Electronic Supplementary Information for Chemical Communications

Direct screening for ribozyme activity in mammalian cells

Yoko Nomura, Hsiao-Chiao Chien and Yohei Yokobayashi*

Materials and Methods

Ribozyme library design and construction

Weinberg et al. [*Nat. Chem. Biol.*, 2015, **11**, 606-610.] reported 676 putative pistol ribozyme sequences identified by genome mining. After extracting the core sequences, 376 unique pistol ribozyme sequences were identified. Each of these unique sequences were divided into five structural elements: P1, L1/L2 (including pseudoknot), J1, P2, and P3. For each structural element, sequences were sorted by frequency, and 5 P1, 7 L1/L2, 5 J1, 5 P2, and 3 P3 sequences were selected (Table S1) for designing synthetic pistol ribozymes based on the frequency found in the natural sequneces and sequence diversity. Total of 2625 (=5×7×5×5×3) ribozyme variants were designed by combinatorially fusing these structural elements. The synthetic ribozymes were named sp-NNNNN with each N corresponding to a specific structural element. The original designations by Weinberg et al. were used for the natural ribozymes.

Five additional bases were added both upstream and downstream of each ribozyme. For the natural ribozymes, these were extracted from the original genome sequences. For the synthetic ribozymes, the most commonly found adjacent sequences associated with the particular P1 (5') and P2 (3') sequences were used.

Each ribozyme sequence was flanked by additional sequences as shown below:

5'agcggccgcgactctagagg[PistolRz]tccggagccataccacatNNNNNNNNNctaagccggaaaggggagac 3'

Green sequences are taken from the 3' UTR of the EGFP expression vector which was used for the reporter gene assay. The underlined sequences denote restriction sites. The 9 bases (N) in orange constitute the barcode that uniquely identifies the associated ribozyme sequence. Barcodes were generated so that every sequence differs from all other sequences by at least two positions. Each ribozyme sequence was randomly assigned four unique barcodes. The constant sequence in blue serves as a binding site for reverse transcription primer.

Total of 12004 sequences (3001 ribozymes with 4 barcodes each) of length 134-170 nt were synthesized on chip by CustomArray Inc., and obtained as ssDNA oligo pool. The oligos were converted to dsDNA by PCR using primers 5sp-f and RTp-r (Table S2). Separately, U6 promoter sequence was PCR amplified from pE22 [Nomura et al., *Chem. Commun.* 2012, **48**, 7215-7217.] using primers U6-f and 5sp-U6-r (Table S2). As 5sp-U6-r contains at its 5' end the 5' constant region of the ribozyme library, the U6 promoter fragment and the ribozyme library were used as templates in another PCR reaction (overlap PCR) using primers U6-f and sp-term-RTp-r (Table S2) to contruct the final library for transfection. The final library structure is as follows:

5'AATTACCCTCACTAAAGGGAGGAGAAGCATGAATTCCCCAGTGGAAAGACGCGCAGGCAAAACGCACCACGTGACGGAGCGTGAC
CGCGCGCCGAGCGCGCGCCAAGGTCGGGCAGGAAGAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGT
TAGAGAGATAATTAGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTA
GTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTCGATTTCTTGGGTTTATATAT
CTTGTGGAAAGGACGCGGGATCCGagcggccgcgactctagagg[PistolRz]tccggagccataccacatNNNNNNNNNctaagc
cggaaaggggagacTTTTTTttgccgtttttttcttccagg 3'

The sequence in purple corresponds to the U6 promoter, with the predicted transcription start site highlighted. Consecutive T's in the 3' end function as Pol III transcription terminator.

