Self-Initiation of Translation of mRNAs Devoid of Translational Initiators in *Escherichia coli*

(chloramphenicol acetyltransferase / leaderless mRNA / Shine-Dalgarno sequence / translation initiation)

V. KOLEV[1], A. BERZAL-HERRANZ[2], I. IVANOV[1]

[1]Department of Gene Regulations, Institute of Molecular Biology, Bulgarian Academy of Sciences, Sofia, Bulgaria

[2]Department of Molecular Biology, Instituto de Parasitologia y Biomedicina "Lopez-Neyra", Consejo Superior Investigaciones Cientificas, Granada, Spain

Abstract. Recent studies have shown that the canonical SD-anti-SD interaction is dispensable for the initiation of translation of certain mRNAs in *Escherichia coli*. In this study the *cat* and *tetR* genes were modified to either destroy complementarity to *E. coli* 16S RNA or completely delete their 5' non-translated regions. Thus a series of *cat*- and *tetR*-derived genes were constructed, cloned under a strong constitutive promoter and expressed in *E. coli* cells. The efficiency of expression was evaluated by the yield of CAT (for the *cat* gene) and cell viability in increasing concentrations of antibiotic (for both *cat* and *tetR* genes). The obtained results show that the mRNAs transcribed from both series of reporter genes (*cat* and *tetR*) were active *in vivo*. Their activity was preserved even in the cases when the length of their 5' non-translated leader sequences was reduced to one nucleotide for the *cat* gene and eight nucleotides for the *tetR* gene. The yield of protein obtained with the latter constructs was detectable and sufficient for bacteria to survive at 50–100 μg/ml chloramphenicol and 20 μg/ml tetracycline, respectively.

Translation initiation in prokaryotes includes base pairing between a purine-rich domain at the 5' non-translated (leader) region of mRNA known as Shine-Dalgarno (SD) sequence and the 3' end of 16S rRNA (McCarthy and Gualerzi, 1990). There are examples, however, showing that this mechanism might be dispensable. The leader sequences of some natural *E. coli* mRNAs are either too short or completely devoid of SD sequences (Porter and Hindley, 1972; Christie and Calendar, 1985; Klock and Hillen, 1986). The Shine-Dalgarno model also cannot explain the initiation of translation driven by sequences derived from non-translated gene regions of tobacco mosaic virus RNA, known as the Ω sequence (Gallie et al., 1987; Gallie et al., 1989; Ivanov et al., 1992a; Ivanov et al., 1995), *Mycoplasma genitalium* (Loechel et al., 1991), phage f1 gene *VII* (Ivey-Hoyle and Steege, 1992), T7 gene *10* (Dunn and Studier, 1983; Olins and Rangwala, 1989), *E. coli dnaG* gene (Smiley et al., 1982), ε sequence (Golshani et al. 1997; Golshani et al. 2000a; Golshani et al. 2000b), etc., which are either devoid of G nucleotides or do not share any homology with the SD consensus sequence (AAGGAGGU). However, the most inconsistent with the SD model is the initiation of translation of some leaderless mRNA such as those of phage λ gene *cI* (Walz et al., 1976), phage P2 gene *V* (Christie and Calendar, 1985), etc.

In our laboratory we have extensively used the chloramphenicol acetyltransferase (*cat*) gene as a reporter for studying the translation initiation efficiency of non-SD sequences in *E. coli* cells. As already mentioned (Ivanov et al., 1992a; Ivanov et al., 1992b; Ivanov et al., 1995) we have always had difficulties in preparing null controls with a zero level of *cat* gene expression. Even the constructs containing a *cat* gene devoid of any specific translational initiator allowed *E. coli* cells to survive at moderate concentrations of chloramphenicol. At the same time other (eukaryotic) reporter genes placed in the expression vectors under identical conditions did not give detectable amounts of gene-specific proteins. The latter suggested that the translation of CAT mRNA in the null constructs might be initiated by some unknown mechanism.

To study the effect of the length and complementarity to 16S rRNA of the 5' non-translated leader sequence on the efficiency of translation of mRNA in *E. coli*, we have mutated in this study two reporter (*cat* and *tetR*) genes to either destroy such complementarity or to shorten their 5' non-translated regions. We have observed an extremely low, but still detectable amount

Received June 4, 2001. Accepted June 7, 2001.

This study is supported by grant CRP/BUL98-07 from the International Centre for Genetic Engeneering and Biotechnology (United Nations Industrial Development Organization).

Corresponding author: Ivan G. Ivanov, Institute of Molecular Biology, Bulgarian Academy of Sciences, "Acad. G. Bontchev", bl. 21, 1113 Sofia, Bulgaria. Tel./Fax: 359 (2) 73 62 27; e-mail: iviv@obzor.bio21.bas.bg.

