Selection of a novel class of RNA–RNA interaction motifs based on the ligase ribozyme with defined modular architecture

Shoji P. Ohuchi[1], Yoshiya Ikawa[2,3] and Yoshikazu Nakamura[1,4,*]

[1]Department of Basic Medical Sciences, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, [2]Department of Chemistry and Biochemistry, Graduate School of Engineering, Kyushu University, 744 Moto-oka, Nishi-ku, Fukuoka 819-0395, [3]PRESTO, Precursory Research for Embryonic Science and Technology, Japan Science and Technology Corporation, Saitama 332-0012 and [4]CREST, Core Research for Evolutional Science and Technology, Japan Science and Technology Corporation, Saitama 332-0012, Japan

ABSTRACT

To develop molecular tools for the detection and control of RNA molecules whose functions rely on their 3D structures, we have devised a selection system to isolate novel RNA motifs that interact with a target RNA structure within a given structural context. In this system, a GAAA tetraloop and its specific receptor motif (11-ntR) from an artificial RNA ligase ribozyme with modular architecture (the DSL ribozyme) were replaced with a target structure and random sequence, respectively. Motifs recognizing the target structure can be identified by *in vitro* selection based on ribozyme activity. A model selection targeting GAAA-loop successfully identified motifs previously known as GAAA-loop receptors. In addition, a new selection targeting a C-loop motif also generated novel motifs that interact with this structure. Biochemical analysis of one of the C-loop receptor motifs revealed that it could also function as an independent structural unit.

INTRODUCTION

Single-stranded RNAs can form defined 3D structures based not only on Watson–Crick base-pairing interactions but also on specific, non-Watson–Crick tertiary interactions, which include additional hydrogen-bonds, hydrophobic interactions, electrostatic (metal-bridged) interactions and base-stacking. In nature, many RNA molecules and motifs exhibit specific functions that require the formation of a specific 3D structure, rather than simply a linear carrier of genetic code information. The most classical examples of such structural, nonprotein-coding RNAs (ncRNAs) are tRNA and rRNA, which play key roles in the central dogma of molecular biology (1). In addition, other structural ncRNAs are also known to play essential roles in the central dogma (1), since several regulatory elements on mRNA, like riboswitches and internal ribosome entry sites (IRESs), also function via their specific 3D structure (2). More recently, several structural ncRNAs have been discovered as specific modulators for intracellular proteins, and a significant number of structural RNAs likely exist within the huge numbers of ncRNAs already found in higher eukaryote genomes (3). Thus, the development of new methods to detect and control structural dynamics of these RNAs is attractive areas of investigation (4–7).

One elegant example of a molecular sensor for RNA dynamics is the P5abc RNA, which is a structural element of the *Tetrahymena* Group I intron ribozyme that acts to physiologically stabilize the core structure of the ribozyme to enhance its activity (8). Since the P5abc acts as an independently folded domain, the isolated P5abc RNA can also activate a truncated ribozyme lacking this region (ΔP5abc ribozyme) *in trans*. The P5abc has three small structural motifs (A-rich bulge, L5b and L5c loops), which independently interact with the ribozyme. These motifs are strictly located at defined distances and orientations, with such a modular architecture enabling the highly specific and extraordinary strong recognition between P5abc RNA and ΔP5abc ribozyme (9–11). Recently Johnson *et al.* have reported a new aspect of P5abc function whereby the P5abc RNA acts as a molecular sensor to discriminate between the native and misfolded conformations of the ΔP5abc ribozyme (12).

Therefore, RNA molecules that specifically recognized and bound a certain conformation of a target structural RNA could be designed in a modular manner. This approach, however, is far from practical since RNA

motifs recognizing local RNA structures are very limited. Therefore, it is important to develop a practical system to select for and identify novel RNA motifs that recognize a given RNA structure.

In the present study, we have applied an artificial ligase ribozyme (designed-and-selected ligase; DSL) for the development of a selection system to generate novel RNA receptor motifs against a target RNA structure within a given structural context. The DSL ribozyme was constructed by using a well-defined and designed self-folding RNA as a structural scaffold. The features of this scaffold include: (i) a GAAA-tetraloop, ii) its receptor motif (11-nt receptor; hereafter 11-ntR) and (iii) stable, coaxially stacked base-triples known as the triple-helical scaffold (THS) found in the *Tetrahymena* ribozyme (13). After experimental validation of the scaffold structure, a catalytic modular unit was installed using *in vitro* selection (14). The resulting DSL ribozyme has a modular architecture that is amenable to module-based engineering, and the *trans*-acting ribozyme has been successfully redesigned by replacing the THS motif with GAAA-loop/11-ntR pair (14).

To develop a DSL ribozyme-based selection system in this study, we have redesigned a shortened derivative of the DSL ribozyme by replacing its GAAA-loop and 11-ntR with a target structure and random sequence, respectively. Motifs recognizing the target structure can be identified by *in vitro* selection based on restoring ribozyme activity.