Sequencing library construction

HEK293 cells were plated on a 96 well plate in the growth medium (10 % FBS [Gibco] and 1× Antibiotic-Antimycotic Solution [Wako 161-23181] in D-MEM (High Glucose) with L-Glutamine and Phenol Red [Wako 044-29765]) one day before transfection (2.6×10^4 cells in 100 μL / well). The dsDNA library as prepared above (100 ng/well) was transfected into HEK 293 cells using Polyfect Transfection Reagent (QIAGEN) according to the manufacturer's instructions (53 wells in total). The growth medium was aspirated 6 h after transfection, and the cells were washed with DPBS(-) (Wako 045-29795) and trypsinized with 0.05% Trypsin-EDTA solution (Wako 202-16931). The collected cells (~2 mL) were suspended in 8 mL fresh growth medium in a 15 mL conical tube and centrifuged at 100 rpm for 3 min. After removing the supernatant, total RNAs were extracted from the cell pellet using Direct-zol RNA MiniPrep kit (Zymo Research, R2050). Isolated RNAs (32 μg in 25 μL) were immediately stored at -80 °C until PAGE purification.

The total RNA sample (12.8 μg) was treated with 4 units of DNase I (NEB M0303S) in 20 μL volume at 37 °C for 10 min, then denatured in RNA Loading Dye 2X solution (NEB B0363S) for 3 min at 95 °C. The denatured RNA sample was applied to 8 % denaturing polyacrylamide gel and separated by electrophoresis. Regions corresponding to the uncleaved (140-176 nt) and cleaved (61 nt) fragments were excised and placed in separate tubes. The gel slices were frozen at -80 °C, crushed, and RNAs were extracted in elution buffer (30 mM NaCl, 30 mM Tris-HCl pH 7.5). The eluted RNAs were precipitated using Quick-Precip Plus Solution (EdgeBio 72641) and dissolved in RNase-free water (23 μL).

The RNAs from the cleaved and uncleaved fractions were separately reverse transcribed with primer R1-503-Pistol (cleaved) or R1-504+2nt-Pistol (uncleaved) (Table S2). Each reaction contained 10 μL purified RNAs, 2 μL primer solution (10 μM), 1 μL dNTP mix (10 mM each), and 1.5 μL water (total 14.5 μL). The mixture was heated for 5 min at 65 °C and cooled on ice. Then, 5X reaction buffer (4 μL), RNase Inhibitor, Murine (NEB, M0314S) (0.5 μL) and MaximaH Minus RT (Thermo Fisher EP0752) (1 μL) were added on ice. The reaction mixture was kept at 65 °C for 30 min then at 85 °C for 5 min to inactivate the enzyme. Reactions without the reverse transcriptase were also prepared as controls. The cDNAs were then combined and used as the template in a PCR reaction (Phusion High-Fidelity PCR Master Mix with HF Buffer, NEB, M0531) with primers AdapterT2new and B2-CGAGTAAT-lsPsTest (Table S2) to attach adapter sequences for deep sequencing. The PCR product was purified by 3 % agarose gel electrophoresis and submitted for deep sequencing.

Deep sequencing data analysis

Deep sequencing was performed on an Illumina MiSeq instrument using MiSeq Reagent Kit v3. Single-end reads were performed. The raw reads were first quality filtered to remove those with >30% of the bases scoring (QS) below 20 which resulted in approximately 19.3 million reads. The filtered reads were then sorted to cleaved and uncleaved reads according to the tag sequence introduced during reverse transcription. After trimming the constant sequences, barcode sequences were counted and the results were exported to CSV format. Total of 11.9 million reads were used for the final analysis. Processing of sequencing data were performed using custom Python scripts. Further analysis of the sequencing data was performed on Microsoft Excel.

Construction of EGFP-ribozyme plasmids

Individual ribozyme sequences (Table S3) were synthesized and cloned in the 3' UTR region of EGFP expression plasmid pEGFP-N1-BspEI [Nomura et al., *ACS Synth. Biol.*, 2013, **2**, 684-689.] using restriction sites XbaI and BspEI. The ribozyme library contained 20-25 identical adjacent sequences on both sides to minimize the context dependence. Each plasmid was sequence verified by Sanger sequencing. The local sequence near the ribozyme insertion site is shown below:

5'GCATGGACGAGCTGTACAAGTAAagcggccgcgactctagagg[PistolRz]tccggagccataccacattgtagagg3'

The upper-case sequence in black is part of the EGFP coding region with the TAA stop codon highlighted.

