Volume 16 Number 7 1988 Nucleic Acids Research

The AGG codon is translated slowly in *E.coli* even at very low expression levels

Fons Bonekamp and Kaj Frank Jensen

Enzyme Division, Institute of Biological Chemistry, University of Copenhagen, Sølvgade 83, DK-1307 Copenhagen K, Denmark

Received January 19, 1988; Accepted March 8, 1988

ABSTRACT

Data are presented which indicate that AGG codons for arginine are translated significantly more slowly than the CGU codons for the same amino acid even when their expression level from the probe is very low. The two types of codons were inserted (three in tandem) on a multicopy plasmid in an artificial leader peptide gene in front of the pyrE attenuator where the frequency of transcription termination is regulated by the degree of coupling between transcription and translation. Transcription of the operon is initiated from the lac-promoter dependent on the concentration of the lac-operon inducer IPTG. At all induction levels it was found that the frequency of transcription past the pyrE attenuator was approximately nine times lower when the AGG codons were present in the leader than with CGT codons present. This shows that AGG codons decouple translation from transcription in the pyrE attenuator region even when the concentration of this codon is not increased significantly relative to that in the unperturbed wild type strain. Thus the results indicate that AGG codons are always slowly translated in Escherichia coli.

INTRODUCTION

No organism uses the genetic code at random. Instead a specific subset of codons is preferred during protein synthesis. In general, the choice of codons becomes more biased towards this subset in highly expressed genes, like genes for the ribosomal proteins. Often the rarely used codons correspond to minor tRNAs and can thus be suspected to be translated more slowly than codons cognate to tRNAs present in high concentrations (1-3). Grosjean and Fiers (4) have suggested that other codons may be slowly translated because the codon-anticodon interaction is too strong for a rapid release of the empty tRNAs, while still other

codons may be avoided because they suffer from infidelity due to a weak anticodon binding. However, as the preferred codons differ between organisms (2,3,5) there is no reason to assume that certain codons are intrinsically "better" than others. Rather, the translation apparatus of an organism and its codon preference have been mutually adjusted during evolution (6,7).

The cause of the biased codon choice and its putative functions are subjects of much debate at present. On one hand, it appears reasonable that the selective use of codons is a question of cell-economy as it should be energetically cheaper for the cell to maintain a high concentration of ternary complexes (aatRNA-EFTu-GTP) for only a subfraction of the tRNAs than for all the tRNAs (7). According to this the preferred codons could be any. On the other hand, but fully compatible with the first explanation, the rare codons may also have a function. We can think of several possibilities: The rare codons -if slowly translated- could for instance have the function of packing more ribosomes on weakly expressed mRNA chains with poor ribosome binding sites and thereby protect these from nucleolytic attacks. Conversely, such codons should be avoided in highly expressed genes in order not to increase the ribosome burden too much. Another function of slow codons may be to act as pause sites for the ribosomes and allow folding of the nascent protein chain before interfering parts are made or the unfolded chain becomes attacked by proteases (8). Moreover in bacteria, the codon composition could contribute to the "fine-tuning" of the coupling between transcription and translation and thus allow the powerful control of regulatory mechanisms like mRNA chain termination or nucleolytic processing of mRNA chains (9,10).

At the heart of this discussion is the question whether all codons are translated at equal or at different rates. There is experimental evidence which suggests that the codons cognate to minor tRNAs are in fact translated more slowly than codons recognised by major tRNA species (7,11-14). However, as all these experiments were carried out with genes present on multicopy plasmids it was argued that the slow translation was seen only because the high gene dosage increased the rare codon

concentration to unphysiologically high values thereby sequestering rare tRNAs normally present in sufficient amounts to allow a fast translation of the codons (7,12).

