Expression of a Small Protein Encoded by the 3' Flanking Sequence of the *Escherichia coli rnpB* Gene

Yool Kim, Kook Han, Jungmin Lee, Kwang-sun Kim, and Younghoon Lee*

*Department of Chemistry and Center for Molecular Design and Synthesis, Korea Advanced Institute of Science and Technology, Daejeon 305-701, Korea. *E-mail: Younghoon.Lee@kaist.ac.kr*

Received April 21, 2007

M1 RNA is the catalytic component of RNase P, a tRNA-processing enzyme in *Escherichia coli*. M1 RNA is produced in the cell by transcription of the *rnpB* gene and subsequent processing at the 3' end. The 3' flanking region of *rnpB* contains repeated sets of overlapping sequences coding for small proteins. The issue of whether these proteins are expressed remains to be established. In this study, we showed the expression of a small protein encoded by the first repeat within the 3' flanking region of *rnpB*. Interestingly, protein expression was increased at lower temperatures. The termination efficiency of *rnpB* terminators was decreased at lower temperatures, suggesting that antitermination is responsible for enhanced protein expression. Moreover, the purified small protein contained M1 RNA, implying a role as a specific RNA-binding protein.

Key Words : M1 RNA, rnpB, RNase P, Antitermination, Small protein

Introduction

RNase P was initially characterized as a tRNA-processing enzyme, which removes the 5' extra sequences from precursor tRNAs to generate mature 5' termini.[1] RNase P is additionally involved in the processing of other non-coding RNAs, such as 4.5 S RNA and tmRNA, and degradation of several *Escherichia coli* mRNAs.[2-5] The RNase P enzymes from diverse organisms contain both essential RNA and protein components.[6,7] *E. coli* RNase P consists of two subunits, M1 RNA comprising 377 nucleotides and C5 protein with 119 amino acids.[8] While M1 RNA alone is the catalytic reaction precursor *in vitro*,[9] both components are essential *in vivo* and for efficient *in vitro* reactions under physiological conditions.[10] The C5 protein plays a crucial role in recognizing substrates and reducing the deleterious effects of mutations in M1 RNA.[11-13]

M1 RNA is synthesized by transcription of the *rnpB* gene and subsequent processing.[14,15] The majority of transcription occurs from the nearest promoter, P-1, with initiation at position 1 of mature M1 RNA, and terminates at an intrinsic terminator (T1) to generate precursor M1 RNA (pM1 RNA) containing an extra 36 nucleotide stretch.[16,17] The 3' flanking region of *rnpB* contains 113 bp repeats that are not perfectly identical, including the sequence coding for the 3' terminal 24 nucleotides of M1 RNA and the intrinsic transcription termination region.[18-20] This unit begins from position +354, and is successively repeated almost 3.5 times. The sequences additionally include three repeated sets of overlapping coding regions, which may generate small proteins. However, the expression patterns and functions of these proteins are currently unknown. To allow protein expression, *rnpB* transcription should run through the terminator regions.

In this study, we examined whether small proteins are expressed from the 3' flanking region. We found that a small protein encoded by the first coding region was expressed and that its expression increased upon lowering of the culture temperature. Moreover, the small protein interacted with RNA, supporting its role as an RNA-binding protein.

Experimental Section

Bacterial strains and plasmids. The *E. coli* K-12 strain, JM109, was used for constructing and maintaining plasmids, while BL21(DE3) was used for expressing recombinant proteins. We employed pLM1, a derivative of pGEM3 that contains the *rnpB* sequence between positions −270 and +1286, as the parental plasmid. Plasmid pLMd23 was a derivative of pLM1 with an internal deletion between positions +57 and +330 in the M1 RNA structural region.[16] The pGEX4T-1 plasmid (Amersham Biosciences) was used as a PCR template for amplifying the GST coding sequence.