MATERIALS AND METHODS

Library construction

Libraries were designed based on the DSL-U5 ribozyme (15). For the construction of GAAA-loop library, whose target motif is GAAA-loop, a synthetic template DNA (Figure 1A) (DSL-GAAA-N19; 5′-AGGGAAGGAAA CTTCCCTGT GTCTTTTTT*GAGCCCC*TCAAT CCNN NNNNNN NNNNNNNNNN NGGATCAATG GGT AGGACCA TCCGTTCCCT AGCAGGGTTC-3′; where N stands for any nucleotide, and BanII site is indicated in italic) was amplified by PCR using *ExTaq* DNA polymerase (Takara-Bio, Japan) with forward and reverse primers (DSL-F3, 5′-<u>TAATACGACT CACTATAAGG</u> GAAGGAAACT TC<u>C</u>CTGTGTC-3′; and DSL-R2R3, 5′-TCTGCCTAAG TGGGCAATGA GACTGGAACC CTGCTAGGGA ACGG-3′, respectively. T7 promoter sequence is underlined). All oligodeoxynucleotides were purchased from Operon Biotechnologies (Japan). For the construction of a C-loop library whose target motif is the C-loop, the above GAAA-loop library was digested with the nonpalindromic restriction enzyme, BanII, and then, the 3′-fragment containing the random sequence was purified by native PAGE and precipitated with ethanol. The fragment was ligated with a synthetic, 5′-fragment containing the C-loop motif (C-50b-F1; 5′-<u>TAATA CGACT CACTATAAGG</u> GAAAGCACTG G<u>TTCG CCAGC TTTCCCTGTG GATTTTT*GAG CC*-3′, where the T7 promoter sequence and the BanII site are underlined and italicized, respectively) in the presence of a splint DNA using T4 DNA ligase (Takara-Bio). The ligated product</u>

was amplified by PCR with C-50b-F1 and DSL-R2R3 as forward and reverse primers, respectively, and the amplified full-length fragment was purified by native PAGE. The PCR product was extracted by phenol/chloroform treatment, precipitated with ethanol and transcribed using T7 RNA polymerase (RiboMAX Large Scale RNA Production System-T7, Promega, WI, USA). After *in vitro* transcription, the template DNA was digested by RQ1 DNase (Promega), and the transcript was extracted by phenol/chloroform treatment, followed by purification on Microcon YM-30 (Millipore, MA, USA) and MicroSpin G-25 columns (GE Healthcare, Uppsala, Sweden) to yield an RNA pool. Throughout the experiment, 1×10^{15} molecules in the RNA pool were used for the *in vitro* selection. Theoretically, 3.6×10^3 copies of 2.7×10^{11} ($= 4^{19}$) variants are included in the initial pool.

In vitro selection

In vitro selection was carried out essentially as described (14). Selection conditions were in 50 mM Tris–HCl (pH 7.5), (Figure 2) 100 mM NaCl and 20 mM MgCl$_2$. Substrate RNAs [S-1 and S-2 (14)] were purchased from Gene Design Inc. (Japan). In order to exchange the guide sequence (complementary to the substrate RNA) of the libraries, and to eliminate undesired mutations at the nonrandomized regions, 5′-fragments of the libraries were exchanged after every third round of the selection as follows. After the first three rounds of selection from the GAAA-loop library, the 5′-fragment was exchanged with DSL-F2 (5′-<u>TAATACGACT CACTATAAGG</u> GAAGG AAACT TCCCTGTGGA TTTTTT*GAGC* C-3′. T7 promoter sequence and BanII site are underlined and italicized, respectively) as described above. After the third and sixth rounds of the selection of C-loop library, the 5′-fragment was exchanged with C-50b-F2 (5′-<u>TAAT ACGACT CACTATAAGG</u> GAAAGCACTG G<u>TTCG CCAGC TTTCCCTGTG TCTTTTT*GAG CC*-3′. The T7 promoter sequence and BanII site are indicated by underline and italic, respectively) and C-50b-F1, respectively. Other conditions for the *in vitro* selection are listed in Table 1, and the detailed selection procedure is described in the Supplementary Data (*in vitro* selection protocol).</u>

Table 1. Conditions for *in vitro* selection

Round	Substrate	Reaction time (h)
GAAA-loop library		
1	S-2	14
2	S-2	2
3	S-2	1
4	S-1	2
5	S-1	1
6	S-1	0.5
C-loop library		
1	S-1	20
2	S-1	20
3	S-1	4
4	S-2	20
5	S-2	4
6	S-2	4
7	S-1	1

Electrophoretic mobility shift assays (EMSA)

For the preparation of uniformly labeled RNAs, *in vitro* transcription was carried out in the presence of [alpha-^{32}P]-GTP. Synthetic single-stranded oligodeoxynucleotide templates annealed with a T7 promoter oligonucleotide were directly used for transcription. After RQ1 DNase treatment, phenol/chloroform extraction and ethanol precipitation, the transcripts were purified by denaturing PAGE, recovered by crush-and-soak method and precipitated with ethanol. The trace amount (0.2–0.4 nM) of the labeled RNAs and the indicated concentrations of unlabeled RNAs were folded separately. Initially, these RNAs were heated in water at 90°C for 2 min, immediately cooled on ice, followed by folding for 10 min at 37°C in the buffer employed for the *in vitro* selection. These separately folded, labeled and unlabeled RNAs were mixed and incubated for >30 min to form dimers. After the incubation, 1 μl of 50% glycerol was added to the 10 μl of the RNA samples, and these samples were run at 4°C on 10% polyacrylamide gels containing 50 mM Tris–borate (pH 8.2) and 20 mM MgCl$_2$. The RNA bands were quantified by FLA-5100 (Fujifilm Life Science, Japan). K_d values were determined as the concentration at which half the RNA molecules form dimers.