To analyse the question we made use of the pyrE attenuator to probe the coupling between transcription and translation. This attenuator, normally located in the intercistronic region of a bicistronic operon, orfE-pyrE, was lifted out of its natural surroundings and inserted in a cloning site early in the lacZ gene on a plasmid (11). The outcome of this manipulation was an artificial operon (Figure 1). The operon is expressed from the lac-promoter and the pyrE attenuator is preceeded by a small artificial leader peptide gene with a cloning region (EcoR1, Pst1) and followed by a pyrE''lacZ gene fusion that can be used to monitor transcription past the attenuator by measuring β-galactosidase (11). Since attenuation is sensitive to variations in the coupling between transcription and translation induced by changes in the UTP pool (11,15-17) we reasoned that alterations in the attenuation frequency also would reflect the codon composition if different codons indeed were translated at different rates (14). Thus the small EcoR1-Pst1 fragment of the cloning region in the leader peptide gene (Figure 1) was substituted by different synthetic DNA fragments with defined and different codon choice at a given position (14). We made four plasmids with three arginine codons at this position, each containing one of the four combinations of AGG and CGT codons (Figure 2). We found that transcription past the pyrE attenuator increased by a factor two each time a rarely used AGG codon, decoded by a minor tRNA, was substituted by a frequently used CGT codon, recognised by a major $tRNA^{Arg}$ species (14). These results indicated that AGG codons are translated more slowly than CGU codons since they apparently reduce the coupling between transcription and translation in the pyrE attenuator region.

The question is now whether the slow translation rate of AGG codons is due to sequestering of a minor tRNA isoacceptor species caused by an unfysiologically high AGG codon concentration, as a result of a high level of transcription from the AGG-codon

containing probe, or whether the AGG codons are translated slowly even in E.coli cells with no plasmids.

MATERIALS AND METHODS

The strain used was S01804 Δ(araD139) Δ(lac)U169 thi pyrB equipped with the episome F'lacI^{q1}Z::Tn5 which overproduces the lac-repressor and therefore is able to turn off transcription from the lac-promoter in the absence of inducers.

The cells were grown exponentially at 37°C with shaking in the (A+B) medium (18) supplemented with glucose (0.2%), casamino acids (0.2%), thiamine (1 μg/ml), uracil (20 μg/ml), kanamycin (30 μg/ml) and ampicillin (50 μg/ml). When OD at 436 nm was 0.15 transcription from the lac-promoter was induced by adding isopropyl-β-D-thiogalactoside (IPTG) at the indicated concentrations. Subsequently, 0.5 ml samples were withdrawn into cold chloramphenicol (0.05 ml, 2 mg/ml) to stop further protein synthesis. After all samples had been taken the cells were disrupted by ultrasonic treatment and the extracts assayed for β-galactosidase activity (19) as previously described (11,14).

Constructions of plasmids pKCL101, pKCL105, pCU201 and pCU202 were previously described (11,14). The plasmid pKCL101 contains a HindIII site in the artificial leader peptide gene as well as an EcoRI site at position zero in the pBR322 part. This EcoRI-HindIII fragment carries the lac-promoter and the translation start for the leader peptide (11). Plasmid pFB101EH was made from pKCL101 by digestion with HindIII and EcoRI. Subsequently, the 5'-protruding ends of the larger fragment were filled in with the Klenow polymerase and ligated.

We interpret differences in the frequency of transcription past the pyrE attenuator when codons in the leader are changed as stemming from variations in the kinetics whereby the ribosomes translate the codons (14). This is because the codon effects are only seen when the intracellular concentration of UTP is normal and not when the cells are grown exponentially with a low UTP pool - a condition where RNA polymerase is expected to move slowly and thus reestablish the coupling between transcription and translation (14). Changes in mRNA stability as a consequence of changing CGT- to AGG codons

between the different constructs were considered unlikely as an explanatory factor since then some mRNA chains should be selectively unstable in the presence of a high UTP pool while stable when the UTP pool is reduced (14). The possibility that the nine nucleotide substitution might give rise to structures that interfere with the folding of the attenuator was considered unlikely since such structures are not obvious from the sequences and since the effects on transcription past the attenuator were simply additive when three CGT codons were changed to AGG codons one by one (14).

There is no direct evidence that substitution of the codon-strings in the plasmids does not interfere with the rate of RNA chain elongation in these regions. However, we consider this unlikely as the major cause for the difference between plasmids pCU201 and pCU202. This is because other codon strings in the leader-peptide gene with the same base-composition as the CGT-string of plasmid pCU201 gave expression levels between the CGT-string and the AGG-string (data not shown).

