Nucleic Acids Research, Vol. 19, No. 19 **5247–5251**

Codon choice and potential complementarity between mRNA downstream of the initiation codon and bases 1471 – 1480 in 16S ribosomal RNA affects expression of *glnS*

Margareta Faxén, Jacqueline Plumbridge[1] and Leif A.Isaksson*
Department of Microbiology, Stockholm University, S106 91 Stockholm, Sweden and [1]Institute de Biologie Physico-Chimique, (URA1139), 13 Rue Pierre et Marie Curie, 750 05 Paris, France

Received July 16, 1991; Revised and Accepted September 9, 1991

ABSTRACT

A cis-acting expression mutation, GAG to GAA, in the third codon of the *glnS* gene is analyzed. Both codons code for glutamic acid but the mutation is known to increase gene expression by four fold. We show that the mutation has an effect only if it is located in the beginning of a gene but not if located internally. Data are presented that suggest that the reason for the increased expression by the mutation is the potential formation of one more base pair between the mRNA and 16S ribosomal RNA. Gene expression varies about 16 fold as the number of potential base pairs within the sequence 1471 – 1480 in 16S RNA increase from two to ten.We also give evidence that supports the idea that the presence of rare codons near the beginning of the mRNA can affect expression.

INTRODUCTION

Although the basic process of translational initiation is well established [1][2], it has long been recognised that the choice of the correct initiation site from amongst all the AUG and GUG codons present on a mRNA is not a trivial problem.The purine rich Shine–Dalgarno sequence is of prime importance but many surveys of translational initiation sites have shown a non random arrangement of nucleotides in the 40 bases surrounding the initiation codon without revealing any convincing rules [3][4][5][6][7][8]. More recently, usually after the analysis of highly expressed proteins, other potential base-pairing interactions between mRNA and 16S rRNA, in addition to the Shine–Dalgarno interaction, have been proposed to explain high translational efficiencies [9][10][11][12][13]. These interactions are proposed to enhance expression by increasing translational initiation. A model has been suggested which incorporates several of these potential interaction sites [14]. Certain of these interactions seem to be better supported by experimental evidence than others. For instance, several groups have evidence favouring an interaction between an AU rich sequence upstream of the initiation codon and sequence around nucleotides 458–466 in 16S rRNA [10][13]. In this report we produce independent evidence supporting an interaction between nucleotides 1471–1480 of 16S rRNA and the sequence downstream of the initiation codon as proposed by Sprengart et al [12].

A series of cis-acting expression mutations has been isolated near the beginning of the glutaminyl tRNA synthetase gene (*glnS*) gene of *Escherichia coli* [15] using a *glnS-lacZ* gene fusion. These mutations are acting either at the transcriptional or translational level to increase gene expression. One of the mutations is located in the third codon in the beginning of the *glnS* gene. It is silent since the codon GAG is altered to GAA and both codons code for glutamic acid. This change of a G to an A increases *glnS* expression four-fold, most likely at the level of translation since the level of mRNA is not affected by the mutation [15].

In this report we describe the further analysis of this regulatory mutation which was performed by inserting the gene sequence into another genetic environment. Our results suggest that the mutation has an effect only if located in the beginning of the gene, but not internally. Data are presented which support the hypothesis of an mRNA anchoring through formation of complementary base pairs between bases downstream of the initiation codon and bases in the region 1469–1483 in 16S ribosomal RNA as proposed by Sprengart et al [12] to explain the strength of the T7 gene 0.3 initiation site. The GAG to GAA mutation at the third codon originally isolated by Plumbridge and Söll [15] most probably gives an increased gene expression because of the formation of an extra base pair to 16S RNA. Both the GAA and GAG codons are frequently used in *E.coli* proteins. We subsequently analyzed some arginine codons (AGG, AGA and CGU) in this same position to test for a contribution to the translational efficiency from the presence of rare or common codons at this site.

* To whom correspondence should be addressed

MATERIAL AND METHODS

Materials

All chemicals were of the highest grade commercially available. Oligonucleotides for *in vitro* mutagenesis were synthesized on a Pharmacia Gene Assembler. Restriction enzymes were either from Promega or BioLabs and the Gene Clean Kit was from Bio Rad 101.

Bacterial strains and plasmids

E.coli MC1061 [16] was the host strain for all plasmids. Cells were grown at 37°C in minimal media supplemented with all aminoacids and 0.15 mg/ml ampicillin [17][18].

