A codon window in mRNA downstream of the initiation codon where NGG codons give strongly reduced gene expression in *Escherichia coli*

Ernesto I. Gonzalez de Valdivia and Leif A. Isaksson*

Department of Microbiology, Stockholm University, S-106 91 Stockholm, Sweden

Received July 2, 2004; Revised and Accepted September 10, 2004

**ABSTRACT**

The influences on gene expression by codons at positions +2, +3, +5 and +7 downstream of the initiation codon have been compared. Most of the +2 codons that are known to give low gene expression are associated with a higher expression if placed at the later positions. The NGG codons AGG, CGG, UGG and GGG, but not GGN or GNG (where N is non-G), are unique since they are associated with a very low gene expression also if located at positions +2, +3 and +5. All codons, including NGG, give a normal gene expression if placed at positions +7. The negative effect by the NGG codons is true for both the lacZ and 3A* model genes. The low expression is suggested to originate at the translational level, although it is not the result of mRNA secondary structure or a lowered intracellular mRNA pool.

**INTRODUCTION**

Translation of most mRNAs in *Escherichia coli* is initiated by anchoring of the mRNA to the 30S ribosomal subunit to form a pre-initiation complex. This anchoring is based on the Shine–Dalgarno (SD) sequence AGGAGGU, or related sequences thereof, and its base pairing to a complementary sequence close to the 3′ end of 16S rRNA. However, mRNAs that lack any obvious SD or SD-like sequence can also be translated at a relatively high efficiency. In fact, the region downstream of the initiation codon is also important for the translation of both *E.coli* and bacteriophage mRNAs (1,2). For instance, a CA repeat sequence in the region downstream of the initiation codon in mRNA increases the level of expression (3). Furthermore, the rare codons AGA/AGG as well as some other minor codons appear to be preferentially used within the first 25 codons in *E.coli* (4). Consecutive AGG or AGAAAGG codons near the 5′ end of the message give significantly lowered gene expression even though these codon sequences further down in the gene had no apparent effect on gene expression (4–7).

Earlier reports have suggested that the region downstream of the initiation codon, referred to as Downstream Box or Downstream Region, influences the translation initiation by complementary base pairing of mRNA to bases 1471–1480 in 16S rRNA (7–9). However, the analysis of mutant sequence variants in both mRNA and the postulated complementary rRNA sequence has refuted this model (10,11). Several reports, however, have shown that the downstream codon context affects the gene expression at the level of translational initiation and/or early elongation (3,7,12–15).

A recent survey of the effects on gene expression by different codons at position +2, immediately following the initiation codon, has shown an effect that varies by ~20-fold. This +2 codon influence on gene expression is the result of an evolutionary adaptation, since codons giving high or low expression tend to be over- or under-represented in natural genes, respectively (15). Also, the cooperative effects by the initiation codon and its flanking regions on translation initiation have been observed (13). Other results indicate that codons further down in the Downstream Region (DR), that follows the initiation codon, can also have strong effects on gene expression. Thus, the codon composition in the early coding region is one of the determinants for the apparent efficiency of gene expression at the translational level (14).

In this study, we have analyzed the effects on gene expression by several codons that are known to influence the expression differently if located at position +2, with the AUG initiation codon being +1. This has been performed by placing them, one at a time, at positions +3, +5 or +7 in the earlier well-characterized DR-A sequence that naturally constitutes the trpL leader (15). By using two different model genes, lacking a canonical SD sequence, we find that most of the analyzed codons that give an inefficient gene expression at position +2 (15) give higher expression if placed a few codons further downstream. The NGG codons AGG, CGG, GGG and UGG are exceptional since they give a consistent and similar low expression also at positions +3 and +5, but not at +7. The results indicate a functional window that involves the first half dozen codons that follow the initiation codon. In this window NGG, including GGG, but not GGN or GNG (where N is non-G), codons give a considerably lowered gene expression.

