# Analysis of the bacterial sequences recognized by RNA-binding proteins and compounds

**Ai Muto**
muto@kuicr.kyoto-u.ac.jp

**Akiyasu C. Yoshizawa**
acyshzw@kuicr.kyoto-u.ac.jp

**Shujiro Okuda**
okuda@kuicr.kyoto-u.ac.jp

**Masumi Itoh**
itoh@kuicr.kyoto-u.ac.jp

**Masahiro Hattori**
hattori@kuicr.kyoto-u.ac.jp

**Minoru Kanehisa**
kanehisa@kuicr.kyoto-u.ac.jp

Bioinformatics Center, Institute for Chemical Research, Kyoto University, Gokasho, Uji, Kyoto 611-0011, Japan

**Keywords**: RNA-binding, RNA secondary structure, riboswitch, aptamer, antitermination

## 1 Introduction

The regulation of gene expression is achieved in the stages of transcription, RNA stabilization, and translation. In bacteria, RNA structures are involved in these regulation stages and are thought to contribute to rapid adaptations to environments. The main RNA-acting elements, by ribosomes and RNA endonucleases, contain consensus sequences in single-stranded regions. Inverted repeats (IRs) in the same RNA molecule may form stem-loop structures, thus protecting from RNases or access of ribosomes. Therefore, the position of recognition sequences in relationto these IRs is important for gene expressions. However, RNAs are not always folded into one definite structure because of widely distributed IRs. In addition to many alternative structures, there are several factors that can affect binding and modify RNA structures in response to the environments.

Recent works have suggested that a variety of RNA structures act as regulators of gene expressions in bacteria. Several RNA-binding proteins and metabolites are identified to be involved in the following processes: protection of RNAs from degradation by RNases, prevention of ribosome binding to mRNA, control of mRNA secondary structure formation via aptamers that permit or prevent translation initiation, and termination or antitermination of transcripts.[1,2] These regulations are not special phenomena and it is important to consider alternative structures of mRNAs and RNA-binding by proteins and metabolites for understanding regulation mechanisms.

In order to search for unknown regulators, we have collected the sequences experimentally confirmed to be recognized by several RNA-binding factors and their homologs. We then surveyed the genes that contain RNA-binding sequences from bacteria genomes.

## 2 Methods

We constructed a dataset of RNA sequences known to be bound by proteins or compounds. The dataset contains 32 sequences (Table 1), which were collected from the literature reporting experimentally confirmed specific RNA-binding proteins or compounds (Table 2).

Table1. The number of experimentally confirmed entries in our dataset.

| RNA binding factors | Recognition RNA sequences | Species |
|---|---|---|
| Proteins (total of 8) | 13 | B.subtilis, E.coli, R.capsulatus, L.lactis etc |
| Compounds (total of 8) | 19 | B.subtilis, E.coli, S.typhimurium |
| | 32 | Total of 8 |

Table2. An example of the entry.

| Entry | |
|---|---|
| species | Bacillus subtilis |
| RNA sequence | 5'acauucggcguuggaauuaaucacauaugaaacagcccauagaucuuagacgauaggggcuaugcgugaaaacagaa-3' |
| Intermolecular gene of the mRNA | hut(histidine utilization) operon |
| Binding factor | HutP |
| Method | Gel-mobility assay |
| Journal | Mol Microbiol. 2000 Mar;35(5):1244-54. PMID: 10712704 |
| comment | The binding is histidine-dependent. |

We surveyed the genes that contain these RNA-binding sequence in the bacterial genomes stored in KEGG GENOME (Release 32.0+/10-02,Oct 04) using BLASTN 2.2.9. The sequences above the threshold of 10e-6 were accepted as homoplogs and a putative result dataset was obtained. We predicted the genes that contain RNA-binding sequences based on their positions on the genomes.

## 3 Results and Discussion

As a query of BLASTN, we used 19 sequences that were long enough in the dataset. The putative result dataset contained 96 sequences. We examined the genomic positions of homologs of two sequences recognized by TPP (thiamine pyrophosphate), because the numbers of these sequences were largest. For one of them we found 25 sequences: 13 in closely related organisms, eight located in homologous genes, and the other four located in the genes of unrelated organisms. For the other we found 22 sequences: 12 in closely related organisms, five located in homologous genes, and the other five located in the genes of unrelated organisms. TPP is known to inhibit the initiation of translation by specific binding to 5'-UTR of the genes that are related to thiamine synthesis in E.coli.[3] Here we found possible TPP binding sites on genes unrelated to thiamine synthesis, and furthermore TPP binding sites on different locations besides 5'-UTR. Therefore, TPP may regulate translation of genes other than thiamine synthesis related genes, and by some uncovered mechanisms besides regulating the access of ribosomes.

## Aknowledgement

This work was supported by grants from the Ministry of Education, Culture, Sports, Science and Technology, the Japan Society for the Promotion of Science, and the Japan Science and Technology Corporation. The computational resource was provided by the Bioinformatics Center, Institute for Chemical Research, Kyoto University.

## References

[1] Stulke J., Control of transcription termination in bacteria by RNA-binding proteins that modulate RNA structures. *Arch Microbiol.* 2002 Jun;177(6):433-40. 2002

[2] Soukup JK, Soukup GA, Riboswitches exert genetic control through metabolite-induced conformational change. *Curr Opin Struct Biol.* 2004 Jun;14(3):344-9.

[3] Winkler W, Nahvi A, Breaker RR. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature.* 2002 Oct 31;419(6910):952-6.