Selections for constituting new RNA–protein interactions in catalytic RNP

Shota Atsumi¹, Yoshiya Ikawa^{1,2}, Hideaki Shiraishi^{1,2} and Tan Inoue^{1,2,*}

¹Graduate School of Science and ²Graduate School of Biostudies, Kyoto University, Kyoto 606-8502, Japan

Received September 16, 2002; Revised October 21, 2002; Accepted November 12, 2002

ABSTRACT

***In vitro* and *in vivo* selection techniques are developed to constitute new RNA–peptide interactions. The selection strategy is designed by employing a catalytic RNP consisting of a derivative of the *Tetrahymena* ribozyme and an artificial RNA-binding protein. An arginine-rich RNA-binding motif and its target RNA motif in the RNP are substituted with randomized sequences and used for the selection experiments. Previously unknown binding motifs are obtained and the newly established interactions have been indispensable for assembling a catalytically active RNP. The method employed in this study is useful for making customized self-splicing intron RNAs whose activity is regulated by protein cofactors.**

INTRODUCTION

According to the RNA world hypothesis, primordial enzymes consisting exclusively of RNA molecules had evolved to use protein molecules by replacing their own structural elements with those of proteins (1–5). Such replacements can be conceived in present RNA–protein complexes (RNPs). For example, the Cyt-18 protein cofactor that assists the self-splicing reaction of group I intron RNA is known to functionally substitute the P5abc RNA element in the RNA (6). Studies comparing RNA/protein compositions between bacterial and mitochondrial ribosomes have also suggested that certain RNA elements of bacterial ribosomes were replaced by protein components in mitochondrial ribosomes (7).

On the basis of that hypothesis and feasible examples, we designed and constructed an artificial catalytic RNP derived from the *Tetrahymena* self-splicing ribozyme (8). In the active form of the *Tetrahymena* intron ribozyme, the P5b and P6 elements are associated, and clamp themselves together (9,10), implying that the P5b and P6 elements can also be clamped by an appropriately designed protein. Accordingly, the original RNA–RNA interaction was replaced by RNA–protein interactions in the designed RNP (Fig. 1). For its RNA component, an inactive derivative of the *Tetrahymena* ribozyme (denoted M12) was devised by replacing the terminal loop in the P5b element and an internal

loop in the P6 element of the wild-type with box B and RRE protein-binding RNA motifs from bacteriophage λ and HIV, respectively (Fig. 1A) (8). For the protein component, two terminal regions consisting of RNA-binding peptide motifs, λ N_{1–19} and HIV Rev_{34–50}, which bind to the target sites, box B and RRE (11), respectively, were joined with a short linker (Fig. 1B). The binding of a protein (termed pep S) facilitated the splicing reaction by the RNA (M12) both *in vivo* and *in vitro*. The protein was responsible for the correct folding of the ribozyme (8).

Several methods have been developed for constructing new RNA-binding proteins or protein-binding RNAs by employing *in vitro* or *in vivo* selection (12–18), mRNA display (19), phage display (20–22) or a genetic screening technique (23). The methods produced several non-natural combinations of arginine-rich RNA-binding peptides and their target RNA motifs (13,14,17–19,24–26). However, no example has been reported in which both of selected RNA and peptide are non-natural.

In the present report, we would like to show a new selection system for constructing non-natural RNA–protein interactions. The system has two distinctive features as follows. Catalytic activity is utilized as an index for the enrichment of target molecules so that the constructed interactions are guaranteed to serve as functional elements in the active 3-dimensional structure of the RNA. And the system is usable for selecting both RNA and peptide components so that it allows reciprocal RNA and peptide selections without employing two systems.

The selection system was designed by using the M12/pep S catalytic RNP. The sites in the RNA and peptide components responsible for the RRE–Rev interaction in the RNP were employed as the targets for the selections. Non-native RNA and peptide sequences that functionally substitute the RRE RNA motif (Fig. 1A) and Rev peptide motif (Fig. 1B), respectively, were established in the respective RNP.

MATERIALS AND METHODS

Plasmids

Plasmids encoding the M12 ribozyme (pTZ-M12IVS) and pep S protein (pSTVpepS) were described previously (8). Plasmids encoding pepSmRev, pep C1 and pep C2 were constructed by site-directed mutagenesis of pSTVpepS. To construct the plasmids encoding RNA fragments for a gel shift

*To whom correspondence should be addressed at Graduate School of Biostudies, Kyoto University, Kyoto 606-8502, Japan. Tel: +81 75 753 3995; Fax: +81 75 753 3996; Email: tan@kuchem.kyoto-u.ac.jp

assay, DNA fragments containing the T7 RNA promoter joined to respective RNA sequences (see Fig. 6A) were prepared by using synthetic oligonucleotides, followed by sub-cloning into pUC18 vectors.

