

# Effects of GUG and AUG initiation codons on the expression of *lacZ* in *Escherichia coli*

A.C. Looman and P.H. van Knippenberg\*

*Department of Biochemistry, State University of Leiden, Wassenaarseweg 64, 2333 AL Leiden, The Netherlands*

Received 2 January 1986

We have replaced the ribosomal binding site (RBS) of the *lacZ* gene of *E. coli* by those of the maturation (A) gene of phage MS2 and that of the *tufA* gene. Both RBSs contain a GUG initiation codon. The expression with the *tufA* RBS is at least 25-fold higher than with the phage RBS. Changing the GUG into AUG results in a 3-fold increase in expression in both cases. In general, higher expression is accompanied by an increase of *lac*-specific mRNA. It is argued that this is a consequence of the more efficient translation of the mRNA.

<i>Site-directed mutagenesis</i>	<i>tufA</i> gene	<i>Phage MS2</i>	<i>Maturation protein</i>	<i>Initiation codon</i>
	<i>Translation efficiency</i>		<i>Ribosomal binding site</i>	

## 1. INTRODUCTION

Several triplets can function in *E. coli* as codons for the initiation of protein synthesis [1,2]. AUG, coding for methionine, is the most abundant triplet used for initiation, but several structural genes start with GUG. In a few cases UUG, AUA and AUU are used in vivo for (re-)initiation of protein synthesis [1,2].

It has been suggested that the use of codons other than AUG is coupled to low expression of the gene in question, or that it is compensated by a more extensive Shine-Dalgarno interaction [1]. In those cases where an AUG initiation codon was altered to another codon by mutation, the expres-

sion of the gene was either abolished or strongly reduced [2,3–6], even when such mutations lead to GUG or UUG [7,8]. On the other hand, at least one highly expressed gene in *E. coli*, i.e. the *tufA* gene, starts with GUG [14].

Recently we described a genetic system in which we can monitor the expression of the *lacZ* gene of *E. coli* as a function of an (inserted) heterologous RBS [9]. In this system we have introduced RBSs which contain a GUG initiation codon and then measured the  $\beta$ -galactosidase fusion protein production upon induction of the *lac* operon. The same was done after the GUG triplets had been changed to AUG by site-directed mutagenesis.

## 2. MATERIALS AND METHODS

### 2.1. Bacterial strains and plasmids

These are listed in table 1. Bacterial growth, cloning procedures, and assay of fusion protein synthesis were as described [9]. [<sup>3</sup>H]RNA was hybridized to M13 ssDNA containing an insert of 1607 bp from *HinfI* 1547–3154 in pKL203, comprising the C-terminal part of the *lacZ* gene and measured as described [9].

\* To whom correspondence should be addressed

**Abbreviations:** Ap, ampicillin; bp, basepair(s); dNTPs, deoxyribose nucleoside 5'-triphosphates; ds, double-stranded; Km, kanamycin; nt, nucleotide(s); <sup>R</sup>, resistance; rATP, ribose adenosine 5'-triphosphate; RBS, ribosomal binding site; rf, replicative form; ss, single-stranded; 1 × SSC, 0.15 M NaCl, 0.015 M Na citrate, pH 7.2

Table 1  
Bacterial strains and plasmids

Number	Structure	Origin or reference
<i>E. coli</i>		
KMBL1164	$\Delta lac-pro, supE, thi$	this laboratory
KMBL1164[F']	$\Delta lac-pro, supE, thi, F' lacI^R, lac-pro$	this laboratory
JM101	$\Delta lac-pro, supE, thi, F' traD36, lacI^R, \Delta lam15, pro$	[10]
Phages and plasmids		
M13MP8	phage M13 containing an insert of part of <i>lacZ</i> and a polylinker sequence	[11]
pKL203	expression vector containing the <i>lacZ</i> gene (except for its translation initiation signals), <i>lacY, A, Ap<sup>R</sup></i>	[9]
pKL313	plasmid pKL203 with phage MS2 maturation protein RBS (GUG initiation codon)	[9]
pKL313M	as pKL313 with AUG initiation codon	this paper
pKL330	plasmid pKL203 with <i>tufA</i> RBS (GUG initiation codon)	this paper
pKL330M	as pKL330 with AUG initiation codon	this paper
pGP82	plasmid containing part of EF-G, EF-TuA, Km <sup>R</sup> and Ap <sup>R</sup>	[12]

## 2.2. Construction of pKL313

pKL313 [9] was constructed by inserting a filled-in *EcoRI* 103 to *RsaI* 144 fragment from the MS2 maturation protein into the *SmaI* site of pKL203 (numbering according to [13]).