Plasmid construction. To construct plasmids expressing GST-ORF fusion proteins, we amplified the GST coding sequence by PCR using pGEX4T-1 as a template, which was cloned into the *Bam*H1 and *Eco*R1 sites of pGEM3 (Promega) to generate pGEM-GST. The DNA fragments spanning positions −200 to +547 and +551 of *rnpB* were amplified using pLM1 as a template for ORF1 and ORF2, respectively. Fragments containing *rnpB* were cloned into the *Hin*dIII and *Bam*HI sites of pGEM-GST to generate *rnpB*-ORF-GST fusion plasmids. For overexpression of ORF-GST fusion proteins, the coding regions were amplified and cloned into the *Xho*I and *Nde*I sites of pET-22(b)+ to generate pET-ORF1-GST and pET-ORF2-GST. The resulting recombinant plasmids were subjected to site-directed mutagenesis for destruction of the stem-loop structure of the *rnpB* transcription terminators, thus generating pET-ORF1m-GST and pET-ORF2m-GST. The control pET-GST plasmid expressing GST protein was constructed using a similar procedure. The primers used for amplification and site-directed mutagenesis are presented in Table 1.

Table 1. Sequences of oligonucleotides used in this study

Primers	Sequences	Description
GST-*Bam*HI(+)	CGGGATCCTCCCCTATACTAGGTTATTG	Amplification of the GST coding sequence
GST-*Eco*RI(−)	CGGAATTCTCAATCCGATTTTGGAGG	
rnpB (−200*Hin*dIII)	CCCAAGCTTACCGATGATGTTGGCGTT	Amplification of *rnpB*
rnpB (+547*Bam*HI)	CGGGATCCTCTTTCTGCCCCTCCAAAATC	Amplification of *rnpB*
rnpB (+551*Bam*HI)	CGGGATCCTTCATCTTTCTGCCCCTCCAA	Amplification of *rnpB*
*Nde*I-ORF1(+)	GGAATTCCATATGATGAATGACTGTCCACGACGCTATACCC	Amplification of the ORF1 coding sequence
*Nde*I-ORF2(+)	GGAATTCCATATGATGACTGTCCACGACGCTATACCCAAAAG	Amplification of the ORF2 coding sequence
*Xho*I-ORF(−)	CCGCTCGAGTCAATCCGATTTTGGAGGATGGT	Amplification of the ORF1 coding sequence
ORF1-mut(+)	GTTTACGTAAGAATCCACTTCGCAGAGTGTTCGCTTTTG	Site-directed mutagenesis to disrupt the terminator hairpins
ORF1-mut(−)	CAAAAGCGAACACTCTGCGAAGTGGATTCTTACGTAAAC	
ORF2-mut(+)	GTTTACGTTAAGACTAGGTTCGGGGGATTCTTGCTTTTGGAG	Site-directed mutagenesis to disrupt the terminator hairpins
ORF2-mut(−)	CTCCAAAAGCAAGAATCCCCCGAACCTAGTCTTAACGTAAAC	
rnpB(−) (+416~+432)	ATTCATCTTTCTGCCCCTCCA	Probe for *rnpB* run-through transcript

Analysis of ORF-GST fusion protein expression. *E. coli* JM109 cells containing rnpB-ORF-GST fusion plasmids were grown to A_{600} of 0.5 at 37 °C. Where necessary, cultures were transferred to 30 or 15, and further grown for 1 h. Cells were harvested, and extracts prepared by sonication after adding the protease inhibitor cocktail (Pharmacia) in PBS (140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.3). Cell extracts were subsequently centrifuged at 25,000 × *g* for 20 min at 4 °C. The supernatant was incubated with glutathione Sepharose 4B beads for 2 h at 4 °C. Beads were washed three times with PBS at 4 °C, resuspended in 2 × SDS gel loading buffer (100 mM Tris-HCl, pH 6.8, 4% w/v SDS, 0.2% v/v bromophenol blue, 20% glycerol, 200 mM dithiothreitol), and boiled for 5 min at 100 °C. Proteins were analyzed by 15% SDS-PAGE, followed by Western blotting using anti-GST and horseradish peroxidase-conjugated anti-mouse IgG antibodies. Immunoreactive proteins were visualized with the ECL Detection system (Amersham Biosciences).