We have tried to estimate the burden of AGG codons on the mRNA chains made under direction of our plasmid pCU202: The leader peptide gene carrying the codon-strings is transcribed from the lac-promoter, and the lacZ start drives translation of the mRNA (Figure 1). It is therefore reasonable to assume that the frequency of transcription and translation of the leader peptide gene is like that of the intact lacZ gene on a similar plasmid without the pyrE-attenuator. Such a plasmid (pKCL11) gives rise to a level of β-galactosidase which is ca. 2000 units/mg protein in the presence of 450 μM IPTG (11). The specific activity of pure β-galactosidase is ≥300000 units/mg (Boehringer catalogue). Thus, the level of β-galactosidase directed by plasmid pKCL11 in the presence of 450 μM IPTG corresponds to less than 0.7% of the total cellular protein. From this we can conclude that the burden of AGG codons from our plasmid under maximal induction conditions is equivalent to a situation where three AGG codons were present per mRNA chain for a 120 kDa polypeptide that makes out 0.7% of the total cellular protein.

We shall compare our results with the data of Robinson et al

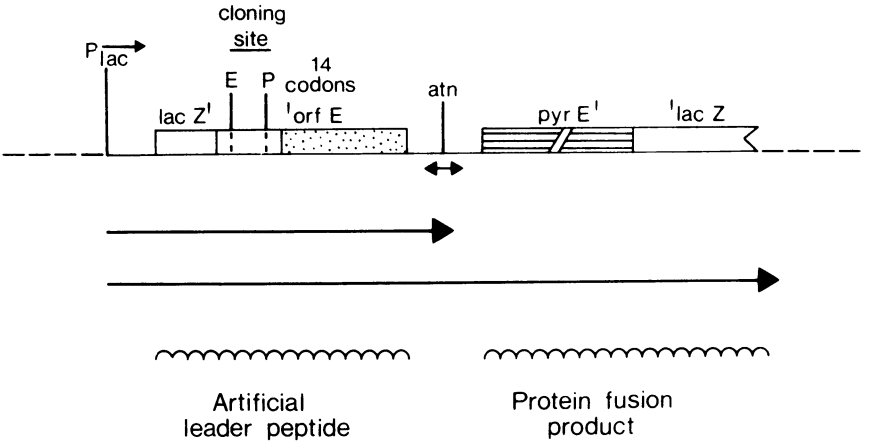

Figure 1: Structure and expression of the pyrE-lacZ gene fusions on plasmids. The plasmids carry ampicillin resistance and their construction is described in refs. 11 and 14. To construct plasmids with defined codon composition (pCU201-204) the EcoRI(E)-PstI(P) fragment was substituted by synthetic oligonucleotide linkers (14). Atn denotes the UTP controlled pyrE attenuator and the arrows below indicate the attenuated and non-attenuated transcripts. Figure 2 gives the sequences of the leader regions.

```
                                                    PstI
pCU201 ------------------ AAT TTT CGT CGT CGT CTG CA ------------------
pCU202 ------------------ AAT TTT AGG AGG AGG CTG CA ------------------
pCU203 ------------------ AAT TTT CGT AGG AGG CTG CA ------------------
pCU204 [ATG ACC ATG ATT ACG AAT TTT CGT AGG CGT CTG CAG GTC GAA TCC ATT GTA
       GCG ACG CAG AAG GCG GCG CTG GCA AAC] TGATTTTTAAGGCGACTGATGAGTCGCCTTT
       TTTTTGTCTGTAGAAAAGTAAGATGAGGAGCGAAGGC ATG AAA CCA TAT ... pyrE
```

Figure 2: Nucleotide sequences of leader region in plasmids pCU201-204. Only pCU201 and pCU202 are used for the present experimental work. Open reading frames for the artificial leader peptides are shown in square brackets. The sequence is shown in full only for pCU204. The areas within the vertical solid lines indicate the synthetic oligonucleotides used for construction (14) while the area inside the broken lines shows the position where the codons are varied. Opposite arrows beneath the sequence of pCU204 show the symmetric pyrE attenuator. The dots below the sequence indicate a potential Pribnow box (see text).

(13) who studied codon effects on the synthesis of chloramphenicol transacetylase from a gene with either four AGG- or four CGT arginine codons present at a given position. The maximal expression level in their experiments corresponds to a situation where four AGG codons are present per mRNA chain for a 25 kDa protein making out 40% of all cellular protein (13).

RESULTS AND DISCUSSION

We made use of the fact that the artificial operon shown in Figure 1 is expressed from the lac-promoter and that transcription initiation at this promoter depends on the concentration of the inducer isopropyl-β-D-thiogalactoside (IPTG) added to the medium, provided that there is enough lac-operon repressor in the cells. The latter was furnished by introduction of the episome F'$lacI^{q1}Z::Tn5$ into the host strain. Only the two plasmids pCU201 with three CGT codons in the leader peptide and pCU202 with three AGG codons at the same position in the leader peptide (Figure 2) were used in these experiments.