Preparation of precursor M12 RNAs and pep S proteins

Preparation of the precursor RNAs and proteins was performed as described (8). All RNAs employed in this study were prepared by *in vitro* transcription with T7 RNA polymerase and [α - 32 P]GTP (27). All proteins for *in vitro* self-splicing assays were synthesized from plasmid pTYB1 expression derivatives in *Escherichia coli* strain ER2566 followed by purification with the IMPACT™ T7 System (New England Biolabs) (28).

In vitro splicing assays

In vitro splicing assays were performed as described with some modifications to the Mg $^{2+}$ concentrations. Here, 32 P-labeled precursor RNAs (10 nM) were dissolved in distilled water, followed by incubation at 80°C for 5 min. After cooling and incubation at 37°C for 1 min, a reaction buffer concentrated 10-fold and protein in a dilution buffer (20 mM Tris-HCl, pH 7.5, 40 mM KCl, 50% glycerol) were simultaneously added at 37°C. Following pre-incubation at 37°C for 5 min, reactions were started with 200 μ M GTP at 37°C in the presence of 40 mM Tris-HCl, pH 7.5, 1.5 mM MgCl $_2$, 80 mM KCl and 2.5% glycerol. Aliquots were removed at specified times. The reactions were terminated by adding an equal volume of stop solution (150 mM EDTA, 70% formamide, 0.25% xylene cyanol) followed by electrophoresis on 5% polyacrylamide denaturing gels. RNAs were quantified with a BioImaging Analyzer (BAS2500; Fuji Film, Japan). To determine the final extents of the splicing reactions, the data were fitted to single exponential curves. Final extents obtained by data fitting were essentially identical to the extents of the reactions after 60 min incubation.

Construction of variant M12 ribozyme pool

We constructed a DNA template for a variant M12 ribozyme pool by ligating two PCR fragments that encode the 5' and 3' halves of the ribozyme that correspond to the 5' exon-P6b loop of the ribozyme and the P6b-3' exon, respectively. By using pTZ-M12IVS encoding the M12 ribozyme as a template, the 5' half fragment was amplified by PCR with primers T7 (TAAACGACTCACTATAGGG-3') and A152 (5'-ACAGCCCCGGTCTCCGATGCNNNNNNNNNNNNCT-

TGGCTGCGTGGTTAGGACCATGTCCGTC-3'; hereafter underlining in primer sequences indicates a *Bsa*I recognition

B Pep S protein and its derivatives

Figure 1. Structure and sequences of M12 ribozyme and pep S protein constituting the catalytic RNP. (A) The secondary structure of the M12 ribozyme derived from *Tetrahymena* group I intron (8). Boxes with broken lines indicate the P5b and P6 regions. Their original sequences forming the GAAA tetraloop-11 nt receptor long-range interaction in the *Tetrahymena* intron (9,10,32) are replaced with λ box B and HIV RRE motifs, respectively, to construct M12 (8). Boxes with solid lines indicate the mutant λ box B (mBoxB) and HIV-RRE (mRRE) motifs lacking peptide-binding ability (8). Regions shaded in gray show the randomized regions of M12 used in *in vitro* selection. The labels N11 and N12 represent the randomized 11 and 12 nucleotides, respectively. Asterisks indicate the mutated positions in some of the selected M12 variants (see Table 1). (B) Sequences of pep S protein and its derivatives. pep C1 and pep C2 were selected as clones that can activate variant ribozymes #1 and #2, respectively (Table 1).

site, 5'-GGTCTC-3'). By using the same plasmid template, the 3' half fragment was amplified by PCR with primers A153 (5'-ACAGCCCGGTCTCGCATCTTGCNNNNNNNNNN-TGCAGTTCACAGACTAAATGTCGGTTCGGG-3') and A18 (5'-GCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGA-3'). Two fragments were digested with *Bsa*I, followed by purification by low melting point agarose gel electrophoresis, and ligated together by using T4 DNA ligase.

The *in vitro* RNA selection

Step 1 (negative selection). After *in vitro* transcription with T7 RNA polymerase, the DNA templates were digested with DNase I and the transcripts were purified on 5% denaturing polyacrylamide gels containing 7 M urea. The purified RNAs were dissolved in the reaction buffer (40 mM Tris-HCl, pH 7.5, 1.5 mM MgCl₂, 80 mM KCl and 200 μ M GTP) followed by incubation at 80°C for 5 min, then at 37°C for 5 min. The procedure was repeated three times. Unreacted RNAs were purified on polyacrylamide denaturing gels.

Step 2 (positive selection). RNAs obtained by step 1 were dissolved in water followed by incubation at 80°C for 5 min. After cooling at 37°C for 1 min, a reaction buffer concentrated 10-fold and protein in a dilution buffer (20 mM Tris-HCl, pH 7.5, 40 mM KCl and 50% glycerol) were simultaneously added at 37°C. Following pre-incubation at 37°C for 5 min, the reactions were started with 200 μ M GTP at 37°C in the presence of 40 mM Tris-HCl, pH 7.5, 1.5 mM MgCl₂, 80 mM KCl and 2.5% glycerol. The circularized intron RNAs were purified by denaturing PAGE.

