

fused to the STIM1 C terminus, which somehow interfere with the intramolecular transition that exposes the Orai1 interacting domain within the C terminus of STIM1 (Muik et al., 2011). Whether the LOV2 domain simply behaves in a passive, sterical manner or actively contributes to the intramolecular transition of STIM1 C terminus is not yet clear. Interestingly, potential dimerization of LOV2 has been reported (Nakasone et al., 2008), and the extent of dimerization may increase upon photo-activation, which in turn could contribute to the intramolecular transition coupling STIM1 C terminus to Orai1.

Close inspection of the fluorescence images and colocalization analysis of LOVS1K and Orai1 presented by Pham et al. (2011) reveals only a partial and ~30% increase in colocalization for regions of interest following light exposure. With respect to this moderate degree of colocalization, there is place for improvement, as the additional fluorescent protein labeling at the C terminus of LOVS1K and Orai1 will clearly reduce affinity of the STIM1/Orai1 interaction. Nevertheless, the amount of LOVS1K recruited was sufficient to induce local as well as global  $\text{Ca}^{2+}$  entry, depending on the number

of blue light exposures applied (Pham et al., 2011). Moreover, repeated activations (80 times over 40 min) led to gradual increase in global  $\text{Ca}^{2+}$  signal.

In conclusion, the probe developed by Pham et al. (2011) represents an important step toward the optogenetics of  $\text{Ca}^{2+}$  channels, as genetically encoded LOVS1K together with endogenous Orai1 will allow for a selective modulation of local and global  $\text{Ca}^{2+}$  entry with the use of light. The classical optogenetic tool, channelrhodopsin-1, is of limited value for this purpose, despite the recent development of a modified version that shows a ~2-fold increase in permeability to  $\text{Ca}^{2+}$  (Kleinlogel, et al., 2011). LOVS1K shall be put to the test initially in nonexcitable cells, such as Jurkat T-lymphocytes and RBL mast cells, as it is expected that LOVS1K-transfected cells will show conditional  $\text{Ca}^{2+}$  entry through endogenous Orai1 channels upon blue light exposure. In excitable cells, such as neurons and muscle cells, and also possibly in vivo models, light-mediated control of  $\text{Ca}^{2+}$  entry will allow for a better understanding of the role of Orai1. Moreover, is tempting to speculate that mutated STIM1 fragments, which interact with Orai1 without activating it,

will provide valuable tools to be able to conditionally knockdown  $\text{Ca}^{2+}$  entry.

#### REFERENCES

- Berridge, M.J., Bootman, M.D., and Roderick, H.L. (2003). *Nat. Rev. Mol. Cell Biol.* 4, 517–529.
- Fahrner, M., Muik, M., Derler, I., Schindl, R., Fritsch, R., Frischauf, I., and Romanin, C. (2009). *Immunol. Rev.* 237, 99–112.
- Hogan, P.G., Lewis, R.S., and Rao, A. (2010). *Annu. Rev. Immunol.* 28, 491–533.
- Kleinlogel, S., Feldbauer, K., Dempski, R.E., Fotis, H., Wood, P.G., Bamann, C., and Bamberg, E. (2011). *Nat. Neurosci.* 14, 513–518.
- Muik, M., Fahrner, M., Schindl, R., Stathopoulos, P., Frischauf, I., Derler, I., Plenk, P., Lackner, B., Groschner, K., Ikura, M., and Romanin, C. (2011). *EMBO J.* 30, 1678–1689.
- Nakasone, Y., Eitoku, T., Zikihara, K., Matsuoka, D., Tokutomi, S., and Terazima, M. (2008). *J. Mol. Biol.* 383, 904–913.
- Pham, E., Mills, E., and Troung, K. (2011). *Chem. Biol.* 18, this issue, 880–890.
- Strickland, D., Moffat, K., and Sosnick, T.R. (2008). *Proc. Natl. Acad. Sci. USA* 105, 10709–10714.
- Wu, Y.I., Frey, D., Lungu, O.I., Jaehrig, A., Schlichting, I., Kuhlman, B., and Hahn, K.M. (2009). *Nature* 461, 104–108.
- Yu, F., Sun, L., Courjaret, R., and Machaca, K. (2011). *J. Biol. Chem.* 286, 8375–8384.

## Finding the Missing Code of RNA Recognition by PUF Proteins

Yu Chen<sup>1</sup> and Gabriele Varani<sup>1,\*</sup>

<sup>1</sup>Department of Chemistry, University of Washington, Seattle, WA 98195, USA

\*Correspondence: varani@chem.washington.edu

DOI 10.1016/j.chembiol.2011.07.001

Pumilio and FBF homology (PUF) proteins represent highly promising candidates for engineering sequence-specific RNA recognition, but were only known to recognize G, A, and U, significantly limiting applications. Two groups (Filipovska et al., 2011; Dong et al., 2011) have now reported the discovery of the cytosine-recognition code for PUF proteins.