The synthetic oligonucleotides were cloned into the linker region of pAB24 and/or pAB26 (Fig.1), which are modified forms of pEZZL$_1$ZT [19][20][21]. This plasmid carries a semisynthetic gene which codes for three identical protein domains corresponding essentially to the antibody binding B domain in protein A from Staphylococcus aureus. The secretion signal sequence which precedes the first domain, as described in the original report, is absent in the constructs used here. The engineered protein domain was originally referred to as Z but it is here denoted as B′ in order to avoid confusion with *lacZ* which is also used in this work. The gene is under the control of an inducible *trc* promoter which is an IPTG inducible variant of the *tac* promoter [22]. Induction was achieved by using 0.4 mM IPTG in the cases when gene constructs based on pAB24 were used. The plasmid pAB26 carries a *lacZ* gene fused to the last domain of the 3B′ construct (Fig.1). In the case of this plasmid β-galactosidase assays were used to determine gene expression. This was carried out in uninduced cells taking advantage of the high background expression of β-galactosidase. β-Galactosidase was measured as described elsewhere [17]. Gene expression from other plasmids was followed by monitoring the gene products by gel electrophoresis. The details concerning constructions of these plasmids are to be published elsewhere.

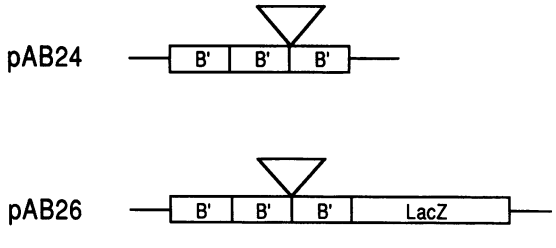

Figure 1. Basic design of plasmids used in this study. The linker region used for insertion of mutant sequences is indicated by the triangle. B′ indicates the modified sequence of the B domain coding part of protein A.

Construction of *in vitro* mutations

Plasmid DNA was prepared according to standard methods [16] and purified using Gene Clean Kit from Bio Rad 101. Cloning of oligonucleotides into the different vectors were accomplished by standard techniques. The sequences of all constructs were verified by using the Sequenase Kit from USB.

Purification and determination of proteins

20 ml cultures of cells with pAB24 derived plasmids were grown to mid log phase, harvested and resuspended in 1ml of sonication buffer (0.1 M tris-HCl, pH 8.5, 0.01 M EDTA, 0.1% Tween 20) and sonicated in three burst of 20 seconds each.The cell debris was centrifugated down and the supernatant was used for purification of protein A by affinity chromatography as described elsewhere [20][21][23]. The purified proteins were analysed by SDS polyacrylamide gel electrophoresis using 15% polyacrylamide.

RESULTS

Location effect of an expression mutation in *glnS*

One of the cis-acting mutations isolated in a *glnS-lacZ* gene fusion by Plumbridge and Söll is a GAG to GAA alteration in the third codon of the sequence in the *glnS* part. Both the wildtype and mutant codons code for glutamic acid but still the mutation gives a 4-fold increase in β-galactosidase activity *in vivo*. Since mRNA levels are unaltered by the mutation its location suggests that it acts at the translational level, possibly in connection with initiation [15]. If so, the effect of the mutation should be seen if the wildtype and mutant seqences were located in the beginning of a gene but not if they were placed further downstream in the same gene.

In order to estimate the impact of the location of the mutation we used a model gene which is derived from the immunoglobulin binding B domain of protein A from Staphylococcus aureus. This domain, originally referred to as Z, but here denoted B′ in order to avoid confusion with *lacZ*, is repeated three times giving the gene product 3B′. A linker is located in the gene between the sequencies coding for the second and third protein domains [19][20][21]. In order to analyze the effect of an internal location of the mutation the sequences comprising bases −3 to +47 of *glnS* were inserted into the linker of the model gene in plasmid pAB24 (Fig.1). The resulting constructs (pJP30 and pJP31) are

pJP30 5' ttcgaaACGATGAGT**GAG**GCAGAAGCCCGCCCGACTAACTTTATCCGTCAGATCGATActagt 3'
pJP31 **A**

pJP34 5' ttcgaaTAGAAACCATGGCTCGAGATAACGTTGTTTACGCTTTGAGGAAATCCACGATGAGT**GAG**GCAGAAGCCCGCCCGActagt 3'
pJP35 **A**

Figure 2. Model genes for studies on location effect by the GAG to GAA mutation (as indicated) in *glnS*. The GAG codon which is mutated in pJP31 and pJP35 is indicated in bold. Sequences derived from *glnS* start at the nucleotide indicated with an arrow. The arbitrary upstream sequence of pJP34 and pJP35 was included to allow sufficient separation of TAG and ATG codons to prevent interferance between terminating and initiating ribosomes. Sequences which originate from the vector are given in lower case letters. Stop-codon, Shine−Dalgarno site and initiation codon are under-lined.