Step 3 (RT-PCR). Purified circular RNAs were reverse transcribed with primer A165a (complementary to the 5' region of the box B motif, 5'-TCTTCAGGGCAC-TTGGTACTGAACGGCTG-3') by ReverTra Ace (Toyobo, Japan) and amplified by PCR with primer A166 (3' region of the box B motif, 5'-ACACGCCGGTCTCGAAGAAGGGCA-TGGCCTTGCAAAGGGTATGG-3') and primer A158 (complementary to the 3' end of intron RNA, 5'-CAC-CGCCGGTCTCACTCCAAACTAATCAATATACTTTC-3'). The DNA was digested with *Bsa*I (in the primer linker sequences) and ligated to the rest of the fragments that were amplified by PCR with the T7 primer/A165b (5'-ACA-GCCCGGTCTCTTCTTCAGGGCACTTGGTACTGAACGGCTG-3') set and the A159 (5'-GCACCGCGGTCTCTG-GAGTACTCGTAAGGTAGCCAAATGCCTC-3')/A18 set, respectively, and digested with *Bsa*I. The resulting ligated DNAs were used as templates for *in vitro* transcription to prepare for the next round of RNA pooling. Steps 1–3 were repeated eight times.

The *in vivo* mutagenic protein selection

Using pSTVpep S as a template, a PCR with primer A200 [5'-GACCTCGGTCTCGCAGCG(XYT)₁₂TCGTGAACGTCA-GCGTGC GGCGGCCGCA-3' where X is a C:A mixture (3:1) and Y is an A:G mixture (1:3)] and A198 (complementary to positions 2398–2410 of pSTV28, 5'-AGGACGGGTCTCTT-CCGGAGTGTATACTGGCTTACTAT-3') was carried out to prepare fragment 1. Using the same plasmid template, a PCR with primer A199 (positions 2415–2434 of pSTV28, 5'-AGGACGGGTCTCTCCGGAAGCGCTGATGTCCGGCGG-

TG-3') and primer A195 (complementary to positions 68–88 of the DNA encoding pep S, 5'-AGAGCCGGTCTCCG-CTGCGGCGCCACGCCCAACCCCG-3') was also carried out to prepare fragment 2. Fragments 1 and 2 were digested with *Bsa*I. A plasmid library of pSTVpep S variants was constructed by ligating *Bsa*I-digested fragments 1 and 2 with T4 DNA ligase, and the plasmid library was amplified on selective agar plates (Invitrogen, Tech-Online, available at <http://www2.lifetech.com/content.cfm?pageid=1419&cfid=578892&cftoken=4691213>). *Escherichia coli* JM109 containing plasmids for the RNAs was transformed with 1 μ g of the plasmid library contents. The cells were incubated in 200 ml of M9 medium containing 400 mg/l lactose, 50 mg/l ampicillin and 50 mg/l chloramphenicol, with shaking at 37°C for 2 days.

RNA-binding gel mobility shift assay

DNA fragments encoding the T7 promoter and RNA fragments (see Fig. 6A) were prepared from the corresponding plasmids by PCR amplification. Using the PCR products as templates, *in vitro* transcriptions were performed with T7 RNA polymerase and [α -³²P]GTP. The protein (5, 20, 40, 80 or 160 nM) and 0.5 nM ³²P-labeled RNA were incubated for 10 min at 4°C in 10 μ l of the binding mixture (20 mM Tris-HCl, pH 7.5, 80 mM KCl, 1.5 mM MgCl₂, 50 μ g/ml tRNA and 10% glycerol). The resulting protein–RNA complexes were resolved and electrophoresed on 20% polyacrylamide/0.5 $\times$ TBE gels containing 1 mM MgCl₂ at 4°C (29). In addition, the *K*_d was defined from the plot of bound/free RNA versus protein concentration.

RESULTS

Selection of Rev-binding RNA motifs in the RNP

To obtain new Rev-binding RNA motifs, we attempted to select the motifs that can replace the function of the RRE in the P6 element of M12 RNA, a derivative of the *Tetrahymena* ribozyme, by employing the *in vitro* selection technique (30). The selection protocol was designed to enrich the RNAs that can perform the splicing reaction if pep S, the RNA-binding protein which bridges P5b and P6, was associated correctly with M12 RNA via a new Rev–RNA interaction (Fig. 2A). To avoid the recovery of the original RRE motif or its homologs, the lengths of the 5' and 3' region sequences were altered from 11 and 13 nt in the RRE to 12 and 11 nt in the designed pool, respectively. Thus the selected M12 variants are designed to contain non-native Rev-binding motifs (8).

Twenty-three nucleotides of the P6 element containing the RRE were randomized (Fig. 1B) and to isolate active variants from the library of 1×10^{12} different RNAs we performed successive negative and positive selections. First, the RNA precursors were incubated in the absence of pep S for the negative selection (Fig. 2A). Next, the unspliced precursors were subjected to positive selection in the presence of pep S that produces the circularized introns due to the activity of the ribozyme. The circularized introns were then amplified by reverse transcription followed by PCR amplification. The selections were repeated to concentrate the active M12 variants.