Abbreviations: *cat* – chloramphenicol acetyltransferase, LB – Luria Bertani, SD – Shine-Dalgarno, *tetR* – tetracycline resistance.

of protein with both series of mutated reporter genes. The protein levels were enough for the cells to survive at chloramphenicol and tetracycline concentrations of 50–100 μg/ml and 20–30 μg/ml, respectively.

Material and Methods

cat and tetR gene modifications

Both *cat* and *tetR* genes were derived from the cloning plasmid pBR325 (Bolivar, 1979) and modified by PCR as shown in Fig. 1A.

The primary structure of all gene constructs was verified after cloning using DNA sequencing kits.

DNA primers for PCR and sequencing were synthesized in a Cyclon 7000 DNA synthesizer (Milligene).

Construction of expression plasmids

The series of *cat* and *tetR* derivative genes were cloned in a pBR322-based expression plasmid under a strong constitutive synthetic promoter. The P_1 promoter (analogue of the T5P25 early promoter) has been originally synthesized by Rommens et al. (1983). In order to be used for serial cloning, a *Xho*I cloning site (missing in the pBR322) was further introduced at the 3' end of this promoter in such a way that the first nucleotide to be transcribed (nucleotide +1 in the corresponding mRNA) would be the first nucleotide of the *Xho*I site (C). The PCR-derived *cat* and *tetR* genes were thereafter cloned between the *Xho*I and *BamH*I sites as shown in Fig. 1B.

When preparing the construct L_6-CAT we have identified clones in which the *Xho*I site was missing. The latter was due to a five-nucleotide spontaneous deletion between the P_1 promoter and the *cat* gene, and this clone was designated as L_1-CAT.

Evaluation of gene expression

CAT assay

E. coli LE392 cells transformed with the series of *cat* gene expression plasmids were cultivated in Luria-Bertani (LB) medium supplemented with 50 μg/ml ampicillin at 37°C until an early stationary phase (A_{590} = 2.0). Bacterial cells (5 ml bacterial suspension) were harvested, washed out with 150 mM NaCl, resuspended in 0.5 ml of TE buffer (10 mM Tris-HCl, pH 7.0, 1 mM EDTA), 0.1% SDS, and sonicated. Clear lysates were obtained by centrifugation, and the concentration of CAT was determined by both enzyme-linked immunosorbent assay (ELISA), using monoclonal CAT antibodies, and CAT assay (Golshani et al., 1997).

Resistance to antibiotics

Resistance to chloramphenicol or tetracycline was measured as follows: *E. coli* cells transformed with *cat* or *tetR* gene expression plasmids were grown overnight in LB supplemented with 50 μg/ml ampicillin and diluted 1 : 10000 with LB (free of antibiotic). Samples of 10 μl were spread onto agar plates containing LB supplemented with increasing concentrations of antibiotic (0–1000 μg/ml chloramphenicol or 0–100 μg/ml tetracycline) and incubated overnight at 37°C. The cells were considered resistant if more than 50 colonies per plate were counted.

CAT and tetR mRNA measurements

The yield of CAT and tetR mRNA in *E. coli* LE392 cells was measured by hybridization. To this end equal amounts of total RNA isolated by standard procedures were loaded on nitrocellulose filters and hybridized with a ^{32}P-labelled oligonucleotide specific for either *cat* or *tetR* gene as already described (Ivanov et al., 1992a).

A.

	***Xho*I**	***Met***
	+1	
R_9-CAT/Tc^R	ctcgagtactc**AAGGAGGT**ttaagctt***ATG***.....	
SD_n-CAT/Tc^R	ctcgagtactac**AGGA**ttaagctt***ATG***.....	
L_{27}-CAT/Tc^R	ctcgagtactctctactattaagctt***ATG***.....	
L_{PolyA}-CAT/Tc^R	ctcgagtaaaaaaaaaaaaaaaaaaaa***ATG***.....	
L_{PolyT}-CAT/Tc^R	ctcgagattttttttttttttttttttt***ATG***.....	
L_8-CAT/Tc^R	ctcgagct***ATG***.....	
L_6-CAT/Tc^R	ctcgag***ATG***.....	
L_1-CAT	c***ATG***.....	

B.

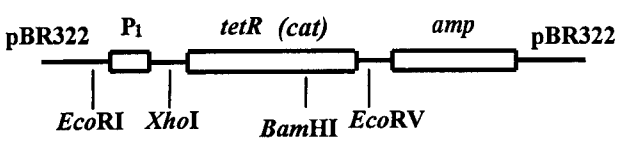

Fig. 1. A) Primary structure of the 5' terminal regions of derivative *cat* and *tetR* gene constructs. Nucleotides shown as capital bold letters correspond to the canonical (either consensus R_9 or *cat* gene native SD_n*) Shine-Dalgarno sequences; nucleotides in italics indicate the initiation Met codon in both *cat* and *tetR* genes; +1 is the first 5' terminal nucleotide in mRNA. B) Schematic diagram of the plasmids for expression of *cat* and *tetR* derivative genes.
*SD_n does not mean the natural 5' non-translated *cat* gene region like in the plasmid pBR325 or Tn9.