**MATERIALS AND METHODS**

*Chemicals and kits*

Chemicals used in this work were of highest quality grade. Restriction enzymes and T4 DNA ligase were from New
England BioLabs and GIFCO BRL. Plasmid DNA was prepared using GFXTM PCR and Gel Band Purification Kit from Amersham Pharmacia Biotech. DNA sequencing was performed by MWG Biotech AG. Immunoglobulin G (IgG)-Sepharose was from Amersham Pharmacia Biotech. The isopropyl-β-d-thiogalactopyranoside (IPTG) and O-nitrophenyl-d-galactopyranoside were purchased from Sigma. The radiolabeled nucleotide [γ-32P]ATP was purchased from Amersham Biosciences.

**Bacterial strains and plasmids**

The Δlac E.coli strain MC1061 (16) was the host strain for the lacZ plasmids (17) and strain XAc was used for the plasmids with the 3A' test system (18) (Table 1).

**Plasmid constructions**

Plasmids were constructed using standard recombinant DNA techniques (16). The plasmid pCMS71 (Figure 1A) (15) is a derivative of pDA3480 (19). The plasmid pEG100, which is a derivative of pCMS71, was used as a starting vector for the insertion of different AUG downstream context sequences, using the restriction sites SwaI/SalI, thus creating the plasmids series pEG1ZZ and pEG2XX. The restriction site SwaI/Csp45I was used to generate the plasmid series pEG4YY. (B) The plasmid pEG1000, with the 3A/2A' reporter system, which is a derivative of pHN109 (18) was used for sub-cloning of AUG downstream context sequences at the restriction sites SwaI/SalI, thus generating the series pEG10ZZ and pEG12XX. The SwaI/Csp45I site in pEG100 was used for sub-cloning of AUG downstream context sequences. Such sub-cloning gave the plasmid series pEG14YY. Both the lacZ and the 3A' reporter genes are under the control of the trc promoter (20).

In the cases of 3A' gene variants, the sequencing primer pHG1 was used. This primer is complementary to a unique sequence (5'-TAC GCA AAC CGC CTC TC-3') 389 bases upstream of the initiation codon of the 3A' gene.

**β-Galactosidase assay and growth conditions**

The plasmid carrying transformants with the lacZ reporter gene (MC1061 with the appropriate plasmid) were grown overnight at 37°C in minimal medium supplemented with all amino acids at recommended concentrations (21) and 100 μg/ml ampicillin. These cultures were used to inoculate into the same fresh medium at 37°C. The exponentially growing cells, at an optical density (OD590nm) of 0.4–0.5, were harvested without IPTG induction. The constructs with each codon were analyzed using four independent cultures, and each culture was assayed for β-galactosidase activity using four samples of the lysed non-induced cells (15). All measurements were carried out using an iEMS Multiscan Microplate Photometer (Labsystems). The SEM ranges from 1 to 20% for different constructs.

**Protein A' assay and growth conditions**

Overnight cultures of plasmid carrying transformants with the 3A' reporter system (XAc with the appropriate plasmid),
in a minimal medium supplemented with all amino acids at recommended concentrations (21) and 100 μg/ml ampicillin, were used to inoculate into the same fresh medium. IPTG (1 mM) was added at an OD$_{590}$nm of 0.2–0.25 to induce the 3A’ gene. The exponentially growing cultures were cooled and the cells were harvested by centrifugation at an OD$_{590}$nm of 0.5, followed by re-suspension in 1 ml of 10x TST buffer (22). The cells were lysed by incubation at 95°C for 10 min, and the cell debris were eliminated by centrifugation. Protein A’ was purified from the supernatant fraction using IgG-Sepharose (Pharmacia) minicolumns and a vacuum minifold system (Promega). The A’ proteins were eluted from the column with 0.1 ml of 0.5 M HAc at pH 3.2. The eluted protein sample was dried in a vacuum Speed-Vac (Techtum). The protein samples were dissolved in sample loading buffer, after denaturation at 95°C for 5 min, and the A’ proteins were separated by SDS–PAGE, 12% (16), followed by scanning quantification. The cultures were analyzed in duplicate.