Figure 2. Selection of artificial Rev-binding RNA motifs. (A) The selection procedure is schematically presented. Open and closed circles indicate 5' and 3' splice sites, respectively. Asterisks indicate internal circularization sites. Parentheses for the excised intron ribozyme, which is subjected to a subsequent circularization reaction, indicate that the corresponding RNA was not isolated. The details of the selection procedure are described in Materials and Methods. (B) Final extent of the splicing reactions by M12, round 8 pool (R8 pool) and the selected clones (10 nM) in the presence (shaded bar) or absence (black bar) of 1 μ M pep S.

After eight selection rounds (R8), the activity of the R8 RNA pool was assayed in the presence of pep S. It was considerably weaker than that of M12-pep S RNP (Fig. 2B). To further concentrate the RNAs in the pool, we carried out positive and negative *in vivo* blue-white colony color assays, in that the color of colonies reflects the efficacy of splicing (31). For screening, the RT-PCR DNAs from the R8 pool were inserted into the β -galactosidase α -fragment gene in the plasmid vector derived from pTZ18U (31); the efficacy of mature lacZ mRNA production should depend on the efficacy of the splicing reaction in *E. coli*. The constructed plasmids were transformed into *E. coli* JM109 cells harboring pep S expression plasmids (8), so that the cells should yield mature lacZ mRNA if a variant M12 ribozyme in the R8 pool can effectively splice out the intron with pep S. Two hundred blue

Table 1. Sequences of the P6 regions of M12 and selected clones

	Sequence
M12	5'-TCCTGGGCGCA-(P6b stem-loop)TGACGGTACAGGA-3'
#1	5'-CCGTACCCTTTC(P6b stem-loop)-GGTAGGCGCGG-3'
#2	5'-CGCATCGCTCTT(P6b stem-loop)-GGAGTGTTCGG-3'
#3*	5'-CCCCGAATAGTG(P6b stem-loop)-CTGTTTCGGGG-3'
#4	5'-CCGTACCCCTC(P6b stem-loop)-TTAGGGAGCGG-3'
#5*	5'-CCGCCACCCCA(P6b stem-loop)-GAGTGTGACGG-3'
#6*	5'-CCGGACCCCGCC(P6b stem-loop)-TGTTGGGCGCG-3'
#7	5'-CTATGCCCTTCT(P6b stem-loop)-ACAGGACATAG-3'
#8*	5'-CCAGCCGTACCC(P6b stem-loop)-TCGTGCGCTGG-3'
#9*	5'-CTGGAGGTGGCT(P6b stem-loop)-TACACTCCAG-3'
#10*	5'-CCATACCTTCC(P6b stem-loop)-AGGAGCCGTGG-3'
#11	5'-CCATATCCACA(P6b stem-loop)-TCGCGGAATGG-3'

Asterisks indicate the clones which contain point mutations in the constant region: #3 (U271C, A299U, insertion of 5'-AUGC-3' between C383 and G384); #5 (U322G, deletion of U377); #6 (U377C); #8 (U271C, U322G); #9 (U322G); #10 (U271C, U357C) (see Fig. 1A). Underlining indicates the regions forming three or more consecutive base pairs between the 5' and 3' termini.

colonies were observed out of ~20 000 colonies. For the negative screening, *E. coli* harboring no pep S expression plasmids were transformed with plasmids encoding the selected RNAs. One hundred white colonies were observed out of 5000 colonies. Following an additional positive screening, 20 colonies were picked for sequencing.

The 11 variants selected *in vivo* (from #1 to #11) were, in the presence of pep S, active *in vitro* (Table 1 and Fig. 2B). We did not observe any apparent conservation of their predicted secondary structures, although G-C or C-G base pairs were considerably dominant in the P6a stem (see Fig. 1A, underlined regions in Table 1 and predicted secondary structures of #1 and #2 in Fig. 6A). The first and last nucleotides of the randomized region were C and G, respectively, without exception (Table 1). In contrast, an A-U base pair was located at the corresponding positions (positions 221 and 252 in Fig. 1A; see also Fig. 6A) in the M12 and the parental *Tetrahymena* ribozyme (Fig. 1A).

Six RNAs (#3, #5, #6, #8, #9 and #10) contained extra point mutations in the constant regions of M12 (see legend to Table 1). The activities of four (#3, #5, #8 and #10) were diminished in the presence of pep S when the mutations were fixed, indicating that some mutations were responsible for activating the pep S-RNA complex (data not shown).

Negative selection was performed to eliminate variants that were active without pep S (Fig. 2B). However, the activity of the selected clones without pep S (basal activity) was much higher than that of the parental M12, suggesting that structural elements related to pep S-dependent activation simultaneously contribute to improving the basal activity. For example, the conserved C-G pair in P6a and/or C-G/G-C-rich P6a duplex (underlined regions in Table 1) might contribute to improving not only pep S-dependent activation but also the basal activity by stabilizing the core structure.

