Potential protein-encoded synthesis of DNA and RNA
Xinmiao Fu*
**HYPOTHESIS**

**Figure 1 | Illustration of the modified central dogma from potential synthesis of DNA and RNA encoded by protein.** A Sequence information transfer (from DNA to RNA to protein) takes place in nature and results in new sequence information from protein to RNA. The bioinformatics analysis results demonstrated here indicate that, although a majority contain 8 PUF repeats, many of the naturally occurring PUF proteins only carry 2 or 3 PUF repeats (Table 1). In particular, the PUF repeats in some PUF proteins are not sequentially connected but discontinuously present as discrete units of one, two or three repeats (as exemplified in Figure 1). These observations indicate that the minimal number of PUF repeats for their stable interaction.

Independent interaction with the RVD loop of one TALE repeat, also strengthening the possibility of an interaction between a TALE repeat and the non-nucleotide base pair. Third, mono- or dinucleotide base pairs, although unable to exist stably in solution, were reported to be stably formed in the hydrophobic cavities of self-assembled cages. Fourth, the apparent dissociation constant between optimized TALE proteins and target DNA was reported to be as low as 0.16 nM. Lastly, TALE repeats were reported to be flexible in conformation to cope with the B-form conformation of the DNA duplex. Together, these observations strongly suggest that mono- or dinucleotide base pairs can be arranged along the corresponding characteristic amino acids of TALE proteins through sequence-specific interactions as described above. As such, independent but arranged base pairs would be easily ligated or condensed as an oligodeoxynucleotide duplex (as illustrated in Figure 1C), by which modular PUF- or condensation agents from mono- or dideoxyribonucleotide base pairs that result in the final transfer of information from protein to RNA and can potentially be achieved using PUF, which contains an RNA-binding domain that comprises multiple PUF repeats and also forms a superhelical structure to bind the RNA molecule within its inner concave surface. Three-dimensional structural studies revealed that each PUF repeat, though not making direct interactions with either the phosphate backbone or the 2'-hydroxyl groups of RNA, recognizes a single RNA base through its three conserved amino acids, two of which make hydrogen bonds or Van der Waals interactions with the edge of an RNA base and a third residue that stacks with the same base and/or the preceding base. The code of RNA-binding specificity by PUF repeats has been decoded and artificially evolved (also illustrated in Figure 1C), by which modular PUF repeats capable of selectively binding specific RNA sequences can be created.

**Figure 1**

**A** Schematic illustration of the synthesis of RNA encoded by PUF proteins. Similarly, the transfer of sequence information from protein to RNA can potentially be achieved using PUF, which contains an RNA-binding domain that comprises multiple PUF repeats and also forms a superhelical structure to bind the RNA molecule within its inner concave surface. Three-dimensional structural studies revealed that each PUF repeat, though not making direct interactions with either the phosphate backbone or the 2'-hydroxyl groups of RNA, recognizes a single RNA base through its three conserved amino acids, two of which make hydrogen bonds or Van der Waals interactions with the edge of an RNA base and a third residue that stacks with the same base and/or the preceding base. The code of RNA-binding specificity by PUF repeats has been decoded and artificially evolved (also illustrated in Figure 1C), by which modular PUF repeats capable of selectively binding specific RNA sequences can be created.

**B** Schematic illustration of the synthesis of a DNA duplex by DNA ligase or condensation agents from mono- or diribonucleotides that are arranged in a sequence-specific manner along the characteristic residues of multiple PUF repeats. The amino acid code of RNA base specificity was adopted from earlier studies.

**C** Schematic illustration of the synthesis of a DNA duplex by DNA ligase or condensation agents from mono- or dideoxyribonucleotide base pairs that are arranged in a sequence-specific manner along the characteristic residues of TALE repeats in a designed TALE protein. The amino acid code of DNA sequence specificity by PUF repeats has been decoded and artificially evolved (also illustrated in Figure 1C), by which modular PUF repeats capable of selectively binding specific RNA sequences can be created.

**Table 1**

<table>
<thead>
<tr>
<th>PUF</th>
<th>Mono- or diribonucleotides</th>
<th>Characteristic residues</th>
<th>Oligodeoxynucleotide duplex</th>
</tr>
</thead>
</table>

**Note:** The figure and table are not included in the natural text representation.
Potentially protein-encoded synthesis of DNA and RNA

Fu

<table>
<thead>
<tr>
<th>Repeat No.</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein No.*</td>
<td>7</td>
<td>38</td>
<td>94</td>
<td>91</td>
<td>167</td>
<td>91</td>
<td>532</td>
<td>4</td>
<td>1024</td>
</tr>
<tr>
<td>Ratio (%)</td>
<td>0.7</td>
<td>3.7</td>
<td>9.2</td>
<td>8.9</td>
<td>16.3</td>
<td>8.9</td>
<td>52</td>
<td>0.4</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 1: Varying number of PUF repeats in naturally occurring PUF proteins.

* A total of 1024 PUF proteins were found in SMART database (http://smart.embl.de/; date of 2013/6/5).

The number of PUF repeats in each protein was counted and a summary of PUF proteins with different repeats is presented here.

PUF = Pumilio/fem-3 mRNA-binding factors; SMART = Simple Modular Architecture Research Tool.

with RNA bases may be as low as 2, or even 1. It follows that the mono- and/or diribonucleotides arranged along the concave surface of a designed PUF protein could be assembled as an oligoribonucleotide (as illustrated in Figure 1C). As such, the oligoribonucleotide sequence is solely determined by the characteristic amino acids of PUF repeats in the PUF protein.