## 2.3. Construction of pKL330

The *tufA* RBS was isolated from plasmid pGP82 [12]. An *EcoRI* 1 to *HpaI* 178 fragment was isolated from a polyacrylamide gel (numbering according to [14]) digested with *HaeIII* and ligated into the *SmaI* site of pKL203. The resulting plasmid pKL330 contained an *HaeIII* 43 to *HpaI* 178 fragment, bearing the *tufA* RBS in frame with the *lacZ* gene of pKL203.

## 2.4. Mutagenesis of GUG initiation codons

The *EcoRI*-*BamHI* fragments of pKL313 and pKL330, comprising the RBSs (cf. fig.1) were recloned into M13MP8 [11]. Two oligonucleotides were used to prime the second strand synthesis: for the DNA with the *tufA* insert: 3'-TATCGGTACAGATTTp-5'; for the MS2 insert: 3'-ACTGGATACGCTCGAp-5' (underlined are the mismatching nucleotides).

The oligonucleotides were incubated with the ssDNA at 95°C in ligase buffer [15] for 5 min after which dNTPs, T<sub>4</sub> DNA ligase, rATP (10 mM) and Klenow fragment DNA polymerase were added

and incubation took place at 14°C overnight. Reaction mixtures were transformed to JM101. From plates containing about 100 plaques, 2 replicas were made on nitrocellulose filters (Schleicher and Schüll BA85) impregnated with 6×SSC. Replicas were washed at room temperature with 0.5 M NaOH, 1.5 M NaCl for 3 min; 3 M NaCl, 0.5 M Tris (pH 7.5) for 2 × 15 min and with 2×SSC for 5 min. The dried filters were baked for 2 h under vacuum at 80°C, and then incubated in hybridization buffer [0.1 mM rATP, 1 mM P<sub>i</sub>, 1 mM PP<sub>i</sub>, 800 mM NaCl, 1×C Denhardt's solution, 0.5% Nonidet P-40, 0.2 µg/ml tRNA, 100 mM Tris (pH 7.5), 6 mM EDTA] for 2 h at room temperature after which the homologous <sup>32</sup>P-labeled oligonucleotides were added and incubation continued at room temperature overnight. Selective washing of the 2 replica filters was for 30 min in 6×SSC at room temperature; 2 × 30 min in 6×SSC, 0.1% SDS and 30 min in 6×SSC at 25 and 30°C for the *tufA* oligonucleotide and 30 and 35°C for the MS2 maturation protein oligonucleotide. After autoradiography, possible positive plaques were selected for sequencing. One out of 8 positive plaques of the MS2 construct and 3 out of 4 of the *tufA* construct had picked up the mutation. The sequences of both wild-type and mutant RBS are shown in fig.1.

### 2.5. Recloning of mutant RBS

The *EcoRI*-*Bam*HI fragments containing the mutant RBS were isolated from M13 rf dsDNA and ligated into the *EcoRI*-*Bam*HI site of pKL203.

