Doklady Biochemistry and Biophysics, Vol. 389, 2003, pp. 118–121. From Doklady Akademii Nauk, Vol. 389, No. 6, 2003, pp. 821–824.

BIOCHEMISTRY, BIOPHYSICS, AND MOLECULAR BIOLOGY

Internal Ribosome Entry Site from Crucifer Tobamovirus Promotes Initiation of Translation in *Escherichia coli*[1]

T. V. Komarova, M. V. Skulachev, P. A. Ivanov, A. G. Klyushin, Yu. L. Dorokhov, and Academician J. G. Atabekov

Received December 15, 2002

mRNA translation efficiency in a prokaryotic system is mainly determined by the complementary interaction of 16S rRNA and the mRNA region located upstream of the start codon of the gene. Blocks of purine residues in 16S rRNA interact with complementary pyrimidine nucleotides in mRNA (Shine–Dalgarno interactions; SD) [1]. Contrary to prokaryotes, eukaryotic mRNAs usually lack SD elements. Translation initiation of most eukaryotic mRNAs proceeds by the ribosome scanning mechanism (hypothesized by Kozak [2, 3]). According to this model, the 40S ribosomal subunit binds to the cap structure on the 5'-terminus of mRNA, and then it scans the 5'-untranslated region (5'UTR) of mRNA until it reaches the initiator codon. Then, the 80S ribosome assembles, and polypeptide synthesis starts.

However, the initiation of translation of a variety of viral and cellular mRNAs proceeds by an alternative mechanism of internal ribosome entry mediated by internal ribosome entry sites (IRESs). It was first shown that, during the translation of encephalomyocarditis virus (EMCV) mRNA, the ribosome binds to some internal region of mRNA (IRES) rather than to the 5'-terminus [4].

Recent studies allowed the identification of two active IRES elements within the genomic RNA of crucifer-infecting tobamovirus (CrTMV) [5, 6]. $IRES_{CP, 148}$ is located upstream of the coat protein (CP) gene, and $IRES_{MP, 75}$ is located upstream of the movement protein (MP) gene. The unique feature of $IRES_{CP, 148}$ is that it is active not only in plant cells, but also in animal cells and yeasts [7].

It was shown that several 5'UTRs can direct the translation of eukaryotic mRNAs in a prokaryotic system [8–10]. In this case, 5'UTRs operate as a ribosome-binding site (RBS), creating SD boundaries with 16S rRNA. It was demonstrated that TMV genomic RNA expresses the 3'-proximal CP gene in an *Escherichia coli* cell-free translational system [11].

Two polypurine tracts (ppt32 and ppt11) were identified within $IRES_{CP, 148}$ (Fig. 1a). In principle, these tracts can form SD boundaries with 16S rRNA (table). This allows us to assume that $IRES_{CP, 148}$ could be another functional RBS of eukaryotic origin.

In this study, we analyzed the translational activity of $IRES_{CP, 148}$ in a prokaryotic system *in vivo*. The analysis was performed using a bicistronic mRNA with $IRES_{CP,148}$ or other test sequences cloned as an intercistronic spacer. The translational activities of $IRES_{CP,148}$, its structural elements and mutants, as well as artificial polypurine sequences, were evaluated in transformed *E. coli*.

Bicistronic gene constructs described in [7] were used as a basis to obtain a set of plasmids containing the T7 promoter, green fluorescent protein (GFP) gene, and firefly luciferase (LUC) gene. The following sequences were cloned as the intercistronic areas of these constructs: the RBS from expression vector pQE30 (Qiagen), $(GAAA)_3$, $(A)_{18}$, $(G)_{16}$, $(U)_{16}$, $IRES_{CP, 148}$, and its elements—ppt32, ppt19, and ppt11 (Fig. 1a). The schemes of the constructs are shown in Fig. 1b.

E. coli strain BL-21 (carrying the T7 polymerase gene under the control of an inducible promoter) was transformed with the described plasmids. Induction of T7 polymerase led to the synthesis of bicistronic GFP-LUC mRNAs. The translational activities of different intercistronic sequences were estimated by the measurement of luciferase activity in lysed *E. coli*. Three to five independent clones were tested for each construct. The mean values and SE bars of the obtained results are shown in Fig. 2. The average activity of the construct containing $IRES_{CP, 148}$ was taken as 100%. PQE30 RBS was used as the positive control, since this sequence provides the maximum number of SD boundaries (table).