Computer analysis of mRNA secondary structure

Analyses of secondary structures were performed as described in (21,24). A nucleotide sequence flanking the initiation codon (5’-CA TGGTAA TAA ATT AAA TAA AAT TTA AAT ATG AAA GCA ATT TTC GTA TCG ACT GG-3’), with the appropriate changes at positions +3, +5 and +7 were analyzed. This flanking sequence is common for the lacZ and protein A’ systems.

Analysis of mRNA pools by northern blotting

The mRNA was extracted using RNeasy Mini Protocol for Isolation of Total RNA from bacteria (RNeasy Mini Kit from QIAGEN). A 4 ml culture in minimal medium, supplemented with all amino acids (21), was induced with 1 mM IPTG at an OD$_{590}$nm of 0.2. The bacteria (Xac with the appropriate plasmid) were then harvested at an OD$_{590}$nm of 0.5. The extracted total RNA was purified and separated by electrophoresis on a 0.8% agarose gel, and transferred by blotting overnight to a Hybond-N nylon membrane (Amersham Life Science). The total RNA was linked to the Hybond-N nylon using ultraviolet light for 5 min. The transferred RNA was hybridized overnight (12–14 h) to the [γ-$^{32}$P]ATP-labeled deoxyoligonucleotide probe ABPO1 (5’-CGTGTTCCTCGTTTATAAGTTAGG-3’) which is complementary to each of the A’ encoding mRNA sequences (16). The radioactive signals on the films were analyzed by scanning (FujiFilm FLA 3000).

RESULTS

Early codons and gene expression

Gene expression is strongly influenced by the nature of the +2 codon that follows the initiation codon (15). In this study, we have extended this analysis to also include positions +3, +5 or +7. The reference sequence for this work is the DR-A (AUG AAA GCA AUU UUC GUA UCG). The first six of the seven codons long DR-A sequence originates from the natural trpL leader sequence, and at position +2 the lysine codon AAA is the one that gives the highest expression values in both lacZ and protein A’ based expression system (15). The codons at positions +2, +3, +5 and +7 were changed one at a time while maintaining the other codons in DR-A. The effects on gene expression by some altered codons in the DR-A are shown in Figure 2. It can be seen, in agreement with an earlier report (15), that AAG gives a lower expression than AAA at position +2 and both the leucine codons CUU and CUC give very low expression. These two codons gave a significantly increased gene expression if located at positions +3, +5 or +7. Apparently the low expression associated with CUU and CUC is valid only for position +2 in the context of DR-A. Other codons (the proline family, the serine family and the threonine codons ACG, ACU and ACA) also gave higher expression at positions +5 and +7 as compared to +2. It was noted, but not analyzed further, that UCG at position +7 in the original DR-A sequence gives lower expression than several other codons at this position (data not shown).

All six arginine codons, that previously had been found to give different effects at position +2 (15), were tested at positions +3, +5 and +7. As can be seen (Figure 3), AGG and CGG were very different from the other four arginine codons since these two gave a remarkably low gene expression at positions +2, +3 and +5. The expression was apparently normal for all

![Figure 2](image-url)
arginine codons if placed at position $+7$. The normal expression for all arginine codons was obtained also if they were placed at position $+11$ (data not shown).

To analyze how other codons with two G-residues influenced the gene expression, the glycine codons (GGG, GGA, GGC and GGU) as well as UGG (tryptophan), GAG (glutamic acid), GUG (valine) and GCG (alanine) were analyzed. All these codons are known to give low expression at position $+2$ (15). As can be seen in Figure 4, all of them except for GGG and UGG gave higher expression if located further downstream in the DR-A sequence. Taken together, all G-rich codons (NGG, GGN and GNG) give low gene expression if located at position $+2$. Also, the NGG codons (AGG, CGG, UGG and GGG) give a similar low expression if located at positions $+3$ and $+5$. All codons with the possible exception of GUG give a high expression if placed at position $+7$. It was noted that one of them (GAG) consistently gave a higher expression if located at position $+3$ compared to the other positions. The exceptional effects by GAG and GUG were not examined further.

