

IMPORTANCE OF SYSTEMATIC IDENTIFICATION OF RNA-BINDING PROTEINS IN A HYPERTHERMOPHILIC ARCHAEON

Akio Kanai^{1,2}

¹Institute for Advanced Biosciences, Keio University, Tsuruoka, Yamagata 997-0017, Japan
Tel: +81-235-29-0524; Fax: +81-235-29-0525; E-mail: akio@sfc.keio.ac.jp

²Department of Environmental Information, Keio University, Fujisawa, Kanagawa 252-8520, Japan

(Received October 26, 2006 Accepted October 30, 2006)

Abstract

Recent findings of huge numbers of non-coding RNAs and accumulating reports of gene regulation at the RNA level support the concept of “the RNA world” at the beginning of life on Earth. So the study of RNAs and their enzymes in a hyperthermophilic archaeon, *Pyrococcus furiosus*, which is believed to be a very ancient organism, may open a new door in the life sciences. We have developed an expression cloning method to classify and identify factors involved in the regulation of RNA metabolism in *P. furiosus*. Here I propose the value of the systematic analysis of regulatory RNAs and their binding proteins.

Keywords: Archaea, *Pyrococcus furiosus*, Expression cloning, DNA/RNA-binding protein, RNA world, Non-coding RNA

Introduction

Approximately 50 years has passed since the establishment of the central dogma of the genetic code—the concept of information flow from DNA to RNA to protein. During this period, this flow has been manipulated thanks to the discovery of reverse transcriptase. RNA is mostly regarded as the information transmitter, and it is widely considered that it has always had this role. However, genome mapping, and especially the recent elucidation of non-coding RNA (untranslated RNA), has seemingly entrenched the concept of RNA as more active functional molecules. Our group has found and characterized non-coding RNAs in a variety of organisms, including mouse (*Mus musculus*) [1], fruit fly (*Drosophila melanogaster*) [2], nematode (*Caenorhabditis elegans*) [3], and a bacterium (*Escherichia coli*) [4]. It is now believed that cells house huge numbers of non-coding RNAs, which may function beyond the central dogma (Figure 1), although the functions of most remain unknown. In addition, the “RNA world” hypothesis, which assumes that genetic information was originally controlled by RNA molecules, apparently renders RNA research more important. Since hyperthermophilic archaeons, especially *Pyrococcus* species, which grow in the deep sea at around 100 °C, are believed to be very ancient organisms, analyzing RNA metabolism in them could bring new insights into the fundamental regulation of

genes [5–7].

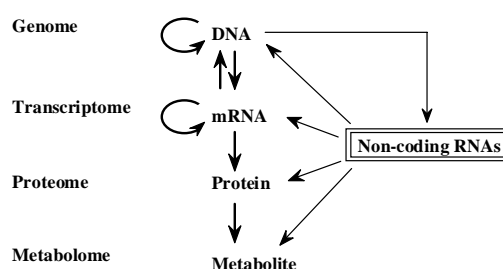


Figure 1. Revised view of the central dogma: non-coding RNAs maintain potential genetic networks beyond the central dogma.

Functional classification of archaeal proteome by expression cloning

Recent progress in genome projects has revealed the complete genomic DNA sequences in many species, from Bacteria to Archaea to Eukarya. However, only half of all proteins deduced from these sequences could be assigned putative cellular roles. The rest are considered to be conserved hypothetical proteins or simply hypothetical proteins. This is mainly because annotations of proteins are made by searching homologies against a limited set of databases of functionally known proteins. Therefore, a systematic way of annotating or analyzing the functions of proteins from the genome level would be valuable for characterizing proteins in the post-genome era. In this respect, the use of an expression cloning method in the test tube [8, 9] or a biochemical genomics approach in yeast [10] may turn out to be very useful. We have developed an efficient, highly sensitive method for identifying new genes present in the hyperthermophilic archaeon *Pyrococcus furiosus* at the genome level. This system has several advantages:

- *P. furiosus* has only about 2000 genes, and the complete genomic nucleotide sequence has been determined (<http://www.genome.utah.edu/>). It is less than half the size of the *E. coli* genome.
- The encoded proteins are mostly heat stable and are easy to handle biochemically.
- Most of the genes involved in nucleic acid metabolism in the Archaea are similar to those found in the Eukarya, but the regulation

mechanisms in the Archaea are simpler.