Lead(II)-induced cleavage

In order to label the 5′-end of RNA, unlabeled transcript was treated with alkaline phosphatase from Calf intestine (Takara-Bio) to dephosphorylate the 5′-end, followed by phosphorylation with [gamma-^{32}P]-ATP using T4 polynucleotide kinase (Takara-Bio). The labeled RNA was purified by denaturing PAGE as described above. The RNA folding and dimerization were carried out as described above in the presence of 50 mM HEPES-OAc (pH 7.5), 100 mM NaOAc and 20 mM Mg(OAc)$_2$ instead of the selection buffer. It was confirmed that the clone #05 ribozyme is also active under this buffer condition (data not shown). Lead(II)-induced cleavage analysis was carried out as described (16).

RESULTS

In order to isolate RNA receptor motifs for the construction of molecular tools to detect and control structural dynamics of given RNAs, we have developed a selection system based on the DSL ribozyme. Although canonical SELEX (systematic evolution of ligands by exponential enrichment) is another method of choice (17), the motifs obtained by SELEX are not easily engineered within different structural contexts without detailed biochemical and structural analyses. In contrast, due to the well-defined architecture of the DSL ribozyme, the relative distance and orientation between the target structure and the isolated motif can be predicted without detailed experimental analyses, and thus, new receptor motifs isolated by this system may be directly employed in different structural contexts.
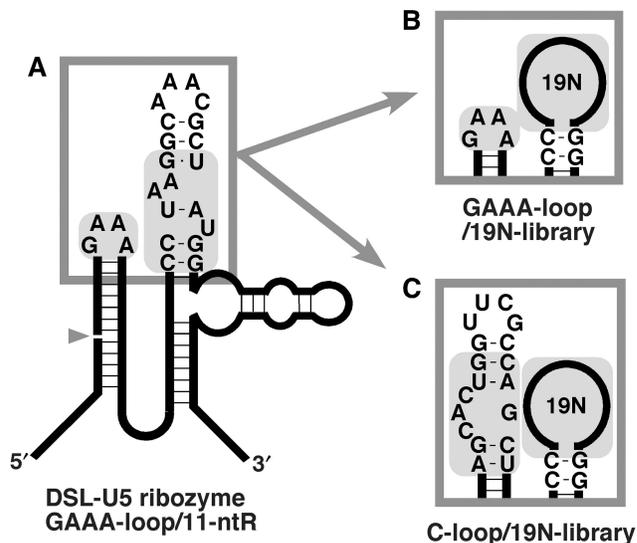

Figure 1. Secondary structures of the parental DSL-U5 ribozyme and its derived libraries. (**A**) The DSL-U5 ribozyme with the GAAA-tetraloop/11-ntR pair essential for the ribozyme activity highlighted in gray. (**B**) The GAAA-loop library with the target GAAA-tetraloop and the randomized nucleotides highlighted in gray. (**C**) The C-loop library with the target C-loop motif (C-50) with neighboring single base pairs and randomized nucleotides highlighted in gray.

Selection of receptor motifs against GAAA tetraloop

Previously, a mutational analysis of the DSL ribozyme showed that the THS motif is less important for ribozyme activity than the GAAA-loop/11-ntR pair. A miniaturized variant lacking the THS motif (DSL-U5; Figure 1A) is as active as the parental DSL ribozyme (14,15). We employed the DSL-U5 variant as a platform for our ribozyme-based selection system. We first carried out the selection of motifs targeting the GAAA-loop (Figure 2). Since there is at least one defined positive control sequence for this target, namely the 11-ntR motif, the GAAA loop is an ideal target to test our experimental design. Fifteen nucleotides constituting the 11-ntR motif as well as a capping stem–loop structure of the DSL-U5 ribozyme were replaced with 19 nt of random sequence (Figure 1B). Although two C-G base pairs adjacent to the catalytic unit are considered to be a part of the 11-ntR motif, they function to stabilize the catalytic unit structure, and thus these two base pairs were excluded from randomization.