**EXPERIMENTAL DESIGN**

**TALE-encoded synthesis of DNA**

To synthesize specific DNA duplexes, the gene encoding multiple TALE repeats plus the N- and C-terminal signals of TALE can be designed and assembled as described previously (as reviewed by Bogdanove et al.). The protein can be expressed and purified as described recently. Mixed 5'-phosphate mono- (4 types) and dideoxyribonucleotides (16 types) are then incubated with the purified TALE protein for a certain length of time. In particular, the binding affinity of the TALE protein to each of these deoxynucleotides can be evaluated by commercial systems based on isothermal titration calorimetry or surface plasma resonance. According to the early reports, the interaction between TALE and short DNA duplexes, the preliminary conditions used here are as follows: TALE at a concentration of approximately 2 μM and the mono- and dideoxyribonucleotides at a concentration of approximately 1 mM.

**dideoxyribonucleotide base pairs due to space hindrance, and thus would not be suitable to catalyze the ligation. An alternative approach to overcome this difficulty is to use chemical agents, such as cyanogen bromide, cyanamide and imidazole, which have been reported to be able to condense mono- and dideoxyribonucleobase base pairs into oligonucleotide duplexes. Since the base pairs are aligned along the characteristic RVDs of the TALE protein, condensation by these agents should be carried out efficiently due to the entropy reduction.**

**PUF-encoded synthesis of RNA**

To synthesize specific single-stranded RNA, the gene encoding multiple PUF repeats can be designed and assembled according to the principles and methods described previously. The PUF protein can be expressed and purified similarly as reported previously. The binding affinity of the PUF protein to various types of mono- and diribonucleotides can be evaluated by commercial systems based on isothermal titration calorimetry or those based on surface plasma resonance. The results of this kind of experiment will determine the approach and conditions that are suitable for the subsequent ligation and condensation. The preliminary conditions used here are as follows: PUF at a concentration of approximately 2 μM and the mono- and diribonucleotides at a concentration of approximately 1 mM.

If polynucleotide phosphorylase is used for ligation, a specific ddiribonucleoside (with a free 3'-terminal hydroxyl group) corresponding to the last two PUF repeats of the PUF protein and mixed nucleoside 3',5'-diphosphates can be added and incubated with the PUF protein. Polynucleotide phosphorylase can be added to the mixture to initiate the addition of mononucleoside 3',5'-diphosphates to the diribonucleoside, which will be further elongated, base by base, to oligoribonucleotides by the enzyme.

A substitute for polynucleotide phosphorylase is RNA ligase, which is known to catalyze the formation of an internucleotide phosphodiester bond between an oligoribonucleotide acceptor with a 3'-terminal hydroxyl (minimal acceptor being trinucleotides) and an oligonucleotide donor molecule with a 5-terminal phosphate (minimal donor being trinucleoside diphosphate, dinucleoside pyrophosphate and mononucleoside 3',5'-biphosphates). For this purpose, a specific triglycerol (with a free 3'-terminal hydroxyl group) corresponding to the last three PUF repeats of the PUF protein serves as the starting primer, and dinucleoside pyrophosphate or mononucleoside 3',5'-biphosphates serve as building blocks. Specifically, if the building block is mononucleoside 3',5'-biphosphate, ATP should be included for ligation; if it is dinucleoside pyrophosphate, ATP is not required.

The third approach to assemble these mono- and diribonucleotides into oligoribonucleotides is to use condensation agents such as cyanogen bromide and multimorinolite. To this end, mixed 5'-biphosphate mono- and diribonucleosides are incubated with the PUF protein for a certain length of time before adding these condensation agents and other crucial chemicals (e.g., metal ions).

**SIGNIFICANCE AND CONCLUSION**

In summary, it is proposed that sequence-specific DNA and RNA molecules can possibly be synthesized according to the characteristic amino acid sequences of designed TALE and PUF proteins, respectively. In terms of thermodynamics, TALE and PUF proteins arrange or fix the free mono- or dinucleotides, leading to a reduction in the entropy of the nucleotides and thus facilitating the subsequent ligation or condensation. This hypothesis is clearly experimentally testable. If proved, it is of interest to further examine whether Trp RNA-binding attenuation proteins and pentatricopeptide repeat proteins, both interacting with RNA in a sequence-specific manner, can serve...
It was proposed recently6 that the third characteristic residue of the PUF repeat, which stacks on and sandwiches successive bound RNA bases, may have played a role during the evolution of such pairing. Therefore, it is of interest to further investigate the evolutionary significance of the other two characteristic residues of the PUF repeat that make hydrogen bonds or Van der Waals interactions with RNA.

ACKNOWLEDGEMENT This work was supported by research grants from the National Natural Science Foundation of China (No. 31100559 and No. 31270804 to X.F.) and the National Basic Research Program of China (973 Program) (No. 2012CB917300 to X.F.).

ABOUT THE AUTHOR Dr. Fu is a protein scientist focusing on the mechanism of protein biogenesis, folding, assembly and degradation, as well as protein evolution. His long-term goal is to elucidate how the genetic information encoded by DNA is accurately transferred to the three-dimensional structure information of proteins, which is usually facilitated by molecular chaperones and folding catalysts in cells.

REFERENCES

Potential protein-encoded synthesis of DNA and RNA

Fu