Analysis of *rnpB* transcription termination. *E. coli* cultures were grown overnight in LB, diluted (1:100) in fresh medium, and further grown to A_{600} of 0.5 at 37 °C. For experiments at lower temperatures, cultures were incubated at 30 °C and 15 °C. Total cellular RNA was prepared from *E. coli* cells containing pLMd23, as described previously.[16] RNA samples were electrophoresed on a 5% polyacrylamide gel containing 7 M urea, electrotransferred to a Hybond-N^+ membrane (Amersham), and hybridized.[21] The probe employed was antisense M1 RNA generated from *Hin*dIII-linearized pLMd23 DNA using T7 RNA polymerase and [α-^{32}P]-CTP or ^{32}P-labeled antisense oligonucleotide that hybridizes specifically to run-through *rnpB* transcripts. Hybridization signals were analyzed quantitatively using an Image analyzer BAS-1500 (Fuji).

Overexpression and purification of ORF-GST fusion proteins. *E. coli* BL21(DE3) cells containing ORF-GST plasmids were grown to A_{600} of 0.5, and induced with 1 mM IPTG for 3 h. Cells were harvested, and extracts prepared by sonication after adding the protease inhibitor cocktail (Pharmacia) in PBS. Cell extracts were centrifuged at 25,000 × *g* for 30 min at 4 °C. The supernatant was loaded onto a glutathione affinity column, and GST fusion protein eluted with 10 mM reduced glutathione (50 mM Tris-HCl, 10 mM reduced glutathione, pH 8.0), according to the manufacturer's instructions (Amersham Biosciences). The GST control protein was purified from BL21(DE3) cells containing pET-GST.

Determination of the M1 RNA content of purified protein fractions. Purified ORF-GST protein (200 μg) was extracted with phenol:chloroform (5:1), and RNA was ethanol-precipitated. The RNA sample was electrophoresed on a 5% polyacrylamide gel containing 7 M urea, and electrotransferred to a Hybond-N^+ membrane (Amersham). Hybridization was performed as described above with the antisense M1 RNA probe. The quantity of M1 RNA was estimated by analyzing the filter with Image analyzer BAS-1500 (Fuji).

Gel mobility shift assay. A gel mobility shift assay for analysis of interactions between ORF-GST and M1 RNA was performed according to the procedure of Lee *et al*.[29] Approximately 0.1 nM RNA internally labeled with [α-^{32}P]-CTP was incubated in standard binding buffer containing 20 mM K-HEPES, pH 8.0, 400 mM NH_4OAc, 10 mM $Mg(OAc)_2$, 0.01% (v/v) Nonidet P-40, and 5% glycerol at 37 °C for 10 min. In addition to ORF-GST protein, GST and MBP-C5 were used as negative and positive controls for M1 RNA binding, respectively. MBP-C5 protein was purified using an established protocol.[22]

Results

The 3' flanking region of *rnpB* contains 113 bp repeats that occur about 3.5 times. Each repeat includes the 3' terminal 24 bp of *rnpB* and an intrinsic terminator (Fig. 1). Sequence analysis of the 3' flanking region reveals the presence of six open reading frames (ORFs) for small proteins, specifically, three overlapping coding regions with two ORFs per region. The overlapping coding region appears three times due to sequence repetition in the 3' flanking region, and contains the putative Shine-Dalgarno sequence.[23] However, for pro-

