

Nested PUF proteins: extending target RNA elements for gene regulation

Kouki Shinoda, Shogo Tsuji, Shiroh Futaki and Miki Imanishi ^{*[a]}

Abstract: RNA-binding proteins recognizing unique sequences within large transcriptomes serve as a powerful tool to control RNA metabolism. Pumilio and fem-3 mRNA-binding factor (PUF) proteins are considered good candidates for such tools because they are typically composed of eight highly homologous repeat segments and can be designed to recognize arbitrary 8-nucleotide (nt) RNA sequences. However, a specific 8-nt RNA sequence is found at multiple sites in various RNAs in the transcriptome, making it difficult to specifically target a single RNA. Designer PUF proteins recognizing longer RNA sequences would achieve more selective binding. Here, we propose an approach for creating 16-repeat PUFs capable of targeting a single, unique mRNA in the transcriptome. Our design is simple and involves either the tandem alignment of two PUF segments or the nesting of one PUF segment in another. Designed 16-repeat PUFs bound to the target RNA sequence without partial recognition derived from the original 8-repeat PUF. Furthermore, based on our strategy, the expression of an endogenous mRNA was selectively and effectively modulated, demonstrating the applicability of 16-repeat PUF proteins for regulating endogenous RNA metabolism.

Introduction

RNA-binding proteins play pivotal roles in regulating gene expression through multiple processes of RNA metabolism that include splicing, transport, translation, and turnover. To provide a better understanding of gene expression mechanisms at the RNA level, designer RNA-binding proteins with tailored sequence specificity have promise to become potent tools in combination with effector domains to manipulate metabolism of targeted RNAs.^[1] Many RNA-binding proteins have a modular structure, consisting of RNA-binding domains and effector domains.^[2] Such proteins should provide a promising platform for molecular tools with sequence-specific functions. However, the difficulty in changing RNA-binding specificity and the limitation in the number of nucleotides recognized by these proteins make it challenging to specify a single, unique binding sequence in a large transcriptome, thus hampering their practical application as tailored RNA modulators. Among various RNA-binding proteins, Pumilio and fem-3 mRNA-binding factor (PUF) proteins and

pentatricopeptide repeat (PPR) proteins provide attractive scaffolds for customized sequence-specific RNA-binding domains because of their modular structures consisting of single nucleotide-recognition units.^[1a, 3]

The RNA-binding domains of PUFs consist of eight highly homologous structural repeats (repeats 1 to 8), flanked by N- and C-terminal repeats (repeats N and C) (Figure 1A).^[4] Each of the eight PUF repeats binds to a single nucleotide, with three amino acids in each repeat responsible for nucleotide recognition. Efforts to establish consensus amino acid sequences in PUF repeats capable of recognizing each nucleobase led to the design of PUF proteins that can specifically target various 8-nucleotide (nt) RNA sequences.^[5] Modulation of cellular events has been

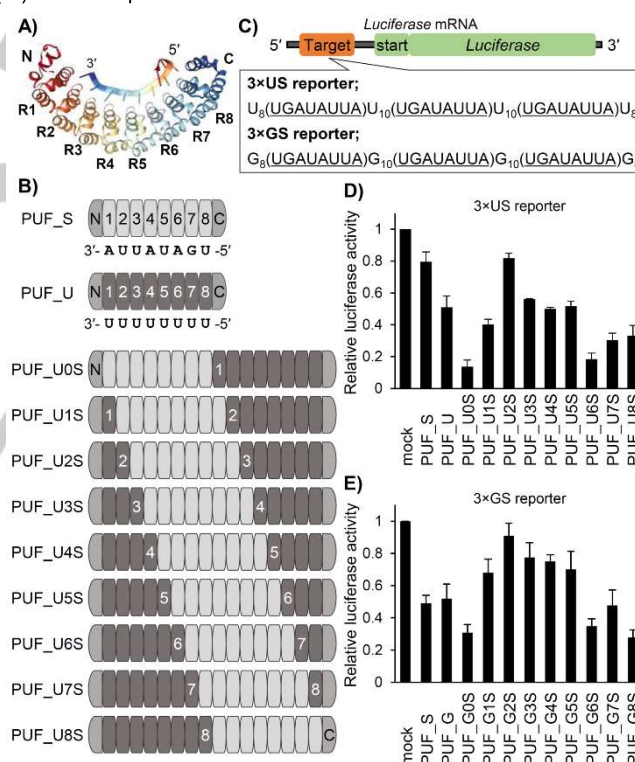


Figure 1. Design of 16-repeat PUFs and their RNA-binding abilities using luciferase reporter assays. (A) Crystal structure of Pumilio 1, a PUF protein, bound to RNA, adapted from a reference [4], PDB ID: 1M8Y. (B) Schematic diagrams of 8-repeat and 16-repeat PUFs. PUF_S and PUF_U are artificial 8-repeat PUFs with eight repeats flanked by N and C terminal regions (repeats N and C). The target RNA sequences of PUF_S and PUF_U are shown below the corresponding repeats. PUF_U0S to PUF_U8S are 16-repeat PUFs in which the eight-repeat of PUF_S have been inserted between two repeats of PUF_U. (C) Design of the luciferase reporter. The target site was inserted in the 5' UTR of the luciferase mRNA. The 3xUS and 3xGS reporters each contain three PUF-binding sites of the PUF_U_nSs and PUF_G_nSs, respectively. (D) and (E) Repressive effects of PUFs on luciferase reporter activity. Luciferase activities were standardized to that of mock (empty vector transfection) for each reporter. At least three replicates were performed. Data are the mean ± S.D.

[a] K. Shinoda, Dr. S. Tsuji, Prof. Dr. S. Futaki, Dr. M. Imanishi
Institute for Chemical Research, Kyoto University
Uji, Kyoto 611-0011 (Japan)
E-mail: imiki@scl.kyoto-u.ac.jp

Supporting information for this article is given via a link at the end of the document. [\(\(Please delete this text if not appropriate\)\)](#)

accomplished by combining these 8-repeat PUFs with various effector domains acting on RNA.^[5a, 6] The visualization of endogenous mRNAs has also been achieved by combining PUFs with fluorescent proteins.^[7] However, cognate binding sites of 8-repeat PUFs theoretically exist once every $4^8 (= 6.5 \times 10^4)$ bases. Considering the number of bases included in most transcriptomes, the potential risk of off-target effects may limit the applicability of these approaches.^[1a, 8] Therefore, the establishment of PUF proteins with the ability to recognize longer RNA sequences should have a considerable impact in this context.

Over the course of studies to develop PUF repeats to specifically recognize cytosine, Filipovska and coworkers designed an artificial PUF with 16 RNA-binding repeats, where a naturally occurring 8-repeat PUF segment was inserted between repeats 5 and 6 of the same PUF.^[5d] Successful recognition of the target 16-nt RNA sequence was demonstrated, although non-negligible binding of this protein was observed to a partially matching RNA sequence containing the target sequence of the original 8-repeat PUF. This pioneering work suggested the feasibility and practicability of the approach to combine two 8-repeat PUFs to expand the recognized RNA sequence. However, no further trials have been reported along this line. In this report, we designed a novel approach for the development of 16-repeat PUFs. Our approach was based on precise studies involving the nesting of an 8-repeat PUF inside another, establishing binding to the target 16-nt RNA sequences with no significant binding to the original 8-nt sequences. This approach was applied to the selective targeting and degradation of endogenous *vascular endothelial growth factor A (VEGFA)* mRNA, yielding a significant reduction in protein expression.

Results and Discussion

Our strategy for obtaining 16-repeat PUFs includes the use of two artificial 8-repeat PUFs, the binding of which has been well confirmed. One was an 8-repeat PUF named PUF_S, which was characterized as binding to 5'-UGAUUUA-3'.^[9] Another was PUF_U, which was designed to target the RNA sequence 5'-UUUUUUUU-3' (Figure 1B and Supplemental Figure S1). The eight repeat of PUF_S was inserted between each pair of repeats of PUF_U. Depending on the insertion position, the engineered 16-repeat PUFs were named PUF_{UnS} ($n = 0-8$), indicating that n repeats of PUF_U were placed at the N-terminal of the 8-repeat PUF_S segment (Figure 1B and Supplemental Table S1). Owing to the employment of PUF_U, of which every repeat recognizes the same nucleotide, the binding abilities of the nested PUFs were analyzed under the same condition using a single, common RNA sequence.

