Supplementary Data

Title: "An *In Vitro*-Selected RNA Receptor for the GAAC Loop: Modular Receptor for Non-GNRA-type Tetraloop"

Authors: Junya Ishikawa, Hiroyuki Furuta, and Yoshiya Ikawa*

* To whom correspondence should be addressed. Tel:+81-92-802-2866; Fax: +81-92-802-2867; Email:yikawa@cstf.kyushu-u.ac.jp

Contents

A) Supplementary Materials and Methods

The sequences of oligo DNA and RNA used for library construction, selection, and chemical modification assay were listed in Table S1, S3, and S6.

1. Library preparation

1.1 PCR amplification of 5'-fragment DNA

The three libraries used in this study vary according to its P1b length, from 6 to 8 bps. 5'-fragments including the P1b region were amplified by PCR. The 100 µl of reaction mixuture contained 5.0 pmol of template (kyi-425 for library{7}, kyi-427 for library{8}, or kyi-429 for library{6}), 50 pmol each of sense and antisense primers (kyi-25 and kyi-431), 10 µl of 10-fold PCR buffer, and 0.125 µM of each dNTP. The PCR reaction was performed with ExTaq DNA polymerase (Takara Bio, Japan) using the following program: 94°C for 2 min followed by 8 cycles of 94°C for 1 min, 55°C for 1 min, and 72°C for 1 min.

1.2 PCR amplification of 3'-fragment DNA

The oligo DNAs containing randomized region were purchased from HSS and purified on 15% non-denaturing PAGE before use. The PCR reaction was performed with ExTaq DNA polymerase (Takara) using the following program: 94°C for 2 min followed by 4 cycles of 94°C for 1 min, 60°C for 1 min, and 72°C for 1 min. The 100 µl of reaction mixture contained 5.0 pmol of template (kyi-432 or kyi-433), 50 pmol each of sense and antisense primers (kyi-434 and kyi-435), 10 µl of 10-fold PCR buffer, and 0.125 µM of each dNTP. The PCR reaction was performed repeatedly until 200 pmol of product DNA was obtained.

Table S1. Oligo DNAs used for library synthesis (N, R, and Y stand for any nucleotides, A or G, and C or T, respectively)

Name	5'-sequence-3'	
3'-fragment DNA containing randomized region		
Kyi-432	TGTGTGGTCTCGGCACCTCARNNNNNNNNNNNNNNNNNNNNNNNNY CAATGGGTAGGACCATCCGTTCCCTAGCAGGTGTCC	Template
Kyi-433	GGACACCTGCTAGGGAACGGATGGTCCTACCCATTGRNNNNNNNNNNN NNNNNNNNNNNNNNYTGAGGTGCCGAGACCACACA	Template
Kyi-434	CGGTCCCTGTGTGGTCTCGGCACCTCA	Sense
Kyi-435	TCTGCCTAAGTGGGCAATGAGACTGGACACCTGCTAGGGAACGG	Antisense
5'-fragment DNA		
Kyi-425	GAAGTAATACGACTCACTATTAGGATCCGAACGGATCCTGTGGATTTTTT GGCACCTCA	Template
Kyi-427	GGAGTAATACGACTCACTATTAGGCATCCGAACGGATGCCTGTGGATTTT TTGGCACCTCA	Template

Kyi-429	GAAGTAATACGACTCACTATTAGATCCGAACGGATCTGTGGATTTTTTGGCACCTCA	Template
Kyi-25	GAAGTAATACGACTCACTATTAG	Sense
Kyi-431	GAAGCTTGAGGTGCCAAAAAA	Antisense

Nucleotides colored by yellow and green represent recognition sites for restriction enzymes, *Bsa*I and *Ban*I.

1.3 Digestion and ligation of the fragments

The resulting DNA fragments were extracted by TE-saturated phenol, precipitated with EtOH, rinsed by 70% EtOH, and dissolved in water. Then, the 5'- and 3'-fragments were digested with BanI and both BanI and BsaI, respectively, followed by purification on 10% non-denaturing PAGE. DNA products were visualized by UV shadowing, cut from the gel, and eluted overnight at 16 ℃ in a buffer [300 mM NaOAc pH 7.5, 0.1% SDS]. The samples were extracted by TE-saturated phenol, precipitated with EtOH, rinsed by 70% EtOH, dissolved in water, and quantitated by UV absorption measurement at 260 nm. The fragments were ligated with T4 DNA ligase (Takara Bio, Japan), and the resulting DNA was purified by ethanol precipitation and used as a template for transcription.

1.4 Transcription

RNA was synthesized enzymatically by in-vitro run-off transcription of the ligated DNA template using T7 RNA polymerase (RNAP; homemade) by incubation for 4 hours at 37 ℃ in a buffer containing 10 mM $MgCl_2$, 10 mM dithiothreitol, 2 mM sperimidine, 40 mM Tris-Cl pH 8.0, 1.0 mM of each NTP, RNase inhibitor (Takara Bio, Japan), and T7 RNAP to a total volume of 200 μl. DNase (RQ1 DNase, promega, USA) was added into the mixture to quench the reaction by digesting the DNA templates during incubation at 37 ℃ for 30 min, followed by purification on 10% denaturing PAGE. RNA products were recovered from gel as essentially mentioned above.

2. *In vitro* selection

In vitro selection was performed as described previously (18). The selection pressure applied in this experiment was catalytic activity of the cisDSL ribozyme of which RNA-RNA ligase activity largely depended on loop/receptor interaction. Therefore, ligation between DSL-based RNA libraries and a biotinylated RNA substrate and the following capture of RNA by using biotin-streptavidin interaction gave enrichment of receptors interacting with the loop from randomized sequences.

2.1 Ligation reaction with biotinylated RNA substrate

Ligation reaction was performed under the following conditions: 50 mM $MgCl_2$, 25 mM KCl and 30 mM Tris-Cl (pH7.5) at 37°C. In order to avoid enrichment of undesired RNA that can be recovered regardless of the loop/receptor interaction, two sets of 5'-biotinylated substrate RNA and their guide sequences, S1 and S2, were alternatively employed in each round.