The cells were grown exponentially at 37°C in a glucose medium supplemented with casamino acids and uracil as well as antibiotics to prevent loss of the F'-episome and the plasmid. When the optical density at 436 nm was 0.15 the inducer IPTG was added to the cultures at different concentrations and subsequently samples were withdrawn into cold chloramphenicol to stop further protein chain elongation and assayed for β-galactosidase activity. The data obtained with pCU201 (3 x CGT) are given in Figure 3A while Figure 3B shows the data obtained for plasmid pCU202 (3 x AGG). Table 1 shows the differential rates of β-galactosidase formation calculated from the data in Figure 3. At 450 μM IPTG transcription of the pyrE''lacZ gene fusion is 9-10 fold lower from plasmid pCU202 with three AGG codons than from pCU201 with three CGT codons at the same position in the leader peptide. This is in good agreement with previously obtained results (14). When the concentration of the lac-operon inducer IPTG was reduced by a factor three (to 150 μM) the rate of synthesis of β-galactosidase from both plasmids decreased in parallel such that the ratio between pyrE''lacZ gene expression from the two plasmids remained constant. This argues against the idea that

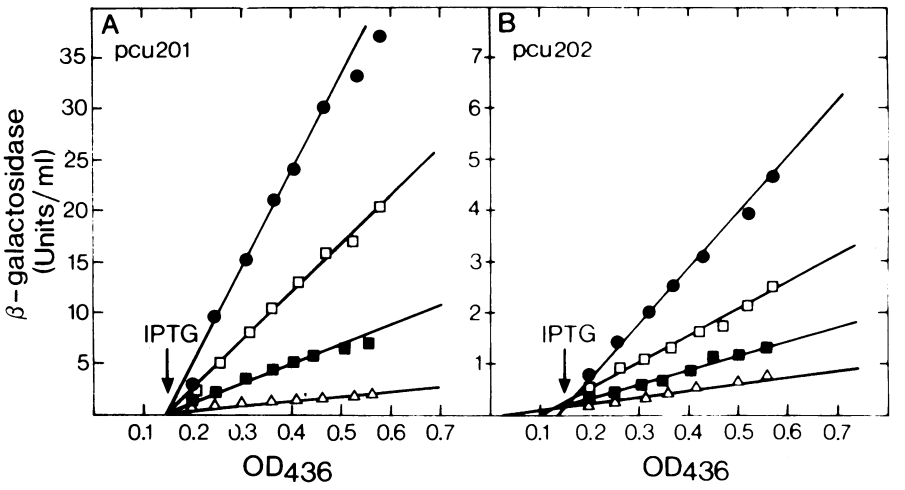

Figure 3: Expression of the pyrE-lacZ gene on plasmids following induction with IPTG. Figure 3A: pCU201 with three CGT codons and Figure 3B: pCU202 with three AGG codons. The cells were grown as described in Materials and Methods and at OD436=0.15 transcription from the lac-promoter was induced by adding IPTG at the following concentrations: 450 μM (●), 150 μM (□), 60 μM (■), and 20 μM (Δ).

the ribosomes work more slowly at AGG codons than at CGU codons because of sequestering of the rare tRNA species that recognize AGG. If this was the case one would expect the ratio (pyrE''lacZ expression from pCU201 divided by pyrE''lacZ expression from pCU202) to fall when the arginine codon concentration is lowered due to a reduced amount of mRNA chain synthesis, since such a reduction is expected to release tRNA "titration". It is assumed that the CGT codons on plasmid pCU201 do not significantly disturb the cell-physiology, since these are present in numerous copies on the native chromosome and are recognized by a major tRNA species (2,4).

When the concentration of IPTG was still further reduced, the levels of β-galactosidase synthesis from the two plasmids began to approach one another (Figure 3, Table 1). Finally, at zero concentration of IPTG, there was very little difference between the two codon plasmids (Figure 4). At first sight this could be interpreted as indicating that the aminoacyl-tRNA which decodes AGG is present in sufficient amounts for a fast translation to

Table 1: Differential rate of synthesis of β-galactosidase from plamids pCU201 and pCU202 at different concentrations of IPTG.