The potential for creating various 16-repeat PUFs by choosing the appropriate nesting positions was demonstrated through evaluation of the RNA targeting abilities of the constructed 16-repeat PUFs using a luciferase reporter assay. A target segment for all 16-repeat PUFs, 5'-UUUUUUUU UGAUUAUUU-3', was designed to contain the RNA-binding sequence of each PUF_{UnS} (Supplemental Table S1). A 3×US reporter containing the above PUF-binding sequence in the

5' untranslated region (UTR), immediately before the start codon, was employed (Figure 1C). The binding of the engineered PUFs to the target sites should prevent the access of ribosomes to the start codon and repress the expression of luciferase.^[6c, 10] The expression vectors for the 16-repeat PUF_{UnS}s and the 3×US luciferase reporter were simultaneously transfected into mammalian cells, and comparable expression levels of the PUF_{UnS} proteins were confirmed (Supplemental Figure S2A).

As shown in Figure 1D, all the 16-repeat PUF_{UnS}s yielded repression (20–85%) of luciferase expression to various extents compared with that of the mock control. Among them, PUF_{U0S} and PUF_{U6S} yielded the most notable effects (~80% repression).^[11] Moreover, similar patterns of repression were observed using an alternative set of 16-repeat PUFs, in which the 8-repeat from PUF_S was inserted into PUF_G (recognizing the 5'-GGGGGGGG-3' sequence), designated as PUF_{GnS} ($n = 0-8$) (Figure 1E and Supplemental Table S2, see also Supplemental Figure S2B for PUF_{GnS} protein expression levels). When analyzed using a 3×GS reporter, the most marked reductions (~70%) were obtained from PUF_{G0S}, PUF_{G6S}, and PUF_{G8S} (Figure 1E).^[11] These results suggest the promise of our nesting approach using PUF insertion positions of $n = 0, 6$, or 8 . Additionally, the obtained levels of repression by these 16-repeat PUFs ($n = 0, 6$, and 8 ; 70–85%) were higher than those of PUF_{U5S} (~50% repression)/PUF_{G5S} (~30% repression), which had the same nesting pattern ($n = 5$) employed by Filipovska and coworkers.^[5d] In our assay, the employment of PUF_U or PUF_G enabled the usage of a single, common reporter to evaluate the binding of each set of the nested PUFs without any concern for the differences in RNA sequences among reporters. This led to the finding of promising insertion positions. Considering that insertions at positions $n = 0$ and 8 represent substantively the same tandem alignment, further evaluation of our nesting approach was conducted using 16-repeat PUFs with insertion positions of $n = 0$ and 6 .

The abilities of the 16-repeat PUF_{U0S} and PUF_{U6S} to recognize the fully matching RNA sequences with no obvious binding to the partially matching RNA sequences were confirmed using electrophoretic mobility shift assay (EMSA). In this assay, the binding of PUF proteins reduces the mobility of the target RNAs, enabling evaluation of the RNA-binding abilities of proteins. Histidine-tagged PUF_{U0S} and PUF_{U6S} proteins, together with the histidine-tagged 8-repeat PUF_S as a control, were expressed in *Escherichia coli* and purified. In marked contrast to the significant binding of PUF_{U0S} and PUF_{U6S} to a fully matching RNA segment [RNA(US)], yielding K_d of ~53 and 60 nM, respectively, no substantive binding of these proteins to the partially matching RNA(CS) sequence was observed (Figure 2A and Supplemental Table S3). Similar binding specificities of PUF_{G0S} and PUF_{G6S} to RNA(GS) over RNA(CS) were also observed (Supplemental Figure S3A). In contrast, as expected, PUF_S showed comparable binding to all RNA segments containing the 8-nt target sequence of PUF_S (Figure 2B and Supplemental Figure S3B).^[12] It is noteworthy that the examined 16-repeat PUFs, in which two 8-repeat PUFs (PUF_U/PUF_S or PUF_G/PUF_S) were simply combined, specifically bound to the fully matching 16-nt RNA without recognizing the partially

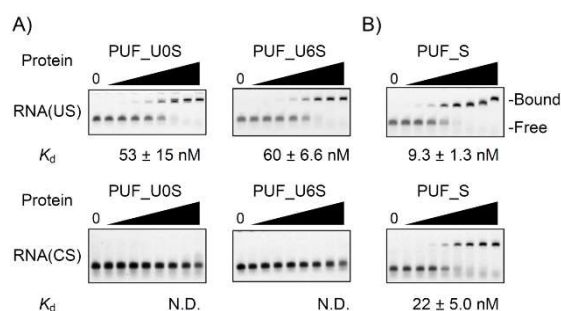


Figure 2. EMSA of PUF_U0S, PUF_U6S and PUF_S to fully matching RNA or partially matching RNA. Representative results for PUF_U0S, PUF_U6S (A) and PUF_S (B) are shown. Upper and lower panels indicate the binding of PUF proteins to RNA(US) and RNA(CS), respectively. Lane 1: protein-free; lanes 2–9: 0.46–1000 nM of PUF_U0S, PUF_U6S or PUF_S; RNA(US): 5'-CCUUUUUUUUUGAUUUUUUUUUUUCC-3'; RNA(CS): 5'-UUCUUUUUUUUUGAUUUUUUUUUUUCC-3' (the target of PUF_S is underlined); K_d indicates the mean ± S.D. (n = 3); N.D.: not determined due to low binding.

matching 8-nt sequence bound by PUF_S. In these 16-repeat PUFs, the 8-repeat derived from PUF_S was lacking in both the additional N- and C-terminal repeats. This may result in almost no binding to the partially matching 8-nt target sequence. These results further support the fidelity of our nesting approach at the $n = 0$ and 6 positions.

Specific control of endogenous mRNA metabolism using tailored RNA-binding proteins is one of the ultimate goals in this field.^[1] Sixteen-repeat PUFs fused with tristetraprolin (TTP), an mRNA decay factor, were thus designed to target the 3' UTR of endogenous VEGFA mRNA, demonstrating effective repression of VEGFA expression (Figure 3). Abil and coworkers reported an 8-repeat PUF (denoted as PUF_V in this report) targeting the 5'-UACAGAU-3' sequence in the 3' UTR of VEGFA mRNA.^[9] Using PUF_V as the core segment, two 16-repeat PUFs, PUF_T (tandem-type) and PUF_I (insertion-type), were designed (Figure 3A and Supplemental Table S4). In PUF_T, an 8-repeat PUF targeting the 8-nt RNA sequence (5'-UAUCUUAA-3') on the 3' side of the PUF_V target site was inserted into the $n = 0$ site of PUF_V. In PUF_I, the PUF_V segment was inserted at the $n = 6$ position of an 8-repeat PUF protein targeting the corresponding 8-nt RNA sequence (5'-AAUAUCUU-3'). Additionally, fusion proteins of PUF with TTP (TTP-PUFs), TTP-PUF_V, TTP-PUF_T and TTP-PUF_I, were constructed and expressed in mammalian cells based on a previous paper.^[9] TTP is a post-transcriptional regulator that promotes mRNA degradation or AU-rich element-mediated translation repression.^[13] Following the binding of PUF to the target mRNA, TTP-PUF is expected to recruit complexes that destabilize mRNA or repress translation, downregulating gene expression.

Treatment with CoCl₂, which is known to mimic hypoxia and induce VEGFA expression,^[14] achieved a ~5-fold VEGFA induction. Although the production of VEGFA was not influenced by the expression of the 8-repeat TTP-PUF_V in accordance with a previous paper,^[9] both the 16-repeat TTP-PUF_V and TTP-PUF_I significantly repressed VEGFA production (Figure 3B). Additionally, no significant effect on VEGFA expression was observed following expression of the TTP-PUF_U0S or TTP-

PUF_U6S fusion proteins, which possess no cognate binding sequences in VEGFA mRNA (Supplemental Figure S4A). Western blotting showed that all the examined TTP-PUF fusion proteins were properly expressed in the cells (Supplemental Figures S4B, C).