RNA-binding proteins (RBPs) play essential roles in regulating every step of RNA maturation and function. Typically, they recognize specific RNA structural features and sequences through a small number of very common RNA-binding modules (Lunde et al., 2007) and are often comple-

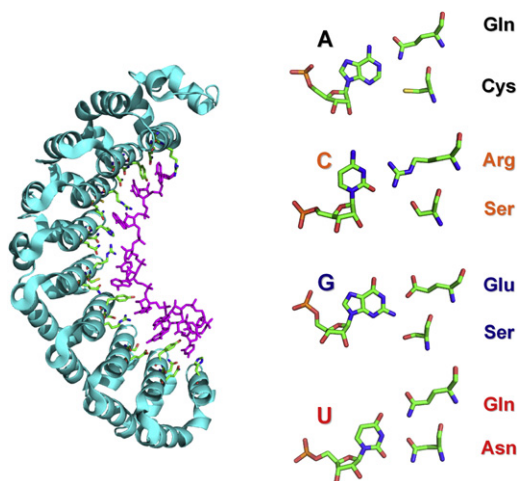
mented by other enzymatic or structural domains that perform additional functional roles. It would be very useful to engineer RBPs with desired RNA-binding specificities for investigating RNA biology and for potential biomedical applications, but this task has proven to be very chal-

lenging (Mackay et al., 2011). The only exception has been the PUF domain, which has been successfully manipulated to target single-stranded RNA in a sequence-specific manner. Thus, this domain represents the most promising candidate for the routine generation of

designer RBPs with required RNA-binding specificity, were it not for one limitation: PUF domains have not been observed to recognize cytosine so far.

PUF proteins are named after their founding members PUMILIO and fem-3 binding factor (FBF). They regulate gene expression by binding to specific sequences in the 3'-UTR of their target mRNAs and promote mRNA degradation and translational repression by recruiting effector proteins to the targeted sequence. All PUF domains contain multiple repeats, typically 8, of the same sequence-specific RNA-binding module, a structurally conserved 36-amino acid repeat. The single-stranded RNA (ssRNA) runs antiparallel to the protein and binds to the inner concave surface generated by the multiple PUF repeats (Figure 1) (Wang et al., 2002). Remarkably, each PUF repeat recognizes one nucleotide using three well-conserved amino acids in a base-specific manner. The amino acid side chain at position 13 in the repeat forms stacking interactions with the aromatic ring of the RNA base, and the Watson-Crick edge of the base is recognized by a specific combination of two amino acids (at positions 12 and 16). Thus, a single recognition code was generated: cysteine and glutamine bind adenine, asparagine and glutamine bind uracil, and serine and glutamate bind guanine (Figure 1). The simplicity of this recognition code allowed switching the specificity of individual repeats from one nucleotide to the other by just mutating only the two amino acids that make specific contacts with the RNA base edge (Cheong and Hall, 2006). Many of the resulting mutant PUF proteins can bind to their cognate RNAs as tightly and specifically as the wild-type protein that they were derived from.

Once proof of principle was provided that specificity could be engineered, different effector domains were attached to the designed PUF domains to engineer new functions. This technology was used to characterize the localization and trafficking of mitochondrial RNA (mtRNA) in single cells (Ozawa et al., 2007), and to engineer artificial splicing factors with desired sequence specificity and activity by combining a designed PUF domain



**Figure 1. RNA Recognition by PUF Proteins**

Shown here is the crystal structure of an engineered PUF domain in complex with target RNA containing cytosine (2YJY.pdb) (Dong et al., 2011). On the right, the RNA recognition code of PUF repeats is illustrated.

with different splicing modulation domains (activator or repressor) (Wang et al., 2009). These artificial splicing factors were shown to modulate splicing of an endogenous human gene, BCL-X.

