Angewandte Chemie
Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information

High-Throughput Mutational Analysis of a Twister Ribozyme

*Shungo Kobori and Yohei Yokobayashi**

anie_201605470_sm_miscellaneous_information.pdf

Experimental Details

Oligonucleotides

The partially doped oligonucleotide **Doped-Twister-R** (Table S2) was purchased from IDT with a "hand-mix" option for the doped bases where a mixture of 97% wild-type and 1% each of the remaining bases was coupled. Other oligonucleotides used in this experiment are listed in Table S2.

In vitro transcription

The dsDNA template used for in vitro transcription of the ribozyme mutants was prepared by annealing and extending **T7-Bio-F** and **Doped-Twister-R** (Table S2) using Phusion High-Fidelity PCR Master Mix with HF Buffer (New England Biolabs). The template was purified by DNA Clean & Concentrator-5 kit (Zymo Research). The sense strand sequence of the resulting dsDNA template is: 5' TAATACGACTCACTATAGGGCCGCCT**AACACTGC CAATGCCGGTCCCAAGCCCGGATAAAAGTGGAGGGGGCGG**CGCGTGGATATG GCACGCAA 3' where the underlined sequence corresponds to the T7 promoter and the bold bases represent the doped positions. Because **T7-Bio-F** is biotinylated at its 5' terminus, the dsDNA template is also biotinylated. The ribozyme mutants were transcribed in vitro in a 25 μL reaction in the transcription buffer (40 mM Tris-HCl pH 8.0, 2 mM spermidine, 10 mM DTT, 6 mM $MgCl_2$; measured pH = 8.1 at 25 °C) in the presence of the dsDNA template (5 pmol), NTPs (2 mM each), RNase Inhibitor, Murine (40 units, New England Biolabs), and T7 RNA polymerase (62.5 units, New England Biolabs) for 2 h at 37 °C. The reaction was terminated by adding 25 μL of 10 mM EDTA and incubated with 45 μL of streptavidin-coated magnetic beads (Streptavidin MagneSphere, Promega) for 20 min on ice with consistent pipetting. The beads were removed and the transcribed RNAs were purified by RNA Clean & Concentrator-5 kit (Zymo Research) and dissolved in 20 μL 10 mM EDTA.

Reverse transcription and template switching reactions

The transcribed and purified RNA mixture (10 pmol) was mixed with 400 pmol of **RT-Twister** (Table S2) and dNTPs (2.82 mM each) in a volume of 23 μL and heated to 72 °C for 3 min and placed on ice. Reverse transcription was initiated by adding 79.5 μL of RT buffer (81.2 mM Tris, pH 8.3, 121 mM KCl, 12.6 mM $MgCl_2$, 8.1 mM DTT, 1 unit/μL of RNase Inhibitor, Murine, 12.6% PEG), 12.6 units/μL of Maxima Reverse Transcriptase (Thermo Fisher Scientific), and 12.6 μM of the **TSO** solution (Table S2). The solution was kept at 42 °C for 1 h, then at 60 °C for 2 min, and finally at 37 °C for 2 h. The reaction was stopped by adding 5 μL NaOH solution (5 M) and heating (95 °C for 3 min).

Some reverse transcriptases are known to add ~3 deoxycytosines (Cs) at the 3' end of the cDNA in a template independent fashion. At some frequency, template switching occurs via hybridization of this extra 3'-CCC of the cDNA with the 3'-rGrGrG of the template switching oligonucleotide (**TSO**) and the cDNA is extended further using **TSO** as the new (switched) template. This mechanism has been used to append adapter sequences to the 3' ends of reverse-transcribed cDNA libraries for deep sequencing applications.[1] The small modification of this adapter attachment step from our previous work[2] which used a splint oligonucleotide and T4 DNA ligase was necessary because the 3' end of the cDNA in this library contains mutations. Splint oligonucleotide-dependent ligation can only be used when the ligase substrate sequences at the ligation junction are fixed.

The cDNAs were first cleaned by Oligo Clean & Concentrator kit (Zymo Research), purified by denaturing PAGE (8 % polyacrylamide, 8 M urea, 19 : 1 acrylamide : bisacrylamide), and recovered by passive elution in 200μL TENa buffer (200 mM NaCl, 10 mM Tris-HCl pH 7.5, 5 mM EDTA) overnight at 4 °C followed by ethanol precipitation using Quick-Precip Plus Solution (Edge BioSystems) according to the manufacturer's instructions. The purified cDNAs were dissolved in 10 μL water. The sequencing library was generated by PCR with primers **PCR-F** and **PCR-R** (Table S2) using Phusion High-Fidelity PCR Master Mix with HF Buffer (New England Biolabs) in a 50 μL reaction containing 5 μL cDNAs and 0.5 μM each primer. The PCR reaction was performed in nine cycles of 10 s at 98 °C followed by 20 s at 72 °C. The PCR products were purified by agarose gel electrophoresis using Zymoclean Gel DNA Recovery Kit (Zymo Research).