Interestingly, although the 8-repeat TTP-PUF_V yielded no significant repression of VEGFA production, EMSAs showed that PUF_V, PUF_T, and PUF_I bound to RNA(VEGFA), which contained their target sequences, with similar dissociation constants (Supplemental Figure S5). The inconsistent results may be attributed to the limitation in the targeting specificity of the 8-repeat TTP-PUF_V, yielding off-target-like binding to sites other than the desired VEGFA mRNA site. In fact, according to an RNA sequence homology search of the human transcriptome using GGRNA, a transcript-oriented search engine for genes and transcripts,^[15] the corresponding sequences for the 16-repeat PUF_T and PUF_I are found only in the 3' UTR of VEGFA mRNA, whereas that for the 8-repeat PUF_V is found at >100 sites in the mRNA sequences for various genes in addition to VEGFA. The binding selectivity of the designed PUF proteins to RNA(VEGFA) was thus examined in the presence of increasing concentrations of competitive RNA containing the target sequence of PUF_V (5'-N₄-UACAGAU-N₁₀-3', where N = A, C, G, or U). Although the shifted band of the 8-repeat PUF_V completely disappeared following a 200-fold addition of the competitor, the same treatment yielded no significant effect on the binding of the 16-repeat PUFs

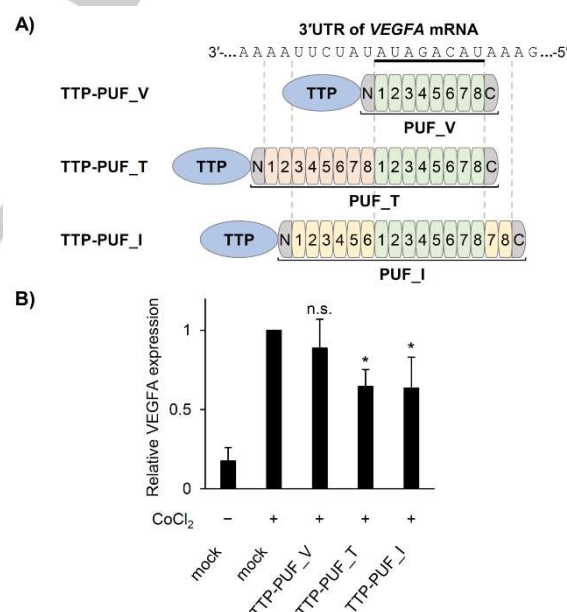


Figure 3. Design of TTP-fused PUFs (TTP-PUFs) targeting the 3' UTR of VEGFA mRNA and repression of endogenous VEGFA by 16-repeat TTP-PUFs. (A) Schematic representation of PUF_V, PUF_T, and PUF_I and their target sequences in the 3' UTR of VEGFA mRNA. (B) Inhibition of hypoxia-induced VEGFA expression in cells transfected with engineered TTP-PUF expression vectors or the empty vector (mock). VEGFA was induced by treatment with CoCl₂ 24 h after transfection, and cells were incubated for an additional 24 h. Secreted VEGFA levels were measured by ELISA. VEGFA expression levels were normalized to the total amount of protein and compared with mock treatment with CoCl₂. Data are the mean ± S.D.; *p < 0.05 (n ≥ 3, t-test).

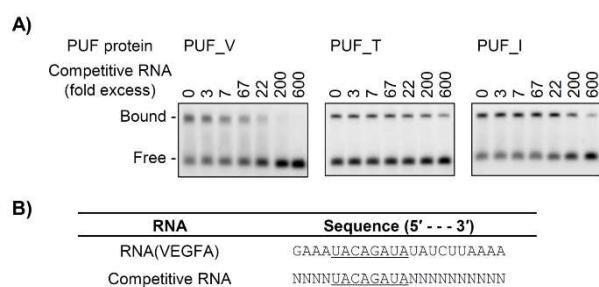


Figure 4. Competitive RNA-binding assay of PUFs targeting *VEGFA* mRNA. (A) Representative results are shown. Competition for the binding of PUFs to fluorescently labeled RNA(VEGFA) was achieved with unlabeled competitive RNA. Fold excesses of competitive RNA are shown above each lane. (B) Sequences of RNA oligos are shown in the table. N = A, U, G, or C. The target sequences of PUF_V are underlined.

(Figure 4). The robust binding of TTP-PUF_T and TTP-PUF_I compared with TTP-PUF_V to the target site was confirmed.

In this study, we created 16-repeat PUFs capable of targeting a single, unique mRNA. The tandem linking of 8-repeat PUFs or the insertion of an 8-repeat PUF between repeats 6 and 7 of another PUF results in successful binding to 16-nt target RNA sequences. In the case of the tandem-linking, two sets of serial 8 repeats are directly connected. The structural stability owing to the undisrupted original alignment may have given rise to the successful construction. Regarding insertion-type 16-repeat PUFs, an 8-repeat PUF is disrupted between repeats 6 and 7 by insertion of another 8-repeat PUF. Adamala *et al.* concatenated single PUF repeats in chain to target desired RNA sequences and showed that repeat 6 can be a template for the canonical modules.^[16] Therefore, repeat 6 may have high structural tolerance for the neighboring repeat, which leads to preferable binding activity of insertion-type 16-repeat PUF.

Conclusions

In summary, we proposed a nesting approach of two 8-repeat PUFs to extend target RNA elements for specific gene regulation. Designed 16-repeat PUFs bound to the long RNA sequences without significant binding to partially matching sequences. Importantly, when fused with TTP, 16-repeat PUFs targeting endogenous *VEGFA* mRNA repressed the production of VEGFA, demonstrating the potential for regulating single-gene expression by the nested PUFs. Although further investigations for the detailed RNA recognition modes are needed, the simple design principle of our approach should ensure easy applicability. We believe that the availability of this strategy for targeting unique RNA sites in the transcriptome should enable powerful techniques in RNA biology and therapeutic applications, thus having a considerable impact on related fields of research.

Experimental Section

Vector construction: PUF-Halo-HA expression vectors for luciferase reporter assay were constructed in a stepwise manner. A receiving vector

available for the Golden Gate Assembly^[9] was constructed as follows. Two kinds of DNA fragments including the flanking N- and C-terminal regions of PUF and the Halo-HA tag coding sequence were amplified by PCR from pET28-GG-PUF in PUF Assembly Kit (Addgene #1000000051) and pFN19 (Promega), respectively, using primer pairs shown in Supplemental Table S5. These fragments were concatenated to pCI vector (Promega) at NheI and NotI sites using Gibson Assembly (NEB), then the BsaI site derived from pCI was removed by point mutation, resulting in pCI-PUF(N-C)-Halo-HA. Expression vectors of 8-repeat PUFs were constructed by PUF Assembly Kit using pCI-PUF(N-C)-Halo-HA as a receiving vector. To create expression vectors of 16-repeat PUFs, the fragments coding the inserted 8-repeat PUFs and the corresponding vector regions were amplified by PCR with each pair of primers (Supplemental Table S6). Subsequently, these fragments were concatenated using Gibson Assembly.

Luciferase reporter vectors were constructed as follows. The DNA sequence including EcoRI and PvuI sites was inserted just before start codon of firefly luciferase *luc2* gene of pmirGLO (Promega). 3×US and 3×GS reporters were constructed by inserting the fragment including each target sequence (Supplemental Table S7) into EcoRI and PvuI sites of the mutated pmirGLO.

PUF expression vectors in *E. coli* were constructed using the same method as the construction of PUF-Halo-HA except using pET28-GG-PUF from PUF Assembly Kit as the receiving vector.

TTP-PUF expression vectors were constructed using the same method as above except using pCMV-TTP(C147R)-GG-PUF from PUF Assembly Kit as the receiving vector.

The amino acids pairs interacting with RNA bases in each repeat and the target RNA sequences of all constructed PUFs are shown in Supplemental Tables S1, S2 and S4.