However, naturally occurring PUF domains have not been observed to recognize cytosine, leaving designers without a recognition code for "C" and therefore limiting the potential target sites. In a significant advance, two groups (Filipovska et al., 2011; Dong et al., 2011) now report using directed evolution methods to select PUF repeat variants that specifically recognize cytosine. Both studies use the yeast three-hybrid system to link the interaction between PUF domains and RNA to a life-death growth selection. A DNA library based on the PUM1 PUF domain containing randomized amino acids at positions 12 and 16 was combined with an RNA, in which the targeted base was cytosine, so that only variants capable of binding to C would survive the selection. In the study by Filipovska et al. (2011), five unique PUF mutants selectively interacted with RNAs containing cytosine, but not with those containing adenine, guanine, or uracil RNA bases; all had an arginine at position 16 and an amino acid with a small or nucleophilic side chain at position 12 (Gly, Ala, Ser, Thr, or Cys). The result of the study by Dong et al. (2011) was more focused: almost all the selected clones contained an SXXXR sequence, perhaps a conse-

quence of more stringent selection conditions and tighter binding, and Y/H/R were found to be the best stacking residues in the cognate repeat to achieve specific binding of cytosine. Once the selected proteins were overexpressed and purified, the directed evolution experiments were proven to have been successful; specificity was re-engineered without the loss of binding affinity. Dong et al. (2011) then went further by engineering a new splicing factor that recognized C-containing target sequences, and demonstrated the specific modulation of alternative splicing of the VEGF-A mRNA, a key regulator of angiogenesis.

The crystal structure of the mutated PUF (SYXXR) protein revealed how cytosine is recognized (Dong et al., 2011). The essential

arginine contacts the O2 and N3 positions of the cytosine, while the serine forms a hydrogen bond with an amino group of the arginine side chain to position it for recognizing C (Figure 1). The cytosine base has to move away slightly from the RNA-binding surface to accommodate the longer arginine side chain. The specific interactions with cytosine are consistent with previous structure-based analysis of protein-RNA interactions, which had already indicated that the guanidinium group of arginine is most frequently used to simultaneously interact with Cyt-N3 and O2.

An additional step was undertaken by Filipovska et al. (2011), who engineered PUFs with 16 RNA-binding repeats to achieve a higher level of RNA sequence discrimination than would be possible with the eight nucleotides that are typically recognized by naturally occurring PUF proteins. The extended PUF bound its cognate 16 nucleotide RNA target in yeast and activated transcription more efficiently than a canonical 8-repeat PUF protein with its cognate RNA. It will now be interesting to establish the structure of this extended PUF protein to examine how it interacts with its extended RNA target. PUF domains adopt a crescent shape (Figure 1), so doubling the number of repeats may not be feasible without distorting the superstructure of the PUF repeats and affecting their RNA binding ability. It may also increase the possibility

of some bases flipping out from the RNA-binding surface, leading to reduced RNA-binding specificity. Nonetheless, this study suggests that it is feasible to use extended PUF repeats to target unique RNA sequences, which would be difficult to achieve with 8-repeat PUF.

About twenty years ago, Yen Choo and Aaron Klug reported the application of engineered zinc finger proteins to bind to a desired DNA sequence and alter gene transcription (Choo et al., 1994). Even if a sequence-specific recognition code for zinc fingers does not quite exist, this study led to many successful applications of engineered zinc finger proteins and nucleases, to the point where we can now simply order a designed zinc finger protein with any targeting DNA sequence specificity from a chemical supplier.

PUF proteins are even more easily adaptable than zinc fingers; the discovery of the cytosine-recognition code clearly provides a straightforward two amino acid tool-kit to design RNA recognition. With the caveat that recognition of sequences longer than eight nucleotides may still be challenging, it should now be possible to simply dial in an RNA sequence to generate a PUF protein with any desired specificity. Applications in research on RNA metabolism are limited only by our ingenuity, and perhaps even biomedical applications of PUF proteins will be possible.

#### REFERENCES

- Cheong, C.G., and Hall, T.M. (2006). *Proc. Natl. Acad. Sci. USA* 103, 13635–13639.
- Choo, Y., Sánchez-García, I., and Klug, A. (1994). *Nature* 372, 642–645.
- Dong, S., Wang, Y., Cassidy-Amstutz, C., Lu, G., Bigler, R., Jezyk, M.R., Li, C., Hall, T.M., and Wang, Z. (2011). *J. Biol. Chem.*, in press. Published online June 8, 2011. 10.1074/jbc.M111.244889.
- Filipovska, A., Razif, M.F.M., Nygård, K.K.A., and Rackham, O. (2011). *Nat. Chem. Biol.* 7, 425–427.
- Lunde, B.M., Moore, C., and Varani, G. (2007). *Nat. Rev. Mol. Cell Biol.* 8, 479–490.
- Mackay, J.P., Font, J., and Segal, D.J. (2011). *Nat. Struct. Mol. Biol.* 18, 256–261.
- Ozawa, T., Natori, Y., Sato, M., and Umezawa, Y. (2007). *Nat. Methods* 4, 413–419.
- Wang, X., McLachlan, J., Zamore, P.D., and Hall, T.M. (2002). *Cell* 110, 501–512.
- Wang, Y., Cheong, C.G., Hall, T.M., and Wang, Z. (2009). *Nat. Methods* 6, 825–830.