shown in Figure 2. As expected (Fig.3a, lanes 1 and 2) formation of protein 3B′ is indistinguishable when the constructs carrying the wildtype or mutant sequences are compared. This suggest that there is no abortive event during translation associated with the GAG codon which would give prematurely aborted translation in the wildtype but not in the case of the GAA mutation. Such prematurely terminated gene products should have the approximate size of the 2B′ domain product and no such product is seen. The results, of course, do not exclude the possibility of differential translational pausing as long as this does not affect expression of the wildtype or mutant model gene.

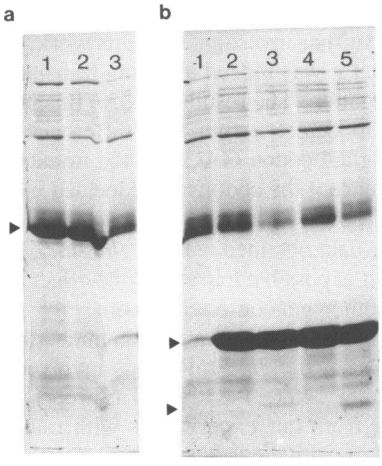

Figure 3. Electrophoretic analysis of gene products from derivatives of pAB24 with different inserted gene sequences. a) Products from plasmid pJP30 (lane 1) and pJP31 (lane 2). Lane 3 shows the result in the absence of gene induction in order to illustrate the background. Samples from equal amounts of cells were applied in all lanes. The arrow indicates the location for protein 3B′. This protein is absent in lane 3; the band slightly above its position in lane 3 consist of IgG. This variable contaminant is released from the affinity column used for purification of gene products. b) Products from uninduced (lane1) and induced (lane2) pJP34. Induced gene products from pJP35 (lane3), pMIN (lane 4) and pMax (lane 5) are shown. The upper arrow indicates the location of the 2B′ protein and the lower arrow indicates 1B′. Protein 3B′ is missing in these samples; the band at its approximate location contains contaminating IgG.

As a comparison, sequences comprising the first 25 bases of the *glnS* gene, but also including an upstream Shine−Dalgarno sequence as well as an in frame termination codon at the end of the second B′ domain were synthesized and inserted into the linker of pAB24. This gave plasmids pJP34 and pJP35 (Fig.2). In these constructs, translation should stop after the second B′ domain but reinitiation should occur at the AUG codon further downstream in the beginning of the third B′ coding region. Any difference in expression between the GAG and the GAA construct should be revealed as a difference in the yield of the third B′ domain giving a 1B′ domain product.

As can be seen (Fig.3b lanes 2 and 3), the yield of the protein 1B′ is low to the extent that it can not be seen in the wildtype but in the mutant case a faint band is noticeable. The result of this experiment which is confirmed below, suggests a higher expression of the construct with GAA compared to the one with GAG. The low yields of 1B′ protein do not only arise from a low expression since the one domain protein has a lower binding affinity to the antibody column than the multimeric domain protein. This gives systematically lowered yields for the 1B′ protein.

The results described above suggest that the GAG to GAA alteration is associated with an increased expression when the altered codon is located in the beginning, but not further downstream, of the gene. This implies some effect on translational initiation.

Mutation in an extended ribosome binding site

It has been suggested that a sequence downstream of the initiation codon of many *E.coli* genes is capable of base-pairing with the 1469−1483 region of 16S RNA [12]. These authors also noted that there was some complementarity of this region with the *glnS*

Table 1. Effects of sequences alterations in mRNA on their potential homology to 16S RNA and gene expression. Complementary bases in mRNA are indicated in bold. Bases outside the complementary region are in lower case letters. The start-codon is underlined. The expression of β-galactosidase from the different mutant sequences are shown.

