25	22	-0.07
	6	117–168	2.27	17	-0.86
	7	92–143	1.72	17	-1.01
	8	51–102	1.32	17	-1.11
APT3	1	101–152	13.5	37	1.77
	2	67–118	11.56	33	1.31
	3	26–77	10.15	32	0.97
	4	76–127	7.97	24	0.45
	5	42–93	5.89	22	-0.05
	6	117–168	2.6	17	-0.84
	7	92–143	1.85	17	-1.02
	8	51–102	1.49	17	-1.11

Table S2. Interaction probability between MLL1 and aptamers aptamers

Interaction propensity and discriminative power were calculated between protein fragments and aptamers from catRAPID *fragment.*

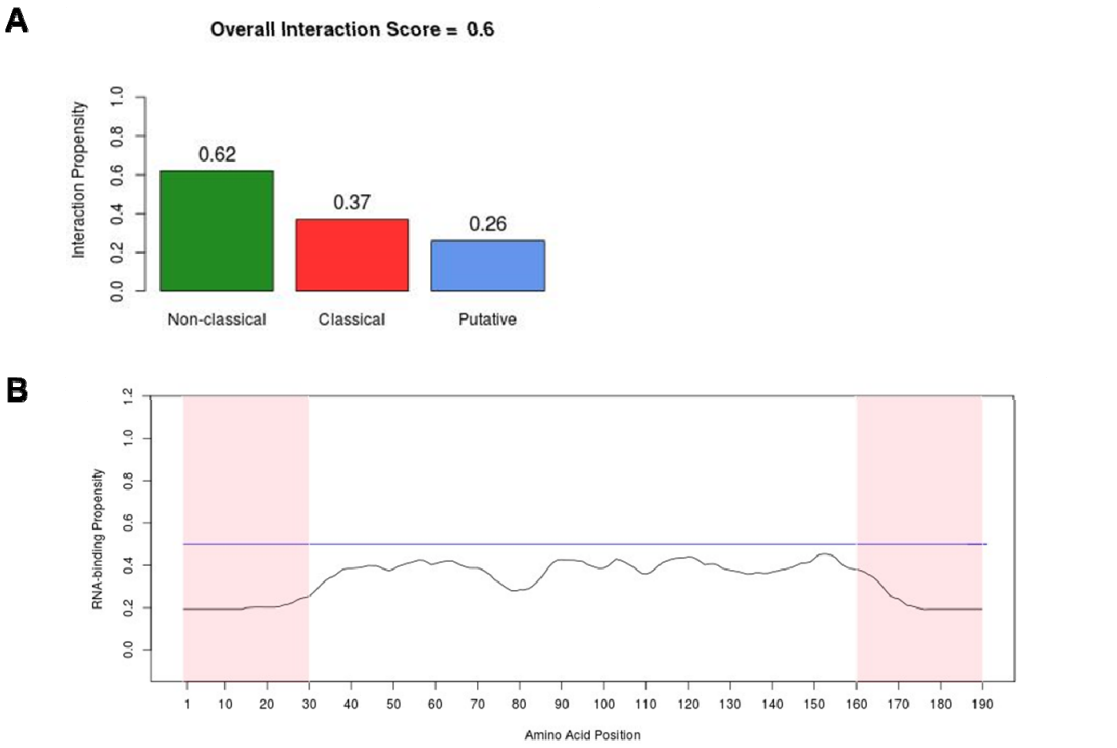
Supplementary Figure Legends

Figure S1. Computational prediction of MLL1-aptamer binding possibilities

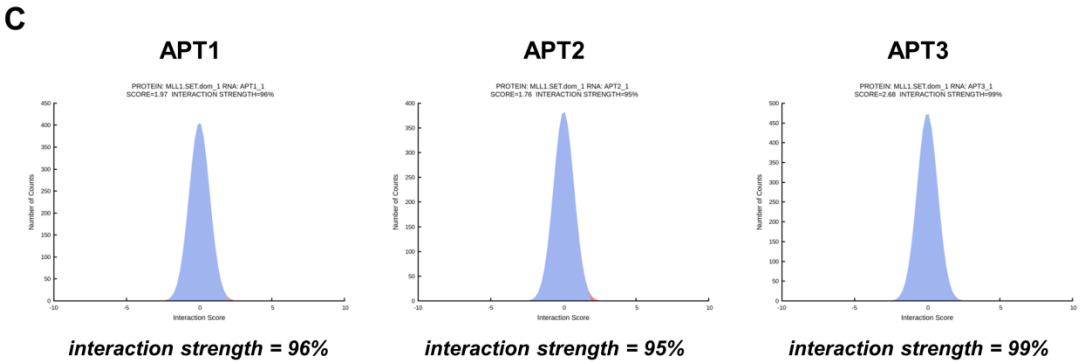
(A) RNA-binding ability of MLL1 protein was evaluated using *cat*RAPID *signature* algorithm. The web server reported the binding probability for the non-classical, classical, and putative RNA-binding protein classes, with an overall interaction score. Prediction score > 0.5 suggests high possibility for RNA binding. **(B)** The profile predicts the amino acid positions prone to bind RNA. **(C)** *cat*RAPID *strength* was used to predict the interaction strength of aptamers with MLL1. The interaction strength is computed using a reference set composed by 100 random protein and 100 random RNA sequences having the same lengths as the molecules under investigation.

63

Figure S1



64



interaction strength = 96% *interaction strength = 95%* *interaction strength = 99%*

65

66

67

Referencese

1. Zuker, M., *Mfold web server for nucleic acid folding and hybridization prediction.* Nucleic Acids Res, 2003. **31**(13): p. 3406-15.
2. Kikin, O., L. D'Antonio, and P.S. Bagga, *QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences.* Nucleic Acids Res, 2006. **34**(Web Server issue): p. W676-82.
3. Agostini, F., et al., *catRAPID omics: a web server for large-scale prediction of protein-RNA interactions.* Bioinformatics, 2013. **29**(22): p. 2928-30.