Volume 307, number 2, 173–176 FEBS 11325 July 1992

Effect of tandemly repeated AGG triplets on the translation of CAT-mRNA in *E. coli*

I. Ivanov[a,b], R. Alexandrova[b], B. Dragulev[a], A. Saraffova[b] and M.G. AbouHaidar[a]

[a]*Department of Botany, University of Toronto, 25 Willcocks Street, Toronto, Ontario, Canada* and [b]*Institute of Molecular Biology, Bulgarian Academy of Sciences, 1113 Sofia, Bulgaria*

Received 28 May 1992

It has been shown that tandems of rare arginine codons AGG have a strong inhibitory effect on translation of mRNA in *E. coli* [5]. This has been explained by the rate-limiting interaction of these codons with the less abundant $tRNA^{AGG}$ [6]. In this study tandemly repeated AGG triplets were introduced into the chloramphenicol acetyltransferase (CAT) gene either upstream of the initiation ATG codon or downstream of it (both in frame and out of frame) and the expression of the modified genes was investigated. We report that the addition of AGG clusters resulted in a substantial inhibitory effect on CAT gene expression independently of their localization in mRNA. This inhibitory effect is explained by a competition of the tandem AGGAGG with the natural Shine–Dalgarno (SD) sequence (consensus AAGGAGGU) for the 3'-end of the 16S small ribosomal RNA (rRNA).

CAT gene; Gene expression; Genetic code; Codon usage; Rare codons; SD-sequence

1. INTRODUCTION

The triplet AGG, coding for arginine (Arg), is one of the most rare codons (due to the low abundancy of the corresponding tRNA) in *E. coli* and other prokaryotes [1,2] and the most frequently occurring in higher eukaryotes [1,3,4]. The high level of expression in *E. coli* (up to 30–40% of the total cellular protein) of numerous eukaryotic genes containing dispersed AGG codons indicates that the scattered AGG codons have no negative effect on the translation of heterologous gene in vivo. In order to study the effect of clusters of AGG codons on gene expression, Robinson et al. [5] constructed plasmids for expression in *E. coli* of the chloramphenicol acetyltransferase (CAT) gene modified by inserting of either 4 tandemly repeated AGG (rare) or the same number of CGT (preferential) Arg codons immediately after the initiaton ATG codon. No difference in the yield of CAT protein has been found when the two manipulated genes were expressed under weak promoters but a threefold difference was noticed when a strong tryptophan (Trp) promoter has been used. The attenuating effect of rare codon tandems on translation of highly expressed genes was explained by Varenne and Lazdunsri [6] by a rate-limiting interaction of AGG codons with the less abundant $tRNA^{AGG}$ and by sharp decrease in the concentration (pool) of charged tRNAs in the bacterial cytoplasm when both sites of the ribosome are occupied with identical rare codons.

Most reports (for review see [7]) tend to explain the inhibitory effect of the AGG triplets on translation by considering them as Arg codons only (i.e. as codons on the mRNA situated in frame downstream of the initiaton AUG codon). However, the interference of the nucleotide sequence $(AGG)_n$ itself (independently of its location in mRNA) with the translation process has not been assessed.

In this paper we investigate the effect of tandemly repeated AGG triplets on gene expression in *E. coli* in relation to their positions (upstream or downstream) of the initiation AUG codon in mRNA.

2. MATERIALS AND METHODS

2.1. *Bacterial plasmids and strains*

The expression plasmid pP_1-CAT was constructed by J. Rommens (Ph.D. Thesis, University of New Brunswick, Canada, 1987) and was used for the construction of the other plasmids shown in Fig. 1. *E. coli* LE392 cells (for phenotype see [8]) were transformed according to Hanahan [9] and used to study the expression of the manipulated CAT genes in vivo.

2.2. *Construction of modified CAT genes*

Synthetic oligonucleotides were prepared by automated gene assemblers (Applied Biosystems, USA and Pharmacia, Sweden) following the manufacture's instructions and were purified by HPLC on a Mono Q column (Pharmacia).

The non-translated (upstream of the initiation ATG codon) region of the CAT gene in vectors $pP_1B(AGG)_2$-CAT, $pP_1B(AGG)_4$-CAT, $pP_1B(GAG)_4$, $pP_1B\Omega$-CAT and $pP_1B(Xho)_2$-CAT was modified by cloning the corresponding oligonucleotides (see Fig. 1) in the *Hin*dIII site either directly or after DNA polymerase I (Klenow fragment)

Correspondence address: M.G. AbouHaidar, Department of Botany, University of Toronto, 25 Willcocks Street, Toronto, Ontario, Canada M5S 3B2. Fax: (1) (416) 978-5878.