2.2 Capturing RNA capable of catalysis by biotin-streptavidin interaction

Ligation mixtures were precipitated with EtOH and dissolved in 0.50-fold SSC buffer [75 mM Sodium Chloride, 7.5 mM Sodium Citrate]. Then, the samples were mixed with streptavidin coated magnetic beads (Promega, USA) suspended in the same buffer just prior to mixing. After 10 minute incubation at room temperature, the beads were washed three times with 0.10-fold SSC buffer and three times with distilled water, and then resuspended in a reverse-transcription (RT) buffer. RT primer (either Kyi-435, -436, or -437) and dNTPs were added to the samples, followed by heating at 64°C for 2 min and annealing at 45°C for 5 min. Dithiothreitol, RNase inhibitor (Takara Bio, Japan), and RTase (TOYOBO, Japan) were added to start RT at 50°C for 1 hour. Reaction was stopped by adding 10-fold alkaline buffer [1.5 N KOH and 0.20 M tris(hydroxymethyl)aminomethane] to degrade RNA and then neutralized with 1.5 N HCl. The resulting cDNA was used as a template for the following 1st PCR.

2.3 Selective amplification of cDNA of ligated product RNA by 1st PCR and regeneration of libraries by 2nd PCR

1st PCR with a sense primer (1st Fw-1 or 1st Fw-2) that has sequence complement to the substrate (S1 or S2) selectively amplified the ligated product, not unreacted RNAs. The PCR product was purified on 2% agarose gel, and the resulting DNA was used as a template for 2nd PCR. The 2nd PCR with a sense primer (kyi-425 or -426 for library{7}, kyi-427 or -428 for library{8}, and kyi-429 or -430 for library{6}) that has T7 RNA promoter sequence re-generated DNA libraries. The resulting DNAs were transcribed into RNA libraries used in the next selection step. Five rounds of selection steps from 2.1 to 2.3 were performed.

Table S2. Conditions used in first selection

Round	Library (pmol)	Biotin-Substrate (pmol/type)	Conditions	Reaction time
1	300	1500/S1	50 mM Mg^{2+}, 25 mM K^{+}, pH 7.5, 37°C	12 hour
2	200	400/S2		1 hour
3	100	200/S1		1 hour
4	100	200/S2		10 min
5	100	200/S1		10 min

Table S3 Oligo RNA and DNA used for cycling the selection (Italic represents ribonucleotide)

Name	5’-sequence-3’	
RNA substrates for reaction with the library DSL ribozyme		
Biotinylated substrate-1	Biotin-*CGUACACGUACUCACGCGUAUACAGUCCAC*	RNA substrate (S1)

Biotinylated substrate-2	Biotin-*ACUUCCGAGCUGUAGAGUUAGCAGCGACAC*	RNA substrate (S2)
Sense primers for 1st selective PCR		
1st Fw-1	CGTACACGTACTCACGCGTATACAGTC	Sense (S1)
1st Fw-2	ACTTCCGAGCTGTAGAGTTAGCAGCGA	Sense (S2)
Sense primers for 2nd regenerative PCR		
Kyi-425	GAAGTAATACGACTCACTATTAGGATCCGAACGGATCCTGTGGATTTTTT GGCACCTCA	Sense (S1)
Kyi-426	GAAGTAATACGACTCACTATTAGGGACCGAACGGTCCCTGTGTCTTTTTT GGCACCTCA	Sense (S2)
Kyi-427	GGAGTAATACGACTCACTATTAGGCATCCGAACGGATGCCTGTGGATTTT TTGGCACCTCA	Sense (S1)
Kyi-428	GAAGTAATACGACTCACTATTAGGCTACCGAACGGTAGCCTGTGTCTTTT TTGGCACCTCA	Sense (S2)
Kyi-429	GAAGTAATACGACTCACTATTAGATCCGAACGGATCTGTGGATTTTTTGG CACCTCA	Sense (S1)
Kyi-430	GAAGTAATACGACTCACTATTAGGACCGAACGGTCCTGTGTCTTTTTTGG CACCTCA	Sense (S2)
Antisense primers for reverse-transcription, 1st and 2nd PCR		
Kyi-435	TCTGCCTAAGTGGGCAATGAGACTGGACACCTGCTAGGGAACGG	Antisense
Kyi-436	TCTGCCTAAGTGGGCAATGAGACTGGACAC	Antisense
Kyi-437	TCTGCCTAAGTGGGCAATGAGACTGG	Antisense

2.4 Selecting RNA of interest by gel separation, cloning, and sequencing

After the 5th round of selection, 1.0 μM (20 pmol) RNA libralies were subjected to ligation reaction with 1.0 μM 5'-FAM labeled substrate RNA and buffer used in selection reaction for 10 hours, and then the mixtures were separated on 9% denaturing PAGE. The ligated products were sliced from the gel and soaked in elution buffer [0.3 M NaOAc (pH 7.5), 0.1% SDS] overnight at 16°C. The RNA was recovered by ethanol precipitation and dissolved in water, then used as a template for reverse transcription with RTase. The resulting cDNA was amplified by PCR with Taq DNA polymerase and the solution was separated on 2% low-melting point agarose gel. The band of interest was recovered and cloned into pGEM T-Vector (Promega, USA) for blue-white screening. 49 clones were sequenced. One dominant and two minor sequences were subjected to secondary structure prediction by using mFold (23).

2.5 Doped selection

Doped-libraries were synthesized by mutating at the rate of 45% or 15% per position into the receptor sequence embedded in the cisDSL_L(GAAC)_R(GAAC). Preparation of the RNA libraries was done as described above. The doped libraries were subjected to selection under the condition listed in Table S4 and S5. After 2 and 4 rounds of selection, libraries were cloned and sequenced. In the case of a mixed library of 15% and 45% libraries, all sequences were same to the original R(GAAC) (12 of 12 and 24 of 24 clones, which were sequenced after 2 and 4 rounds of selection respectively, were same to the original R(GAAC) sequence). In contrast to this rapid enrichment, the 45% library after 2 rounds of selection had variants (Figure S1 A).

Table S4. Conditions used in 15% and 45% mix doped selection

Round	Library (pmol)	Biotin-Substrate (pmol/type)	Conditions	Reaction time
1	1000	1500/S1		18 hour
2	300	450/S2	50 mM Mg^{2+}, 25 mM K^{+},	1 hour
3	300	450/S1	pH 7.5, 37°C	5 min
4	300	450/S2		5 min

After 2nd and 4th rounds, the original sequence was recovered (All 12 clones sequenced after 2nd, and all 24 after 4th had the original sequence.).