After six rounds of the selection/amplification cycle based on ligation activity, the resulting RNA pool showed activity comparable to the parental DSL-U5 ribozyme (data not shown). Hence, we determined the sequences of clones randomly picked from the pool after round 6 of the selection (Figure 3). Out of 28 clones sequenced, 13 clones were highly similar (or even identical) to the 11-ntR motif. Fourteen additional clones were either identical, or highly similar to the previously reported C7.2 motif, which is an artificially selected receptor for the GAAA-loop (18). Together, these results demonstrate the proof-of-concept of this *in vitro* selection technology to generate receptor motifs against a given target RNA structure.
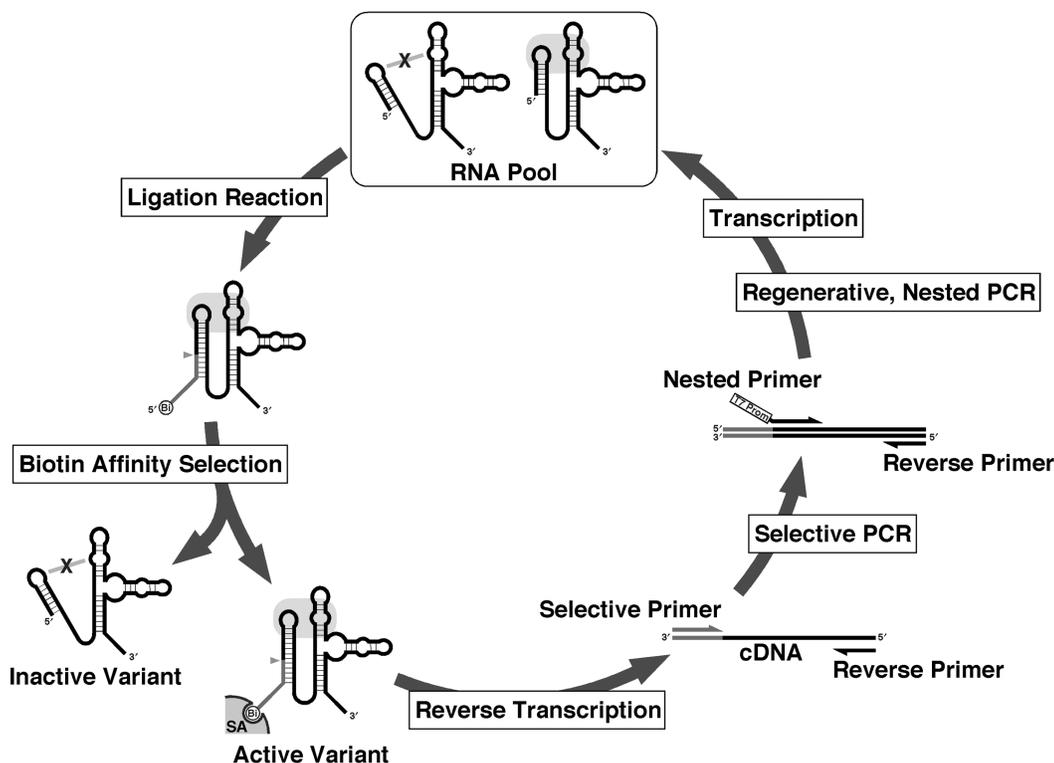
Figure 2. Schematic diagram of the *in vitro* selection. The RNA pool is incubated with the substrate RNA with 5′ biotin moiety (a gray line with encircled 'Bi'). Only the active variants with a desired RNA–RNA interaction (the right variant in the RNA pool) can react and be covalently ligated with the substrate. Thus, the active variants can be selectively recovered by affinity purification using streptavidin beads ('SA'). After the reverse transcription, cDNAs encoding the active variants are selectively amplified by PCR using the selective primer with sequence identical to the substrate, followed by PCR using the nested primer with T7 promoter sequence ('T7 prom') to regenerate the enriched DNA pool for the next round of the selection.

Selection of receptor motifs against the C-loop motif

Next, we attempted to generate novel motifs recognizing an RNA structure whose natural receptor is completely unknown. We chose the C-loop motif from 23S rRNA of *Haloarcula marismortui* (C-50) as the first target (Figure 1C). The C-loop motif is a class of asymmetric internal loops that locally increases the helical twist between neighboring stems, and typically, they do not directly interact with other RNA structures (19).

To construct a library to isolate C-loop receptor motifs, the GAAA tetraloop of the GAAA-loop library was replaced with a C-loop motif with a neighboring single base pair, and the free end of the motif was capped with a UUCG tetraloop (Figure 1C). After seven rounds of selection, the ligation activity of the RNA pool was clearly detectable, suggesting restoration of ribozyme activity by binding to the new C-loop motif (Figure 4). Clones were randomly picked from the pool and their sequences determined (Table 2). Interestingly, 15 clones shared an apparent consensus sequence (underlined nucleotides in Table 2) and can be classified to three subgroups: Group 1, the major variant, consists of clone #05 (nine isolates) and its derivative clone #09; Group 2, the next frequent variant, consists of clone #10 (three clones) and its derivative clone #03; and Group 3 is clone #11.

Next, we analyzed the ligation activity of these clones. Under the selection conditions we employed, all clones,

except for clone #09, showed efficient ligation activity (Table 2). Although the activities were one order of magnitude lower than the parental DSL-U5 ribozyme, their reaction rates are still 10^4-fold higher than the reported ligation efficiency of the nonenzymatic, template-dependent reaction under similar conditions (20).