Comment	Sequence	Complementary bases	Expression (units)
16S RNA (1483-1469)	3' **A G U A C U U A G U G U U U C** 5'	-	-
pJP34 wt	5' au**G A G U G A G G C A G A A G C**c 3'	7	24 ±3
pJP35 mut	au**G A G U G A A G C A G A A G C**c	8	93 ±7
pMin	au**G A C C G A G G G U G U U U C**c	2	9.0 ±1
pMax	au**G A G U G A A U C A C A A G C**c	10	140 ±11
pAGA	au**G A G U A G A G C A G A A G C**c	6	27 ±3
pAGG	au**G A G U A G G G C A G A A G C**c	5	7.0 ±3
pCGT	au**G A G U C G U G C A G A A G C**c	5	13 ±3

gene and that the mutation GAG to GAA increased the potential base-paring interactions from 8 to 9. This raises the possibility that the explanation for the increased expression obtained by GAA compared to GAG could be a more efficient anchoring of the mRNA to 16S RNA during initiation.

In order to test for this possibility mutations were constructed such that the potential base pairing with the indicated region of 16S RNA should be low (pMin) or high (pMax). These sequences shown in (Table 1), were inserted into pAB24 and gene expression was analyzed. The result was that in the case of pMin no expression could be seen, but in the case of pMax formation of the 1B′ indicator protein was significant (Fig.3b, lanes 4 and 5, respectively).

The expression of the 1B′ coding gene with the *glnS* initiation signal is too low to be measured from sequences permitting only low efficiency initiation so another vector was used. For these experiments a derivative of the model gene, pAB26, which carries 3B′ fused to *lacZ* was used (Fig.1) (Björnsson and Isaksson, manuscript in preparation). The mutant sequences which were used for the experiment described above and examined in plasmids pJP34 and pJP35 (Fig.2) were inserted into pAB26. The expression levels of the corresponding fused genes, measured as β-galactosidase activity, are shown in Table 1. First, it can be seen that the GAA mutation indeed gives a higher expression than the original codon GAG. Second, the constructs pMin and pMax with a minimal and maximal base pairing potential to the indicated 16S RNA sequence, respectively, give very different expression levels since the latter construct gave about 16 times higher expression. A good correlation between the number of potential base pairs to the 16S RNA and β-galactosidase activity is thus seen for these four constructs. This suggests that gene expression is dependent on base pairing between sequences downstream of the inititiation codon and bases close to the 3′ end of 16S RNA. It also suggests that the original finding of the increased expression of the *glnS-lacZ* fusion by the GAG to GAA alteration [15] is explained, at least in part, by an increased binding to 16S RNA by the mutant mRNA.

It is conceivable that even though an altered complementarity to 16S RNA is the major reason for the variation in expression levels other factors could also contribute. One such complication could be the fact that codons are translated at different speeds. This could give ribosomal queuing which would block translation initiation and thereby gene expression. The contribution from differences in translation speed is illustrated (Table 1) by a comparison between the constructs which carry AGG, AGA or CGU as the third codon. All these codons code for arginine. However, the first two are extremely rare in *E.coli* and should be translated at a low speed in contrast to the CGU codon which is very abundant and which should be translated at a high speed [24]. As can be seen, AGG and and CGU can potentially form the same five base pairs to 16S RNA but the CGU construct gives about a twofold higher expression. On the other hand, the slow AGA codon sequence, which can form one more base pair, gives a twofold higher expression then the CGU sequence. Thus, for the *glnS* mRNA the formation of the extra base-pair gives a major enhancing effect but the nature of the codon also plays a role.

DISCUSSION

Glutaminyl-tRNA synthetase from *E.coli* is the only synthetase known to mis-aminoacylate non-cognate tRNAs. Although this mischarging tendency is enhanced by mutations [25][26], it is inherent to the wild-type enzyme and can be observed when the enzyme is present in excess compared to $tRNA^{Gln}$ [27]. For this reason it is important for the cell to control the level of glutaminyl-tRNA synthetase to prevent the introduction of translational errors via tRNAs mischarged with glutamine. Expression of the *glnS* gene is metabolically regulated, like other components of the translational apparatus [18]. It should be noted though that neither transcription nor translation of *glnS* is maximal since mutations increasing both these steps were isolated [15]. One of these mutations is the GAG to GAA change in the third codon of *glnS* which has been further investigated in this work. Our initial aim was to verify that it was indeed affecting translational initiation rather than elongation. The results confirm this hypothesis since we were able to detect a four-fold difference in expression between the original codon GAG and the mutated GAA codon containing sequence if the codons were located in the beginning, but not further downstream, in the gene (see Table 1). This confirms earlier data obtained using a *glnS-lacZ* fusion [15] but also shows that it is possible to transfer the effect of the mutation in *glnS* into another genetic environment.