For improving the basal activity, it is possible to postulate that the disrupted RNA-RNA interaction between P5b and P6 in M12 (8) was regenerated in the selected RNAs. To test this hypothesis, the P5b loop of two clones (#1 and #5) was replaced with a UUCG tetraloop which cannot participate in the RNA-RNA interaction (9,32,33). The basal activity of the

Figure 3. *In vitro* splicing reaction by the RNPs. Splicing reaction of (A) M12, (B) two selected variants, (C) mutants of the selected variants and (D) the *Tetrahymena* ribozyme and the mutants of M12 with or without proteins. The autoradiograms of the splicing reactions with 10 nM 32 P-labeled precursor RNA in the presence or absence of 1 μ M pep S or its derivatives. Abbreviations of RNAs are as follows: M12 (mboxB) and M12 (mRRE), the M12 mutants having the mutated box B and RRE motif, respectively (see Fig. 1A); #1 (mboxB) and #2 (mboxB), the derivatives of ribozymes #1 and #2 possessing the mutated box B motifs (see Fig. 1A). Abbreviations of proteins are as follows. S, pep S; C1, pep C1; C2, pep C2; Sm, pep SmRev. Abbreviations of the splicing reaction products are as follows: C-I, circularized intron; pre, precursor RNA; L-I, linear intron; LE, the ligated exons.

UUCG mutants was the same as that of the parental clones, and addition of pep S did not enhance the activity of the mutants (data not shown), indicating that the loop region of P5b is not interacting with P6 in the selected variants.

It can also be postulated that the stem region of P5b interacts with the selected P6 sequences to raise the basal activity. However, this seems unlikely because the two regions are isolated far from each other in the 3-dimensional model of the M12 RNA (8).

Selection of arginine-rich peptide motifs in RNP

To obtain new arginine-rich peptide motifs which specifically recognize the selected RNA motifs in the #1 and #2 M12 variants (Table 1; see also Fig. 6A), we performed *in vivo* selection experiments by employing a library containing variant pep S. Twelve amino acid residues in the Rev motif were randomized with mixed sequences consisting of arginine, serine, asparagine or histidine (Fig. 1B) (17). For

the selection, the plasmids were constructed for expressing M12 variants and the pep S library (8). The template for the #1 or #2 variants RNA was inserted into the β -galactosidase α -fragment gene in a vector derived from pTZ18U (31) and the variant pep S expression library was constructed in plasmids derived from pSTV28. Thus, in M9 medium, the growth rate of *E. coli* cells harboring two plasmids for an RNA and a peptide should reflect the level of β -galactosidase expression, depending on the efficacy of the splicing reaction *in vivo* (8).

Fast growing cells were isolated from 1×10^6 transformants after one round of selection. Ten clones were picked randomly from each pool containing either #1 or #2 RNA and their sequences were compared. Two characteristic arginine-rich sequences were identified for the peptides from the two pools. One peptide from the pool containing #1 RNA and another one from the pool containing #2 RNA were named pep C1 and pep C2 (from #1 and #2 RNA, respectively; Fig. 1B). No site-specific conservation was observed for their arginine residues

Figure 4. Splicing reactions by the M12 ribozyme or its selected variants in the presence of various concentrations of the proteins. The final extent of the splicing reactions by M12 and ribozymes #1 and #2 with three proteins are shown in (A), (B) and (C), respectively.

(Fig. 1B). To see whether the growth rate actually reflected the splicing activity, we conducted blue–white colony color assays. The cells harboring the plasmids encoding both the RNA and selected protein gave dark blue colonies, whereas pale blue colonies were produced for the cells harboring the plasmids encoding only RNA (data not shown).

Selected new RNA–protein interactions in self-splicing reaction

The splicing reaction of the precursor RNAs containing #1 or #2 M12 variant RNA was performed with the selected protein *in vitro* (Figs 3–5). Under optimal conditions, the final extent of the splicing reaction by #1 or #2 with pep C1 or pep C2, respectively, was comparable to that by M12 with pep S, indicating that the new interactions are substituting for the RRE–Rev interaction (Fig. 4). Interestingly, both pep C1 and

Figure 5. Time courses of the splicing reactions by the RNPs. The splicing reactions with 10 nM 32 P-labeled M12 and ribozyme #1 or #2 with or without 1 μ M proteins are shown in (A), (B) and (C), respectively.

pep C2 were able to enhance the activity of the M12 ribozyme containing the Rev motif (Fig. 3A).

To see whether newly selected RNA motifs in the #1 and #2 RNAs (Table 1 and Fig. 6A) are physically associated with the RNA-binding motifs in the peptide, the splicing reactions of the two RNAs were attempted in the presence of a pep S variant that is unable to bind to RRE; four critical residues in the Rev peptide were mutated to abolish the binding (pep SmRev; see Fig. 1B). As shown in Figure 3C, the Rev peptide motif is conceivably responsible for enhancing the splicing of #1 and #2 RNA with pep S.