Analyses of a novel C-loop receptor motif

In order to see whether the newly isolated motifs can act as independent structural units as intended, we examined the ability to place the most abundant sequence (hereafter #05 receptor) into a different structural context. We transplanted the #05 receptor into TectoRNA, an artificial RNA architecture developed by Jaeger and colleagues (21,22), since its self-dimerization properties are suitable to examine the modularity of the selected motif. The original TectoRNA (construct 1 of ref. 22), in which the GAAA-loop/11-ntR pair was connected by a linker helix of a suitable length, forms a homodimer in a concentration-dependent manner. The dissociation constant (K_d) for the dimerization can be determined by the titration of unlabeled RNA in the presence of trace amount of labeled RNA on EMSA. Under our assay conditions, the K_d value of the original TectoRNA with GAAA-loop/11-ntR pair was 34.5 ± 3.4 nM (data not shown). The K_d value is one order of magnitude higher than that in the original report (22), probably due to the differences of EMSA conditions

```
Library
    CC NNNNNNNNNNNNNNNNNNNN GG

Original DSL-U5
    CC TAAGGC--AAAC--GCTAT GG

11-nt-motif family
01  CC TAAACCGGCAGGTCGGCAT GG
06  CC TAACCTGCTTAGGCAGTAT GG
08  CC TAACCTGTGAGTACAGTAC GG
09  CC TAAGCACAGTAGGGTGTAT GG
14  CC TAAGGCGCGTGGGCGCTAT GG
19  CC TAAAGCGTTGGTGCGCTAT GG
20  CC TAAAGCGTATCCGCGCCAT GG
22  CC TAAGGCTCTGAGGGGCTAT GG
24  CC TAAGGCCAAAATTGGCGAT GG
25  CC TACACGCATATCGGCGCAT GG
28  CC TACAGCGGCAGGCCGCCAT GG
29  CC TACTTACCAGCTGGTATAC GG
32  CC TAAGGCGTGTATGTGCTAT GG

C7.2-motif family
02  CC TGTACAGCTGCTTGCTGAC GG
03  CT TGTACACCTCATGGGTGAT GG
04  CC TGTAGCGCTATT-GCGTAT GG
05  CC TGTAGCGCTCTTTGCGTAT GG
07  CC TGGACACGCAATGCGTGAT GG
10  AC TAGAGCCGTTAAGCGGTAT GG
11  CC TGTAGCTCTCCGTGAGTAT GG
12  CC TGTAGCTGCTCTGCAGTAT GG
17  CC TGGAGCAGAGTTACTGTAT GG
21  CC TGAAGCGCCAAGGGCGTAT GG
23  CC TGTGAGACTTAGGTCTTAT GG
26  CC TGTAGCTGCTGTGCAGTAT GG
30  CC TGTAGCGCGTTAGGCGTAT GG
31  CC TGTAGCGCGCGTTGCGTAT GG

others
27  CC CGCAACGCATCGTGGCGTG GG
```

Figure 3. Clones obtained after the sixth round of selection from the GAAA-loop library. Consensus primary sequences and base pairings are highlighted in pink and purple, respectively. Point mutations at nonrandomized positions are indicated in red.

between Jaeger's and ours that is based on the condition employed in the *in vitro* selection.

First, we grafted the C-loop motif and the #05 receptor into the corresponding positions of TectoRNA (Figure 5A). EMSA of this construct showed an apparent reduction in mobility in a concentration-dependent manner (Figure 5C). The mobility of low and high concentrations of the construct was close to those of the original TectoRNA, and the degree of the mobility change is consistent with a biphasic dimerization model with typical fast exchange kinetics ($R^2 = 0.956$–0.998), suggesting that the construct dimerized, as is the case for the original TectoRNA (22). Its K_d value, determined as the kinetic equilibrium, was 168 nM (Figure 5C).

In the library design, the free end of the stem neighboring the C-loop motif is capped with a UUCG tetraloop. To exclude the possibility that the selected motif might recognize this cap structure rather than the C-loop,

Table 2. Clones obtained from the round 7 pool of C-loop library

Clone	Sequence[a]	k_{obs} [h^{-1}][c]	FE [%][b]
Library	NNNNNNNNNNNNNNNNNNNN	ND	ND
DSL-U5	TAAGGCAAACGCTAT	0.34 ± 0.02	27 ± 5
#05 (x9)[c]	<u>GTGGCGATGCGAGCAGCAA</u>	0.05 ± 0.01	13 ± 2
#09	<u>GTGGCGATGCGAGCAGCAG</u>	ND	ND
#10 (x3)[c]	<u>GTGGCGATTAGAGTAGCAA</u>	0.02 ± 0.00	18 ± 3
#03	<u>GTGGCGATTAGAGTAGGAA</u>	0.06 ± 0.01	10 ± 3
#11	<u>GTGGCGAGCAAGTTAGTAA</u>	0.04 ± 0.00	15 ± 2

[a]Sequences originating from the randomized region are shown. None of the clones has point mutations within the nonrandomized regions. Consensus sequences are underlined.
[b]The reaction was carried out in the presence of 2 μM of the substrate RNA (S-1) under the same conditions as the *in vitro* selection. The k_{obs} and FE values are calculated based on data fitting with a single exponential curve.
FE, final extent of the reaction; ND, not detectable.
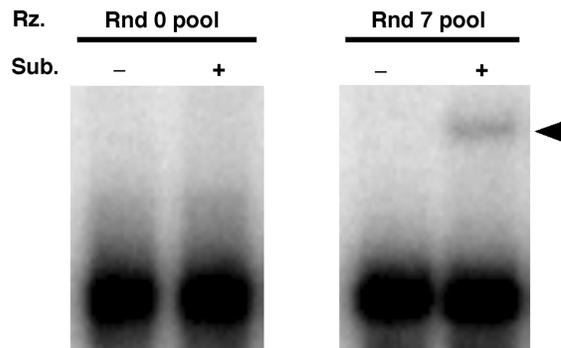[c]Number of clones sharing the same sequence.

Figure 4. Ligation reaction of the RNA pool from the C-loop library. Uniformly [^{32}P]-labeled RNAs from round 0 or round 7 pool were incubated with or without substrate RNA (2 μM of S-1) in the buffer employed for the *in vitro* selection at 37°C for 1 h. The reactions were terminated by adding sample-loading buffer (14), and these samples were separated on denaturing-PAGE. The ligated product is indicated with an arrowhead.

we redesigned the TectoRNA construct to remove the UUCG-loop (Figure 5B). This new construct also dimerized with affinity comparable to the construct with the UUCG-loop (K_d value of 264 nM as the kinetic equilibrium) (Figure 5D). As control experiments, substitution of the C-loop with 5 bp or the #05 receptor with 11-ntR was shown to abolish the self-dimerization ability (see Figure S1, Supplementary Data). These results indicate that the selected #05 receptor mainly recognizes the target C-loop motif and can be placed within a different structural context as intended.