The data obtained allowed us to estimate the relative efficiency of the tested sequences as RBSs in *E. coli* (Fig. 2). $(G)_{16}$ and $(U)_{16}$ displayed almost no activity and can be used as negative controls providing the background level of the luciferase activity. $IRES_{CP,148}$ turned out to be able to provide efficient initiation of translation in *E. coli*; however, its activity is was two to three times lower than the activity of pQE30 RBS.

[1] This article was submitted by the authors in English.

Moscow State University, Vorob'evy gory, Moscow, 117234 Russia

1607-6729/03/0304-0118$25.00 © 2003 MAIK "Nauka/Interperiodica"

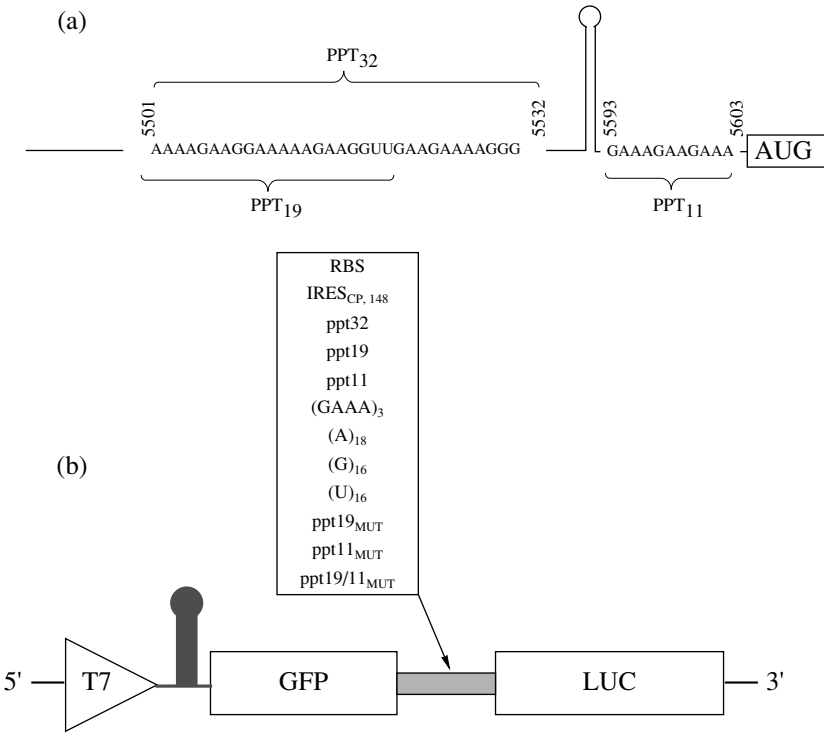

Fig. 1. (a) A scheme of $IRES_{CP,148}$. Polypurine tracts (ppt32, ppt19, ppt11) and the start codon are indicized. Numbers correspond to the nucleotide positions in CrTMV genomic RNA. (b) A scheme of bicistronic constructs GFP-LUC. The names of different intercistronic spacers are indicated inside the bar. All constructs were cloned under the control of the T7 promoter into pBluescript SK+ plasmid (Stratagene) and contained the genes of green fluorescent protein (GFP) and firefly luciferase (LUC). The sequence between the T7 promoter and GFP contains the inverted repeat that forms a stable secondary structure in RNA.

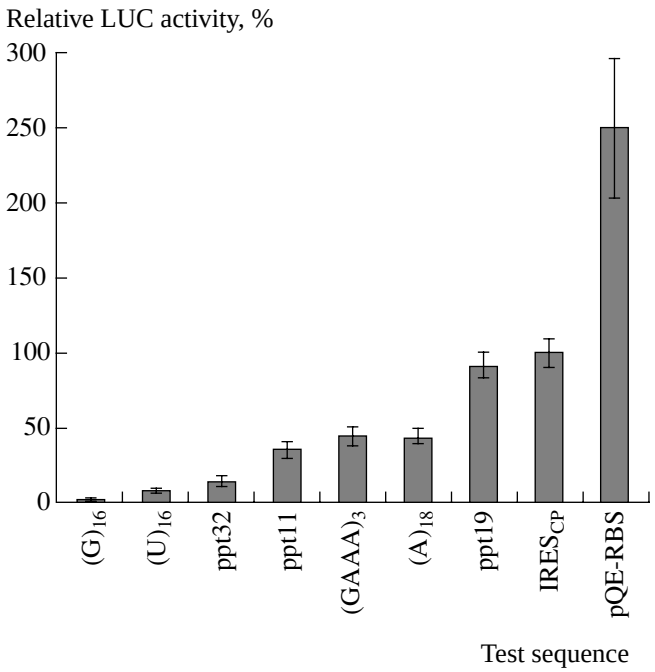

Fig. 2. Mean values of luciferase activity measurements in *E. coli* cells transformed with GFP-LUC constructs that contained different sequences upstream of LUC AUG. The $IRES_{CP,148}$ activity was taken as 100%.