Luciferase reporter assay: 2×10⁴ of NIH3T3 cells were seeded in 48-well plates day before transfection. Transfection was performed with Lipofectamine 3000 (Thermo Fisher Scientific). Transfection mixtures contained 120 ng PUF-Halo-HA expression vector, 80 ng pCI vector and 1 ng reporter vector. For mock transfection, pCI vector was used instead of PUF-Halo-HA expression vector. 24 h after transfection, cells were lysed by Passive Lysis Buffer (Promega). Firefly luciferase and Renilla luciferase activities were measured using Dual-Luciferase Reporter Assay System (Promega) with measurements taken on GloMax-Multi Detection System (Promega). Firefly luciferase activity was normalized by Renilla luciferase activity. At least three biological replicates were performed.

Purification of PUF proteins: PUF proteins were expressed in *E. coli* BL21 (DE3) (Nippon gene). The bacterial cells were lysed by sonication in buffer A (20 mM NaH₂PO₄, 500 mM NaCl and 40 mM imidazole, pH 7.4). The lysates were centrifuged and the supernatants were purified by Ni-NTA HisTrap FF column (GE Healthcare) chromatography using buffer A and buffer B (20 mM NaH₂PO₄, 500 mM NaCl and 250 mM imidazole, pH 7.4). Fractions were concentrated using AmiconUltra-0.5 (Merck Millipore) and transferred to buffer C (20 mM HEPES, pH 7.4, 100 mM KCl, 1 mM EDTA and 1 mM DTT). The purity of PUF proteins was confirmed by SDS-PAGE and coomassie blue staining. 50% (v/v) glycerol was added to purified proteins and stored at -80°C for further analysis (EMSA and competitive binding assay).

Electrophoretic Mobility Shift Assay (EMSA): RNA oligos labeled with 5,6-FAM at the 5'-end were obtained from FASMAC. The sequences of the RNA oligos for EMSA are shown in Supplemental Table S3. 5 nM labeled RNA and three-fold serially diluted concentrations of purified PUF

proteins (0–1 μM) were mixed in the reaction buffer containing 10 mM HEPES, pH 7.4, 50 mM KCl, 0.5 mM EDTA, 2 mM DTT, 0.2 U/ μL RNasin Plus RNase Inhibitor (Promega), 0.1 mg/mL BSA, 0.1 mg/mL Yeast tRNA (Ambion), 0.01% Tween20, 2% Ficoll. The mixture was incubated for 1 h at 25°C. Free RNA and bound RNA were separated by gel electrophoresis using 5% non-denaturing polyacrylamide gels in 0.5 $\times$ TBE at 4°C. After electrophoresis, fluorescently labeled RNAs were detected using Typhoon FLA 9000 (GE Healthcare). The peak intensity for bound and free labeled RNA was measured using ImageJ. The fractions of bound RNA were plotted against the protein concentrations. For calculating the equilibrium dissociation constants (K_d), the plotted data were fit to the following equation using Kaleida Graph:

$$\theta = \frac{([P] + [R] + K_d) - \sqrt{([P] + [R] + K_d)^2 - 4[P][R]}}{2[R]}$$

, where [P] is the total protein concentration, [R] is the total RNA concentration, and θ is the fraction of RNA bound to protein. The mean and standard deviation are obtained from three independent experiments.

Western blotting assay: Cells were lysed by RIPA buffer, and cell lysates were subjected to SDS-PAGE. The proteins were detected using rat anti-HA (Roche; 3F10), mouse anti-FLAG antibody (Sigma Aldrich; F3165) or mouse anti- β -actin antibodies (Sigma Aldrich; AC-74) as primary antibodies, and anti-rat IgG (Beckman Coulter; 732664) or anti-mouse IgG (Amersham; NA931) antibody as HRP-labeled secondary antibodies. The bands were visualized using Amersham ECL Prime (GE Healthcare).

VEGFA repression assay: 2×10^5 of Flp-In 293 (Thermo Fisher Scientific) cells were seeded in 24-well plates one day before transfection. 500 ng TTP-PUF expression vectors or pCI were transfected using Lipofectamine 3000. 24 h after transfection, the medium was changed to fresh DMEM (+10% FBS) containing 500 μM CoCl₂. The supernatant was collected 24 h after medium exchange. VEGFA expression levels were measured by ELISA following the instruction of human VEGF DuoSet kit (R&D Systems). The cells were lysed by RIPA buffer. After centrifugation, the total protein concentrations were measured by Pierce BCA Protein Assay Kit (Thermo Fisher Scientific). VEGFA expression levels were normalized by total protein concentrations. At least three biological replicates were performed.

Competitive binding assay: Competitive RNA (5'-NNNNUACAGAUANNNNNNNNN-3') was obtained from FASMAC. Protein concentrations used in the competition assay were 120, 120 and 200 nM of PUF_V, PUF_T and PUF_I, respectively, which gave the < 50% of the protein-bound labeled RNA. Each purified PUF protein was mixed with 5 nM labeled RNA(VEGFA) (Supplemental Table S3) and varying concentrations of non-labeled competitive RNA (0 to 3 μM) in the same reaction buffer for EMSA. After incubation for 1 h at 25°C, the mixtures were loaded on 5% non-denaturing polyacrylamide gels and separated. Finally, the gels were visualized using Typhoon FLA 9000.

RNA sequence homology search: The numbers of transcripts that contain the sequence perfectly matched to the target sequence of PUF_V, PUF_T or PUF_I were obtained using GGRNA (<http://GGRNA.dbcls.jp/>). In this search, we considered transcript variants of the same gene as one gene and excluded the uncharacterized transcript.

Acknowledgements

We kindly thank Huimin Zhao for plasmids to construct PUFs. This work was supported by JST/CREST (JPMJCR14W3) and partially by JSPS KAKENHI (16H03281) (M.I.).

Keywords: RNA recognition • protein engineering • gene expression • gene technology