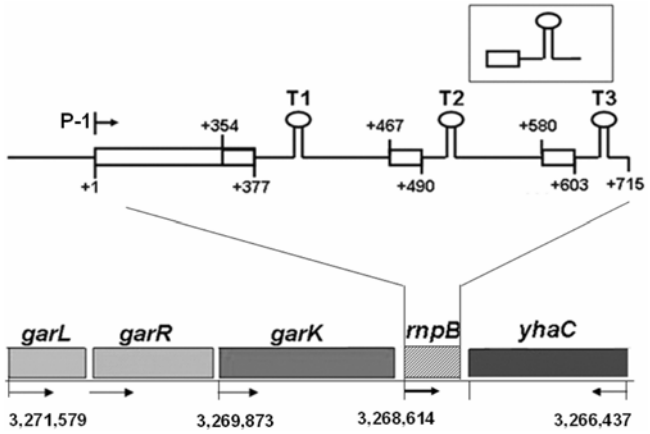

Figure 1. Schematic representation of the *rnpB* gene structure. The rectangle indicates the structural sequence of M1 RNA. Numbers in the upper figure refer to positions in the *rnpB* gene. Numbers in the lower figure indicate positions of the *E. coli* MG1655 genome. The 113 bp repeat sequence starts at position +354, and contains the 3' terminal 24 nucleotides of M1 RNA. The repeat sequence is schematically presented in the inset.

tein translation, transcription from the *rnpB* promoter should run through the terminators, T1 and T2, for the first coding region, and all three terminators (T1, T2, and T3) for the second and third coding regions. It appears that the first overlapping region has higher potential to code for small proteins than the second or third regions. ORF1 in the first overlapping region is initiated from position +431, while ORF2 starts from position +435. To resolve whether these small proteins are expressed, we constructed ORF-GST fusion plasmids where the GST coding region was fused to each ORF in phase. If proteins are expressed through the *rnpB* transcription unit, ORF-GST fusion proteins should be generated (Fig. 2). Accordingly, expression of fusion proteins from cells containing ORF-GST fusion vectors was examined. Bacterial cell lysates were pulled down with glutathione-Sepharose beads, and proteins analyzed by Western blot with an anti-GST antibody. We observed expression of ORF1-GST, but not ORF2-GST (Fig. 3), suggesting that a small protein encoded by ORF1 is expressed.

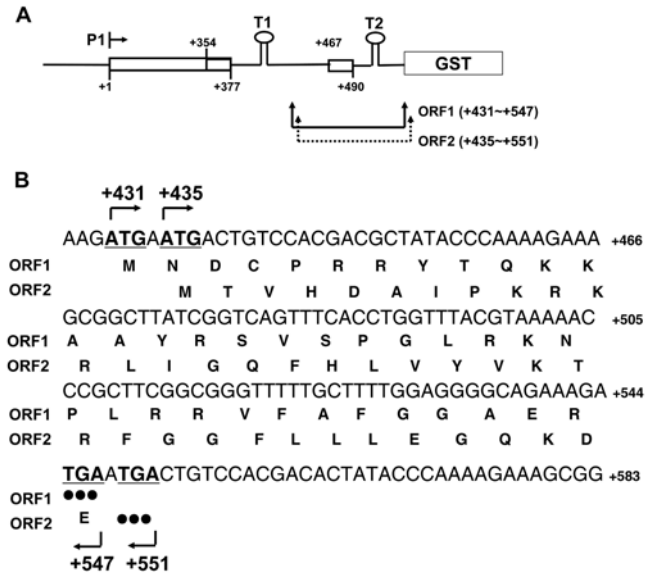

Figure 2. Schematic representation of ORF-GST fusion constructs. (A) Expression of ORF1-GST and ORF2-GST fusion proteins by *rnpB* transcription. (B) Putative amino acid sequences of ORF1 and ORF2 are shown.