The stability (half-life) of CAT mRNA was determined after double-block of transcription by rifampicin and naldixic acid as described earlier (Gigova et al., 1989).

Results

The *cat* gene is routinely used as a reporter for studying efficiency of transcription and translation in both prokaryotes and eukaryotes. In *E. coli* the *cat* gene is expressed to moderate levels under its natural promoter and SD sequence and is capable of reaching extremely high expression levels (up to 50% of the total bacterial protein) when placed under a strong promoter and strong SD sequence (Chen et al., 1994).

Expression of modified cat genes

To study the effect of length and nucleotide composition of the 5' non-translated region on initiation of translation of CAT mRNA in *E. coli* cells, the wild-type *cat* gene was mutated as shown in Fig. 1A. Two of the constructs, carrying a consensus (R_9) and the *cat* gene natural (SD_n) SD sequences, were used for evaluation of the upper levels of *cat* gene expression. In both constructs the SD sequence was placed 8 nucleotides before the initiation (AUG) codon as a part of a 27 nucleotide-long non-translated region. Three of the constructs, L_{27}-CAT, L_8-CAT and L_6-CAT bore neutral (with no homology to the 16S rRNA) leader sequences of 27, 8 and 6 nucleotides, respectively. Two constructs were designed to carry 20 nucleotide-long homopolymeric tracks of either poly-A (L_{PolyA}-CAT) or poly-T (L_{PolyT}-CAT) in a 27 nucleotide-long non-translated leader. The L_1-CAT was obtained accidentally (see above) and bore one nucleotide only before the initiation (AUG) codon. In our study this construct is considered as leaderless.

All *cat* gene derivative constructs were placed under the strong constitutive promoter P_1 (Fig. 1B) and the level of expression was measured by both CAT assay and cell viability at different concentrations of chloramphenicol.

Taking into consideration that minor changes in gene structure could sometimes interfere with the efficiency of transcription and stability of mRNA, both half-life and yield of mRNA were also measured. Our results show that whereas the stability of mRNA originating from the different gene constructs was almost the same (estimated half-life of about 60–65 seconds), its yield was variable. Because of that, for a precise evaluation of the efficiency of translation, the yield of CAT was normalized to that of corresponding mRNA. Table 1 shows that the protein yield was very much dependent on the length and nucleotide composition of the 5' non-translated region of CAT mRNA. For instance, when the natural SD sequence (SD_n) was substituted for a stronger one (R_9) preserving the same distance to the initiation codon and the same length of the leader sequence, the protein yield decreased five times. Surprisingly, it did not reach the zero level even upon

Table 1. Expression of cat and tetR derivative genes in E. coli LE392.

Construct	Relative protein yield[a]		Relative mRNA yield[b]	Protein yield related to mRNA[c]		Antibiotic resistance
	ELISA	Enzymatic		ELISA	Enzymatic	(μg/ml)
R_9-CAT	1.000	1.000	1.00	1.000	1.000	>800
SD_n-CAT	0.220	0.200	0.90	0.244	0.222	>800
L_{27}-CAT	0.040	0.035	0.88	0.045	0.040	>600
L_{polyA}-CAT	0.015	0.012	0.85	0.018	0.014	>200
L_{PolyT}-CAT	0.125	0.118	0.90	0.139	0.131	>600
L_8-CAT	0.014	0.012	0.78	0.018	0.015	>100
L_6-CAT	0.010	0.011	0.75	0.013	0.014	>50
L_1-CAT	0.006	0.006	0.70	0.008	0.008	>50
R_9-Tc^R	–		1.00	–		>100
SD_n-Tc^R	–		0.95	–		>100
L_{27}-Tc^R	–		0.90	–		>40
L_{PolyA}-Tc^R	–		0.85	–		>30
L_{PolyT}-Tc^R	–		0.90	–		>40
L_8-Tc^R	–		0.80	–		>20
L_6-Tc^R	–		0.78	–		>12

[a] Average values of CAT yield determined by ELISA or enzymatic CAT assay from three independent experiments for each construct are normalized to that of the referent construct R_9-CAT. The real yield taken as 1.000 in the table corresponds to 80–85 mg per liter of bacterial culture (cell density of about 4 A_{590} units per ml).

[b] Average values of mRNA yield determined by hybridization of total cellular RNA with *cat* or *tetR* gene-specific DNA probes from three independent experiments are normalized to that of the referent construct R_9-CAT.

[c] To normalize the yield of protein to that of mRNA, the values from the second column are divided by those in the third column.