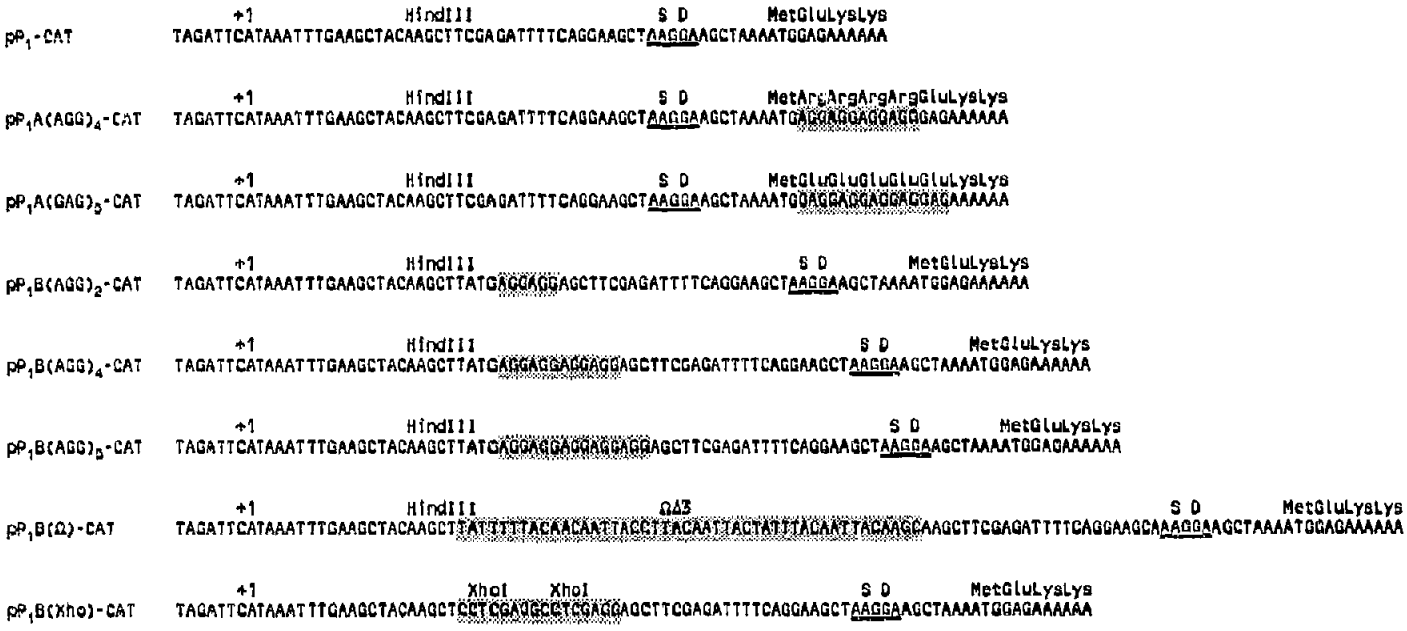

Fig. 1. Primary structures of the 5'-end of modified CAT genes. P_1: synthetic constitutive promoter [16] (partial sequence). (+1) represents the first nucleotide transcribed into mRNA. The natural SD sequence of the CAT gene is underlined. Shaded areas represent new sequences added to the original CAT gene in the plasmid pP_1-CAT. $\Omega\Delta3$ represents a derivative of the 5' non-coding region (omega) of TMV-RNA. In the two series of plasmids $pP_1A..$ and $pP_1B..$ the new added sequences are localized either after (A) or before (B) the initiation ATG codon.

treatment. To avoid fragment concatenation non-phosphorylated oligonucleotides were used.

For construction of the plasmids $pP_1A(AGG)_4$-CAT and $pP_1A(GAG)_5$-CAT, CAT genes were modified by insertion of the corresponding additional codons (AGG or GAG) in frame downstream of the initiation ATG codon by site directed mutagenesis. Instead of these codons, the oligonucleotides used for mutagenesis contained 20 more nucleotides (10 at each side of the initiation ATG codon). All enzymatic reactions, isolation of DNA fragments and site-directed mutagenesis were performed according to standard protocols [8].

Modified CAT genes were cloned in pP_1-CAT. Positive clones were selected by either DNA [10] or RNA [11] colony hybridization using ^{32}P-labelled oligonucleotides and the primary structure of the modified CAT genes was verified by DNA sequencing [12].

2.3. *CAT assay*

E. coli LE 392 cells transformed with the different plasmids shown in Fig. 1 were cultivated at the same time in LB (Luria–Bertani medium) inoculated with overnight cultures and grown at 37°C until the early stationary phase (8 h). Samples of 4 A_{590} units of cells were sonicated in 1 ml of 50 mM potassium phosphate buffer, pH 7.5, 1 mM EDTA and aliquots of 10 µl of the clear lysates were used for spectrophotometric determination of the CAT assay as described [13]. For each reaction usually three samples were run in parallel and the average values are presented in Table 1.