Next, we investigated whether pep C1 and pep C2 can assist the splicing of #1 and #2 RNAs by using their box B peptide-binding motifs in the P5b regions (Fig. 1A). Neither peptide was able to assist the splicing of the variant #1 and #2 RNA

Figure 6. RNA-protein gel mobility shift assay. (A) Predicted secondary structures of RNA fragments used for a gel shift assay. RNAs are derived from the P6 region of the M12 ribozyme or its variant ribozymes 1 and 2. Small letters indicate extra nucleotides for stabilizing P6a stems. Asterisks indicate a completely conserved C-G base pair among the 11 selected sequences (see Table 1). Secondary structures of the fragments are predicted by using Zucker's mfold program (42). (B-E) Autoradiograms of the electrophoreses on native gels with the RNA-binding proteins (160 nM) and 32 P-labeled protein-binding RNAs (0.5 nM). (b) fg-RRE; (c) fg-1; (d) fg-2; (e) fg-mRRE. S, pep S; C1, pep C1; C2, pep C2; Sm, pep SmRev. Kd indicates the dissociation constant between RNA fragments and proteins that were determined by varying protein concentrations. Bars for K_d values indicate that the constants were too high to be determined by gel shift assay.

with mutations in P5b box B loops (Figs 1A and 3C). The fact that the selected peptides were incapable of enhancing the activity of two M12 variants with a disrupted P6 RRE or P5b box B indicate that both peptide-binding sites are employed for facilitating the splicing reaction (Figs 1A and 3D).

The Hill coefficient was determined by Hill's plot as the amount of the active RNP related proportionally to the final extent of the reaction at each protein concentration (Fig. 4) (8). The coefficient was approximately 1.0 for both pep C1 and pep C2 (data not shown), indicating that one RNA molecule associates with one protein.

RNA-binding gel mobility shift assay of the selected proteins

To detect the interaction between the selected RNA motifs and the peptide motifs, we attempted a gel mobility shift assay based on the RNA motifs (parental RRE or newly obtained protein-binding motifs) and the peptide motifs (Rev or newly obtained RNA-binding motifs) (Fig. 6). RNA fragments consisting of the P6a stem, RRE or newly obtained peptide-binding motifs (#1 and #2) and the P6b stem-loop were prepared as fg-RRE, fg-#1 or fg-#2, respectively (Fig. 6A). The assay revealed that the affinity between fg-#1 and pep C1 or fg-#2 and pep C2 was comparable to that between fg-RRE and pep S (Fig. 6B-D). However, pep C1 and pep C2 had a 1.5-fold weaker association than pep S with fg-RRE (Fig. 6B). pep SmRev (Fig. 1B) did not bind to fg-#1 or fg-#2, suggesting that arginine residues of the Rev motif are critical for the interaction with the RNA motifs. A RNA fragment containing a mutant RRE motif with no affinity for Rev peptide also showed no affinity for pep C1 or pep C2 (Fig. 6E), suggesting that the Rev and the newly selected C1 and pep C2 peptides share the same mechanism for recognizing the RRE element of RNA.

DISCUSSION

New arginine-rich RNA-binding motifs and corresponding peptide-binding motifs in RNA, which replace the Rev and RRE motifs, respectively, were obtained by selection experiments employing a catalytic RNP consisting of M12 RNA derived from the *Tetrahymena* ribozyme, and pep S, which is a designed protein component. Newly established RNA-peptide interactions effectively regulate and facilitate the splicing reaction both *in vivo* and *in vitro*. The affinities between the peptide motifs and the corresponding RNA motifs were comparable to that between the Rev and RRE.

Arginine-rich motifs

The sequences and structures are divergent for natural arginine-rich motifs that specifically interact with respective parts in RNAs. The sequences of newly obtained motifs in pep C1 and pep C2 were arginine-rich (67%) but their sequences were divergent (Fig. 1B), suggesting that a variety of arginine-rich sequences are usable for substituting natural arginine-rich motifs such as Rev.

pep C1 and pep C2 were tolerant to sequence variation of the corresponding binding sites in RNA. The peptides able to bind to the binding sites of #1 and #2 RNA were also able to bind to RRE (Fig. 6B). Moreover, each peptide can activate non-cognate parter RNAs (Fig. 3B), indicating that the C1 and C2 peptides have cross-affinity for the #1 and #2 RNA motifs even though cognate C1-#1 and C2-#2 combinations were selected via two steps of RNA and protein selection. Such tolerance in the RNA-protein interactions has been observed for the interaction between natural arginine-rich motifs and

their target RNA motifs. For example, the RNA-binding domain of Tat protein from JDV can recognize two different TAR RNA sites from HIV and BIV (29). The protein adopts different conformations on the two RNAs by employing different amino acids that specifically recognize the binding sites (29). Conversely, it has been reported that one RNA motif recognizes two peptide fragments: namely, the RRE motif has been shown to recognize Rev as well as artificial arginine-rich peptides by switching its RNA conformation (18,25,26). Furthermore, it is interesting to note that a variety of arginine-rich motifs are found in three related viruses (HIV, BIV and JDV) that are likely to have evolved rapidly (11). This suggests that an ancestral form of the arginine-rich peptides had readily evolved to form a variety of arginine-rich RNA-binding motifs.