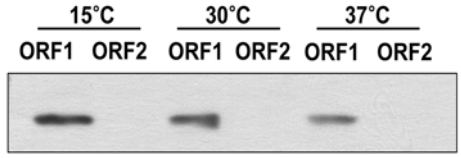

Figure 3. Expression of ORF-GST fusion proteins in *E. coli*. JM109 cells containing ORF-GST fusion plasmids were grown to A_{600} of 0.5 at 37°C, and the temperature was shifted to 30°C or 15 °C. Lysates from cells before and after the temperature shift were prepared, and used for pull-down of ORF-GST fusion proteins with glutathione Sepharose 4B beads. Proteins were electrophoresed on a 15% SDS-polyacrylamide gel, and immunoblotted using an anti-GST antibody. ORF1 and ORF2 represent proteins obtained from cells containing the ORF1-GST and ORF2-GST fusion plasmids, respectively.

Expression of ORF1 requires transcription through T1 and T2. Since antitermination is favorable at lower temperatures, possibly due to the melting of terminator stems by cold shock proteins.[24-26] We examined ORF1-GST expression at 30°C and 15°C. Interestingly, higher expression was evident at lower temperatures. To determine whether this elevated expression is related to antitermination, we examined alterations in termination efficiency at T1 at lower temperatures. For this purpose, we employed pLMd23, which generates a 147 nt T1-terminated transcript.[20] Upon incubation of cells

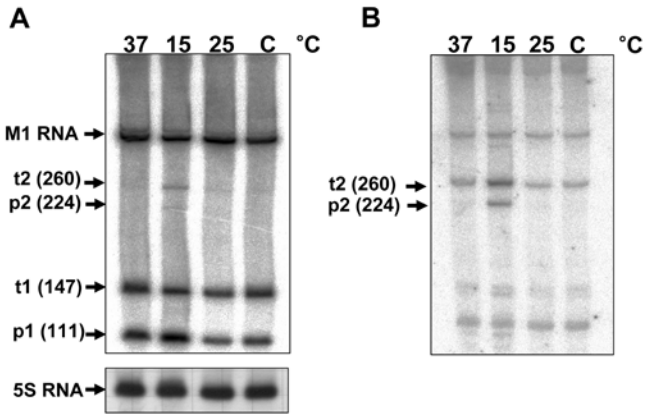

Figure 4. Antitermination at *rnpB* terminators. *E. coli* JM109 cells containing pLMd23 were grown to A_{600} of 0.5 at 37°C, and the temperature was shifted to 25°C or 15°C for 1 h. Total cellular RNAs were prepared before and after the temperature shift, and analyzed by Northern blot using the anti-M1 RNA probe (A) and the antisense oligonucleotide probe containing the *rnpB* sequence between positions +416 and +432 (B). Lane C represents total cellular RNA isolated from cells withdrawn from the same culture before the temperature shift, and grown for a further 1 h. Primary transcripts are designated t1 and t2, according to the termination position. Processed RNAs from primary transcripts terminated at T1 and T2, denoted p1 and p2, respectively. The expected RNA sizes are specified in parentheses.