3. RESULTS AND DISCUSSION

The CAT gene is an appropriate model for studying the effect of AGG triplets on translation because of the following: (a) it is devoid of AGG codons [14]; (b) this gene can be expressed in *E. coli* under strong promoters to extremely high levels reaching 30–40% of total cellular protein [5,15] (J. Rommens, unpublished data); (c) there are sensitive methods to quantitate the yield [13].

To study the effect of the location in mRNA of tandemly repeated AGG triplets on CAT gene expression in *E. coli*, the gene was modified as follows (see Fig. 1): (a) four AGG codons were inserted in frame downstream of the initiation ATG codon. This construct ($pP_1A(AGG)_4$-CAT) resembles the construct pMG 802 of Robinson et al. [5]; (b) four GAG (preferential Glu

Table 1

Yield of CAT protein and stability of mRNA obtained from vectors containing modified CAT genes and expressed in *E. coli* LE 392.

Plasmid	Yield[a] of CAT protein (units/ mg protein)	Yield of CAT relative to control[b]	Free energy (ΔG) of mRNA (kJ)[c]
pP_1-CAT	8480	1.00	−406.9
$pP_1A(AGG)_4$-CAT	820	0.10	−400.2
$pP_1A(GAG)_5$-CAT	700	0.08	−399.3
$pP_1B(AGG)_2$-CAT	2800	0.33	−393.9
$pP_1B(AGG)_4$-CAT	2020	0.24	−376.3
$pP_1B(AGG)_5$-CAT	1120	0.13	−368.0
$pP_1B\Omega\Delta3$-CAT	10800	1.27	−333.7
$pP_1B(Xho)_2$-CAT	10600	1.25	−414.4

[a] Average values of 3 independent experiments determined as in section 2.

[b] Ratio between the yield found for the corresponding construction and that of the control plasmid pP_1-CAT.

[c] The free energy is calculated on the basis of the first 300 nucleotides of the mRNA.

codons in *E. coli*) were inserted in frame downstream of the initiation ATG codon (construct $pP_1A(GAG)_5$-CAT). These codons along with the natural GAG (the second after ATG) codon in the CAT gene form a sequence containing four AGG triplets. However, these triplets are out of frame to become GAG and code for Glu instead of Arg (Fig. 1); (c) in three constructs ($pP_1B(AGG)_2$-CAT, $pP_1E(AGG)_4$-CAT and $pP_1B(AGG)_5$-CAT) two, four and five AGG triplets were inserted respectively upstream from the natural SD site of the CAT gene.

The modified CAT genes were expressed under the strong constitutive promoter P_1 (synthetic analogue to the T5P25 early promoter [16]), and the data from the CAT assay are presented in Table I. The level of expression of all modified CAT genes described above was lower (4–10 times) as compared with that of the control construct (pP_1-CAT). The results from the CAT assay were also confirmed by polyacrylamide gel electrophoresis of total bacterial proteins (data not shown).

Our data support the results of Robinson et al. [5] showing that the insertion of 4 Arg (AGG) codons next to initiation ATG codon in the translated region of CAT gene has a strong inhibitory effect on its expression (Table I). However, we found that this effect occurred equally well when the tandemly repeated AGG triplets were inserted out of frame (and coded for Glu instead of Arg). Moreover, even when they were introduced in the non-translated region of the CAT gene upstream of the natural SD site. The insertion of increasing numbers of AGG triplets ($(AGG)_2$, $(AGG)_4$ and $(AGG)_5$) in the 5′-terminal non-coding region of the CAT gene resulted in a gradual decrease in the efficiency of its expression (Table I).

To explain the data obtained in this study we propose the following hypothesis. Apart from its coding function, the tandem of triplets AGGAGG resembles the major part of the bacterial SD consensus sequence (AAGGAGGU, as defined by Scherer et al. [17]). Our recent studies on the expression of genes in vectors containing multiple SD sequences showed that the level of gene expression decreased dramatically when two and particularly three tandemly repeated SD sequences were inserted in front of the initiation codon ([18] and unpublished data).

This phenomenon as well as the inhibitory effect of clusters of AGG codons might be explained by a competitive (with the functional SD sequence) interaction of mRNA with the 3′-end of 16S rRNA. This interaction and the dissociation of the resulting inefficient (false) complex is time-consuming and delays the formation of the functional SD-16S rRNA complex.

To check this hypothesis two more modifications of the CAT gene were prepared. In vector $pP_1B\Omega$-CAT a synthetic $\Omega\Delta3$ sequence (a modified 5′-end non-translated region of the tobacco mosaic virus RNA [19]) was inserted. As shown recently, this sequence can initiate translation of mRNA in *E. coli* [20]. Since it is devoid of any guanosines (G's) it cannot be a competitor of the functional SD sequence for the 3′-end of 16S rRNA. In plasmid pP_1 $(Xho)_2B$-CAT a 16 nucleotides long fragment (representing a dimer of an *Xho*I linker CCTCGAGG) was inserted. This fragment contains two AGG triplets separated by a 5 nucleotide spacer (Fig. 1). The new CAT gene constructs gave CAT protein at an even higher yield than the starting construct pP_1-CAT (Table I).