#### pKL313 (MS2 maturation protein RBS)

5'-AAUUGUGAGCGGAUAACAUUUCACACGGAUU

CCCAAUUCCAUCCUAGGAGGUUGACCUGUG CGA  
A arg

GCU UUU AGU GGG GAU CCC  
ala phe ser gly asp pro

#### pKL330 (*tufA* RBS)

5'-AAUUGUGAGCGGAUAACAUUUCACACGGAUU

CCCCGUAUUGAAGCCCGUGGUAAUAGCCUAAG

GGUUAUACCAAAGUCCCGUGCUCUCUCCUGAAGGG

GAGAGCACUAUAGUAAGGAUAUAGCCGUG UCU  
A ser

AAA GAA AAA UUU GAA CGU ACA AAA CCG  
lys glu lys phe glu arg thr lys pro

CAC GUU GGG GAU CCC  
his val gly asp pro

Fig.1. Nucleotide (nt) sequences of the constructed plasmids and mutants. Nucleotides are numbered from the 5'-end of the mRNA. Shine-Dalgarno sequences are underlined. The amino acid sequence is given for the coding part of the mRNA. The sequence inserted in pKL203 is in italics. pKL313: MS2 maturation protein sequence is inserted into vector pKL203 from nt 37-77. Site of the initiation codon mutation in pKL313M is nt 63. Restriction sites for *Eco*RI are at nt 29 and for *Bam*HI at nt 80. pKL330: *tufA* sequence is inserted into vector pKL203 from nt 37-171. Site of the initiation codon mutation is at nt 133. Restriction sites for *Eco*RI are at nt 29 and for *Bam*HI at nt 174. At nt 114 our sequence is in disagreement with the sequence given in [14] where these authors give an A (overbar).

Red transformants of KMBL1164 on MacConkey lactose-ampicillin plates containing the insertion of RBS and correctly oriented promoter fragment were designated pKL313M and pKL330M.

To measure induced expression of the *lacZ* fusions, plasmids were transformed to KMBL1164[F'].

All constructions were sequenced after insertion of the *EcoRI*-*Bam*HI fragment into M13MP8.

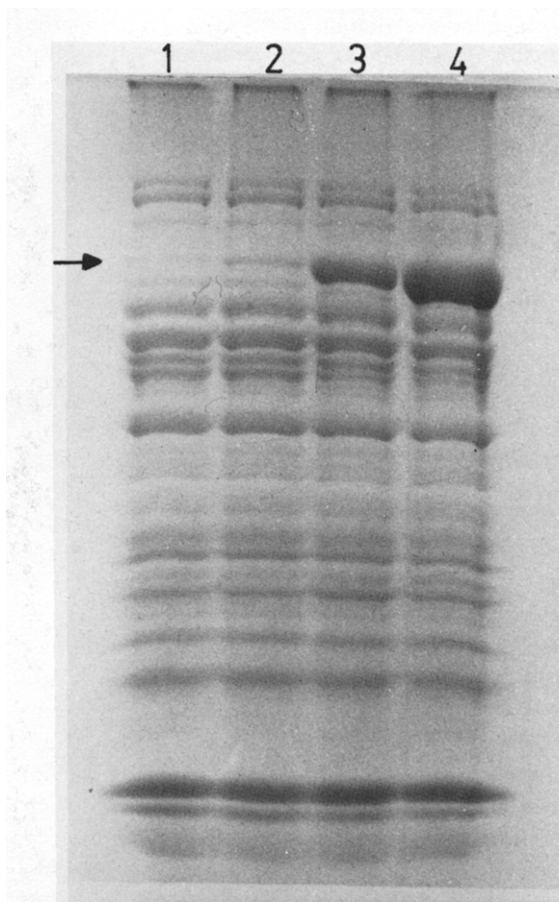


Fig.2. Coomassie brilliant blue-stained protein gel of mutant and wild-type RBSs. Lysates were made from logarithmically growing cells in minimal medium. Plasmid-coded fusion protein production in KMBL1164[F'] was induced for 45 min with 1 mM isopropyl thiogalactopyranoside prior to determination. Protein was made soluble by freeze-thawing in the presence of lysozyme, DNase and RNase [9]. Lanes: 1, pKL313 (GUG); 2, pKL313M (AUG); 3, pKL330 (GUG); 4, pKL330M (AUG). The arrow indicates the position of the fusion protein.

Table 2  
Fusion protein and mRNA levels

Plasmid	Origin RBS	Codon	% protein	% mRNA
pKL313	MS2 maturation protein	GUG	0.2	0.1
pKL313M	MS2 maturation protein	AUG	0.5	0.1
pKL330	<i>tufA</i>	GUG	5.6	0.7
pKL330	<i>tufA</i>	AUG	15.6	2.6

% protein: % protein per total trichloroacetic acid-precipitable protein in the lysate determined by immunoprecipitation of  $^{35}\text{S}$ -labeled fusion protein [9]. % mRNA: % [ $^3\text{H}$ ]mRNA hybridized to a *lac*-specific DNA fragment containing M13 ssDNA per total input of RNA [9]. Determinations were performed after 45 min induction with 1 mM isopropyl thiogalactopyranoside in strain KMBL1164[F']

### 3. RESULTS

Fig.1 illustrates the nucleotide sequences of the constructs containing the heterologous RBSs in the *lacZ* gene and also indicates the position of the G→A mutations which alter the GUG initiation codons to AUG codons.