Concentration of IPTG (μM)	Differential rate of β-gal. (Units/mlxOD)		pCU201/pCU202 after corr.[a]
	pCU201	pCU202	
450	93.8	10.9	9
150	48.1	5.3	11
60	20.3	2.8	10
20	5.0	1.4	8
0	1.2	0.8	-

a) The correction is made by subtracting the uninduced value (0.8) obtained also with the lac-promoterless plasmid pFB101EH in the absence of IPTG (Figure 4). Exactly the same uninduced levels were found with pCU202 and with plasmid pKCL105 (11) which contains a stop-codon 31 residues upstream of the symmetric pyrE attenuator (not shown).

occur in plasmid-less strains, but that the tRNA concentration is so small that just a minute increase in the concentration of AGG codons over that present in the native E.coli mRNAs is sufficient to sequester enough of the tRNA population to make the translation of AGG codons slow.

There is, however, another explanation for the results. This is if a weak and IPTG independent transcription was initiated from a position downstream from the lac-promoter after the region where the codons are varied. The pyrE operon does in fact contain a candidate for such a weak promoter (20, see Figure 2) and in the native pyrE operon of E.coli (20, and unpublished results) as well as in Salmonella typhimurium (21) a weak and non-pyrimidine regulated transcription past the attenuator does indeed take place. Such a transcription should be subtracted from the total level of transcription of the pyrE''lacZ fusion since it would not be influenced by codon differences in the artificial pyrE leader peptide.

To test this possibility we took a similar plasmid pKCL101 (11) and deleted the lac-promoter as well as the translational start for the leader peptide by cutting out an EcoRI-HindIII fragment covering this region, as described in Materials and Methods. The resulting plasmid pFB101EH gave rise to a level of β-galactosidase formation that was very close to that obtained

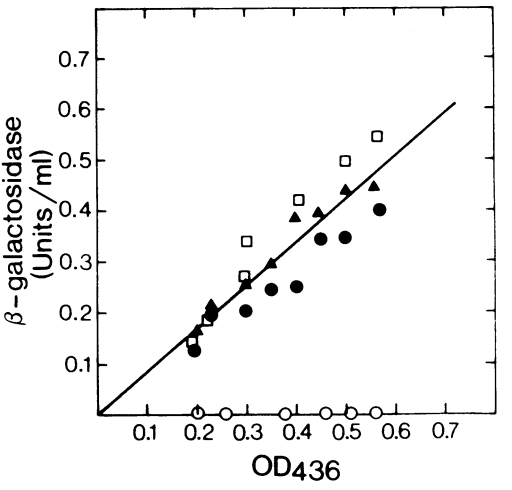

Figure 4: Expression of the pyrE-lacZ gene on plasmids without lac-operon inducers. The plasmid were: (□) pCU201 with three CGT codons; (●) pCU202 with three AGG codons; (▲) the lac-promoter-less plasmid pFB101EH, and (○) pBR322.

with pCU202 (3 x AGG) thus supporting the idea that at least the major part of the low uninduced level of enzyme formation is not caused by transcription from the lac-promoter. Thus this level should be subtracted from the levels obtained with the codon plasmids.

When this correction of the measured values is made it appears that the level of β-galactosidase synthesis from pCU201 with three CGT codons in the artificial pyrE leader peptide gene is always approximately nine times higher than that seen with pCU202 which contains three AGG codons at the corresponding position in the leader, regardless of the induction level. Thus, we must conclude that AGG codons are translated more slowly than CGU codons even at very low mRNA concentrations since the AGG codons cause decoupling of translation from transcription in the pyrE attenuator region. It is not possible to conclude whether the AGG codons are translated slowly due to a low V_{max} of the translation step at these codons or because this process has a high K_m for the tRNA relative to its concentration in the cell. It is even possible that the binding of the cognate tRNA is tight, but that there is not enough tRNA to saturate the AGG

codons also without plasmids in the cells. After all, the concentration of AGG codons derived from our probe at the highest induction level may very well be small compared to the total cellular concentration of AGG codons (see MATERIALS AND METHODS). Thus, the data indicates only that AGG codons are slow codons also when present in normal mRNAs originating from the chromosome of plasmid-less Escherichia coli.