Table S5. Conditions used in 45% doped selection

Round	Library (pmol)	Biotin-Substrate (pmol/type)	Conditions	Reaction time
1	300	450/S1	50 mM Mg^{2+}, 25 mM K^{+},	18 hour
2	300	450/S2	pH 7.5, 37°C	12 hour

After 2nd round, sequencing was performed. (Figure SI. Sequences obtained from doped selection).

3. Dissociation constant (K_d) determination

K_d values for tectoRNA hetero-dimer formation were derived from non-linear fitting with the following equation (A) as previously reported (46).

$$K_d = \frac{[M1]\cdot[M2]}{[D]} = \frac{([M1]_0 - f\cdot[M1]_0)\cdot[M2]}{f\cdot[M1]_0} = \frac{(1-f)\cdot[M1]_0\cdot[M2]}{f\cdot[M1]_0} = \frac{(1-f)\cdot[M2]}{f}$$

$$\therefore \quad f = \frac{[M2]}{K_d+[M2]} \qquad \text{(A)}$$

, where [M1] and $[M1]_0$ are concentrations of free and initial labeled RNA, [M2] is concentration of the free unlabeled RNA, [D] is dimer concentration, and *f* is fraction of labeled-RNA dimerized that

can be calculated from band intensities. Note that [M2] was calculated by $[M2] = [M2]_0 - f \cdot [M1]_0$, where $[M2]_0$ equal to the concentration of the unlabeled RNA added to the solution. K_d values correspond to the average of two independent experiments.

4. Chemical probing experiments with DMS and CMCT

4.1 Chemical modification

DMS modification experiments were performed essentially as described previously (24,25). RNA (100 pmol for tectoRNA, 20 pmol for cisDSLs and intron ribozymes) dissolved in 45 uL of distilled water was heated at 80°C for 3 min, then snap-cooled on ice for 2 min. Immediately afterward, the RNA was allowed to fold at appropriate temperature (30°C, 37°C, and 50°C for tectoRNAs, cisDSLs, and *Tetrahymena* intron ribozymes respectively) for 30 min by adding 5.0 uL of 10-fold Tris-buffer [10 mM Tris-OAc (pH7.5) and 15 mM $Mg(OAc)_2$ or 0.10 mM EDTA (pH8.0) at final concentration]. Then 1.0 uL of 20% DMS in EtOH was added, followed by incubation for appropriate duration (5 min for tectoRNA and cisDSLs, 1 min for intron ribozymes). Modification reactions were quenched by adding 10 uL of 6-fold stop solution [400 mM DTT and 1.8 M NaOAc (pH7.0)] and 180 uL of cold EtOH. For negative controls, DMS was added after adding the stop solution and EtOH. RNAs were precipitated and dissolved in water. Note that because the DMS is highly toxic and a suspected carcinogen, appropriate precautuions should be taken for handling, storage, and disposal. Modification experiments with 1-cyclohyxyl-3-(2-morpholinoethyl)carbodiimide metho-p-toluenesulfonate (CMCT) were carried out following the protocol for DMS modification with minor changes. The reaction solutions contains 2.0 μM tectoRNA, 20 mM CMCT, 50 mM K_3BO_3 (pH 8.0), and 15 mM $MgCl_2$ or 1.0 mM EDTA. Reaction was carried out at 30°C for 15 min, and stopped by ethanol precipitation.

4.2 Detection of modified bases by reverse transcription

The chemically modified RNA was reverse-transcribed into cDNA. Reaction solutions containing 1.5 pmol of 5'-IR700-dye labeled primer DNA, 1.5 pmol of the modified RNA, and 1 mM dNTPs were heated at 65°C for 3 min, prior to incubation at 45°C for 5 min. 5-fold RT buffer (TOYOBO, Japan), ReverTra Ace (TOYOBO, Japan), and RNase inhibitor (Takara Bio, Japan) were added to the sample tubes. Reaction was performed in 10 uL of final volume at 45°C for 1 hour. 10-fold alkaline buffer [1.5N KOH, 0.2 M tris(hydroxymethyl)aminomethane] was added to degrade the template RNA at 65°C for 20 min. The samples was neutralized with 1.5N HCl and precipitated by ethanol. The RNA was dissolved in loading dye [60% formamide, 0.01% Bromophenol Blue, and 0.10 mM EDTA], was heated at 96°C for 2 min, prior to electrophoresis using DNA sequencer (4300 DNA Analyzer; Li-COR). Semi-Automated Footprinting Analysis software (SAFA) was used to quantitate band intensities at single nucleotide resolution.[4] Each band was assigned by comparing the modified sample's lane to sequencing ladders run on the same gel. Modification intensities of each nucleotide

were normalized by dividing it by sum of intensity of the corresponding lane, and normalized values were averaged from two independent experiments.

Table S6. Oligo DNA and RNA used for assays (Italic represents ribonucleotide)

Name	5'-sequence-3'	
FAM-substrate-1	6FAM-CGTACACGTACTCACGCGTATACA*GUCCAC*	cisDSL assay
5'-FAM substrate for Tet intron	6FAM-*GGCCCUCUAAAAA*	*Tetrahymena* assay
Dye-labeled primer for Tet intron	IR700-GAGAAGAATACATCTTCCCCGACCG	Reverse-transcription
Dye-labeled primer for tectoRNA and cisDSL	IR700-GGATAACAATTTCACACAGG	Reverse-transcription

5. *In vivo* splicing assay

In vivo splicing activity of the intron was estimated by blue/white colony color as described previously (47). The DNA template for the *Tetrahymena* intron RNA was inserted into a β-galactosidase α-fragment gene in the pTZ18U vector in order to direct production of mature mRNAs following splicing. *E. coli* JM109 strain was employed for transformation, followed by plating on LB agar plates containing ampicillin (50 μg/ml), IPTG (isopropylthiogalactoside, 150 μg/ml), and X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside, 40 μg/ml). Single colonies were picked from the plates, streaked on LB-plates, incubated at 37°C for 12 hours, and then stored at 4°C for further color development. Colony color was checked 12 h and one week after streaking. The blue color indicates the active splicing, and white inactive.