EGFP reporter gene assay

EGFP reporter gene assay was performed essentially as previously reported.[Nomura et al., *ACS Synth. Biol.*, 2013, **2**, 684-689.] Briefly, HEK293 cells (2.6×10^4 cells in 100 μL growth medium per well) were plated in a 96 well plate one day before transfection. Fifty nanograms of an EGFP-ribozyme plasmid (or pEGFP-N1-BspEI as the "empty" vector), 10 ng of pCMV-mCherry (mCherry expression vector to normalize for transfection efficiency) and 50 ng pUC19 (to adjust the total DNA amount for transfection) were co-transfected using 1 μL Polyfect (QIAGEN) reagent. After 3.5 h incubation in a CO_2 incubator, the supernatant was replaced with 100 μL of fresh growth medium. Two days later, the medium was removed, the cells were washed with 100 μL of DPBS(-), and 150 μL of fresh DPBS(-) was added to each well. After ~10 min incubation in a CO_2 incubator, EGFP (484 nm excitation, 510 nm emission, 10 nm bandwidth) and mCherry (587 nm excitation, 610 nm emission, 10 nm bandwidth) fluorescence intensities were by a microplate reader (Infinite M100PRO, Tecan). Background fluorescence (untransfected cells) was first subtracted from the raw fluorescence values. Then, the EGFP fluorescence (background-subtracted) was divided by the mCherry fluorescence to account for the variations in transfection efficiency. This EGFP level was furhter normalized by that of the "empty" vector (pEGFP-N1-BspEI).

Table S1 Structural elements used to design synthetic pistol ribozymes (5' to 3')

P1	L1/L2	J	P2	P3
ACUC…GAGU	GUCUGGGC…UUAGGCCCAGAGCG	AUAAAUA	GUAA…UUAG	UCCCGGCAGCUGCCGGGG
GCUC…GAGG	GUCUGAGC…UUAAGCUCAGAGCG	AUAAAUAG	GGCU…AGUC	UCUUGGUCGGUAUUACCGGCUGGGG
AGUG…CACU	GUCAUAGC…UUAAGCUAUGAGCG	GUAAAUA	GACA…UGUA	GCAGGGUUCUUCCCUGC
AGCC…GGCU	GACUAAGC…UUAGGCUUAGAGCG	UUAAAUA	GGUG…CACA	
CGUC…GACG	GACUAUGC…UUAGGCAUAGUGCG	AUAAAC	GUUA…UAAC	
	GUUAGGGC…UUAAGCCCUAAGCG			
	GUGGCGGC…UUAAGCCGCCAGCG			

Table S2 Oligonucleotide sequences

Name	Sequence (5' to 3')
5sp-f	AGCGGCCGCGACTCTAGAGG
RTp-r	GTCTCCCCCTTTCCGGCTTAG
U6-f	AATTACCCTCACTAAAGGGAGGAGAAGC
5sp-U6-r	CCTCTAGAGTCGCGGCCGCTCGGATCCCGCGTCCTTTCCA
sp-term-RTp-r	CCTGGAAGAAAAAACGGCAAAAAAAAGTCTCCCCCTTTCCGGCTTAG
R1-503-Pistol	ACACGACGCTCTTCCGATCTTATCCTCTGTCTCCCCCTTTCCGGCTTAG
R1-504+2nt-Pistol	ACACGACGCTCTTCCGATCTAGAGTAGACAGTCTCCCCCTTTCCGGCTTAG
AdapterT2new	AATGATACGGCGACCACCGAGATCTACACACACTCTTTCCCTACACGACGCTCTTCCGATC
B2-CGAGTAAT-lsPsTest	CAAGCAGAAGACGGCATACGAGATCGAGTAATGTGACTGAGAATCCGGAGCCATACCAC

Table S3 Ribozyme sequences[a]