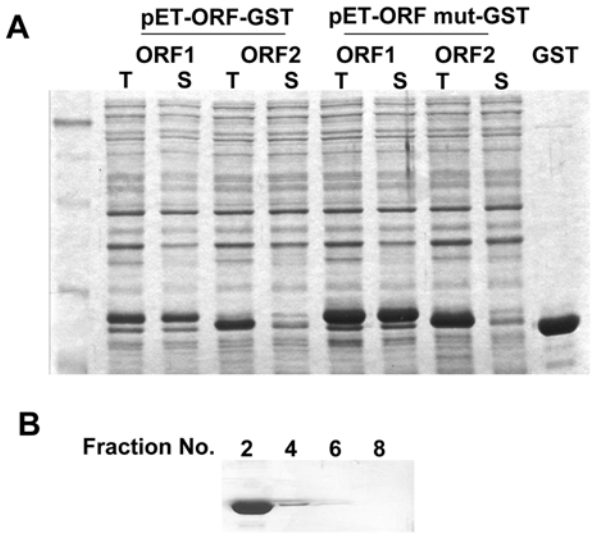

Figure 5. Overexpression of ORF-GST fusion proteins. (A) *E. coli* BL21(DE3) cells containing P^{T7}-ORF-GST transcription fusion constructs were grown to A_{600} of 0.5 at 37 °C, and protein synthesis induced by adding 1 mM isopropyl (β-D-thiogalactoside (IPTG). Total cells (T) or soluble fractions (S) were analyzed by 15% SDS-PAGE. GST control protein purified from cells containing pET-GST. (B) GST-ORF1 was purified using glutathione-Sepharose 4B affinity chromatography. Elution was performed with 10 mM glutathione. Each fraction was analyzed by 15% SDS-PAGE. Fraction numbers are shown above each lane.

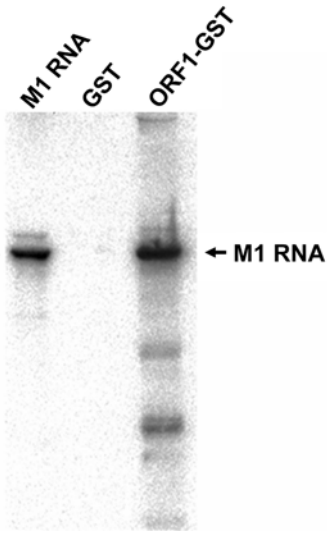

Figure 6. Analysis of RNA bound to ORF1 protein. Purified proteins were extracted with phenol, and RNA samples were ethanol-precipitated. RNAs bound to ORF1 were separated on a 5% polyacrylamide gel containing 7 M urea, and subjected to Northern analysis. ^{32}P-labeled antisense RNA was used as a probe. GST purified from cells containing pET-GST was additionally employed. Total cellular RNA from JM109 cells containing pLM1 was used as the positive control (M1 RNA).

containing pLMd23 at 15°C, we observed increased levels of a 260 nt T2-terminated transcript resulting from run-through transcription at T1. Our results suggest that anti-termination is responsible for the high expression of ORF1 at low temperatures (Fig. 4).

As an initial step to establishing the possible function of ORF1, we examined whether M1 RNA bound to the purified ORF1-GST protein. Significant quantities of ORF1-GST fusion proteins were required for the experiment. We constructed expression plasmids, pET-ORF1-GST and pET-ORF2-GST, where transcription is initiated from the T7 promoter, as well as derivatives (pET-ORF1m-GST and pET-ORF2m-GST) whereby the *rnpB* intrinsic terminators, T1 and T2, were disrupted by substituting nucleotides in some triplets with others encoding the same amino acids. Higher levels of protein were generated from pET-ORF1m-GST and pET-ORF2m-GST than pET-ORF1-GST and pET-ORF2-GST, respectively, suggesting that some transcription from the T7 promoter by T7 RNA polymerase terminates at T1 and T2 (Fig. 5). While both ORF1-GST and ORF2-GST were expressed in cells containing the plasmids, only ORF1-GST was present as soluble protein. ORF2-GST was also expressed, but the protein aggregated into inclusion bodies. Consequently, ORF1-GST was purified and analyzed for its M1 RNA content. The purified ORF1-GST protein contained M1 RNA (about 0.01% in the mole ratio), but not control GST protein (Fig. 6), suggesting that ORF1 functions as a M1 RNA-binding protein. However, we did not observe *in vitro* interactions between ORF1-GST and M1 RNA (data not shown). It is possible that binding of ORF1-GST to M1 RNA requires additional cellular factors. Alternatively, this discrepancy may be due to the different efficiencies of ribonucleoprotein assembly *in vivo* and *in vitro*.