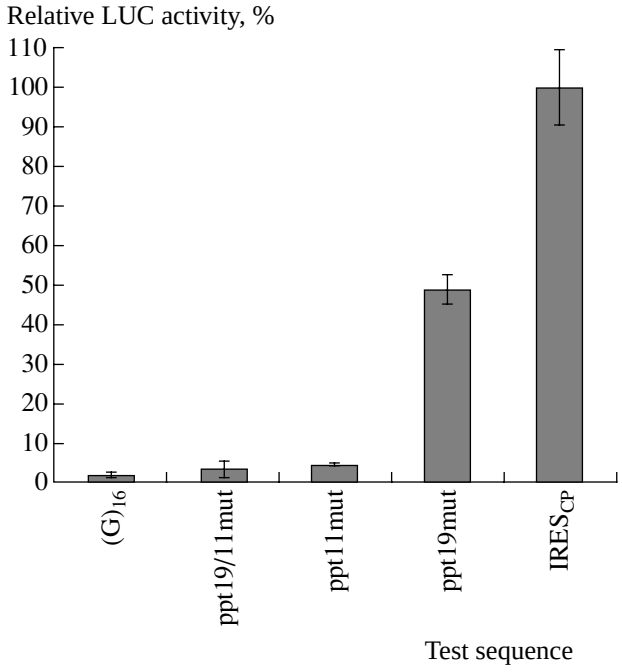

Fig. 3. Mean values of luciferase activity measurements in *E. coli* cells transformed with GFP-LUC constructs that contained $IRES_{CP,148}$ or its mutant variants as intercistronic spacers. The wild-type $IRES_{CP,148}$ activity was taken as 100%.

Nucleotide sequences of intercistronic spacers in GFP-LUC constructs. Polypurine tracts (ppt19 and ppt11) within $IRES_{CP, 148}$ are underlined. Nucleotides presumably responsible for SD interactions are boldfaced and underlined. The luciferase start codon is italicized. Nucleotide substitutions in $IRES_{CP, 148}$ mutant variants are lowercased. The additional box contains the sequence of the 16S rRNA 3'-terminus. Nucleotides responsible for SD interactions are boldfaced

Sequence name	Nucleotide sequence located upstream of LUC AUG
$IRES_{CP, 148}$	CAGCATTTAAAGCGGTTGACAACTTTAAAAGAAGGAAAAAGAAGGTTGAAGAAAAGGGTGTAGTAAG
	TAAGTATAAGTACAGACCGGAGAAGTACGCCGGTCCTGATTCGTTTAATTTGA**AAG**A**AG**AAACC***ATG***
$ppt19_{MUT}$	CAGCATTTAAAGCGGTTGACAACTTTAAA**c**G**c c**G**c c g c**AAG**c g c**GTTGAAGAAAAGGGTGTAGTAAG
	TAAGTATAAGTACAGACCGGAGAAGTACGCCGGTCCTGATTCGTTTAATTTGA**AAG**A**AG**AAACC***ATG***
$ppt11_{MUT}$	CAGCATTTAAAGCGGTTGACAACTTTAAAAGAAGGAAAAAGAAGGTTGAAGAAAAGGGTGTAGTA
	AGTAAGTATAAGTACAGACCGGAGAAGTACGCCGGTCCTGATTCGT**T**T**A**ATTT**G**c g c**GA**c**GACC***ATG***
$Ppt19/11_{MUT}$	CAGCATTTAAAGCGGTTGACAACTTTAAA**c**G**c c**G**c c g c**AAG**c g c**GTTGAAGAAAAGGGTGTAGTA
	AGTAAGTATAAGTACAGACCGGAGAAGTACGCCGGTCCTGATTCGT**T**T**A**ATTT**G**c g c**GA**c**GACC***ATG***
ppt11	AGCTTGAATTCTTGA**AAG**A**AG**AAACC***ATG***
ppt19	AGCTTGAATTCAAAAGAAGGAAA**AAG**A**AGGT**TAACC***ATG***
ppt32	AGCTTGAATTCTTAAAAGAAGGAAAAAGAAGGTT**GA**A**G**AA**A**AGGGAACC***ATG***
$(GAAA)_3$	AGCTTGAATTCTT**G**A**A**A**G**AA**A**GAAAAACC***ATG***
RBS of pQE30	AGCTTGAATTCAT**TAA**A**GAGG**A**GA**AATTAACC***ATG***
$(A)_{18}$	AGCTTGAATTCAA**AA**AA**A**AAAA**A**AAAAAACC***ATG***
$(T)_{16}$	AGCTTGAATTCTTTT**T**TTTTTTT**T**TT**T**AACC***ATG***
$(G)_{16}$	AGCTTGAATTCGGGGGGGG**GG**G**GG**G**G**AACC***ATG***
3'-terminus of 16S rRNA	**3'-AUUCCUCCACUA**GGUUGGCGUCCAAGGGGAUGCCAAUGGAACAA-5'