- [1] a) Z. Abil, H. Zhao, *Mol Biosyst* **2015**, *11*, 2658–2665; b) J. P. Mackay, J. Font, D. J. Segal, *Nat Struct Mol Biol* **2011**, *18*, 256–261; c) Y. Chen, G. Varani, *FEBS J* **2013**, *280*, 3734–3754.
- [2] B. M. Lunde, C. Moore, G. Varani, *Nat Rev Mol Cell Biol* **2007**, *8*, 479–490.
- [3] T. M. Hall, *Curr Opin Struct Biol* **2016**, *36*, 116–121.
- [4] X. Wang, P. D. Zamore, T. M. Hall, *Mol Cell* **2001**, *7*, 855–865.
- [5] a) S. Dong, Y. Wang, C. Cassidy-Amstutz, G. Lu, R. Bigler, M. R. Jezyk, C. Li, T. M. Hall, Z. Wang, *J Biol Chem* **2011**, *286*, 26732–26742; b) X. Wang, J. McLachlan, P. D. Zamore, T. M. Hall, *Cell* **2002**, *110*, 501–512; c) C. G. Cheong, T. M. Hall, *Proc Natl Acad Sci U S A* **2006**, *103*, 13635–13639; d) A. Filipovska, M. F. Razif, K. K. Nygard, O. Rackham, *Nat Chem Biol* **2011**, *7*, 425–427; e) Z. T. Campbell, C. T. Valley, M. Wickens, *Nat Struct Mol Biol* **2014**, *21*, 732–738.
- [6] a) Y. Wang, C. G. Cheong, T. M. Hall, Z. Wang, *Nat Methods* **2009**, *6*, 825–830; b) A. Cooke, A. Prigge, L. Opperman, M. Wickens, *Proc Natl Acad Sci U S A* **2011**, *108*, 15870–15875; c) J. Cao, M. Arha, C. Sudrik, D. V. Schaffer, R. S. Kane, *Angew Chem Int Ed Engl* **2014**, *53*, 4900–4904; d) R. Choudhury, Y. S. Tsai, D. Dominguez, Y. Wang, Z. Wang, *Nat Commun* **2012**, *3*, 1147.
- [7] a) H. Yoshimura, A. Inaguma, T. Yamada, T. Ozawa, *ACS Chem Biol* **2012**, *7*, 999–1005; b) T. Yamada, H. Yoshimura, A. Inaguma, T. Ozawa, *Anal Chem* **2011**, *83*, 5708–5714; c) T. Ozawa, Y. Umezawa, *Methods Mol Biol* **2011**, *714*, 175–188; d) T. Ozawa, Y. Natori, M. Sato, Y. Umezawa, *Nat Methods* **2007**, *4*, 413–419; e) T. Yamada, H. Yoshimura, R. Shimada, M. Hattori, M. Eguchi, T. K. Fujiwara, A. Kusumi, T. Ozawa, *Sci Rep* **2016**, *6*, 38910.
- [8] Y. Wang, Z. Wang, T. M. Hall, *FEBS J* **2013**, *280*, 3755–3767.
- [9] Z. Abil, C. A. Denard, H. Zhao, *J Biol Eng* **2014**, *8*, 7.
- [10] J. Cao, M. Arha, C. Sudrik, A. Mukherjee, X. Wu, R. S. Kane, *Nucleic Acids Res* **2015**, *43*, 4353–4362.
- [11] PUF_C, which does not have the binding site in both 3 $\times$ US and 3 $\times$ GS reporter mRNA, showed no significant effects on luciferase expression, suggesting that 16-repeat PUFs specifically bound to each reporter (Supplemental Figure S6).
- [12] K_d values of PUF_S were higher than those previously reported in a fluorescence polarization assay (0.59 $\pm$ 0.10 nM).^[9] This might be because of differences in the detection method. Furthermore, we added tRNA to the reaction buffer to reduce sequence-independent interactions, which could have resulted in the higher K_d values. A slight increase of band shift at high concentration is maybe because of weak non-specific interactions of excess PUF_S.
- [13] a) S. A. Brooks, P. J. Blackshear, *Biochim Biophys Acta* **2013**, *1829*, 666–679; b) M. Y. Qi, Z. Z. Wang, Z. Zhang, Q. Shao, A. Zeng, X. Q. Li, W. Q. Li, C. Wang, F. J. Tian, Q. Li, J. Zou, Y. W. Qin, G. Brewer, S. Huang, Q. Jing, *Mol Cell Biol* **2012**, *32*, 913–928; c) T. M. Franks, J. Lykke-Andersen, *Genes Dev* **2007**, *21*, 719–735.
- [14] a) C. Osera, J. L. Martindale, M. Amadio, J. Kim, X. Yang, C. A. Moad, F. E. Indig, S. Govoni, K. Abdelmohsen, M. Gorospe, A. Pascale, *RNA Biol* **2015**, *12*, 1121–1130; b) J. P. Piret, D. Mottet, M. Raes, C. Michiels, *Ann N Y Acad Sci* **2002**, *973*, 443–447.
- [15] Y. Naito, H. Bono, *Nucleic Acids Res* **2012**, *40*, W592–596.
- [16] K. P. Adamala, D. A. Martin-Alarcon, E. S. Boyden, *Proc Natl Acad Sci U S A* **2016**, *113*, E2579–E2588.

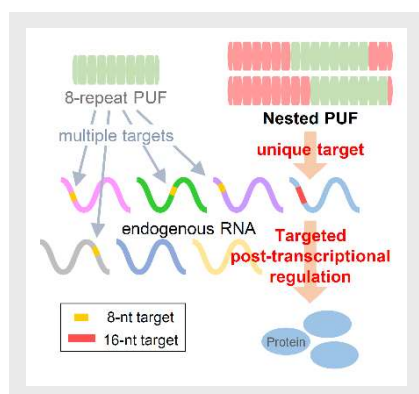
FULL PAPER

Entry for the Table of Contents (Please choose one layout)

Layout 1:

FULL PAPER

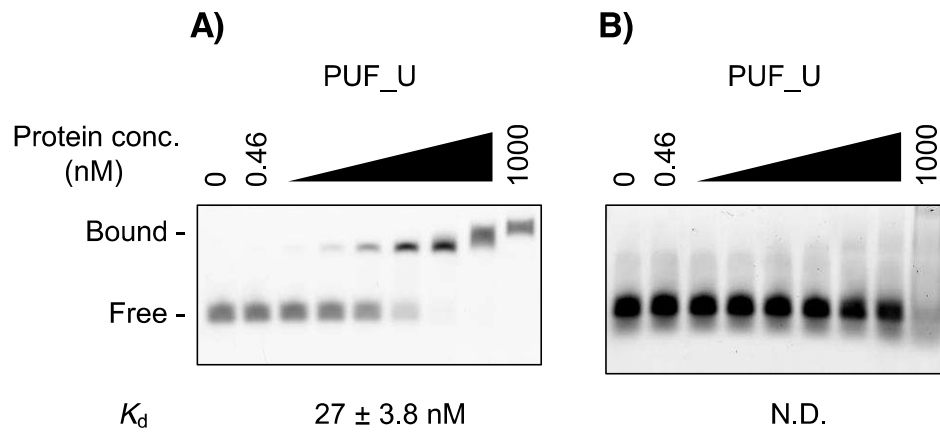
RNA-binding proteins recognizing unique sequences within large transcriptomes serve as a powerful method to control RNA metabolism. Here, tailored RNA-binding proteins, nested PUFs, were engineered by uniting two 8-repeat PUFs. Using 16-repeat PUFs, an endogenous RNA metabolic process was successfully controlled.



Kouki Shinoda, Shogo Tsuji, Shiroh Futaki and Miki Imanishi*

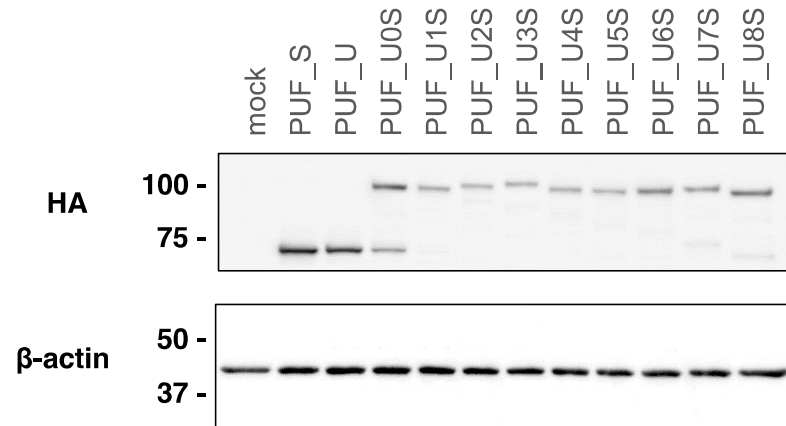
Page No. – Page No.

Nested PUF proteins: extending target RNA elements for gene regulation

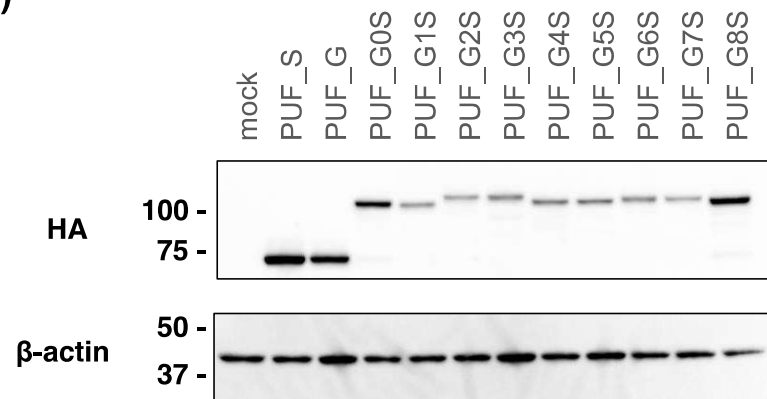


Supplemental Figure S1. Specific binding of PUF_U to the target RNA. (A) A representative EMSA result of PUF_U with the RNA fragment containing target sequence is shown. RNA sequence: 5'-UUGGUUUUUUUUUGGACAUGUAUAUAGU-3' (the target sequence of PUF_U is underlined). (B) A representative EMSA result of PUF_U with the RNA fragment containing non-target sequence is shown. RNA sequence: 5'-UUCCCCCCCCCUGAUAUUACCCCCCCCCUU-3'. K_d indicates the mean $\pm$ S.D. ($n = 3$); N.D.: not determined due to low binding.