A possible explanation for a translation initiation specific effect is that this region of the *glnS* mRNA is involved in secondary interactions with the 1469−1483 region as proposed by Sprengart et al [12]. Data from our experiments with mutant sequencies clearly showed a positive correlation between the level of gene expression and the number of potential base pairs that could be formed between the mRNA and bases in this region of 16S RNA. These results suggest that the interaction between a region of mRNA immediately downstream of the initiation codon and region 1469−1483 indeed is involved in anchoring of *glnS* mRNA and affect initiation efficiency as suggested earlier [12].

An additional explanation of the effect of the GAG to GAA replacement could be that the inherent translation rate of the GAA codon was greater than that of the GAG codon. In fact, the translation rate through GAA codons has been found to be three times faster than that through GAG [28]. As discussed by these authors this kind of difference could produce ribosome queuing or stalling near the translational initiation site making it inaccessible to newly initiating ribosomes. There is indeed circumstantial evidence that this is a regulatory mechanism employed by *E.coli* since there is a preferential occurrence of rare codons in the first 25 codons of *E.coli* genes [29][30]. It should be noted that rare codons are not synomynous with slowly translated codons, and some frequent codons are more slowly translated than others [24][28][29][31]. Neither is there a direct correlation between tRNA abundance and codon translation rate [24][31]. In order to study the effect of slowly translated codons in the beginning of a gene mutant constructs were made which differed only in the nature of the third codon where the arginine codons AGA, AGG and CGU were compared. The first two are extremely rare whereas the last codon is commonly found in highly expressed genes. Moreover it seems clear that AGG and AGA are translated more slowly than CGU [29][30][32][33]. The results reported here indicate that the speed of translation is involved in determining the efficiency of initiation as well.

Our data illustrate the potential importance of two regulatory mechanisms in the expression of the *glnS* gene. They provide clear supporting evidence that the interaction between 16S rRNA and the 'downstream' box, proposed on the basis of phage T7 gene 0.3 expression [12], is important for the expression of a more modest *E.coli* gene, *glnS*. Our results show a good but not total correlation between binding potential to 16S RNA and gene

expression. This is not surprising if gene expression is controlled not only by base pairing potential to 16S RNA but by codon choice as well. In fact, our data also demonstrate that altering the translation rate in the vicinity of the initiation site by insertion of rare, slowly translated codons, can modulate translational initiation. Indeed, the fourfold effect of the GAG to GAA mutation could be a combination of increased base pairing with 16S rRNA and an enhanced translation rate. Furthermore, the occurence of non-standard base pairings and effects on stacking interactions between the mRNA and 16S RNA could complicate the picture. Finally it should be mentioned that the two mechanisms discussed here do not necessarily describe all the potential ways of controlling *glnS* expression since they do not totally explain all of the available regulatory mutants in the beginning of the *glnS* gene [15].

ACKNOWLEDGEMENTS

We thank Asgeir Björnsson for allowing us to use some of his vectors prior to their publication. Jacqueline Plumbridge is pleased to acknowledge receipt of a short-term EMBO fellowship. This work was supported by grants from the Natural Science Reasearch Counil (NFR) to L.A. Isaksson.