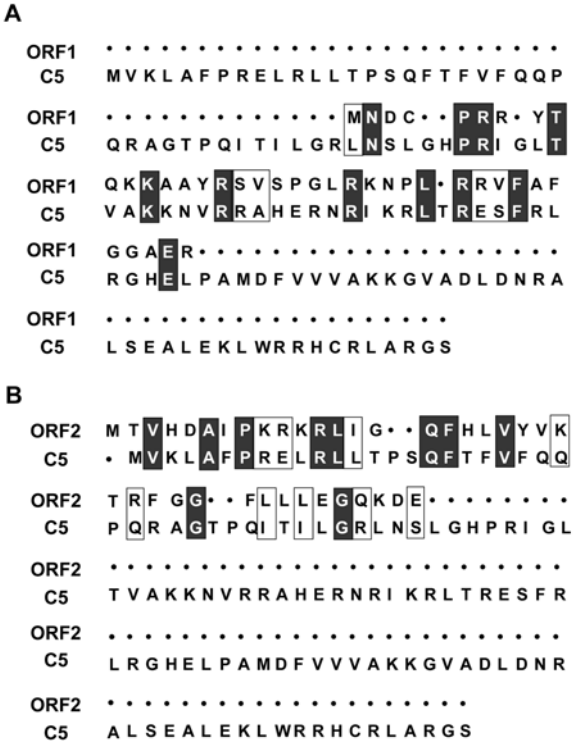

Figure 7. Sequence alignment of ORF1 (A) and ORF2 (B) with C5 protein. Identical matches of amino acids are indicated by shading, and conservative substitutions are boxed.

Discussion

We observed ORF1 expression from the first set of overlapping coding sequences in the 3' flanking region of *rnpB*. ORF1 is expressed only when transcription runs through both the T1 and T2 terminators of *rnpB*. Moreover, ORF1 expression is higher at lower temperatures, possibly due to elevated antitermination at the *rnpB* terminators resulting from an increase in run-through transcription at T1.

No ORF2 expression was evident from the first set of overlapping coding sequences. ORF2 may not be efficiently translated due to the longer distance between the Shine-Dalgarno sequence and start codon, compared to that of ORF1.[27] Alternatively, ORF2 may be expressed as an insoluble protein, and only soluble proteins are detected with our system using GST pull-down assays. This theory is supported by the observation that most ORF2 proteins expressed from the T7 promoter are insoluble. The over-expressed ORF2 protein might have been be insoluble because soluble ORF1 was not available in the cell, as exemplified in the case where MhpE is soluble only when soluble MhpF was available.[28] Notably, ORF1 and ORF2 are homologous with the N-terminal and C-terminal halves of C5, respectively (Fig. 7). This may explain why ORF1 does not bind to M1 RNA *in vitro*, but contains M1 RNA *in vivo*. Both ORF1 and ORF2 mimic C5 protein in terms of M1 RNA binding. In cells, ORF1 may bind to M1 RNA by interacting with endogenous ORF2 expressed at the basal level. Since C5 protein interacts with RNA molecules other than M1 RNA,[29,30] it is additionally possible that ORF1, together with ORF2, binds to various types of RNA.

The biological relevance of small proteins encoded in the 3' flanking region of *rnpB* is unclear. Since lower temperatures are associated with elevated protein expression, their function may be related to the response of bacterial cells upon cold shock. For example, these proteins may modulate the catalytic activity of M1 RNA by promoting the formation of an active RNA conformation at lower temperatures. However, further experiments are essential to confirm the roles of these proteins in cells.

Acknowledgement. This work was supported by the Korea Research Foundation Grant (KRF 2004-015-C00330).