The AGG and CGG codons at position $+6$ gave intermediate results as compared to positions $+5$ and $+7$ (data not shown). The $+6$ position therefore seems to be located at the border of the codon window described here and codons at this position were not analyzed further. For the $+4$ location, there is one suggestion that AGG is connected with low expression of the lambda gene $int$ (25). Likewise, the glycine codon GGG that gives low expression at positions $+3$ and $+5$ also appears to give low expression if located at position $+4$ (14). These two cases thus resemble the negative influence we find for the four NGG codons at positions $+3$ and $+5$. Since it appeared likely that the $+4$ codon location would give similar effects as $+3$ and $+5$, any further analysis of the $+4$ codon location was not performed.

The question was asked whether the strongly negative effect by an early CGG or AGG codon is merely the result of a base sequence in the mRNA or whether the bases need to constitute a codon triplet. For this reason, one base was inserted and one was deleted to give a local shift of the codon reading frame. As summarized in Table 2, both the $+1$ and the $-1$ shifts of the CGG and AGG triplets gave a much higher gene expression...
Table 2. Effect on gene expression by arginine codons and local shift variants

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Downstream region</th>
<th>lacZ (Miller units)</th>
<th>Frame</th>
</tr>
</thead>
<tbody>
<tr>
<td>pCM S40</td>
<td>5′-aau AUG AAA gca aau UUC gua ucg acu-3′</td>
<td>0.99 ± 0.04</td>
<td>0</td>
</tr>
<tr>
<td>pEG 211</td>
<td>5′-aau AUG AAA gca aau CGU gua ucg acu-3′</td>
<td>1.41 ± 0.03</td>
<td>0</td>
</tr>
<tr>
<td>pEG 214</td>
<td>5′-aau AUG AAA gca aau CGG gua ucg acu-3′</td>
<td>0.14 ± 0.004</td>
<td>0</td>
</tr>
<tr>
<td>pEG 245</td>
<td>5′-aau AUG AAA gca aauGG c gua ucg acu-3′</td>
<td>3.53 ± 0.13</td>
<td>−1</td>
</tr>
<tr>
<td>pEG 246</td>
<td>5′-aau AUG AAA gca aau gCG Guu ucg acu-3′</td>
<td>2.20 ± 0.14</td>
<td>+1</td>
</tr>
<tr>
<td>pEG 215</td>
<td>5′-aau AUG AAA gca aau AGA gua ucg acu-3′</td>
<td>1.20 ± 0.02</td>
<td>0</td>
</tr>
<tr>
<td>pEG 247</td>
<td>5′-aau AUG AAA gca aauAGAc gua ucg acu-3′</td>
<td>0.83 ± 1.1</td>
<td>−1</td>
</tr>
<tr>
<td>pEG 248</td>
<td>5′-aau AUG AAA gca aau cAG Aua ucg acu-3′</td>
<td>2.47 ± 0.28</td>
<td>+1</td>
</tr>
<tr>
<td>pEG 216</td>
<td>5′-aau AUG AAA gca aau AGG gua ucg acu-3′</td>
<td>0.13 ± 0.02</td>
<td>0</td>
</tr>
<tr>
<td>pEG 249</td>
<td>5′-aau AUG AAA gca aau A GG c gua ucg acu-3′</td>
<td>2.52 ± 0.42</td>
<td>−1</td>
</tr>
<tr>
<td>pEG 250</td>
<td>5′-aau AUG AAA gca aau cAG Guu ucg acu-3′</td>
<td>1.31 ± 0.18</td>
<td>+1</td>
</tr>
</tbody>
</table>

Insertion or deletion of one base was carried out in order to move the indicated arginine codons from zero frame to the +1 or −1 frame, respectively. Non-induced β-galactosidase values are presented as relative expression where 1.0 stands for 234 ± 7.6 U (46) as found for the DR-A variant (first line). Each value represents the mean value for at least four independent measurements.