The level of expression of the *lacZ* gene in the various constructs was investigated after 45 min induction with IPTG. No fusion proteins were produced in the absence of the inducer. This control also supports the fact that the putative second promoter in the *tufA* gene [14] is not included in the construction.

Fig.2 is a photograph of a stained SDS slab gel after electrophoresis of complete lysates. We have shown earlier [9] that determination of expression using  $\beta$ -galactosidase activity of fusion proteins can be hazardous, due to the instability of the enzymes. However, since the fusion proteins are not degraded during thermal inactivation [9], their amount is a good measure of the level of expression. Table 2 shows the amounts of fusion proteins as measured by the previously described immunological technique [9]. Both methods demonstrate clearly that there is a large difference in fusion protein production between the constructs containing the RBS of MS2 A protein (lanes 1,2 in fig.2) and those containing the *tufA* RBS (lanes 3,4). More quantitatively, table 2 shows that the *tufA* RBS is roughly 25–30-fold more active than the MS2 RBS, in either the GUG or AUG mode.

In both constructs the level of expression is increased approx. 3-fold when the GUG codon is

changed to AUG. As table 2 demonstrates, the higher level of expression of the *tufA* vs MS2 A construct is accompanied by a higher level of *lac*-specific mRNA production. The 3-fold increase in protein production observed by alteration of GUG into AUG in the *tufA* construct is similarly matched by an increase in mRNA production. In the experiment of table 2 no increase in mRNA production is seen when the phage RBS is mutated to AUG. In other experiments, however, using a strain that lacks the *lac* repressor (constitutive in KMBL1164), the level of mRNA increased also by a mutation of GUG into AUG in this construct (not shown).

### 4. DISCUSSION

The tremendous difference in efficiency with which the two GUG-containing RBSs allow expression of *lacZ* is not readily explained. Although the mRNA levels approximately parallel the amount of protein produced (table 2), it is very unlikely that this is due to a difference in promoter activity since the transcription initiation site is some 40 nucleotides removed from the region where the nucleotide sequences are at variance. In agreement with our previous results [9], we find that low expression of the *lacZ* gene (MS2 A protein RBS), is paralleled by a low amount of *lacZ* specific mRNA and that high expression (*tufA* RBS) is accompanied by a high level of the mRNA. It is our contention that the levels of mRNA in these constructs are determined by the efficiency with which these RNAs are used for translation. The structure of the RBS influences the rate of initiation of transla-

tion and this in turn has an effect on either the rate of synthesis of the mRNA or on the rate of degradation, or on both processes [16–18]. In any event, it is unlikely that the level of mRNA is independently determined by an effect of the nucleotide sequence in the RBS on transcription. This is especially clear when one compares the mRNA levels in the constructs containing the *tufA* gene RBS in the AUG and GUG mode (table 2). It is entirely reasonable that this one nucleotide alteration has a primary effect on the rate of translation initiation and that as a result of this the mRNA is either synthesized more rapidly and/or protected better against degradation.

Because of the effects of alterations in the RBS on the level of mRNA it would be incorrect to quantify the translation efficiency of the mRNAs in terms of protein produced per mRNA. This would only make sense if the mRNA level were a parameter independent of translation. In fact, the efficiency of an RBS could already be estimated from measuring the amounts of mRNA, since these reflect the efficiency of translation.

The two GUG-containing RBSs are very similar in the extent of their Shine-Dalgarno sequence and in the spacing between this sequence and the GUG codon (fig.1). Nevertheless, in their original context these sites appear quite different in efficiency. The MS2 A protein is present at a very low level after phage infection [19] and its expression is also low from an individually cloned gene [20]. In contrast, expression of the genes for EF-Tu, especially *tufA* [21,22], is very high. As our studies show, the ability to direct initiation of protein synthesis is in this case an inherent property of the RBS. Using a computer method [23], we have analyzed the potentiality of the initiation regions of both constructs to form secondary structures. The MS2 A RBS has more possibilities to base pair the Shine-Dalgarno region and the initiation codon with other regions of the mRNA than has the *tufA* RBS (not shown). Whether this may explain the large difference in efficiency is a matter of speculation.