B) Supplementary Figures

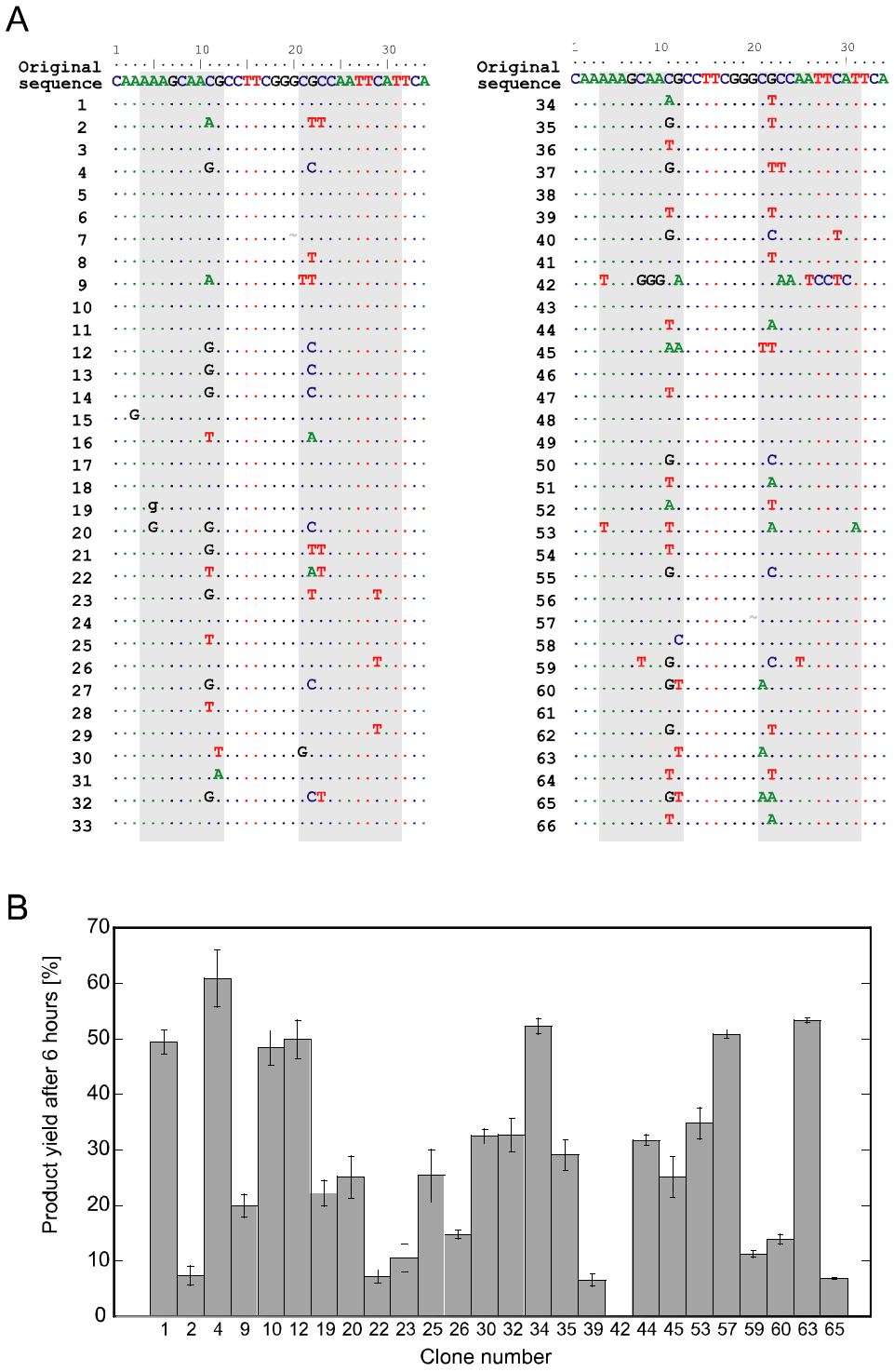


Figure S1. Sequences and catalytic activity of clones obtained from a 45% doped library after 2 rounds of selection. (A) Doped nucleotides were shaded. Dots and wavy line represent nucleotides identical to those of the original sequence, and a gap. (B) Some of selected clones were subjected to ligation reaction for 6 h under the conditions used in selection experiments.