Name	Sequence (5' to 3')
env-842	ACTAAACTCGATACGGCGAGTATAAATAGGACTTTAAGCCGTAAGCGTTCCCGCCGCAGGGCGGGAGGTAGTCAATCC
sp-51343	AAGCACGTCGTCTGGGCGACGGTAAATAGGTGTTAGGCCCAGAGCGGCAGGGTTCTTCCCTGCGTCACATTAAT
env-854	GTAACACTCGACTAGGCGAGTATAAACAGGTGTCAAGCCTAGTGCGTTACCCACTCAAAAAGAGTGGGTAGGTCACATAGAT
sp-15143	AGAAAACTCGACTATGCGAGTATAAATAGGTGTTAGGCATAGTGCGGCAGGGTTCTTCCCTGCGTCACATTAAT
sp-16143	AGAAAACTCGTTAGGGCGAGTATAAATAGGTGTTAAGCCCTAAGCGGCAGGGTTCTTCCCTGCGTCACATTAAT
sp-11141[b]	AGAAAACTCGTCTGGGCGAGTATAAATAGGTGTTAGGCCCAGAGCGTCCCGGCAGCTGCCGGGGGTCACATTAAT
sp-11251[b]	AGAAAACTCGTCTGGGCGAGTATAAATAGGTTATTAGGCCCAGAGCGTCCCGGCAGCTGCCGGGGGTTAACTAATA
sp-56441[b]	AAGCACGTCGTTAGGGCGACGTTAAATAGGTGTTAAGCCCTAAGCGTCCCGGCAGCTGCCGGGGGTCACATTAAT
sp-17341	AGAAAACTCGTGGCGGCGAGTGTAAATAGGTGTTAAGCCGCCAGCGTCCCGGCAGCTGCCGGGGGTCACATTAAT
sp-55443	AAGCACGTCGACTATGCGACGTTAAATAGGTGTTAGGCATAGTGCGGCAGGGTTCTTCCCTGCGTCACATTAAT
sp-33222	AATAAAGTGGTCATAGCCACTATAAATAGGGCTTTAAGCTATGAGCGTCTTGGTCGGTATTACCGGCTGGGGGTAGTCAAAAA
sp-15441	AGAAAACTCGACTATGCGAGTTTAAATAGGTGTTAGGCATAGTGCGTCCCGGCAGCTGCCGGGGGTCACATTAATT
env-839	AGAACAGTCGTTTGAGCGACTATAAATAGGACTTTAGGCTCAAAGCGTTCCCATCTACATAAGGTGGGAGGTAGTCGGCTA
env-814	TAATAAGTGGTCATAGCCACTATAAACAGGGCTTTAAGCTATGAGCGTTCCCCGTCAGATTTGACGGGTAGGTAGTCAAAAC
sp-27251	ATGAAGCTCGTGGCGGCGAGGATAAATAGGTTATTAAGCCGCCAGCGTCCCGGCAGCTGCCGGGGGTTAACTAATA
env-423	TTATGACTCGTCCCAGCGAGTATAAATAGGACATTAGGCTGGAGCGGTACATTTGTGCGTTGTCAAACC
sp-43522	TATAAAGCCGTCATAGCGGCTATAAACGGCTTTAAGCTATGAGCGTCTTGGTCGGTATTACCGGCTGGGGGTAGTCAAAAA
env-168	CGCTGACCCGTTCCGGCGGGTATAAACAGGGCATTAGGCCGGAAGCGTGGGCGAGATTTCGCCCAGTTGTCGAAAA
sp-12513	AGAAAACTCGTCTGAGCGAGTATAAACGTAATTAAGCTCAGAGCGGCAGGGTTCTTCCCTGCGTTTAGCAGTT
env-874	GCCGCGGGCGTCTGGACGCCTAAAACCGCGTTAAGTCCAGAGCGGGCACCGGGCGTATCGGTGCCGTCGTGATCTC
sp-32553	AATAAAGTGGTCTGAGCCACTATAAACGTTATTAAGCTCAGAGCGGCAGGGTTCTTCCCTGCGTTAACTAATA
sp-32513	AATAAAGTGGTCTGAGCCACTATAAACGTAATTAAGCTCAGAGCGGCAGGGTTCTTCCCTGCGTTTAGCAGTT
env-873	GCCGCGGGCGTCTGGACGCCCTAAAACCGCGTTAAGTCCAGAGGCGGGCGACCGGGCGTATCGGTGCCGTCGTGATCTC

[a]The bases are color-coded by structural elements as depicted in Figure 2.

[b]Inactive versions of these ribozymes were constructed by mutating the underlined GC to TA.