In both cases alteration of the initiation codon from GUG to AUG results in an increase of expression, without affecting the secondary structure of the mRNA. This indicates that the nature of the initiation codon indeed co-determines the efficiency of the RBS. Recently, Reddy et al. [24] showed a similar increase in expression when going from

an UUG→GUG→AUG initiation codon in the adenylate cyclase gene.

## ACKNOWLEDGEMENTS

We thank Mr E. Vijgenboom for plasmid pGP82, Mr J. Marugg and co-workers for preparation of the oligonucleotides and Mr B. Berkhout for technical advice on M13 mutagenesis and sequencing.

## REFERENCES

- [1] Kozak, M. (1983) *Microbiol. Rev.* 47, 1–45.
- [2] Gren, E.J. (1984) *Biochimie* 66, 1–29.
- [3] Dunn, J.J., Buzash-Pollert, E. and Studier, F.W. (1978) *Proc. Natl. Acad. Sci. USA* 75, 2741–2745.
- [4] Singer, B.S., Gold, L., Shinedling, S.T., Colkitt, M., Hunter, L.R., Pribnow, D. and Nelson, M.A. (1981) *J. Mol. Biol.* 149, 405–432.
- [5] Busby, S. and Dreyfus, M. (1983) *Gene* 21, 121–131.
- [6] Dreyfus, M., Kotlarz, D. and Busby, S. (1985) *J. Mol. Biol.* 182, 411–417.
- [7] Wulff, D.L., Mahoney, M., Shatzman, A. and Rosenberg, M. (1984) *Proc. Natl. Acad. Sci. USA* 81, 555–559.
- [8] Munson, L.M., Stormo, G.D., Niece, R.L. and Reznikoff, W.S. (1984) *J. Mol. Biol.* 177, 663–683.
- [9] Looman, A.C., De Gruyter, M., Vogelaar, A. and Van Knippenberg, P.H. (1985) *Gene* 37, 145–154.
- [10] Messing, J. (1979) *Recombinant DNA Technical Bulletin*, NIH publication no.79–99, vol.2, no.2, 43–48.
- [11] Messing, J. and Vierra, J. (1982) *Gene* 19, 269–276.
- [12] Van der Meide, P.H., Kastelein, R.A., Vijgenboom, E. and Bosch, L. (1983) *Eur. J. Biochem.* 130, 409–417.
- [13] Kastelein, R.A. (1981) PhD Thesis, University of Leiden, The Netherlands.
- [14] Yakota, T., Sugisaki, H., Takanami, M. and Kaziro, Y. (1980) *Gene* 12, 25–31.
- [15] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning. A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
- [16] Fischer, R.F., Das, A., Kolter, R., Winkler, M.E. and Yanofsky, C. (1981) *J. Mol. Biol.* 152, 397–409.
- [17] Grundstrom, T. and Normark, S. (1985) *Mol. Gen. Genet.* 198, 411–415.

- [18] Singer, P. and Nomura, M. (1985) *Mol. Gen. Genet.* 199, 543–546.
- [19] Fiers, W., Contreras, R., Duerinck, F., Haegeman, G., Merregaert, J., Min Jou, W., Raeymaekers, A., Volkaert, G., Ysebaert, M., Van der Kerkhove, J., Nolf, F. and Van Montagu, M. (1975) *Nature* 256, 273–278.
- [20] Remaut, E., De Waele, P., Marmenaut, A., Stanssens, P. and Fiers, W. (1982) *EMBO J.* 1, 205–209.
- [21] Pedersen, S., Reeh, S.V., Parker, J., Watson, R.J., Friesen, J.D. and Fiil, N.P. (1976) *Mol. Gen. Genet.* 144, 339–343.
- [22] Van Der Meide, P.H., Vijgenboom, E., Talens, A. and Bosch, L. (1983) *Eur. J. Biochem.* 130, 397–407.
- [23] Zuker, M. and Stiegler, T. (1981) *Nucleic Acids Res.* 9, 133–148.
- [24] Reddy, P., Peterkofsky, A. and McKenney, K. (1985) *Proc. Natl. Acad. Sci. USA* 82, 5656–5660.