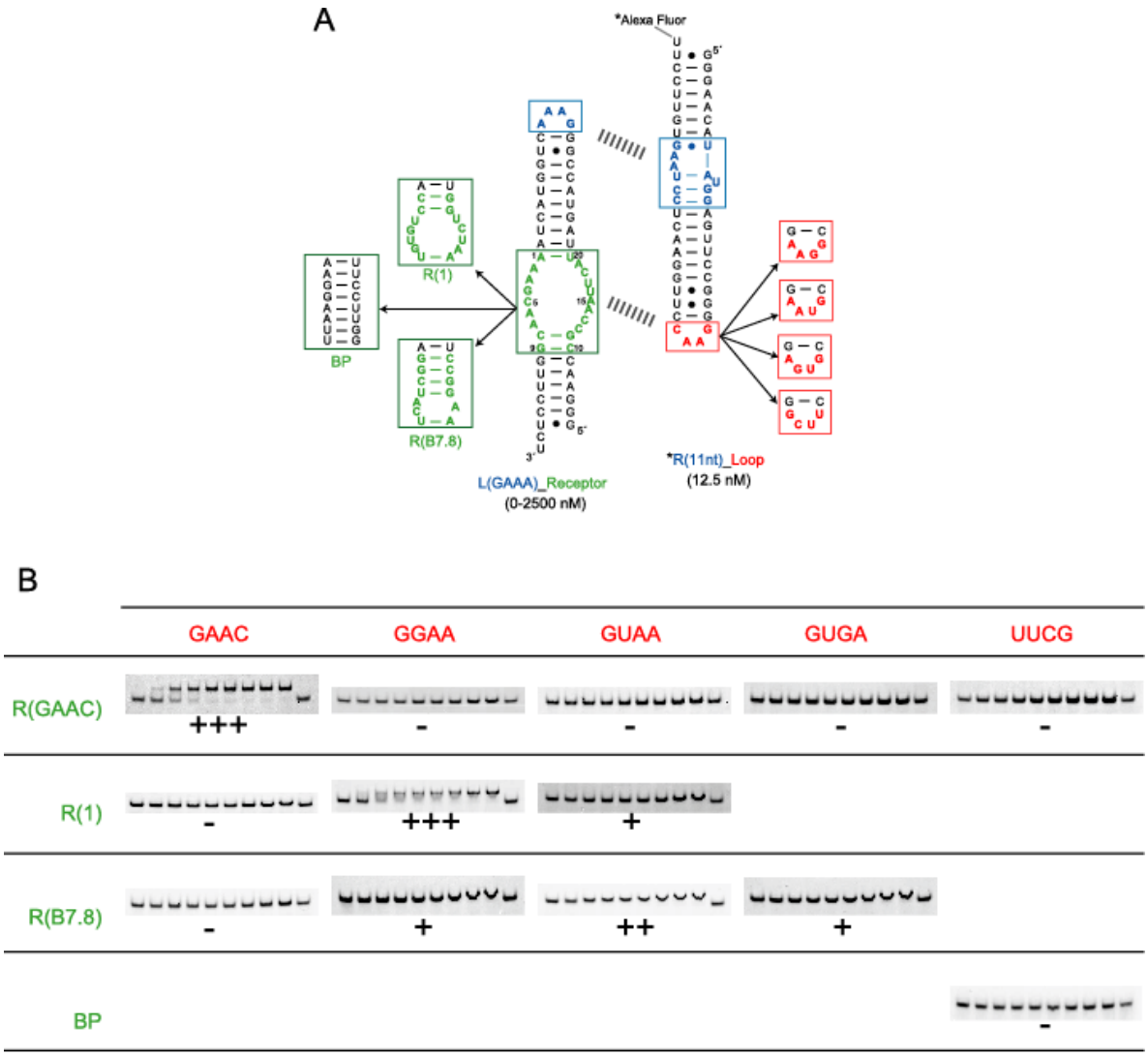


Figure S2. Orthogonality between GNRA/receptor and GAAC/R(GAAC) interacting motifs. (A) Secondary structures of unlabeled tectoRNA with various receptors, L(GAAC)_Receptor, and fluorescently labeled tectoRNA with various loops, *R(11nt)_Loop, assembled in various combinations of loops and receptors. (B) Representative gel images of tectoRNA of various combinations which was assembled and run under the conditions used in Figure 4. Concentrations of tectoRNAs of each lane were also same as in Figure 4B. The degree of assembling was scored using the following scale: (+++) assembled at low concentration; (++) assembled at medium concentration; (+) assembled at high concentration; (-) assembling-deficient. Assembling, gels, and electrophoresis conditions are the same given in materials and methods.

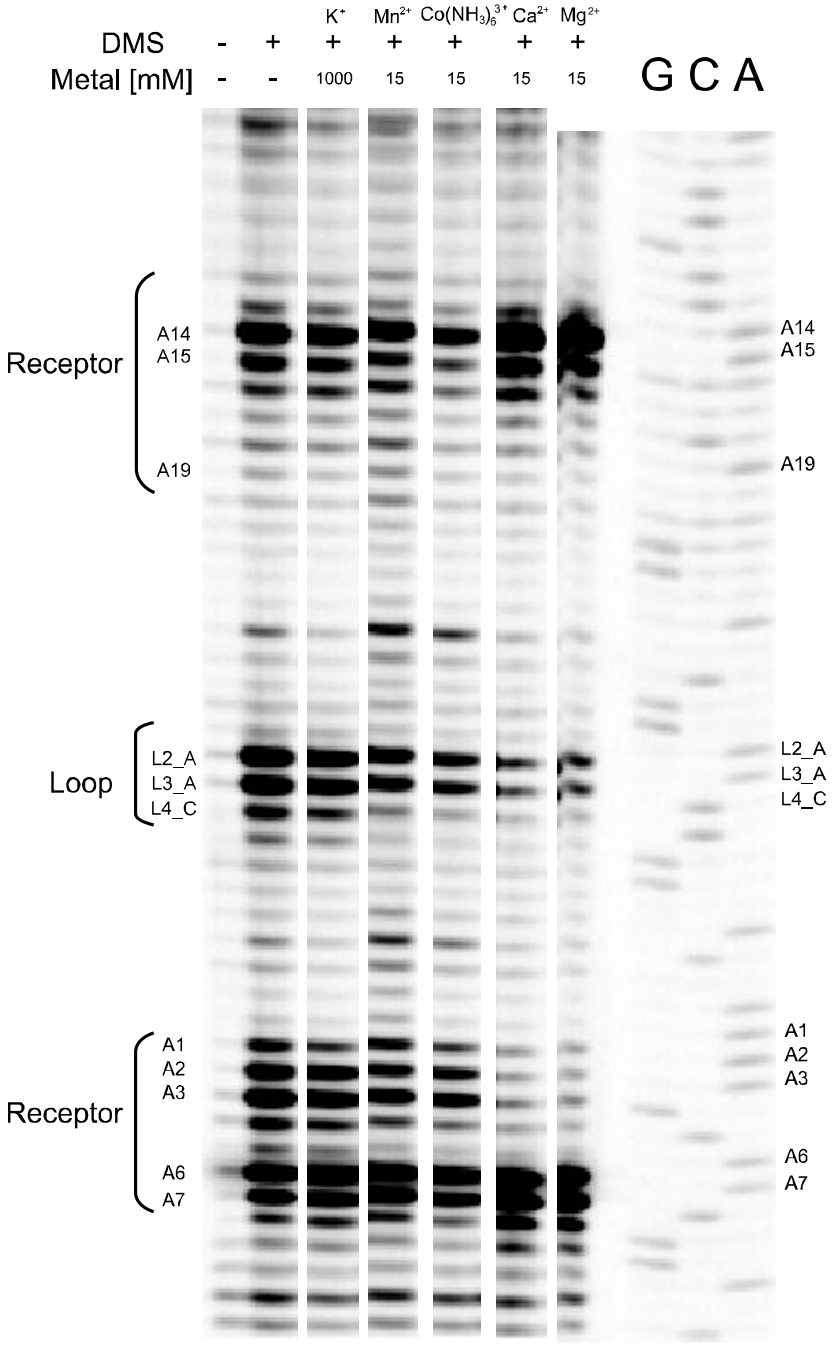


Figure S3. DMS modification of tectoRNA with various metals. A self-dimerizing tectoRNA, tecto_L(GAAC)_R(GAAC) were subjected to DMS modification under the conditions used in gel shift assay except for the presence of various metals including 15 mM Mg^{2+}, Ca^{2+}, Mn^{2+}, $Co(NH_3)_6^{3+}$, or 1000 mM K^+. The bands were assigned by comparing them with sequencing ladder.