![Figure 5](image)

**Figure 5.** Influence on the expression of the 3A frame by different codons at positions +2, +3, +5 and +7. Plasmids and codons are arranged in correspondence with the presentation in Figure 2. Values are given as molar ratios.

and AGA gave, as expected, high expression even in the zero frame. Both the out-of-frame variants of AGA gave a high expression.

**No contribution to the codon effect by transcriptional polarity, mRNA secondary structure or mRNA pools**

Most of the codons mentioned above have earlier been analyzed at position +2 for their effects on gene expression using both a lacZ expression system and a 3A frame test gene (15). The lacZ gene carries transcription polarity signals that possibly could trigger premature termination of transcription, thus giving a reduced level of gene expression. Such signals are not present in the semi-synthetic 3A frame gene (18). For this reason, the codon effects on gene expression were analyzed using the 3A test gene system, also in a plasmid that carries a 2A′ internal standard gene (18). As can be seen in Figure 5, the codon effects observed for the 3A′ system were quite similar to those found for the β-galactosidase expression system. In particular, the NGG codons gave a very low expression at positions +2, +3 and +5, but not at position +7. The position-dependent negative effect by NGG, as found also in the 3A′ test system, indicates that the codon effect on gene expression is not the result of an abortive event like transcriptional polarity. Furthermore, the similar results obtained on gene expression by NGG codons in the DR, using two totally different model genes, suggest a general effect by these codons in the early coding sequence of mRNA rather than being connected with the choice of test gene.

Lowered gene expression by +2 codons could not be explained by the involvement in mRNA secondary structures (13,15). Using gene variants examined here, computer analyses were carried out for secondary structures (24) involving a nucleotide sequence of about 20 nt flanking the initiation codon on each side. Such analysis did not indicate any differences in mRNA secondary structures that could explain the observed effects, notably by the NGG codons, on gene expression described above (data not shown).

Intracellular mRNA levels do not explain the different gene expression effects that are associated with the nature of the +2 codons (13,15,18,26). The question raised here was whether the low gene expression associated with an early AGG, in contrast to AGA, was the result of different translational efficiency or a difference in the mRNA pools. As shown in
domain in the 3A deoxyoligonucleotide probe, i.e. complementary to each A presented for the sake of comparison. Values are given as molar ratios. Protein data from Figure 5 are quantified on gels. Protein data from Figure 5 are presented for the sake of comparison. Values are given as molar ratios.

Figure 6. Influence by AGA and AGG arginine codons on the gene expression at the mRNA and protein levels. The indicated codons were placed at the positions +5 or +7 in the 3A' model gene. A [γ-32P]ATP-labeled deoxyoligonucleotide probe, i.e. complementary to each A' encoding domain in the 3A' test gene and 2A' control gene, was used to identify and quantify the respective mRNAs on gels. Protein data from Figure 5 are presented for the sake of comparison. Values are given as molar ratios.

Figures 5 and 6, the protein expression values for AGG at position +5 are ~10 times lower than that for AGA at this position or as compared to both AGA and AGG at position +7. These protein expression values are in good agreement with the results obtained for AGA or AGG at positions +5 or +7 in the corresponding situation using the β-galactosidase expression system (Figures 3 and 4). As can be seen in Figure 6, the relative levels of 3A'/2A' mRNAs were quite similar in the cases of all tested gene variants having AGG or AGA at positions +5 or +7. The low expression connected with AGG at +5 is thus manifested at the protein, but not at the mRNA, level and a low mRNA pool is not the reason for the low gene expression that is associated with this codon. Instead, the data suggest that the AGG effect on gene expression is the result of inefficient translation of the mRNA. This implication is also true for CGG, UGG and GGG at position +5 since the low gene expression associated with these codons at position +5 is manifested at the protein, rather than at the mRNA, level (data not shown).