Our strategy for the systematic identification of DNA/RNA-binding proteins is described in Figure 2. We made a genomic DNA expression library of *P. furiosus* [5]. Briefly, after the *P. furiosus* genomic DNA was prepared, partially digested DNA fragments (about 7 kb average size) were ligated into a pRSET A plasmid vector. Although the vector contained the T7 RNA polymerase promoter sequence, *P. furiosus* genes were expressed in *E. coli* without induction by T7 RNA polymerase. Next, we prepared protein pools (one pool consisting of 30 independent colonies of *E. coli* in the library). These protein pools were heat-treated to kill endogenous proteins from *E. coli*, reducing the background noise and revealing the DNA/RNA-binding activities of proteins derived from *P. furiosus*. Because the genome of *P. furiosus* is about 2 megabases long, screening of 1200 clones (40 pools of 30 clones) should cover the whole genome. Thus, half a day is enough to screen a genome of this size.

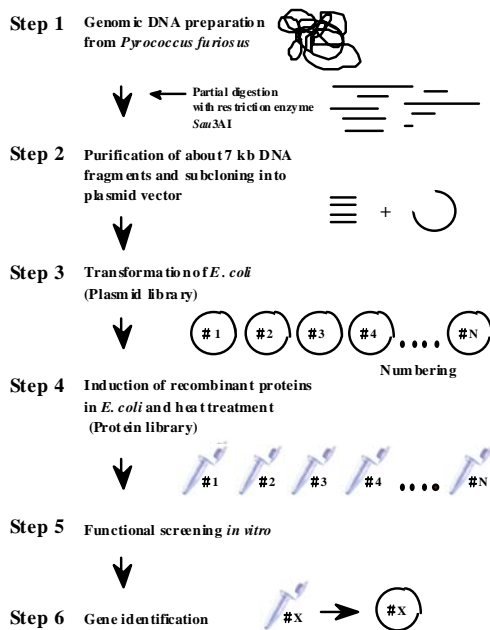


Figure 2. Expression cloning of *P. furiosus* genes to identify protein function at the proteome level.

RNA secondary structures and their binding proteins

After making the *P. furiosus* genomic expression library, we screened the library to isolate novel genes for DNA/RNA-binding proteins by using a series of artificial probes such as r(A-U)₁₀, r(G-C)₁₀, d(A-T)₁₀, and d(G-C)₁₀. We detected about 50 DNA/RNA-binding activities, and isolated and characterized one gene product, named FAU-1 (*P. furiosus* AU-binding protein-1), which is able to bind the r(A-U)₁₀ sequence [5]. Although FAU-1 showed

no significant homology with any protein whose functions are well known at either the nucleotide or the amino acid level, we found weak homology with the RNase E/RNase G protein family at the N-terminus (25%–30% identity). No ribonuclease activity was found, however, in the purified FAU-1 protein fractions, suggesting that FAU-1 is a novel RNA-binding protein. To determine the most suitable RNA sequence for recognition by FAU-1, we performed *in vitro* selection experiments (SELEX analysis) with RNA ligands and found that FAU-1 binds specifically to an AU-rich sequence in a loop region of a possible RNA ligand [5].

In recent years, it has been well accepted that RNA secondary structures, including stem-loop structures, are involved in many stages of gene regulation, such as transcription, splicing, translation, and degradation. In particular, a stem-loop RNA structure located near the translation start AUG codon appears to be the key regulator for translation. In the next round of screening, we therefore used an oligoribonucleotide probe with a specific RNA secondary structure (a stem-loop RNA oligo containing an AUG sequence in the loop region), and isolated the gene encoding thymidylate synthase (*Pf*-Thy1) as an RNA-binding protein [7]. *Pf*-Thy1 also bound to the stem-loop structure located near the translational start codon AUG in its own mRNA. *In vitro* translation tests using *E. coli* lysate indicated that the stem-loop structure of *Pf*-Thy1 mRNA might work as a translational repressor (Figure 3). Also, *Pf*-Thy1 inhibits the *in vitro* translation system. This evidence