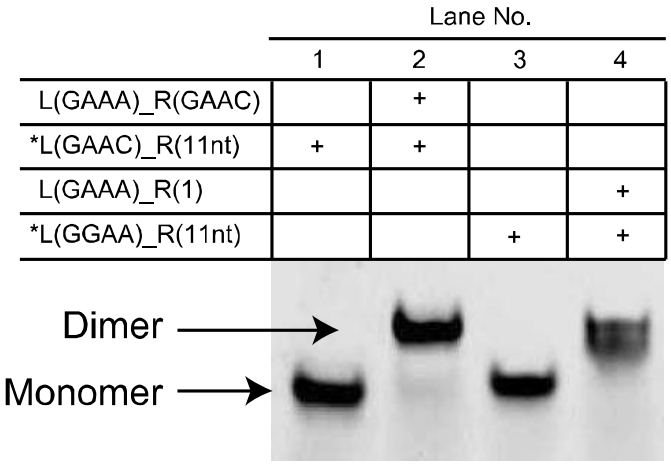


Figure S4. Ca^{2+}-dependent assembly of hetero-dimeric tectoRNA. Asterisks denote Alexa Fluor 488 labeled RNAs. Lane 1 and 3 contain 0.10 μM labeled RNA only, and lane 2 and 4 contain both 0.10 μM labeled and 2.0 μM unlabeled RNA. Gel and electrophoresis conditions are the same given in materials and methods except for the presence of Ca^{2+} instead of Mg^{2+}.

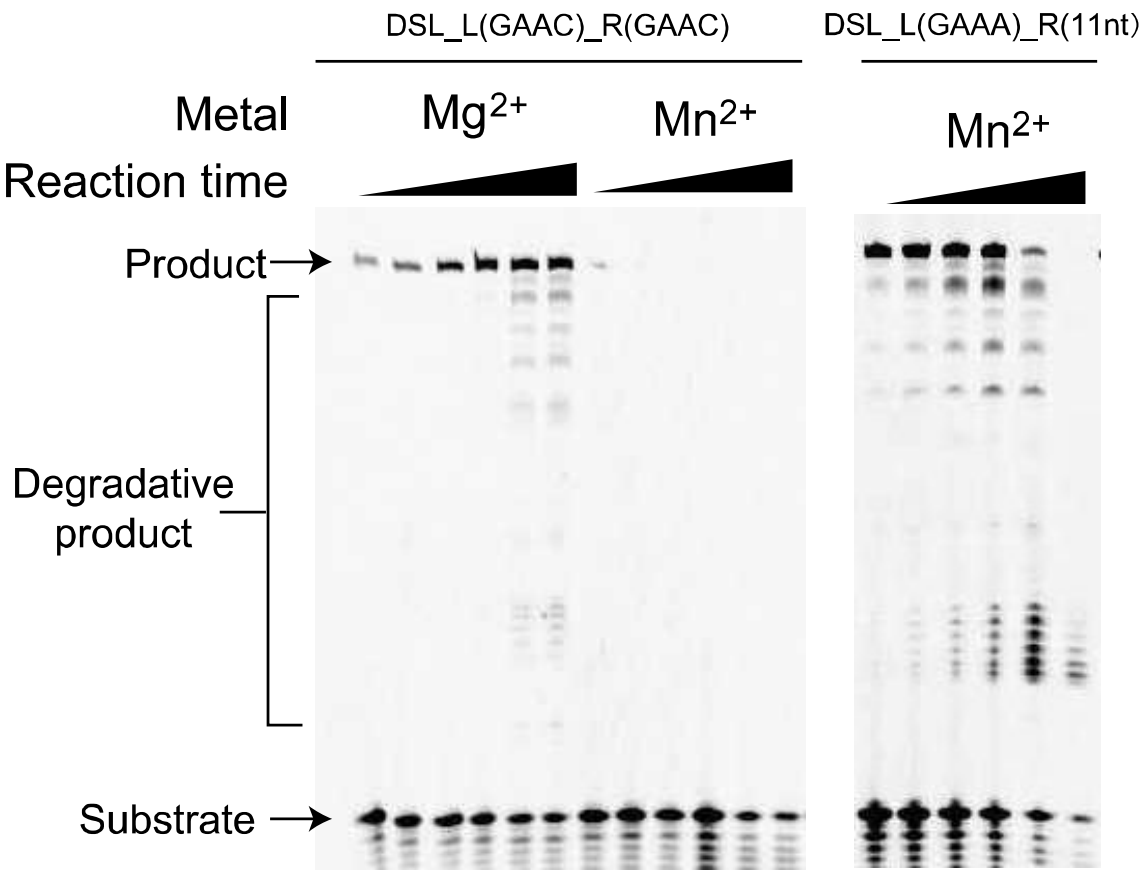


Figure S5. Effect of divalent cations (Mg^{2+} or Mn^{2+}) on ligation activity of cisDSL ribozymes, DSL_L(GAAC)_R(GAAC) and DSL_L(GAAA)_R(11nt). Reaction conditions are the same as described in materials and methods except for the presence of 50 mM Mn^{2+} instead of Mg^{2+}. In the presence of 50 mM Mn^{2+} ion, enhanced degradation of RNA samples incubated for longer duration was observed.