Theoretically, the variations in the yield of CAT protein might be due to more or less favourable changes in the secondary structure of mRNA. Secondary structure of mRNAs transcribed from the modified CAT genes (Fig. 1) was studied by a standard computer programme [21] and the free energies obtained are presented in Table I. More noticeable destabilization of the mRNA secondary structure was found for the construct $pP_1B(AGG)_5$-CAT and $pP_1B\Omega$-CAT. The increased productivity of the latter only could be explained by a destabilizing effect of the $\Omega\Delta3$ sequence on mRNA. However, the increased free energy of the construct $pP_1B(AGG)_5$-CAT (ΔG = −368 kJ) did not correlate with the decreased yield of CAT protein, nor the ΔG values of other CAT constructs (Table I).

The data presented in this paper do not exclude an eventual negative effect of tandems of rare AGG codons on the translation of mRNA in *E. coli* when they are in frame and code for Arg [5,6]. However, our results demonstrated clearly that tandemly repeated AGG triplets have strong inhibitory effect on gene expression independently of their localization in mRNA (upstream or downstream) in or out of frame with the initiation ATG codon. This finding is important and should be taken into consideration when eukaryotic genes (where AGG is a preferential Arg codon and the AGGAGG tandems are frequently occurring) are designed for expression in bacteria and a high yield of recombinant protein is expected.

Acknowledgements: We thank Dr. J. Rommens, Department of Genetics, Hospital for Sick Children, Toronto, Canada for providing us with the CAT expression plasmid pP_1-CAT. This work was supported in part by grants from the NSERC and from the Ontario Ministry of Food and Agriculture to M.G. AbouHaidar and from the Ministry of Industry and Technology, Bulgaria to I.G. Ivanov.

REFERENCES

[1] Grantham, R., Gautier, C., Jacobzone, M. and Mercier, R. (1981) Nucleic Acids Res. 9, 143–173.
[2] Grosjean, H. and Fiers, W. (1982) Gene 18, 199–209.
[3] Bennetzen, J.L. and Hall, B.D. (1982) J. Biol. Chem. 257, 3026–3031.
[4] Hastings, K.E.M. and Emerson, C.P. (1983) J. Mol. Evol. 19, 214–218.
[5] Robinson, M., Lilley, R., Little, S., Emtage, J.S., Yarranton, G., Stephens, P., Millican, A., Eaton, M. and Humphreys, G. (1984) Nucleic Acids Res. 12, 6663–6671.

[6] Varenne, S. and Lazdunski, C. (1986) J. Theor. Biol. 210, 99–110.
[7] deBoer, H.A. and Kastelein, R.A., in: Maximizing Gene Expression (W. Resnikoff and L. Gold, Eds.), Butterworth Publ., Stoneham, MA, 1986 pp. 225–285.
[8] Sambrooke, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular cloning: A Laboratory Manual, Cold Spring Harbor, New York.
[9] Hanahan, D. (1983) J. Mol. Biol. 166, 557–580.
[10] Grunstein, M. and Hogness, D. (1975) Proc. Natl. Acad. Sci., USA 78, 6091–6095.
[11] Ivanov, I. and Gigova, L. (1986) Gene 46, 287–290.
[12] Hsiao, K.-C. (1991) Nucleic Acids Res. 19, 2787.
[13] Show, W.V. (1975) Methods Enzymol. 43, 737–755.
[14] Alton, N.K. and Vapnek, D. (1979) Nature 282, 864–869.
[15] Bennett, A.D. and Show, W.V. (1983) Biochem. J. 215, 29–38.
[16] Rommens, J., MacKnight, D., Pomeroy-Cloney, L. and Jay, E. (1983) Nucleic Acids Res. 11, 5921–5940.
[17] Scherer, G., Walkinshaw, M., Arnott, S. and Morre, D. (1980) Nucleic Acids Res. 8, 3894–3907.
[18] Alexciev, K., Uscheva, A., Pavlova, M., Yavachev, L. and Ivanov, I. (1989) Int. J. Biochem. 21, 987–996.
[19] Gallie, D.R., Sleat, D.E., Watts, J.W., Turner, P.C. and Wilson, T.M. (1988) Nucleic Acids Res. 16, 883–893.
[20] Gallie, D.R. and Kado, C.I. (1989) Proc. Natl. Acad. Sci. USA 86, 129–132.
[21] Devereux, J., Haeberly, P. and Smithies, O. (1984) Nucleic Acids Res. 12, 387–397.