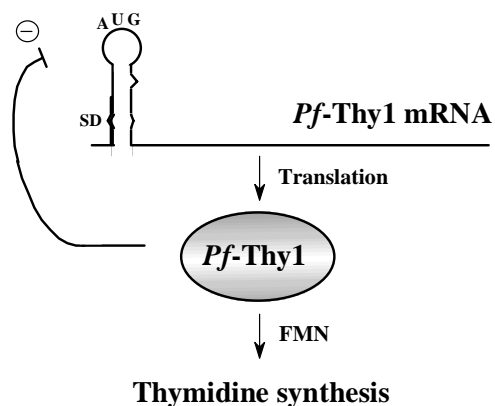


Figure 3. A model for autoregulation of *Pf*-Thy1 mRNA translation.

A stem-loop RNA secondary structure is located around the translation start AUG codon of *Pf*-Thy1 mRNA and acts as an inhibitory regulator of translation through a Shine–Dalgarno (SD)-like sequence in the stem region. Addition of *Pf*-Thy1 into the *in vitro* translation system also inhibits translation. These results suggest that thymidylate synthases of this class regulate their own translation. Flavin mononucleotide (FMN) is required for the thymidine synthetase activity of the enzyme.

strongly suggests that *Pf*-Thy1 controls its own

mRNA as an RNA-binding protein. This finding is consistent with the fact that another thymidylate synthase, ThyA, acts as an RNA-binding protein against its own mRNA [11, 12], although it belongs to a different class of thymidylate synthases. Now we are advancing our expression cloning method using a variety of oligoribonucleotide probes to pick up novel RNA-binding proteins in the Archaea.

Analyzing specific RNA secondary structures in the untranslated region (UTR) of mRNAs may help us understand a new RNA-protein network system possibly involved in gene regulation at the RNA level. One of these basic secondary structures is the stem-loop. Some stem-loop RNA structures in mRNAs, or “RNA domains”, are functional *cis*-elements such as riboswitches, internal ribosome entry sites (IRES), or selenocysteine-insertion sequences (SECIS). In the post-genome era, it is extremely important to summarize these RNA domains at the genome level, through prediction of RNA secondary structures and mapping of the positions of these structures in transcripts. Based on these analyses and our strategy for gene identification, the systematic identification of RNA-binding proteins that bind to specific RNA secondary structures will contribute to the mapping of RNA–protein interaction networks at the genome level

It is also true that classical non-coding RNAs such as rRNAs and tRNAs work with various types of RNA-binding proteins. So far, experimental and bioinformatics approaches have predicted and identified a huge amount of non-coding RNAs. However, there are few reports of the functional analysis of these non-coding RNAs. As described above, some RNA-binding proteins specifically regulate their target RNA domains in the protein-coding transcripts. It is possible that secondary structures found in non-coding RNAs interact with certain RNA-binding proteins. These could be identified by searching for common secondary structures among non-coding RNAs. For example, mapping of conserved RNA secondary structures predicts thousands of functional non-coding RNAs in the human genome [13].

Conclusions

This review introduces a systematic methodology for characterizing RNA-binding proteins. The methodology can identify new functional proteins because the cloning strategy is based on biological, chemical, or physical characteristics, not on the primary sequence of amino acids or on homology. It further makes it possible to add new functional properties (such as DNA/RNA-binding, kinase, protease, or other enzymatic capabilities) to already annotated proteins. In addition, functional classification using a bioinformatics approach is becoming important for revealing RNA molecules [14] and unknown proteins [15]. Collaborative structural analysis of RNA and RNA-binding proteins is valuable [16, 17]. The combination of molecular

biology, biochemistry, bioinformatics, and structural biology at the whole genome level is very useful for creating a new biology in the post-genome era. This is an ideal strategy of “systems biology”.

Acknowledgments

I would like to thank my many collaborators in the research.