DISCUSSION
In our studies described here, we have analyzed the seven codon long DR-A sequence (15) that has been placed in the beginning of a lacZ or 3A' based reporter gene. The first six codons of this sequence correspond to the first codons of the natural trpL sequence. The DR-A sequence has been analyzed earlier and it is known to give a high gene expression, being up to 40% compared to the corresponding sequence with an SD-like sequence upstream of the AUG initiation codon (14). The gene variants analyzed here do not carry any SD-like sequence, since a strong SD sequence would partly hide the contribution from the downstream codons we are interested in. The influence of a particular DR sequence to give a high or a low gene expression is in most cases qualitatively similar irrespective of an upstream ‘strong’ or ‘weak’ SD sequence (16). The nature of the +2 codon in DR-A is important for the influence on gene expression. Here, we have extended such studies with the emphasis on codon locations at positions +3, +5 or +7 in DR-A. Representative codons were changed one at a time to give variants with an altered codon at these positions.

We find that most codons that give a low gene expression at position +2 are associated with a higher gene expression if located further down in the mRNA at positions +3, +5 or +7. The NGG codons, including GGG, represent remarkable exceptions to this finding since they also have a strong negative effect on gene expression at positions +3 and +5. The expression is normal in the case of codons at position +7. Both CGG and AGG code for arginine but UGG is the tryptophan codon. Furthermore, the low expression is not found for the other arginine codons AGA, CGA, CGU and CGC, but it is seen for the GGG glycine codon. The other glycine codons GGA, GGU and GGC also give a negative effect on gene expression at position +2 but not at positions +3, +5 or +7. Thus, as indicated by the specific effects of the individual codons in the arginine and the glycine codon families, the effect on gene expression is not correlated with the nature of the corresponding amino acid in the nascent peptide. If the CGG and AGG codons were altered to give a +1 or −1 local shift, their negative effects disappeared. Thus, the negative effect by CGG or AGG at position +5 is the result of its function as a codon and not the consequence of two neighboring G-residues as such in the mRNA. The finding that GGG and UGG also give low gene expression suggests that the determinant for the negative effect on gene expression is the codon signature NGG. Such position-dependent negative effect is not connected with the other G-rich codons of the type GGN and GNG (where N is non-G).

A number of rare codons, including AGA and AGG, have been found to be over-represented among the first 25 positions in mRNA even though a systematic study of their effects on gene expression was not performed (4). There are two examples in the literature that an individual AGG (25) or GGG (14) codon at an early position can give a negative effect on gene expression. As it appears from our studies here this reflects a general property of early NGG codons to give a low gene expression. The codon combination of AGA AGG at an early position, and also a consecutive row of AGG codons later in a gene have been found to decrease gene expression (4,25,27), although any effect by a single AGG codon was not defined. The reason for low gene expression was in one case attributed to peptidyl-tRNA drop-off (27,28). In such a case, an insufficient pool of free cognate tRNA can be the result (4,6,25,29–31). Peptidyl-tRNA drop-off at an early NGG is also a very plausible reason for the observed low gene
expression reported here. However, in preliminary experiments using a plasmid encoding tRNA^{Arg4}, which decodes both AGA and AGG (6), we did not find any significantly increased gene expression in the case of a +5 AGG codon (data not shown). This indicates that the concentration of AGG decoding tRNA is not a limiting factor even though this possibility certainly is not ruled out.

Both AGA and AGG are decoded by tRNA^{Arg4}, but AGG is also decoded by tRNA^{Arg2}. Thus, the combination of AGA and tRNA^{Arg4} is not enough to give low expression. The question remains as to what extent AGG lowers the gene expression during decoding by tRNA^{Arg2} or by tRNA^{Arg4}. Similarly, the glycine codon GGG is decoded by tRNA^{Gly1} and tRNA^{Gly2}. The tRNA^{Gly2} also decodes GGA. In this case, it appears that a combination of GGA and tRNA^{Gly1} is not a determinant for low expression. The question remains whether the negative effect by GGG is the result of decoding by tRNA^{Gly1} or by tRNA^{Gly2}. It appears likely that the nature of the decoding tRNA is also a determinant for the observed low expression at early codons.

For some NGG codons, low gene expression values have been reported before (4–6,25–27). However, this finding was based on the analysis of tandems consisting of 2–9 low usage and/or rare codons. Based on our results, we suggest that a single NGG codon is enough to give the negative effect observed in this study, provided that the codon is located in the early coding region.