Sequencing and data analysis

The libraries were sequenced on an Illumina HiSeq 2500 sequencer using SR Rapid Cluster Kit v2-HS and Rapid SBS Kit v2-HS (92 cycles, single-end) with 30% PhiX control by OIST DNA Sequencing Section. The sequencing data were analyzed with custom Perl scripts unless otherwise noted. First, the read quality was checked by NGS QC Toolkit v2.3[3] and low quality reads with a Phred quality score under 20 were removed. Reads that include N were also removed by FASTX-toolkit. The reads were sorted as twister ribozyme based on the first six sequenced bases in the undoped region at the 3' end of the ribozyme. Each ribozyme read records the identity of the mutant from the doped region, and was classified as cleaved or uncleaved based on the identity of the six bases upstream of the expected cleavage site.

Gel-based ribozyme assay

Individual ribozyme mutants tested were first cloned in a plasmid and sequence verified. The plasmids were used to prepare in vitro transcription templates by PCR using Q5 High-Fidelity 2X Master Mix (New England Biolabs). In vitro transcription reactions were

performed under the same conditions as the library construction in 10 μL scale. The transcribed RNAs were separated by 8 % denaturing PAGE and stained by SYBR Gold (Thermo Fisher Scientific). The gels were photographed and analyzed using SE-6100 LuminoGraph I (ATTO).

Applicability of the method to ribozymes of different sizes

Applicability of our method to other ribozymes mainly depends on the size of the ribozyme sequence targeted for mutation and the total number of sequencing reads. For a ribozyme with a length of N bases (targeted for mutagenesis), there are $3^i \times {}_N\mathrm{C}_i$ possible variants with exactly i mutations compared to the wild-type (e.g. wild-type is i=0, single mutants are i=1, double mutants are i=2, etc.).

If the doped bases are statistically mutated to any one of the three bases other than the wild-type with a probability of m per position, the probability (f_i) of a DNA template encoding a specific mutant with exactly i mutations can be expressed as follows.

$$f_i = (1\text{-}m)^{N-i} \times (m/3)^i$$

Therefore, the expected average number of reads per mutant with exactly i mutations (R_i) can be expressed as:

$$R_i = R_{Total} \times f_i = R_{Total} \times (1\text{-}m)^{N-i} \times (m/3)^i$$

where R_{Total} is the total number of sequencing reads.

Given R_{Total} defined by the currently available sequencing technology (e.g. ~2×10^9 using an Illumina HiSeq 4000 instrument), one can vary m and N to estimate the technical limits of our method. If we set an arbitrary threshold of $R_i = 500$ and $R_{Total} = 2 \times 10^9$, R_2 (number of expected average read counts for each double mutant) is greater than the threshold for N up to 379 bases when $m = 0.01$. This size is sufficiently large to cover most known ribozymes.

Higher order mutants (triple mutants, etc.) are less amenable to our method due to the exponentially increasing number of possible mutants, but coverage may be possible for shorter sequences. For example, $R_3 > 500$ can be achieved for $N = 59$ ($m = 0.05$).