In order to further analyze the interaction between the C-loop and #05 receptor motifs, we carried out chemical footprinting by using lead(II)-induced RNA cleavage (Figure 6). In the presence of sufficient concentration of other divalent cations, lead ions (Pb^{2+}) induce the cleavage of the phosphate backbone of RNAs at nonbase pairing, solvent accessible sites (16). Phosphates around the C-loop motif were cleaved under monomeric conditions but protected under dimeric conditions (in the presence of a concentration of unlabeled RNA well above the K_d value), indicating that these phosphates, originally located at the

surface of the RNA structure, became solvent-inaccessible upon dimerization (Figure 6, see Figure S2, Supplementary Data). This observation supports the physical interaction between the loop and #05 receptor under dimeric conditions. In contrast, several residues in the #05 receptor were cleaved efficiently under dimeric conditions but not under monomeric conditions (Figure 6, see Figure S2,

Supplementary Data). The opposite effects of lead(II)-induced cleavage on the C-loop and its #05 receptor may suggest that the interaction between the C-loop and #05 receptor is likely to be an induced-fit type of recognition accompanying conformational rearrangement of the #05 receptor rather than a lock-and-key type recognition. Because there is no obvious sequence complementarity
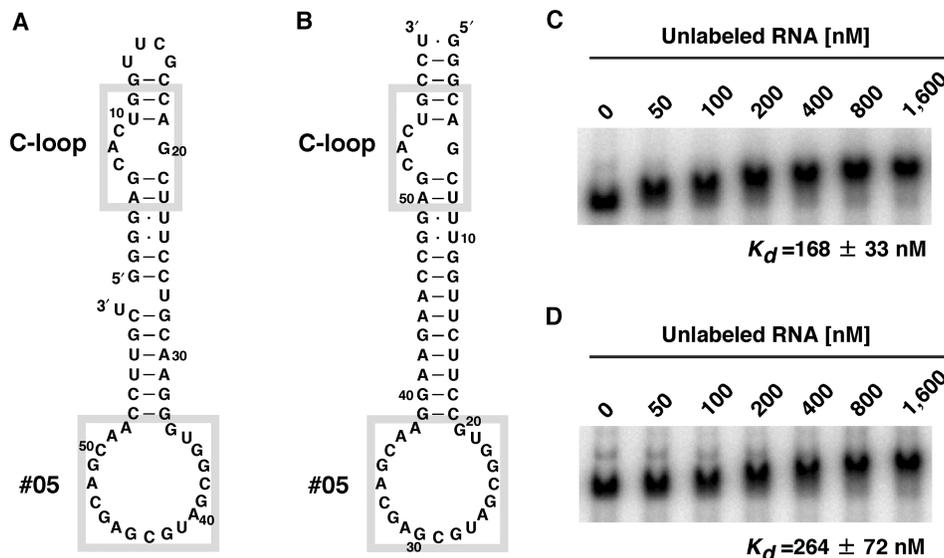

Figure 5. Electophoretic mobility shift assays (EMSA). (**A** and **B**) Secondary structures of TectoRNA-derived, homodimer-forming constructs. The UUCG tetraloop capping the end of the stem in the construct shown in (A) is eliminated in the construct shown in (B). The target C-loop and the obtained #05 receptor motifs are highlighted with gray boxes. (**C** and **D**) Respective autoradiograms of the EMSA experiments for constructs shown in (A) and (B). From left to right, 0, 50, 100, 200, 400, 800 and 1600 nM of unlabeled RNA were added.
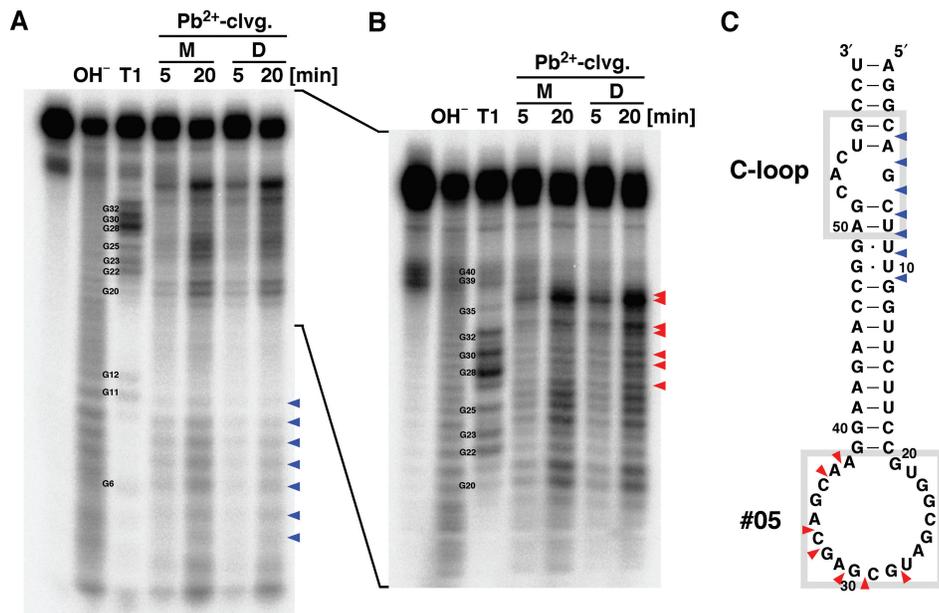