We do not imply that the slow rate of translation as a function of the codon composition of a gene leads to less protein product per se. The point is that the codon constallation will affect the coupling between transcription and translation. Thereby it might contribute to the control of gene expression by the polarity effects on transcription, whenever a transcription terminator is positioned in such a way that its capacity for mRNA chain termination could be influenced by the ribosomes.

Our conclusion may seem to be in conflict with the results of Robinson et al (13). These authors observed a lower level of protein synthesis from a gene for chloramphenicol transacetylase when four AGG codons were present in the reading frame than from a gene containing four CGT codons at the same position. This effect was only observed at high levels of transcription, thus indicating sequestering of a minor tRNA as the cause of the phenomenon. However, the experiments done by Robinson et al (13) were not designed to reveal translational rate differences at the specific codons. Thus, in the absence of induction, the synthesis of the individual protein chains may well have taken more time when AGG codons were present in the mRNAs than when CGU codons were present instead. Furthermore, the chloramphenicol transacetylase protein comprised more than 40% of all cellular protein under inducing conditions. In this case there could be other reasons for the observed codon effects (13) than an increased steptime for the ribosomes at the AGG codons. This could for instance be frame-shifts or ribosome drop-off at these codons.

As outlined in MATERIALS AND METHODS, the maximal expression level of the AGG codons in our experiments is likely to be more than 200 fold lower than in the experiments made by Robinson et

al (13). Thus our data and the data of Robinson et al (13) are compatible under the assumption that AGG codons are translated slowly at low expression levels, but that sequestering of the cognate tRNA becomes significant at very high levels of expression and causes (for instance) frame-shifts. This phenomenon, rather than an increased steptime for the ribosomes at specific codons, will obviously lead to the formation of fewer intact protein chains per time unit.

ACKNOWLEDGEMENTS

We appreciate the help from Claus Jeppesen in the design and construction of pFB101EH and the technical assistance by Lise Schack. The work was supported by grants from the Danish Natural Science Foundation.

REFERENCES

1. Ikemura, T. (1981) J.Mol.Biol. 151, 389-409.
2. Ikemura, T. (1982) J.Mol.Biol. 158, 573-597.
3. Ikemura, T. (1985) Mol.Biol.Evol. 2, 13-34.
4. Grosjean, H. and Fiers, W. (1982) Gene 18, 199-209.
5. Bennetzen, J.L. and Hall, B.D. (1982) J.Biol.Chem. 257, 3026-3031.
6. Bulmer, M. (1987) Nature 325, 728-730.
7. Kurland, C.G. (1987) Trends Biochem. Sci. 12, 126-128.
8. Varenne, S., Buc, J., Lloubes, R. and Lazdunski, C. (1984) J.Mol.Biol. 180, 549-576.
9. Yanofsky, C. (1981) Nature 289, 751-758.
10. Jensen, K.F., Bonekamp, F. and Poulsen, P. (1986) Trends Biochem. Sci. 11, 362-365.
11. Bonekamp, F. Clemmesen, K., Karlström, O. and Jensen, K.F. (1984) EMBO J. 3, 2857-2861.
12. Petersen, S. (1984) EMBO J. 3, 2895-2898.
13. Robinson, M., Lilley, R., Little, S., Emtage, J.S., Yarranton, G., Stephens, P., Millican, A., Eaton, M. and Humphreys, G. (1984) Nucl.Acids Res. 12, 6663-6671.
14. Bonekamp, F., Andersen, H.D., Christensen, T. and Jensen, K.F. (1985) Nucl. Acids Res. 13, 4113-4123.
15. Jensen, K.F., Neuhard, J. and Schack, L. (1982) EMBO J. 1, 69-74.
16. Jensen, K.F., Fast, R., Karlström, O. and Larsen, J.N. (1986) J.Bacteriol. 166, 857-865.
17. Poulsen, P. and Jensen, K.F. (1987) Mol.Gen.Genet. 208, 152-158.
18. Clark, D.J. and Maaløe, O. (1967) J.Mol.Biol. 23, 99-112.
19. Miller, J.H. (1972) Experiments in Molecular Genetics published by Cold Spring Harbor Laboratory Press, NY.
20. Poulsen, P., Jensen, K.F., Valentin-Hansen, P., Carlsson, P. and Lundberg, L.G. (1983) Eur.J.Biochem. 135, 223-229.
21. Neuhard, J., Stauning, E. and Kelln, R.A. (1985) Eur.J.Biochem. 146, 597-603.