The translational activities of three polypurine tracts (ppt32, ppt19, and ppt11), structural components of $IRES_{CP, 148}$, were also studied. As seen from Fig. 2, the activities of ppt19 and the full-length $IRES_{CP, 148}$ were equal; however, the activity of ppt11 and artificial polypurine sequences ($(GAAA)_3$, $(A)_{18}$) are relatively low. In general, this correlates with the number of SD links that can be created by the corresponding sequence (table).

The very low efficiency of the ppt32 translation is noteworthy (Fig. 2). This sequence includes the ppt19 that exhibits the highest homology to pQE30 RBS. This phenomenon may be explained by the fact that ppt19 within ppt32 is situated distantly from AUG. Possibly, in this case, a 5'-proximal part of ppt32 is responsible for SD-interactions (table).

Three mutant variants of $IRES_{CP, 148}$ were cloned as intercistronic spacers to study the contribution of each polypurine tract into the overall translational activity of $IRES_{CP, 148}$ ($ppt19_{MUT}$, $ppt11_{MUT}$, and $ppt11/19_{MUT}$) (table). The activities of these GFP-LUC constructs are shown in Fig. 3. These data indicate that the key $IRES_{CP,148}$ element for prokaryotic translation is ppt11. It is the closest to the initiator-codon polypurine tract, and mutations in ppt11 reduced the $IRES_{CP,148}$ activity by more than ten times. However, the activity of ppt11 alone is more than two times lower than that of $IRES_{CP, 148}$ (Fig. 2). This allows us to suppose that some other element of $IRES_{CP, 148}$ is playing an enhancing role for its translatlonal activity.

Mutations in ppt19 decreased the $IRES_{CP, 148}$ activity by two times. Therefore, ppt19 is the element within the $IRES_{CP, 148}$ that stimulates the translation initiation provided by the ppt11 interaction with 16S rRNA.

The data obtained allow us to conclude that CrTMV $IRES_{CP, 148}$ is able to provide initiation of translation in *E. coli* by functioning as an RBS. Artificial polypurine sequences $(GAAA)_3$ and $(A)_{18}$, as well as structural elements of $IRES_{CP,148}$ (polypurine tracts), are also able to promote translation in prokaryotes. The translational activity of $IRES_{CP,148}$ is mainly determined by ppt11, but ppt19 stimulates the $IRES_{CP, 148}$ functioning as an RBS.

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research (project nos. 99-04-49119 and 02-04-48308).

REFERENCES

1. Shine, J. and Dalgarno, L., *Proc. Natl. Acad. Sci. USA*, 1974, vol. 71, p. 1341.
2. Kozak, M., *Microbiol. Rev.*, 1983, vol. 47, pp. 1–45.
3. Kozak, M., *Gene*, 1999, vol. 234, pp. 187–208.
4. Pelletier, J. and Sonenberg, N., *Nature*, 1988, vol. 334, pp. 320–325.
5. Ivanov, P.A., Karpova, O.V., Skulachev, M.V., *et al.*, *Virology*, 1997, vol. 232, pp. 32–43.
6. Skulachev, M.V., Ivanov, P.A., Karpova, O.V., *et al.*, *Virology*, 1999, vol. 263, pp. 139–154.
7. Dorokhov, Y.L., Skulachev, M.V., Ivanov, P.A., *et al.*, *Proc. Natl. Acad. Sci. USA*, 2002, vol. 99, pp. 5301–5306.
8. Gallie, D.R. and Kado, C.I., *Proc. Natl. Acad. Sci. USA*, 1989, vol. 86, pp. 129–132.
9. Ivanov, I.G., Alexandrova, R.A., Dragulev, B.P., *et al.*, *Eur. J. Biochem.*, 1992, vol. 209, pp. 151–156.
10. Loechel, S., Inamine, J.M., and Hu, P.C., *Nucl. Acids Res.*, 1991, vol. 19, pp. 6905–6911.
11. Glover, J.F. and Wilson, T.M.A., *Eur. J. Biochem.*, 1981, vol. 122, pp. 485–492.