Knight and Landweber have shown that the sequences for arginine codons (GCN and AGR; N, any nucleotide, R, purine) are over-represented at arginine-binding sites in RNAs (34). For example, 38% of the RRE nucleotides theoretically encode the arginine whereas the statistical probability is 28% for a base in a random sequence to be in an arginine codon. In the selected clones, 36% of the nucleotides at the binding sites were GCN or AGR (data not shown). If the over-representation of GCN or AGR was due to the base composition, sequences such as CGN or GAR should also be equally represented. However, CGN and GAR comprised only 23%. Thus our data support the hypothesis that amino acids can interact specifically with RNA sequences that contain their cognate codons (34).

Protein-dependent allosteric ribozyme

We successfully obtained alternative forms of self-splicing intron RNA whose activities were regulated by a protein component. The resulting allosteric ribozyme could be a useful tool in the field of molecular biology and medicine, as it has been shown that the *Tetrahymena* ribozyme is usable for gene therapy (35–41). For example, the present method is applicable for developing new ribozymes whose activity is regulated by a variety of protein cofactors so that the ribozyme whose activity is regulated by a viral protein, such as HIV Rev or Tat, may be obtained via *in vivo* or *in vitro* selection.

ACKNOWLEDGEMENTS

We thank members of the Inoue laboratory for their helpful advice. This work was supported by Grants-in-Aids for Scientific Research on Priority Areas (T.I.) and the Encouragement of Young Scientists (Y.I.) from the Ministry of Education, Science, Sports and Culture, Japan.

REFERENCES

- Gilbert, W. (1986) The RNA world. *Nature*, **319**, 618.
- Cech, T.R. and Golden, B.L. (1998) Building a catalytic active site using only RNA. In Gesteland, R.F., Cech, T.R. and Atkins, J.F. (eds), *The RNA World*, 2nd Edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 321–349.
- Lambowitz, A.M. and Perlman, P.S. (1990) Involvement of aminoacyl-tRNA synthetases and other proteins in group I and group II intron splicing. *Trends Biochem. Sci.*, **15**, 440–444.
- Lambowitz, A.M., Caparara, M.G., Zimmerly, S. and Perlman, P.S. (1998) Group I and group II ribozymes as RNPs: clues to the past to the guides to the future. In Gesteland, R.F., Cech, T.R. and Atkins, J.F. (eds), *The RNA World*, 2nd Edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 451–485.
- Schimmel, P. and Kelley, S.O. (2000) Exiting an RNA world. *Nature Struct. Biol.*, **7**, 5–7.
- Mohr, G., Caparara, M.G., Guo, Q. and Lambowitz, A.M. (1994) A tyrosyl-tRNA synthetase can function similarly to an RNA structure in the *Tetrahymena* ribozyme. *Nature*, **370**, 147–150.
- O'Brien, T.W. (2002) Evolution of a protein-rich mitochondrial ribosome: implications for human genetic disease. *Gene*, **286**, 73–79.
- Atsumi, S., Ikawa, Y., Shiraishi, H. and Inoue, T. (2001) Design and development of a catalytic ribonucleoprotein. *EMBO J.*, **20**, 5453–5460.
- 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.
- Cate, J.H., Gooding, A.R., Podell, E., Zhou, K., Golden, B.L., Kundrot, C.E., Cech, T.R. and Doudna, J.A. (1996) Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science*, **273**, 1678–1685.
- Draper, D.E. (1995) Protein-RNA recognition. *Annu. Rev. Biochem.*, **64**, 593–620.
- Bartel, D.P., Zapp, M.L., Green, M.R. and Szostak, J.W. (1991) HIV-1 Rev regulation involves recognition of non-Watson-Crick base pairs in viral RNA. *Cell*, **67**, 529–536.
- Giver, L., Bartel, D., Zapp, M., Pawul, A., Green, M. and Ellington, A.D. (1993) Selective optimization of the Rev-binding element of HIV-1. *Nucleic Acids Res.*, **21**, 5509–5516.
- Giver, L., Bartel, D.P., Zapp, M.L., Green, M.R. and Ellington, A.D. (1993) Selection and design of high-affinity RNA ligands for HIV-1 Rev. *Gene*, **137**, 19–24.
- Franklin, N.C. (1993) Clustered arginine residues of bacteriophage lambda N protein are essential to antitermination of transcription, but their locale cannot compensate for boxB loop defects. *J. Mol. Biol.*, **231**, 343–360.
- Jensen, K.B., Green, L., MacDougall-Waugh, S. and Tuerk, C. (1994) Characterization of an *in vitro*-selected RNA ligand to the HIV-1 Rev protein. *J. Mol. Biol.*, **235**, 237–247.
- Harada, K., Martin, S.S. and Frankel, A.D. (1996) Selection of RNA-binding peptides *in vivo*. *Nature*, **380**, 175–179.
- Harada, K., Martin, S.S., Tan, R. and Frankel, A.D. (1997) Molding a peptide into an RNA site by *in vivo* peptide evolution. *Proc. Natl Acad. Sci. USA*, **94**, 11887–11892.
- Barrick, J.E., Takahashi, T.T., Ren, J., Xia, T. and Roberts, R.W. (2001) Large libraries reveal diverse solutions to an RNA recognition problem. *Proc. Natl Acad. Sci. USA*, **98**, 12374–12378.
- Friesen, W.J. and Darby, M.K. (1997) Phage display of RNA binding zinc fingers from transcription factor IIIA. *J. Biol. Chem.*, **272**, 10994–10997.
- Blancafort, P., Steinberg, S.V., Paquin, B., Klinck, R., Scott, J.K. and Cedergren, R. (1999) The recognition of a noncanonical RNA base pair by a zinc finger protein. *Chem. Biol.*, **6**, 585–597.
- Danner, S. and Belasco, J.G. (2001) T7 phage display: a novel genetic selection system for cloning RNA-binding proteins from cDNA libraries. *Proc. Natl Acad. Sci. USA*, **98**, 12954–12959.
- Jain, C. and Belasco, J.G. (1996) A structural model for the HIV-1 Rev-RRE complex deduced from altered-specificity rev variants isolated by a rapid genetic strategy. *Cell*, **87**, 115–125.
- Ye, X., Gorin, A., Frederick, R., Hu, W., Majumdar, A., Xu, W., McLendon, G., Ellington, A. and Patel, D.J. (1999) RNA architecture dictates the conformations of a bound peptide. *Chem. Biol.*, **6**, 657–669.
- Gosser, Y., Hermann, T., Majumdar, A., Hu, W., Frederick, R., Jiang, F., Xu, W. and Patel, D.J. (2001) Peptide-triggered conformational switch in HIV-1 RRE RNA complexes. *Nature Struct. Biol.*, **8**, 146–150.
- Zhang, Q., Harada, K., Cho, H.S., Frankel, A.D. and Wemmer, D.E. (2001) Structural characterization of the complex of the Rev response element RNA with a selected peptide. *Chem. Biol.*, **8**, 511–520.
- Milligan, J.F., Groebe, D.R., Witherell, G.W. and Uhlenbeck, O.C. (1987) Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res.*, **15**, 8783–8798.
- Chong, S., Mersha, F.B., Comb, D.G., Scott, M.E., Landry, D., Vence, L.M., Perler, F.B., Benner, J., Kucera, R.B., Hirvonen, C.A. et al. (1997) Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. *Gene*, **192**, 271–281.