References

1. Li, Y.; Altman, S. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 13213.
2. Komine, Y.; Kitabatake, M.; Yokogawa, T.; Nishikawa, K.; Inokuchi, H. *Proc. Natl. Acad. Sci. USA* **1994**, *73*, 1912.
3. Bourgaize, D. B.; Fournier, M. J. *Nature* **1987**, *325*, 281.
4. Bothwell, A. L.; Garber, R. L.; Altman, S. *J. Biol. Chem.* **1976**, *251*, 7709.
5. Robertson, H. D.; Altman, S.; Smith, J. D. *J. Biol. Chem.* **1972**, *247*, 5243.
6. Kurz, J. C.; Fierke, C. A. *Curr. Opin. Chem. Biol*. **2000**, *4*, 553.
7. Frank, D. N.; Pace, N. R. *Annu. Rev. Biochem.* **1998**, *67*, 153.
8. Altman, S.; Baer, M.; Guerrier-Takada, C.; Vioque, A. *Trends Biochem. Sci.* **1986**, *11*, 515.
9. Guerrier-Takada, C.; Gardiner, K.; Marsh, T.; Pace, N. R.; Altman, S. *Cell* **1983**, *35*, 849.
10. 10. Sakano, H.; Yamada, S.; Ikemura, T.; Shimura, Y.; Ozeki, H. *Nucleic Acids Res.* **1974**, *1*, 355.
11. Peck-Miller, K. A.; Altman, S. J. *Mol. Biol.* **1991**, *221*, 1.
12. Lumelsky, N.; Altman, S. *J. Mol. Biol.* **1988**, *202*, 443.
13. Guerrier-Takada, C.; Lumelsky, N.; Altman, S. *Science* **1989**, *246*, 1578.
14. Lee, Y.; Ramamoorth, R.; Park, C.-U.; Schmidt, F. J. *J. Biol. Chem.* **1989**, *264*, 5098.
15. Sakamoto, H.; Kimura, N.; Shimura, Y. *Proc. Natl. Acad. Sci. USA* **1983**, *80*, 6187.
16. Kim, S.; Kim, H.; Park, I.; Lee, Y. *J. Biol. Chem.* **1996**, *271*, 19330.
17. Lee, Y. M.; Lee, Y.; Park, C.-U. *Korean Biochem. J.* **1989**, *22*, 276.
18. Adhya, S.; Gottesman, M. *Annu. Rev. Biochem*. **1978**, *47*, 967.
19. Platt, T. *Annu. Rev. Biochem*. **1986**, *55*, 339.
20. Yagar, T. D.; von Hippel, P. H. In *Escherichia coli and Salmoella typhymurium: Cellular and Molecular Biology*; Neidhardt, F. C., Ed.; American Society for Microbiology: Washington, DC., 1987; pp 1241-1275.
21. Kim, S.; Lee, Y. *FEBS Lett.* **1997**, *407*, 353.
22. Kim, M.; Park, B. H.; Lee, Y. *Biochem. Biophys. Res. Commun*. **2000**, *268*, 118.
23. Shine, J.; Dalgarno, L. *Nature* **1975**, *254*, 34.
24. Phadtare, S.; Severinov, K. *Nucleic Acids Res.* **2005**, *33*, 5583.
25. Phadtare, S.; Inouye, M.; Severinov, K. *J. Biol. Chem.* **2002**, *277*, 7239.
26. Bae, W.; Xia, B.; Inouye, M.; Severunov, K. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 7784.
27. Dalboge, H.; Carlsen, S.; Jensen, E. B.; Christensen, T.; Dahl, H. H. *DNA* **1988**, *7*, 399.
28. Lee, S. J.; Ko, J.; Kang, H. Y.; Lee, Y. *Biochem. Biophys. Res. Commun.* **2006**, *346*, 1009.
29. Lee, J. H.; Kim, H.; Ko, J.; Lee, Y. *Nucleic Acids Res.* **2002**, *30*, 5360.
30. Kim, K. S.; Ryoo, H.; Lee, J. H.; Kim, M.; Kim, T.; Kim, Y.; Han, K.; Lee, S. H.; Lee, Y. *Bull. Korean Chem. Soc.* **2006**, *27*, 699.