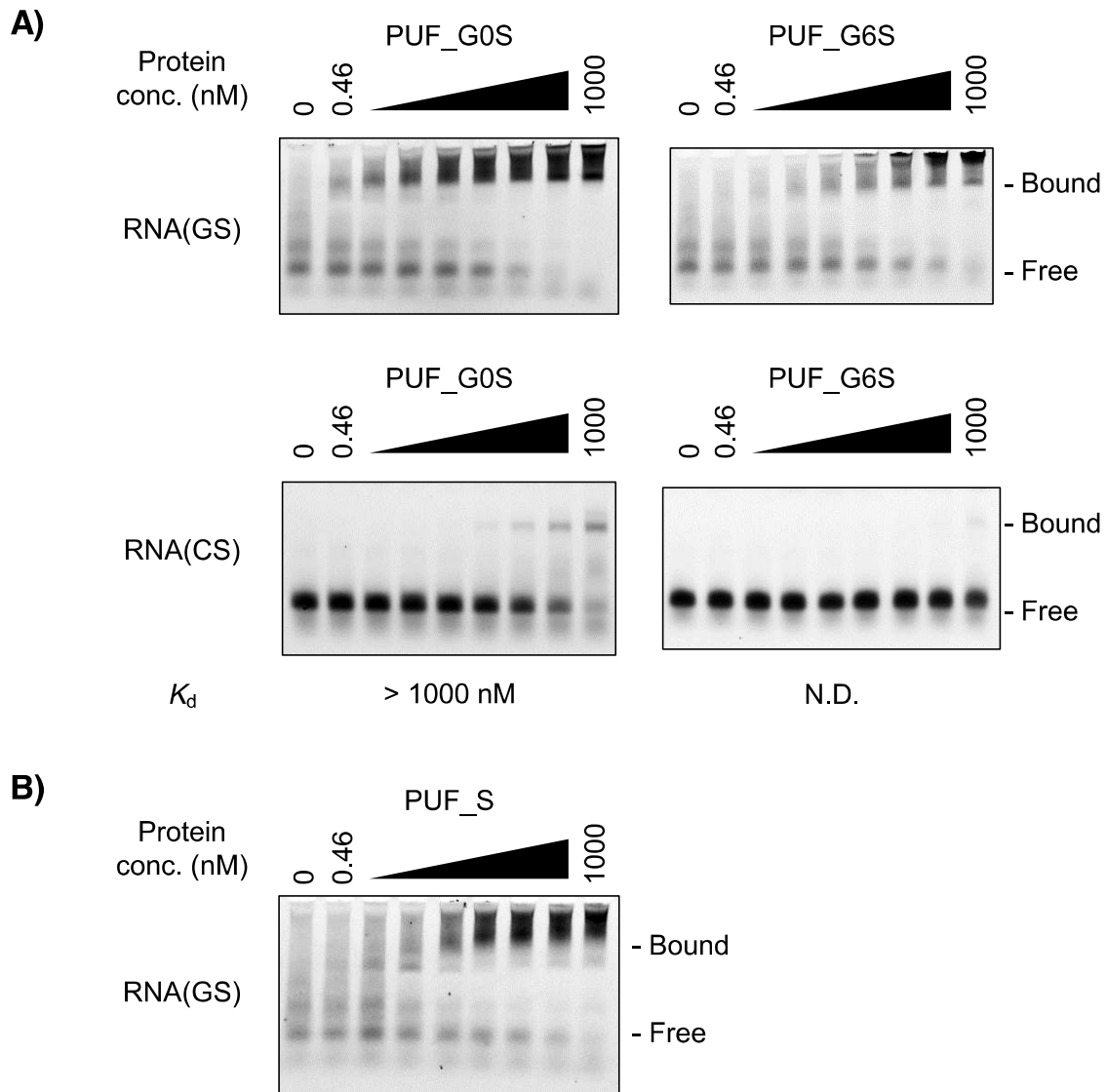
A)



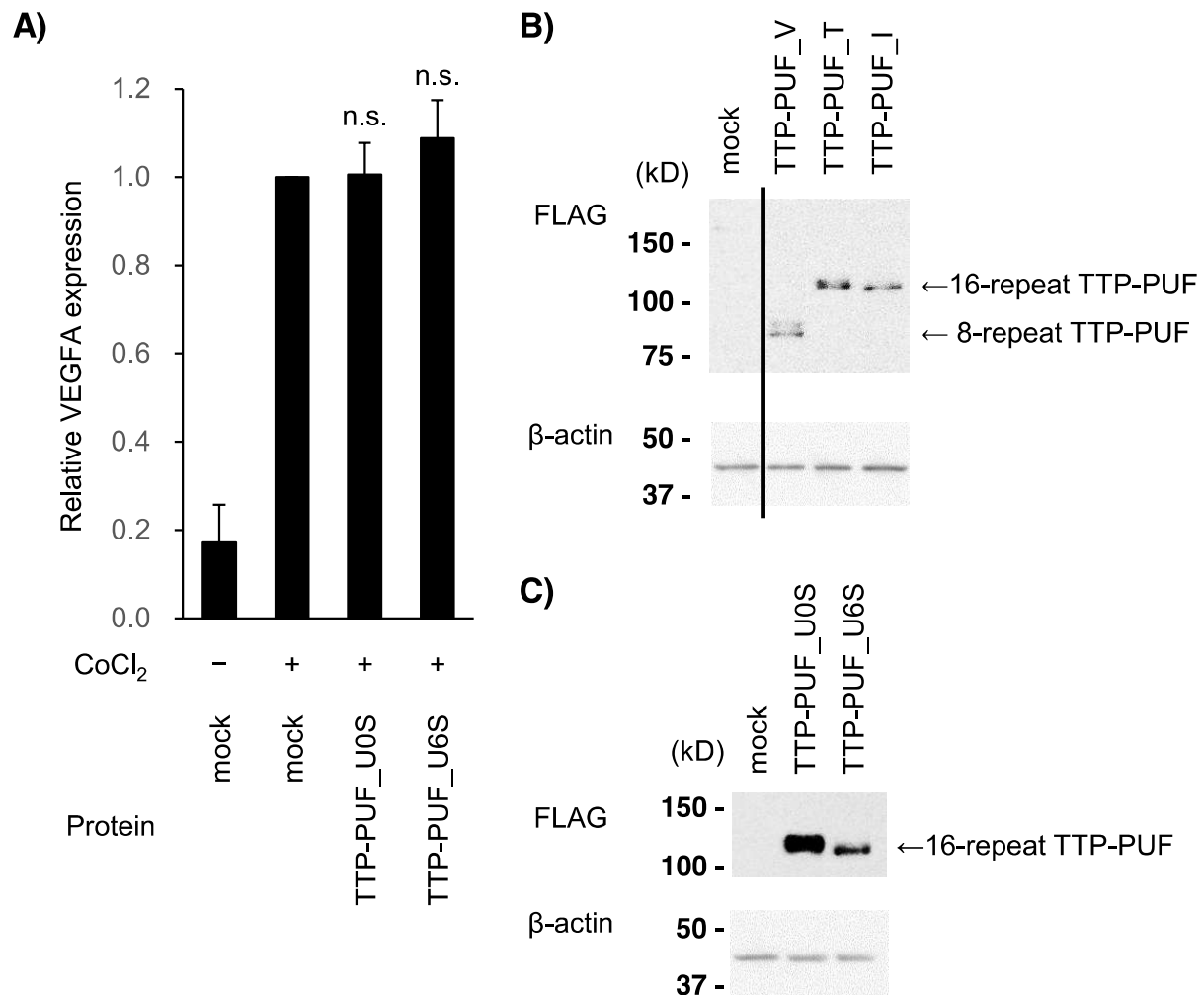
B)