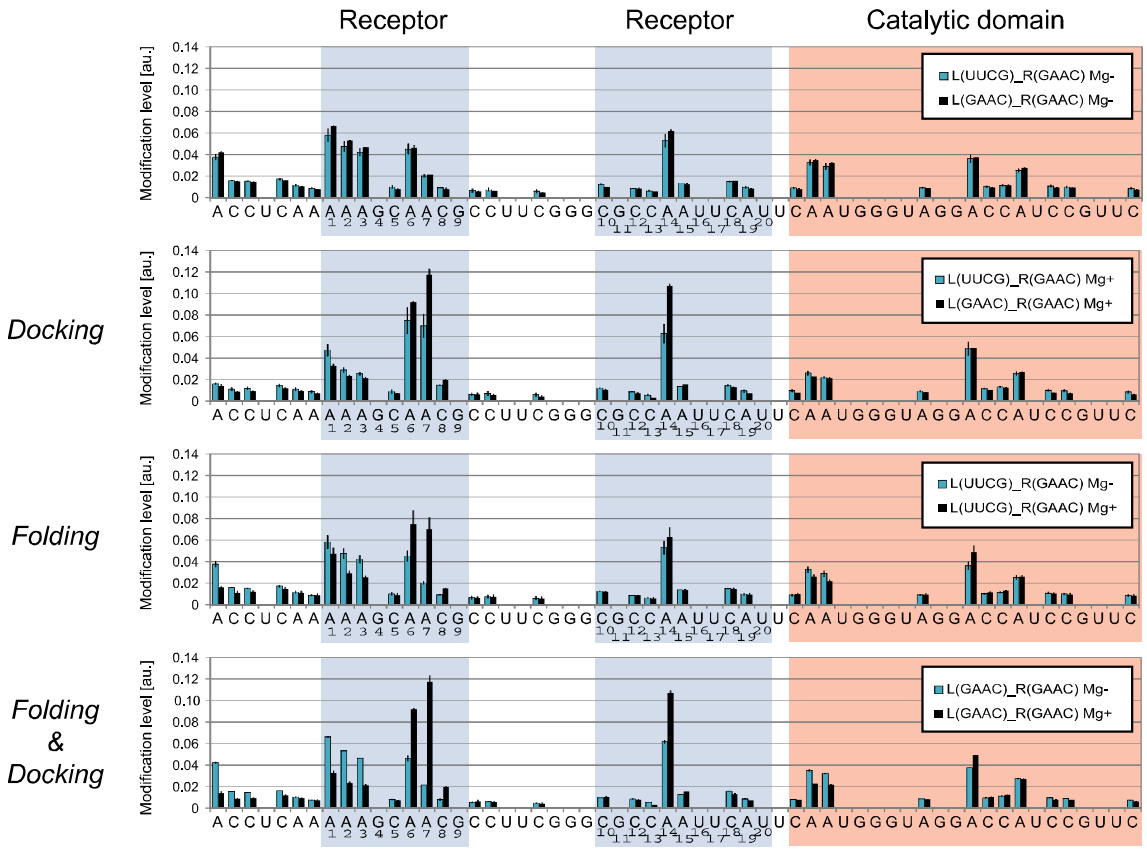


Figure S6. DMS modification levels of the mutated receptors embedded in the cisDSL context with (black bar) or without (blue bar) Mg^{2+}. Receptor regions and catalytic domain were highlighted by light blue and red.

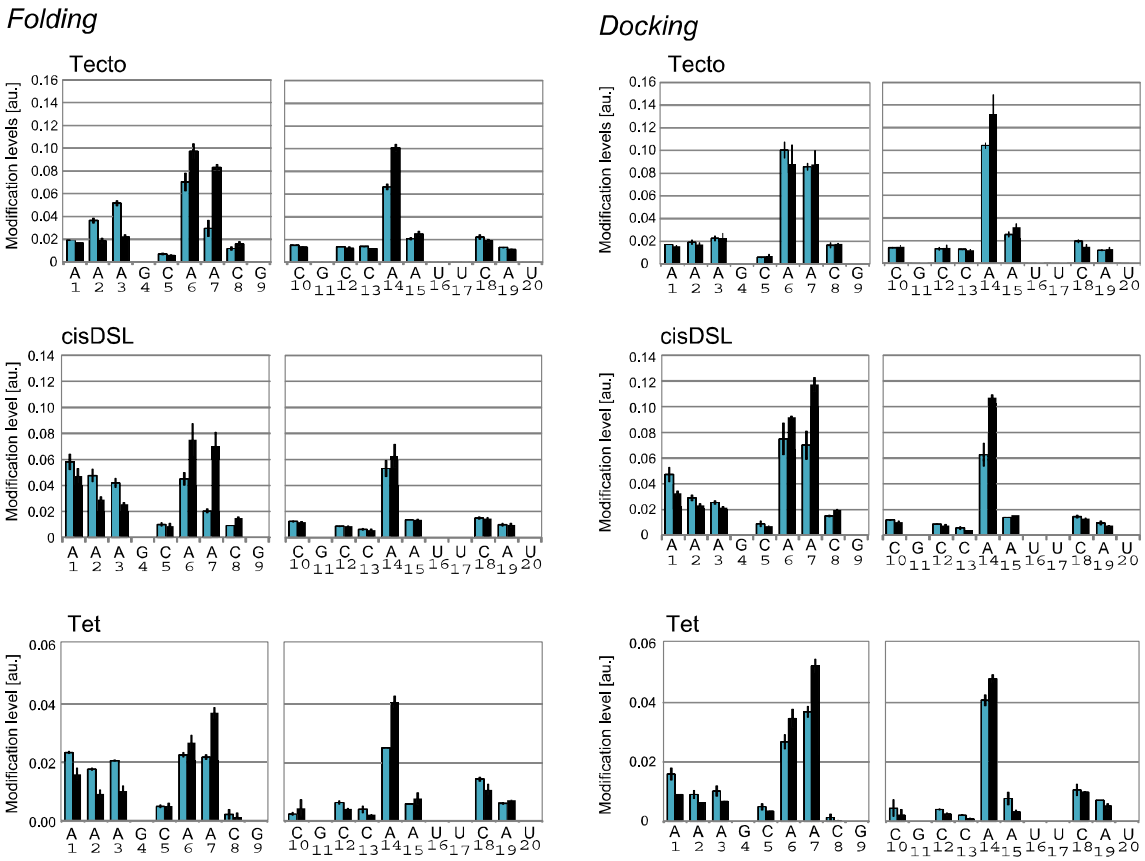


Figure S7. Comparison of DMS modification of the tectoRNA, the cisDSL, and the *Tetrahymena* ribozymes. *Folding* (left) and *docking* (right) effect on modification levels of the receptor within the three contexts were shown to for the sake of comparison.