References

1. Numata, K., Kanai, A., Saito, R., Kondo, S., Adachi, J., Wilming, L. G., Hume, D. A., Hayashizaki, Y., and Tomita, M. Identification of putative noncoding RNAs among the RIKEN mouse full-length cDNA collection, *Genome Res* 13, 1301-1306 (2003).
2. Inagaki, S., Numata, K., Kondo, T., Tomita, M., Yasuda, K., Kanai, A., and Kagayama, Y. Identification and expression analysis of putative mRNA-like non-coding RNA in *Drosophila*, *Genes Cells* 10, 1163-1173 (2005).
3. Watanabe, Y., Yachie, N., Numata, K., Saito, R., Kanai, A., and Tomita, M. Computational analysis of microRNA targets in *Caenorhabditis elegans*, *Gene* 365, 2-10 (2006).
4. Yachie, N., Numata, K., Saito, R., Kanai, A., and Tomita, M. Prediction of non-coding and antisense RNA genes in *Escherichia coli* with gapped Markov model, *Gene* 372, 171-181 (2006).
5. Kanai, A., Oida, H., Matsuura, N., and Doi, H. Expression cloning and characterization of a novel gene that encodes the RNA-binding protein FAU-1 from *Pyrococcus furiosus*, *Biochem J* 372, 253-261 (2003).
6. Sato, A., Kanai, A., Itaya, M., and Tomita, M. Cooperative regulation for Okazaki fragment processing by RNase HIII and FEN-1 purified from a hyperthermophilic archaeon, *Pyrococcus furiosus*, *Biochem Biophys Res Commun* 309, 247-252 (2003).
7. Kanai, A., Sato, A., Imoto, J., and Tomita, M. Archaeal *Pyrococcus furiosus* thymidylate synthase 1 is an RNA-binding protein, *Biochem J* 393, 373-379 (2006).
8. Stukenberg, P. T., Lustig, K. D., McGarry, T. J., King, R. W., Kuang, J., and Kirschner, M. W. Systematic identification of mitotic phosphoproteins, *Curr Biol* 7, 338-348 (1997).
9. King, R. W., Lustig, K. D., Stukenberg, P. T., McGarry, T. J., and Kirschner, M. W. Expression cloning in the test tube, *Science* 277, 973-974 (1997).
10. Martzen, M. R., McCraith, S. M., Spinelli, S. L., Torres, F. M., Fields, S., Grayhack, E. J., and Phizicky, E. M. A biochemical genomics approach for identifying genes by the activity of their products, *Science* 286, 1153-1155 (1999).
11. Chu, E., Koeller, D. M., Casey, J. L., Drake, J. C., Chabner, B. A., Elwood, P. C., Zinn, S., and Allegra, C. J. Autoregulation of human thymidylate synthase messenger RNA translation by thymidylate synthase, *Proc Natl Acad Sci U S A* 88, 8977-8981 (1991).
12. Chu, E., Voeller, D., Koeller, D. M., Drake, J. C., Takimoto, C. H., Maley, G. F., Maley, F., and Allegra, C. J. Identification of an RNA binding site for human thymidylate synthase, *Proc Natl Acad Sci U S A* 90, 517-521 (1993).
13. Washietl, S., Hofacker, I. L., Lukasser, M., Huttenhofer, A., and Stadler, P. F. Mapping of conserved RNA secondary structures predicts thousands of functional noncoding RNAs in the human genome, *Nat Biotechnol* 23, 1383-1390 (2005).
14. Sugahara, J., Yachie, N., Sekine, Y., Soma, A., Matsui, M., Tomita, M., and Kanai, A. SPLITS: a new program for predicting split and intron-containing tRNA genes at the genome level, *In Silico Biology* 6, 0039 (2006).
15. Fujishima, K., Imoto, J., Kanai, A., and Tomita, M. A new method for characterizing functionally-unknown proteins using specific amino acid frequency and periodicity at the proteome level, *Genome Informatics* 14, 526-527 (2003).
16. Okada, K., Takahashi, M., Sakamoto, T., Kawai, G., Nakamura, K., and Kanai, A. Solution structure of a GAAG tetraloop in helix 6 of SRP RNA from *Pyrococcus furiosus*, *Nucleosides Nucleotides Nucleic Acids* 25, 383-395 (2006).

17. Okada, K., Matsuda, T., Sakamoto, T., Muto, Y., Yokoyama, S., Kanai, A., and Kawai, G. (1)H, (13)C and (15)N resonance assignments of the 2'-5' RNA ligase-like protein from *Pyrococcus furiosus*, *J Biomol NMR* **online**

publication

02 February 2006

(doi:10.1007/s10858-005-5581-8) (2006).