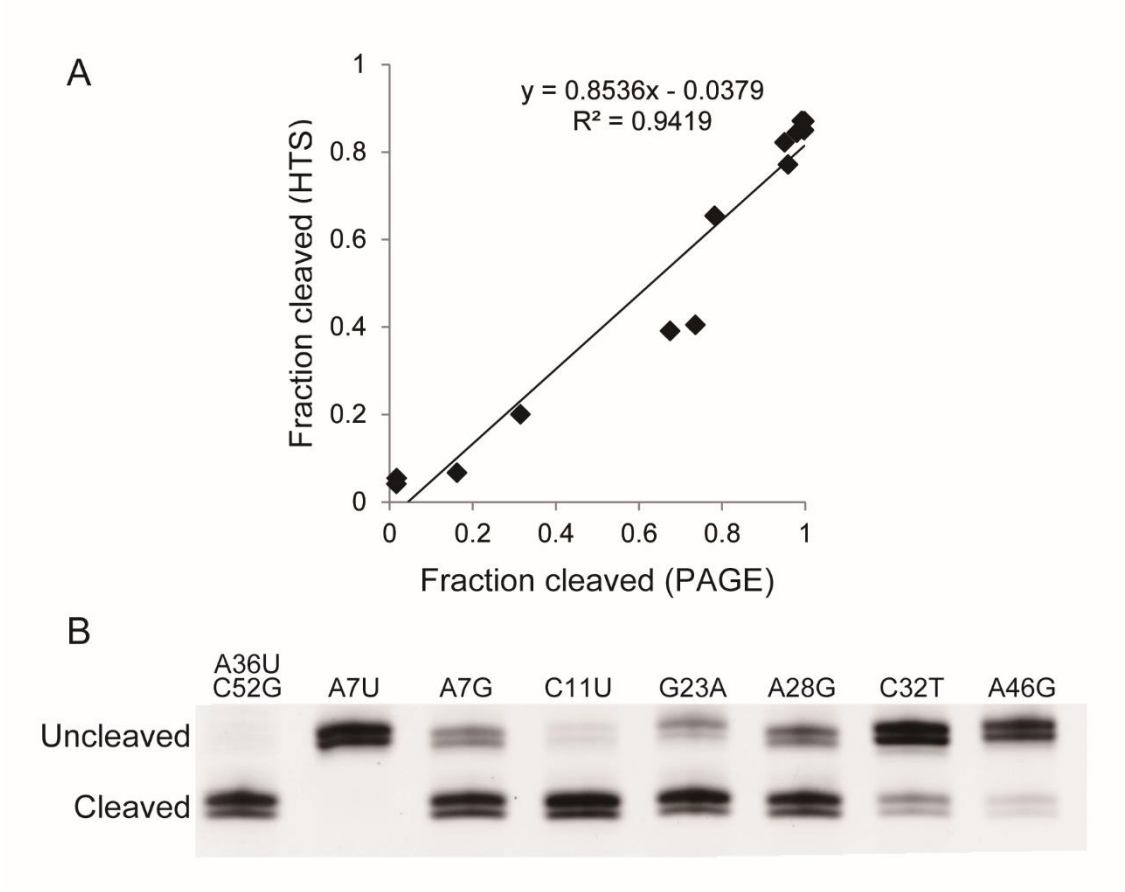

Figure S1. Validation of the sequencing-based ribozyme assay. (A) Correlation of "fraction cleaved" (FC) values obtained by deep sequencing (HTS) and conventional gel-based assay (PAGE) of selected ribozyme mutants. The original data for this plot are provided in Table S3. (B) A representative gel image of individual ribozyme assay based on PAGE. As seen in this case, run-off in vitro transcription using T7 RNA polymerase sometimes yields transcripts with an extra (template independent) base at the 3' terminus.[4]

Table S1 Summary of deep sequencing results.

Number of mutations	Total mutants	Reads	Average reads per mutant	S.D.
0 (wild-type)	1	6,193,111	–	–
1 (single mutants)	144	13,469,920	93,541	43,248
2 (double mutants)	10,152	13,861,619	1,365	727
3 (triple mutants)	466,992	8,470,262	18.2	17.9

Table S2 Oligonucleotides used in this study.

Name	Sequence (5' – 3')	Notes
T7-Bio-F[a]	Biotin-TAATACGACTCACTATAGGGCCGC	Template preparation.
Doped-Twister-R[a]	TTGCGTGCCATATCCACGCG**CCGCCCCCTCCAC TTTTATCCGGGCTTGGGACCGGCATTGGCAG TGTT**AGGCGGCCCTATAGTGAGTCGTATTA	Template preparation. Doped bases shown in bold.
RT-Twister[b]	AACACTCTTTCCCTACACGACGCTCTTCCGATCT TTGCGTGCCATATCCACGCG	RT primer.
TSO[c]	CAAGCAGAAGACGGCATACGAGATrGrGrG	Template switching oligonucleotide. rG indicates RNA bases
PCR-F[b]	AATGATACGGCGACCACCGAGATCTACACTCTTT CCCTACACGACGCTCTTCCGATCT	PCR primer.
PCR-R[c]	AAGCAGAAGACGGCATAC	PCR primer..

[a] Purchased from IDT.
[b] Purchased from Macrogen Japan Co.
[c] Purchased from FASMAC Co.

Table S3 Summary of selected ribozyme activity measurements by HTS and PAGE.

	HTS			PAGE[a]
Mutation	Uncleaved reads	Cleaved reads	Fraction cleaved	Fraction cleaved
WT	1,108,635	5,084,476	0.82	0.95
A7U	117,840	6,682	0.05	0.02
A7G	45,910	31,113	0.40	0.74
A10U	84,302	577	0.04	0.02
C11U	14,071	47,324	0.77	0.96
G23A	49,416	93,281	0.65	0.78
A28G	42,054	26,928	0.39	0.68
C32U	67,233	16,787	0.20	0.32
A46G	67,045	4,783	0.07	0.16
A36U, C52U	65	435	0.87	1.00
A36U, C52G	44	296	0.87	0.99
A38C, C52G	38	205	0.84	0.98
A38C, C52U	55	309	0.85	1.00

[a] The values are averages of two independent measurements.

References

[1] S. Picelli, O. R. Faridani, Å. K. Björklund, G. Winberg, S. Sagasser, R. Sandberg, *Nat. Protoc.* **2014**, *9*, 171-181.

[2] S. Kobori, Y. Nomura, A. Miu, Y. Yokobayashi, *Nucleic Acids Res.* **2015**, *43*, e85.

[3] R. K. Patel, M. Jain, *PLoS ONE* **2012**, *7*, e30619.

[4] a) J. F. Milligan, D. R. Groebe, G. W. Witherell, O. C. Uhlenbeck, *Nucleic Acids Res.* **1987**, *15*, 8783-8798; b) C. Kao, M. Zheng, S. Rüdisser, *RNA* **1999**, *5*, 1268–1272.

Supporting Data: The original Excel file is available from http://yokobayashilab.net/data.html or from the corresponding author.

Complete matrix of the relative activity of all single and double mutants