The involvement of the ribosome-binding region of the mRNA in a secondary structure is known to lower the translation initiation efficiency, and thereby gene expression (32). However, computer analyses (24) of the NGG gene variants analyzed here suggest that the low gene expression associated with these codons is not explained by the involvement in secondary structures.

All the gene variants analyzed in this study have the same promotor region. Therefore, the frequencies of transcription initiation should be similar. However, the effects on mRNA degradation by the altered DR-A codons are not known making it conceivable that the observed low gene expression values for certain codons could be the result of a decreased mRNA pool. This is, however, not the case since our estimates of mRNA levels for AGG gene variants, using the 3A′ system with 2A′ as an internal control, suggest that the relative mRNA levels are similar for AGG at positions +5 and +7, being similar as for AGA at these locations. This is despite the fact that the protein expression is much lower for AGG at position +5 compared with position +7, or in comparison with the high protein expression found for AGA at both positions +5 and +7. Therefore, the observed decreased gene expression by AGG at position +5 should be the result of low translational efficiency rather than a decreased mRNA pool. This is also true for CGG, UGG and GGG. Since the mRNA sequences are the same, starting with DR-A for the both the lacZ and 3A′ model systems used here, the location effect observed for NGG originates at the translational level and it is associated with the DR-A sequence and not with the nature of the rest of the model gene.

The presence of an infrequent codon could lead to a slow translation in comparison with common codons (33,34). However, rare codons are not necessarily slowly translated codons, and some frequent codons are more slowly translated than others (4,35–37). Binding of ribosomes to mRNA can protect against its degradation by RNase (38,39), and chemical footprinting analysis of mRNA–ribosome complexes indicates that at least 13 codons, and possibly as many as 20 codons, in the mRNA are covered by a ribosome upon binding (40). This would make it unlikely that a second ribosome could initiate while the first ribosome is pausing at some early slowly translated codon. Thus, an early slowly translated codon should give low gene expression. It is known that the three arginine codons CGA, CGU and CGC are decoded exclusively by the tRNA^{Arg2}, which carries inosine (I) as the wobble base in the anticodon. The I–A base pair seems to be less stable than I–U and I–C during the anticodon–codon interaction, thus giving slow and inefficient decoding of CGA (41,42). This I–A wobbling deficiency at CGA is significant enough to be revealed as a negative effect on decoding of a following stop codon at the ribosomal A-site, giving a lowered read-through of the stop codon (43,44). One would expect that a slow and inefficient decoding of CGA, as compared with CGU and CGC, in particular at an early location, should have a specific negative effect on gene expression, but this is not found. It is therefore puzzling to note that CGA gives as high expression as CGU and CGC. These two codons should be decoded faster by tRNA^{Arg2} than is CGA.

If the NGG codons in the DR-A region are translated slowly down to, and including position +5, this would give extended pausing and thus low gene expression. However, pausing at position +7 should also interfere with ribosome loading at the translational start site even if the effect could possibly be smaller. Alternatively, if the codon is rapidly translated at position +7, a high gene expression should be the result. In this case, the question must be raised as to why NGG should be decoded faster at position +7 than at position +5. As a more likely alternative to differences in speeds of translation one should consider differences in efficiency of translation, such as peptidyl-tRNA drop-off (45), thereby giving a reduced gene expression.

Our results suggest that the NGG codons AGG, CGG, UGG and GGG at positions +2, +3 and +5, but not further down at position +7, lower the gene expression at the translational level. The effect is not seen for GGN or GNG codons (where N is non-G) or if the reading frame of the NGG triplet is altered. The functional codon window that we observe in the early coding region is thus the result of a specific NGG codon effect.

ACKNOWLEDGEMENTS

We thank Dr Magnus Stenström and Dr Haining Jin for the gifts of plasmids and helpful discussions. This work was supported by a grant from the Swedish Research Council to L.A.I.

REFERENCES