29. Smith,C.A., Calabro,V. and Frankel,A.D. (2000) An RNA-binding chameleon. *Mol. Cell*, **6**, 1067–1076.
30. Williams,K.P. and Bartel,D.P. (1996) *In vitro* selection of catalytic RNA. In Eckstein,F. and Lilley,D.M.J. (eds), *Catalytic RNA*. Springer-Verlag, Berlin, Germany, pp. 367–382.
31. 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.
32. Costa,M. and Michel,F. (1995) Frequent use of the same tertiary motif by self-folding RNAs. *EMBO J.*, **14**, 1276–1285.
33. 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.
34. Knight,R.D. and Landweber,L.F. (1998) Rhyme or reason: RNA-arginine interactions and the genetic code. *Chem. Biol.*, **5**, R215–R220.
35. Sullenger,B.A. and Cech,T.R. (1994) Ribozyme-mediated repair of defective mRNA by targeted, trans-splicing. *Nature*, **371**, 619–622.
36. Campbell,T.B. and Cech,T.R. (1995) Identification of ribozymes within a ribozyme library that efficiently cleave a long substrate RNA. *RNA*, **1**, 598–609.
37. Campbell,T.B., McDonald,C.K. and Hagen,M. (1997) The effect of structure in a long target RNA on ribozyme cleavage efficiency. *Nucleic Acids Res.*, **25**, 4985–4993.
38. Ayre,B.G., Kohler,U., Goodman,H.M. and Haseloff,J. (1999) Design of highly specific cytotoxins by using trans-splicing ribozymes. *Proc. Natl Acad. Sci. USA*, **96**, 3507–3512.
39. Kohler,U., Ayre,B.G., Goodman,H.M. and Haseloff,J. (1999) Trans-splicing ribozymes for targeted gene delivery. *J. Mol. Biol.*, **285**, 1935–1950.
40. Watanabe,T. and Sullenger,B.A. (2000) Induction of wild-type p53 activity in human cancer cells by ribozymes that repair mutant p53 transcripts. *Proc. Natl Acad. Sci. USA*, **97**, 8490–8494.
41. Watanabe,T. and Sullenger,B.A. (2000) RNA repair: a novel approach to gene therapy. *Adv. Drug Deliv. Rev.*, **44**, 109–118.
42. Zuker,M., Mathews,D.H. and Turner,D.H. (1999) Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide. In Barciszewski,J. and Clark,B.F.C. (eds), *RNA Biochemistry and Biotechnology*. Kluwer Academic, Boston, MA, pp. 11–43.