REFERENCES

1. Gold, L. and Stormo, G. (1987) In Neidhardt, F.C. (ed) *Escherichia coli and Salmonella typhimurium Cellular and molecular biology*, ASM, Vol **2**, pp 1302–1307.
2. Kozak, M. (1983) *Microbiol. Rev.* **47,** 1–45.
3. Gold, L., Pribnow, D., Schneider, T., Shinedling, S., Singer, B.S. and Stormo, G. (1981) *Ann. Rev. Microbiol.* **35,** 365–403.
4. Scherer, G.F.E., Walkinshaw, M.D., Arnott, S. and Morré, D.J. (1980) *Nucleic Acids Res.* **8,** 3895–3908.
5. Stormo, G.D., Schneider, T.D. and Gold, L.M. (1982) *Nucleic Acids Res.* **10,** 2971–2996.
6. Gren, E.J. (1984) *Biochimie.* **66,** 1–29.
7. Schneider, T.D., Stormo, G.D., Gold, L. and Ehrenfeucht, A. (1986) *J. Mol. Biol.* **188,** 415–431.
8. Dreyfus, M. (1988) *J. Mol. Biol.* **204,** 79–94.
9. Petersen, G.B., Stockwell, P.A. and Hill, D.F. (1988) *EMBO J.* **7,** 3957–3962.
10. Olins, P.O. and Rangwala, S.H. (1989) *J. Biol. Chem.* **264,** 16973–16976.
11. Gallie, D.R. and Kado, C.I. (1989) *Proc. Natl. Acad. Sci. USA.* **86,** 129–132.
12. Sprengart, M.L., P.Fatscher, H. and Fuchs, E. (1990) *Nucleic Acids Res.* **18,** 1719–1723.
13. McCarthy, J.E.G., Sebald, W., Gross, G. and Lammers, R. (1986) *Gene.* **41,** 201–206.
14. Firpo, M.A. and Dahlberg, A.E. (1990) In McCarthy, J.E.G. and Tuite, M.F. (eds) *Post-Transcriptional Control of Gene Expression*, Springer-Verlag Berlin. Heidelberg. Vol **H49**, pp. 185–195.
15. Plumbridge, J. and Söll, D. (1989) *Mol Gen Genet.* **216,** 113–119.
16. Sambrook, J., Fritsch, E.F and Maniatis, T. (1989) *Molecular Cloning, a Laboratory Manuel*. Cold Spring Harbor University Press. Cold Spring Harbor.
17. Miller, J.H. (1972) *Experiments in Molecular Genetics*. Cold Spring Harbor Laboratory Press. Cold Spring Harbor.
18. Neidhardt, F.C., Bloch, P.L., Pedersen, S. and Reeh, S. (1977) *J. Bacteriol.* **129,** 378–387.
19. Björnsson, A. and Isaksson, L.A. (1988) *Nucleosides & Nucleotides.* **7,** 565–569.
20. Nilsson, B., Moks, T., Jansson, B., Abrahmsén, L., Elmblad, A., Holmgren, E. and Henrichson, C. (1987) *Protein Engineering.* **1,** 107–113.
21. Abrahmsén, L., Moks, T., Nilsson, B. and Uhlén, M. (1986) *Nucleic Acids Res.* **14,** 7487–7500.
22. Amann, E., Ochs, B. and Abel, K.J. (1988) *Gene.* **69,** 301–315.
23. Neuhoff, V., Stamm, R. and Eibl, H. (1985) *Electrophoresis.* **6,** 427–448.
24. Bonekamp, F., Dalböge, H., Christensen, T. and Jensen, K.F. (1989) *J. Bacteriol.* **171,** 5812–5816.
25. Inokuchi, H., Hoben, P., Yamao, F., Ozeki, H. and Söll, D. (1984) *Proc. Natl. Acad. Sci. USA.* **81,** 5076–5080.
26. Perona, J.J., Swanson, R.N., Rould, M.A., Steitz, T.A. and Söll, D. (1989) *Science.* **246,** 1152–1154.
27. Swanson, R., Hoben, P., Summer-Smith, M., Uemura, H., Watson, L. and Söll, D. (1988) *Science.* **242,** 1548–1551.
28. Sørensen, M.A., Jensen, K.F. and Pedersen, S. (1990) In McCarthy, J.E.G. and Tuite, M.F. (eds) *Post-Transcriptional Control of Gene Expression*, Springer-Verlag Berlin. Heidelberg. Vol **H49**, pp. 207–216.
29. Chen, G.F.T. and Inouye, M. (1990) *Nucleic Acids Res.* **18,** 1465–1473.
30. Inouye, M. and Chen, G. (1990) In McCarthy, J.E.G. and Tuite, M.F. (eds) *Post-Transcriptional Control of Gene Expression*, Springer-Verlag Berlin. Heidelberg. Vol **H49**, pp. 217–225.
31. Curran, J.F. and Yarus, M. (1989) *J. Mol. Biol.* **209,** 65–77.
32. Bonekamp, F. and Jensen, K.F. (1988) *Nucleic Acids Res.* **16,** 3013–3024.
33. Chen, K.S., Peters, T.C. and Walker, J.R. (1990) *J.Bacteriol.* **172,** 2504–2510.