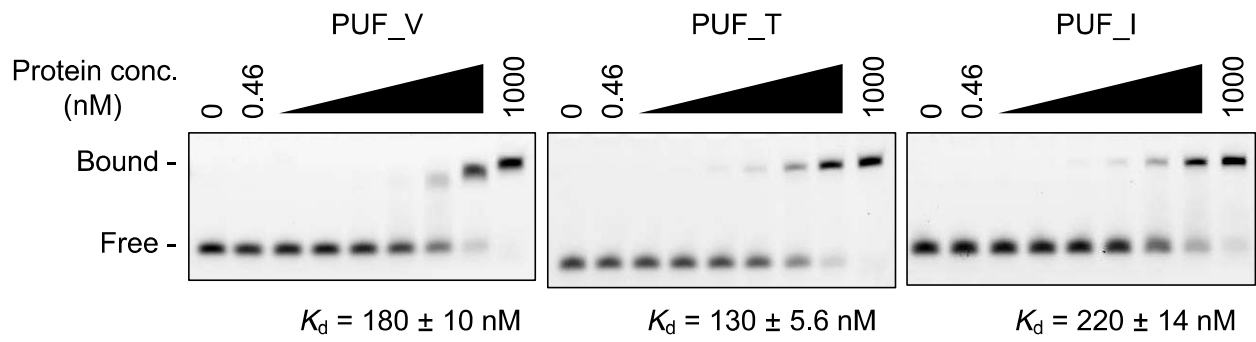
Supplemental Figure S2. Expression of PUFs in NIH3T3 cells. (A) PUF_S, PUF_U, PUF_U0S to PUF_U8S and (B) PUF_S, PUF_G, PUF_G0S to PUF_G8S were analyzed by western blotting assay.



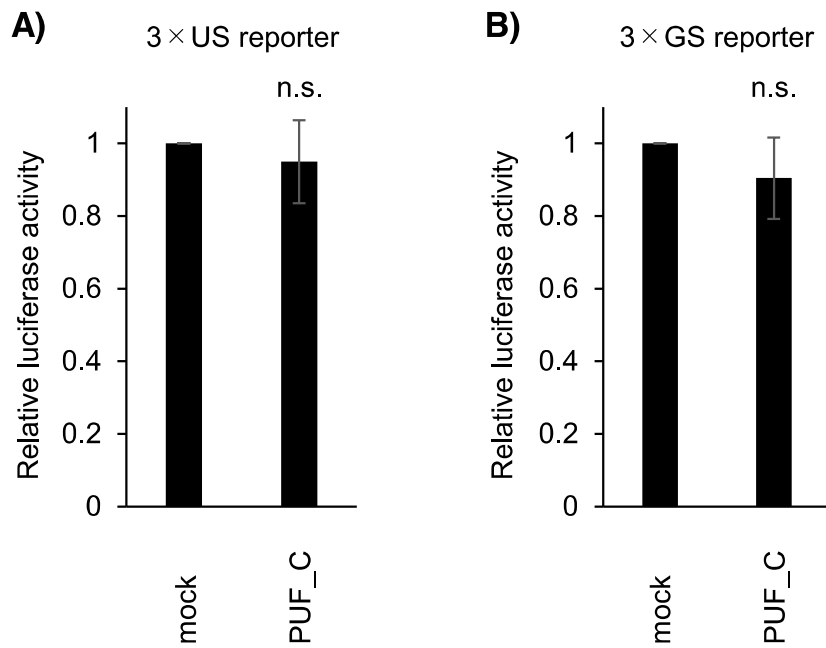
Supplemental Figure S3. EMSA of PUF_G0S, PUF_G6S and PUF_S to fully matching RNA or partially matching RNA. (A) Representative results of PUF_G0S and PUF_G6S are shown. Upper and lower panels indicate the binding of PUF proteins to RNA(GS) and RNA(CS), respectively. (B) A representative result of PUF_S with RNA (GS) is shown. Lane 1: protein free; lanes 2–9: 0.46–1000 nM of PUF_G0S, PUF_G6S or PUF_S; RNA(GS): 5'- UUGGGGGGGGUGAUUAUUAGGGGGGGGUU-3' (the target of PUF_S is underlined). K_d indicates the mean $\pm$ S.D. ($n = 3$); N.D.: not determined due to low binding. K_d for binding to RNA(GS) was not calculated because of the smear band patterns of RNA(GS), which probably resulted from structured RNA containing the G-rich sequence.



Supplemental Figure S4. Effects on VEGFA expression by TTP-PUFs having no binding target. (A) No effects on VEGFA expression by TTP-PUF_U0S or U6S, that do not have any target sequences in 3'UTR of *VEGFA* mRNA. n.s.: not significant. Data are the mean $\pm$ S.D. (n = 3). (B) and (C) Confirmation of TTP-PUF expression by western blotting assay.



Supplemental Figure S5. EMSA of PUF domains, PUF_V, PUF_T and PUF_I, targeting *VEGFA* mRNA to RNA(VEGFA). Representative results are shown. K_d indicates the mean $\pm$ S.D. (n = 3).



Supplemental Figure S6. Repressive effects of non-target PUF on luciferase reporter activity. No significant effects were observed on luciferase activities of (A) 3 × US reporter and (B) 3 × GS reporter by PUF_C. The target RNA sequence of PUF_C is 5'-CCCCCCCC -3', which is not included in each reporter mRNA. Luciferase activities were standardized to that of mock (empty vector transfection) for each reporter. At least three replicates were performed. n.s.: not significant. Data are the mean $\pm$ S.D.

Supplemental Table S1.

The amino acids interacting with RNA bases and target RNA sequences are shown. The amino acid pairs and the recognition bases of PUF_S are shown in red. Nt and Ct mean N-terminal and C-terminal, respectively.

Protein	Amino acids pair in repeats		Target RNA sequence
	Ct	Nt	5'-----3'
PUF_S	8NQ-7SE-6CQ-5NQ-4CQ-3NQ-2NQ-1CQ		UGAUAUUA
PUF_U	8NQ-7NQ-6NQ-5NQ-4NQ-3NQ-2NQ-1NQ		UUUUUUUU
PUF_U0S	8NQ-7NQ-6NQ-5NQ-4NQ-3NQ-2NQ-1NQ-8NQ-7SE-6CQ-5NQ-4CQ-3NQ-2NQ-1CQ		UUUUUUUUUGAUAUUA
PUF_U1S	8NQ-7NQ-6NQ-5NQ-4NQ-3NQ-2NQ-8NQ-7SE-6CQ-5NQ-4CQ-3NQ-2NQ-1CQ-1NQ		UUUUUUUUUGAUUUUAU
PUF_U2S	8NQ-7NQ-6NQ-5NQ-4NQ-3NQ-8NQ-7SE-6CQ-5NQ-4CQ-3NQ-2NQ-1CQ-2NQ-1NQ		UUUUUUUGAUUUUAUU
PUF_U3S	8NQ-7NQ-6NQ-5NQ-4NQ-8NQ-7SE-6CQ-5NQ-4CQ-3NQ-2NQ-1CQ-3NQ-2NQ-1NQ		UUUUUUGAUUUUAUUU
PUF_U4S	8NQ-7NQ-6NQ-5NQ-8NQ-7SE-6CQ-5NQ-4CQ-3NQ-2NQ-1CQ-4NQ-3NQ-2NQ-1NQ		UUUUUGAUUUUAUUUU
PUF_U5S	8NQ-7NQ-6NQ-8NQ-7SE-6CQ-5NQ-4CQ-3NQ-2NQ-1CQ-5NQ-4NQ-3NQ-2NQ-1NQ		UUUUGAUUUUAUUUUU
PUF_U6S	8NQ-7NQ-8NQ-7SE-6CQ-5NQ-4CQ-3NQ-2NQ-1CQ-6NQ-5NQ-4NQ-3NQ-2NQ-1NQ		UUUGAUUUUAUUUUUU
PUF_U7S	8NQ-8NQ-7SE-6CQ-5NQ-4CQ-3NQ-2NQ-1CQ-7NQ-6NQ-5NQ-4NQ-3NQ-2NQ-1NQ		UUGAUUUUAUUUUUUU
PUF_U8S	8NQ-7SE-6CQ-5NQ-4CQ-3NQ-2NQ-1CQ-8NQ-7NQ-6NQ-5NQ-4NQ-3NQ-2NQ-1NQ		UGAUUUUAUUUUUUUU

Supplemental Table S2.