Figure 6. Lead(II)-induced RNA cleavage analysis. (**A** and **B**) Autoradiograms of the cleavage analysis of the construct shown in Figure 4B. Monomeric (without the unlabeled RNA) and dimeric (with 1 μM of the unlabeled RNA) conditions are indicated by M and D, respectively. OH⁻ and T1 correspond to alkaline treatment and digestion with RNase T1, respectively. Protected or cleaved residues under dimeric condition are indicated with blue and red arrowheads, respectively. (**C**) Mapping of the cleavage positions on the secondary structure of the construct. Positions more efficiently cleaved in the monomer than in the dimer are indicated with blue arrowheads, and the positions more efficiently cleaved in the dimer than in the monomer are indicated with red arrowheads.

between these two sequences, the #05 receptor is likely to recognize the C-loop motif by specific, non-Watson–Crick tertiary interactions. However, it cannot be excluded at present that the receptor binding may cause the structural change of the C-loop motif. Therefore, structural and physicochemical studies of the RNA–RNA interaction are necessary to determine whether the receptor motif recognizes the native structure of the C-loop motif or not.

DISCUSSION

In this study, we have developed a selection system that enables the identification of novel RNA motifs that interact with a target RNA structure within a desired structural context. After successful selection against the GAAA tetraloop, used as a proof-of-concept model, the work aimed to generate novel receptor motifs against the C-loop motif. Although the C-loop is not considered as an RNA–RNA interaction motif, we successfully identified a novel class of receptor motifs against it.

In the model selection experiment against the GAAA-loop, we isolated two previously known GAAA-loop receptors, a naturally occurring 11-ntR and an artificial C7.2 motif generated previously by Costa and Michel, via *in vitro* selection based on the activity of T4 *td* Group I intron ribozyme (18). Interestingly, their selection also generated a third class of GAAA-loop receptors (C7.34 motif), which was not obtained in the present work [Note that clone #27 has CCC/GGG base pairings similar to the C7.34 motif but is expected to form the secondary structure different from the consensus structure of the C7.34 motif family (18,23,24)]. Possible explanations for not having obtained a C7.34-like motif include differences in experimental conditions, library design (i.e. the length and the position of the random sequences introduced), and most importantly, the difference of the parental ribozymes as platforms for the two selection systems. The two ribozymes may have distinct structural contexts to the common target motif (GAAA-loop). For example, if steric hindrance around the GAAA-loop differs between the two ribozymes, it may provide selective pressure since one receptor motif may be bulky, requiring much more void space around the GAAA-loop while the other receptor motif may be very compact. The ability of the ribozyme to maintain activity despite differences in orientation and/or physical affinity between the GAAA-loop and its receptor might also be possible factor(s). These factors may result in different selective pressures, under which different motifs can be adapted and isolated even though they are independent structural units. Thus, parallel and/or sequential selections based on different contexts (for example, refs. 18,25–28) may be required to produce motifs truly independent from the structural context.

Importantly, the experiment aiming to generate novel receptor motifs against the C-loop motif, which is not considered as an RNA–RNA interaction motif, also identified novel receptor motifs. The most abundant sequence (#05) was further investigated to determine if it can function within a different structural context. By grafting the target/receptor motif pair into a previously reported RNA architecture (TectoRNA), structural and functional independency of the new receptor motif was clearly demonstrated. Therefore, we believe that RNA motifs isolated via this selection system can be directly employed for RNA engineering, such as the design of artificial RNA architectures (29) or novel molecular tools for desired target RNAs including structured ncRNAs, regulatory mRNA elements, as well as RNA components of large, complicated ribonucleoprotein complexes like the ribosome and the spliceosome.

After submission of this manuscript, we became aware of a recent study that warrants mention. Geary *et al.* (30) have reported the isolation of artificial receptor motifs against the GGAA-tetraloop by the selection method based on TectoRNA dimerization.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. Gesteland,R.F., Cech,T.R. and Atkins, J.F. (eds) (2006) *The RNA World*, 3rd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
2. Batey,R.T. (2006) Structures of regulatory elements in mRNAs. *Curr. Opin. Struct. Biol.*, **16**, 299–306.
3. Mattick,J.S. and Makunin,I.V. (2006) Non-coding RNA. *Hum. Mol. Genet.*, **15**, R17–R29.
4. Sosnick,T.R. and Pan,T. (2003) RNA folding: models and perspectives. *Curr. Opin. Struct. Biol.*, **13**, 309–316.
5. Onoa,B. and Tinoco,I. Jr. (2004) RNA folding and unfolding. *Curr. Opin. Struct. Biol.*, **14**, 374–379.
6. Bokinsky,G. and Zhuang,X. (2005) Single-molecule RNA folding. *Acc. Chem. Res.*, **38**, 566–573.
7. Furtig,B., Buck,J., Manoharan,V., Bermel,W., Jaschke,A., Wenter,P., Pitsch,S. and Schwalbe,H. (2007) Time-resolved NMR studies of RNA folding. *Biopolymers*, **86**, 360–383.
8. van der Horst,G., Christian,A. and Inoue,T. (1991) Reconstitution of a group I intron self-splicing reaction with an activator RNA. *Proc. Natl Acad. Sci. USA*, **88**, 184–188.
9. Murphy,F.L. and Cech,T.R. (1993) An independently folding domain of RNA tertiary structure within the *Tetrahymena* ribozyme. *Biochemistry*, **25**, 5291–5300.
10. Murphy,F.L. and Cech,T.R. (1994) GAAA tetraloop and conserved bulge stabilize tertiary structure of a group I intron domain. *J. Mol. Biol.*, **236**, 49–63.
11. Naito,Y., Shiraishi,H. and Inoue,T. (1998) P5abc of the *Tetrahymena* ribozyme consists of three functionally independent elements. *RNA*, **4**, 837–846.
12. Johnson,T.H., Tijerina,P., Chadee,A.B., Herschlag,D. and Russell,R. (2005) Structural specificity conferred by a group I RNA peripheral element. *Proc. Natl Acad. Sci. USA*, **102**, 10176–10181.