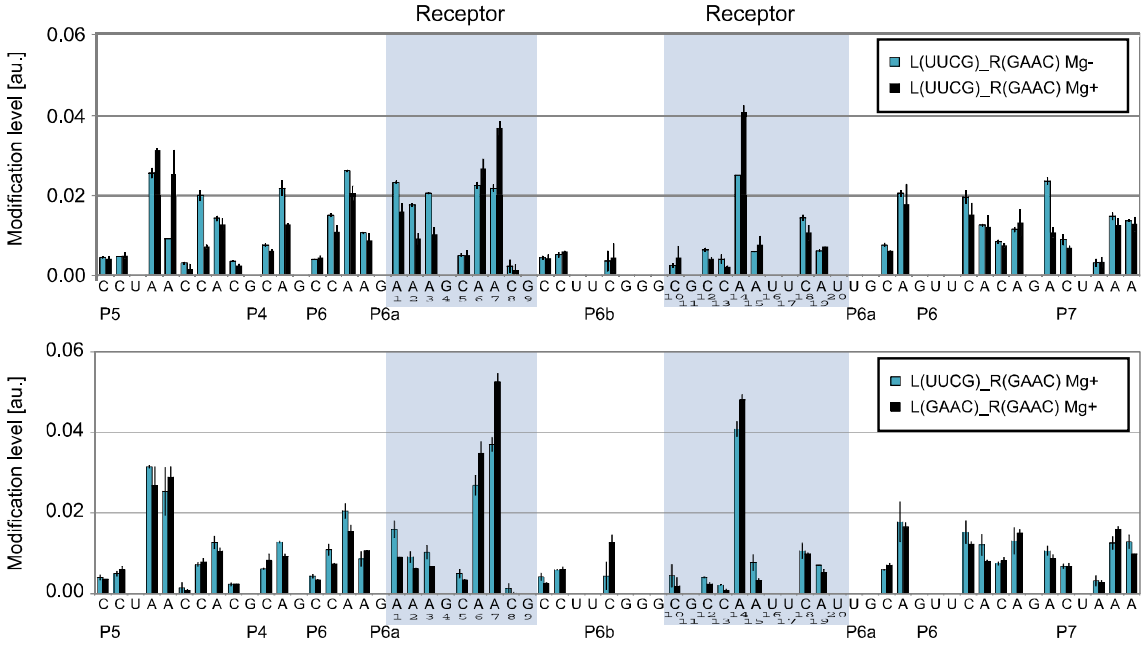
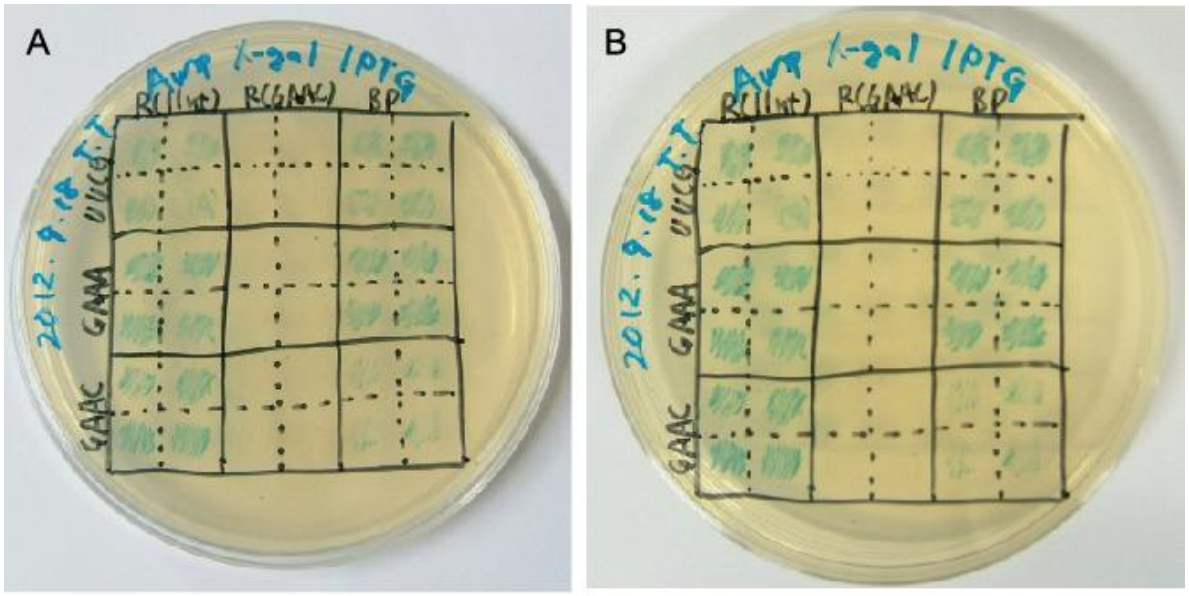


Figure S8. DMS modification levels of the mutated receptors embedded in the *Tetrahymena* ribozyme context with (black bar) or without (blue bar) Mg^{2+}.



C

Tetrahymena construct Loop_Receptor	Colony color (1 day after streaking)	Colony color (a week after streaking)
L(UUCG)_R(11nt)	Pale blue	Pale blue +
L(UUCG)_R(GAAC)	White	White
L(UUCG)_BP	Pale blue	Pale blue +
L(GAAA)_R(11nt) (Wild-type)	Blue	Blue
L(GAAA)_R(GAAC)	White	White
L(GAAA)_BP	Pale blue +	Blue
L(GAAC)_R(11nt)	Pale blue +	Blue
L(GAAC)_R(GAAC)	White	White
L(GAAC)_BP	Pale blue	Pale blue

Figure S9. *In vivo* splicing assay. Cultures of *E. coli* JM109 carrying various plasmids as indicated. Single colonies were picked from a plate of transformants and streaked on LB agar plates containing ampicillin, X-gal, and IPTG, and grown at 37°C for 12 hours (A), and further incubated at 4°C for a week (B).

C) Supplementary References

18. Ohuchi,S.P., Ikawa,Y. and Nakamura,Y. (2008) Selection of a novel class of RNA–RNA interaction motifs based on the ligase ribozyme with defined modular architecture. *Nucleic Acids Res.,* **36,** 3600-3607.

23. Zuker,M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.,* **31,** 3406-3415.

24. Kladwang,W., Cordero,P. and Das,R. (2011) A mutate-and-map strategy accurately infers the base pairs of a 35-nucleotide model RNA. *RNA,* **17,** 522-534.

25. Tijerina,P., Mohr,S. and Russell,R. (2007) DMS footprinting of structured RNAs and RNA-protein complexes. *Nat. Protoc.,* **2,** 2608-2623.

47. Novikova,I.V., Hassan,B.H., Mirzoyan,M.G. and Leontis,N.B. (2011) Engineering cooperative tecto-RNA complexes having programmable stoichiometries. *Nucleic Acids Res.,* **39,** 2903-2917.

48. Williamson,C.L., Desai,N.M. and Burke,J.M. (1989) Compensatory mutations demonstrate that P8 and P6 are RNA secondary structure elements important for processing of a group I intron. *Nucleic Acids Res.,* **17,** 675-689.