The amino acids interacting with RNA bases and target RNA sequences are shown. The amino acid pairs and the recognition bases of PUF_S are shown in red. Nt and Ct mean N-terminal and C-terminal, respectively.

Protein	Amino acids pair in repeats		Target RNA sequence
	Ct	Nt	5'-----3'
PUF_S	8NQ-7SE-6CQ-5NQ-4CQ-3NQ-2NQ-1CQ		UGAUAUUA
PUF_G	8SE-7SE-6SE-5SE-4SE-3SE-2SE-1SE		GGGGGGGG
PUF_G0S	8SE-7SE-6SE-5SE-4SE-3SE-2SE-1SE-8NQ-7SE-6CQ-5NQ-4CQ-3NQ-2NQ-1CQ		GGGGGGGGUGAUAUUA
PUF_G1S	8SE-7SE-6SE-5SE-4SE-3SE-2SE-8NQ-7SE-6CQ-5NQ-4CQ-3NQ-2NQ-1CQ-1SE		GGGGGGGGUGAUUUAG
PUF_G2S	8SE-7SE-6SE-5SE-4SE-3SE-8NQ-7SE-6CQ-5NQ-4CQ-3NQ-2NQ-1CQ-2SE-1SE		GGGGGGUGAUUUAGG
PUF_G3S	8SE-7SE-6SE-5SE-4SE-8NQ-7SE-6CQ-5NQ-4CQ-3NQ-2NQ-1CQ-3SE-2SE-1SE		GGGGGUGAUUUAGGG
PUF_G4S	8SE-7SE-6SE-5SE-8NQ-7SE-6CQ-5NQ-4CQ-3NQ-2NQ-1CQ-4SE-3SE-2SE-1SE		GGGGUGAUUUAGGGG
PUF_G5S	8SE-7SE-6SE-8NQ-7SE-6CQ-5NQ-4CQ-3NQ-2NQ-1CQ-5SE-4SE-3SE-2SE-1SE		GGGUGAUUUAGGGGG
PUF_G6S	8SE-7SE-8NQ-7SE-6CQ-5NQ-4CQ-3NQ-2NQ-1CQ-6SE-5SE-4SE-3SE-2SE-1SE		GGUGAUUUAGGGGGG
PUF_G7S	8SE-8NQ-7SE-6CQ-5NQ-4CQ-3NQ-2NQ-1CQ-7SE-6SE-5SE-4SE-3SE-2SE-1SE		GUGAUUUAGGGGGGG
PUF_G8S	8NQ-7SE-6CQ-5NQ-4CQ-3NQ-2NQ-1CQ-8SE-7SE-6SE-5SE-4SE-3SE-2SE-1SE		UGAUUUAGGGGGGGG

Supplemental Table S3.

Sequences of RNA oligos for EMSA

RNA	Sequence (5' - -3')
RNA(US)	CCUUUUUUUUUGAUUUUUUUUUUCC
RNA(GS)	UUGGGGGGGGUGAUUUAGGGGGGGGUU
RNA(CS)	UCCCCCCCCCUGAUUUACCCCCCCCUU
RNA(VEGFA)	GAAAUACAGAUUAUCUUAAAA

Supplemental Table S4.

The amino acids interacting with RNA bases and target RNA sequences are shown. The amino acid pairs and the recognition bases of PUF_V are shown in red. Nt and Ct mean N-terminal and C-terminal, respectively.

Protein	Amino acids pair in repeats		Target RNA sequence
	Ct	Nt	5'-----3'
PUF_V	8NQ-7CQ-6SR-5CQ-4SE-3CQ-2NQ-1SQ		UACAGUA
PUF_T	8NQ-7CQ-6SR-5CQ-4SE-3CQ-2NQ-1SQ-8NQ-7CQ-6NQ-5SR-4NQ-3NQ-2CQ-1SQ		UACAGUAUAUCUAA
PUF_I	8CQ-7CQ-8NQ-7CQ-6SR-5CQ-4SE-3CQ-2NQ-1SQ-6NQ-5CQ-4NQ-3SR-2NQ-1NQ		AAUACAGUAUAUCUU

Supplemental Table S5.

Primer pairs for constructing pCI-PUF(N-C)-Halo-HA vector for luciferase reporter assay.

Sequence (5' - -3')	
PUF construction	fwd: ATACGACTCACTATAGGCTAGCGCCACCATG
	rev: CGCTACCTCCGAATTCCCCTAAGTCAACACCGTTCTTCATGTAG
Halo-HA tag	fwd: GAATTCGGAGGTAGCGGC
	rev: TATCATGTCTGCTCGAAGCGGCCGCTTAAGCGTAATCTGGAACATCGTATGGGTATGGCTCGCCGAAATCTC

Supplemental Table S6.

Primer pairs for constructing 16-repeats PUFs by Gibson Assembly

position	Vector Sequence (5' - -3')		Insert Sequence (5' - -3')	
0	fwd	CATATAATGGAATTTTCCCAAGACCAGC	fwd	CGGGAGATTGCCGGACATATAATGGAATTTTCCCAAGAC
	rev	TCCGGCAATCTCCCGCAG	rev	GGGAAAATTCCATTATATGGGGTCGGATCTTATGCATG
1	fwd	GCTGCCTACCAACTCATG	fwd	TCAATGAAATCCTCCAGCATATAATGGAATTTTCCCAAGAC
	rev	CTGGAGGATTTTCAATTGAAGAC	rev	GAGTTGGTAGGCAGCGGGTCGGATCTTATGCATG
2	fwd	CACGTCCTGTCATTGGCA	fwd	AGAACGGATTTCGAGGTCATATAATGGAATTTTCCCAAGAC
	rev	ACCTCGAATCCGTTCTGC	rev	CCAATGACAGGACGTGGGGTCGGATCTTATGCATG
3	fwd	CATGTCTTGAAGTGTGTG	fwd	CGGGAAGTAGATGGCCATATAATGGAATTTTCCCAAGAC
	rev	GCCATCTAGTTCCCGAAC	rev	ACACACTTCAAGACATGGGGTCGGATCTTATGCATG
4	fwd	CAGGTATTTGCCTTATCCAC	fwd	CGATGCGTTTAAAGGGACATATAATGGAATTTTCCCAAGAC
	rev	TCCCTTAAACGCATCGATG	rev	GATAAGGCAAATACCTGGGGTCGGATCTTATGCATG
5	fwd	CACACAGAGCAGCTTGTAC	fwd	TTAGAGGAGCTTCACCAGCATATAATGGAATTTTCCCAAGAC
	rev	CTGGTGAAGCTCCTCTAAATAG	rev	CAAGCTGCTCTGTGTGGGGTCGGATCTTATGCATG
6	fwd	AATGTACTTGTATTGAGTCAGC	fwd	GCAGAAATCCGAGGCCATATAATGGAATTTTCCCAAGAC
	rev	GCCTCGGATTTCTGCTAC	rev	GACTCAATAACAAGTACATTGGGTTCGGATCTTATGCATG
7	fwd	GCCTTATACACCATGATGAAGGAC	fwd	GACGGTCCCCACAGTCATATAATGGAATTTTCCCAAGAC
	rev	ACTGTGGGGACCGTCGTT	rev	ATCATGGTGTATAAGGCGGGTCGGATCTTATGCATG
8	fwd	CACATCGCAACTCTTCGTAAG	fwd	GCATAAGATCCGACCCCATATAATGGAATTTTCCCAAGAC
	rev	GGGTCGGATCTTATGCATG	rev	GAAGAGTTGCGATGTGGGGTCGGATCTTATGCATG

Supplemental Table S7.

DNA sequences inserted in 5'UTR of pmirGLO for luciferase reporter assay

Reporter	Sequence (5' - -3')
3×GS	aattcGGGGGGGGTGATATTAGGGGGGGGGTGATATTAGGGGGGGGGTGATATTAGGGGGGGGcgat
3×US	aattcTTTTTTTTTGATATTATTTTTTTTTTTGATATTATTTTTTTTTTTGATATTATTTTTTTTTcgat