13. Ikawa,Y., Fukada,K., Watanabe,S., Shiraishi,H. and Inoue,T. (2002) Design, construction, and analysis of a novel class of self-folding RNA. *Structure*, **10**, 527–534.

14. Ikawa,Y., Tsuda,K., Matsumura,S. and Inoue,T. (2004) *De novo* synthesis and development of an RNA enzyme. *Proc. Natl Acad. Sci. USA*, **101**, 13750–13755.

15. Ikawa,Y., Matsumoto,J., Horie,S. and Inoue,T. (2005) Redesign of an artificial ligase ribozyme based on the analysis of its structural elements. *RNA Biol.*, **2**, 137–142.

16. Kirsebom,L.A. and Ciesiolka,J. (2005) Pb^{2+}-induced cleavage of RNA. In Hartmann,R.K., Bindereif,A., Schön,A. and Westhof,E. (eds), *Handbook of RNA Biochemistry*, Vol. 1. Wiley-VCH, Weinheim, Germany, pp. 214–228.

17. Toulme,J.J., Darfeuille,F., Kolb,G., Chabas,S. and Staedel,C. (2003) Modulating viral gene expression by aptamers to RNA structures. *Biol. Cell*, **95**, 229–238.

18. Costa,M. and Michel,F. (1997) Rules for RNA recognition of GNRA tetraloops deduced by *in vitro* selection: comparison with *in vivo* evolution. *EMBO J.*, **16**, 3289–3302.

19. Lescoute,A., Leontis,N.B., Massire,C. and Westhof,E. (2005) Recurrent structural RNA motifs, isostericity matrices and sequence alignments. *Nucleic Acids Res.*, **33**, 2395–2409.

20. Rohatgi,R., Bartel,D.P. and Szostak,J.W. (1996) Kinetic and mechanistic analysis of nonenzymatic, template-directed oligoribonucleotide ligation. *J. Am. Chem. Soc.*, **118**, 3332–3339.

21. Jaeger,L. and Leontis,N.B. (2000) Tecto-RNA: one-dimensional self-assembly through tertiary interactions. *Angew. Chem. Int. Ed. Engl.*, **39**, 2521–2524.

22. Jaeger,L., Westhof,E. and Leontis,N.B. (2001) TectoRNA: modular assembly units for the construction of RNA nano-objects. *Nucleic Acids Res.*, **29**, 455–463.

23. Ikawa,Y., Naito,D., Aono,N., Shiraishi,H. and Inoue,T. (1999) A conserved motif in group IC3 introns is a new class of GNRA receptor. *Nucleic Acids Res.*, **27**, 1859–1865.

24. Ikawa,Y., Nohmi,K., Atsumi,S., Shiraishi,H. and Inoue,T. (1999) A comparative study on two GNRA-tetraloop receptors: 11-nt and IC3 motifs. *J. Biochem.*, **130**, 251–255.

25. Atsumi,S., Ikawa,Y., Shiraishi,H. and Inoue,T. (2001) Design and development of a catalytic ribonucleoprotein. *EMBO J.*, **20**, 5453–5460.

26. Atsumi,S., Ikawa,Y., Shiraishi,H. and Inoue,T. (2003) Selections for constituting new RNA-protein interactions in catalytic RNP. *Nucleic Acids Res.*, **31**, 661–669.

27. Saksmerprome,V., Roychowdhury-Saha,M., Jayasena,S., Khvorova,A. and Burke,D.H. (2004) Artificial tertiary motifs stabilize *trans*-cleaving hammerhead ribozymes under conditions of submillimolar divalent ions and high temperatures. *RNA*, **10**, 1916–1924.

28. Juneau,K. and Cech,T.R. (1999) *In vitro* selection of RNAs with increased tertiary structure stability. *RNA*, **5**, 1119–1129.

29. Jaeger,L. and Chworos,A. (2006) The architectonics of programmable RNA and DNA nanostructures. *Curr. Opin. Struct. Biol.*, **16**, 531–543.

30. Geary,C., Baudrey,S. and Jaeger,L. (2007) Comprehensive features of natural and *in vitro* selected GNRA tetraloop-binding receptors. *Nucleic Acids Res.*, **36**, 1138–1152.