Volume 17 Number 24 1989 Nucleic Acids Research

Expression of tetanus toxin fragment C in *E.coli*: high level expression by removing rare codons

A.J.Makoff*, M.D.Oxer, M.A.Romanos, N.F.Fairweather and S.Ballantine[1]

Departments of Molecular Biology and [1]Protein Chemistry, Wellcome Biotech, Langley Court, Beckenham, Kent BR3 3BS, UK

Received October 20, 1989; Revised and Accepted November 20, 1989

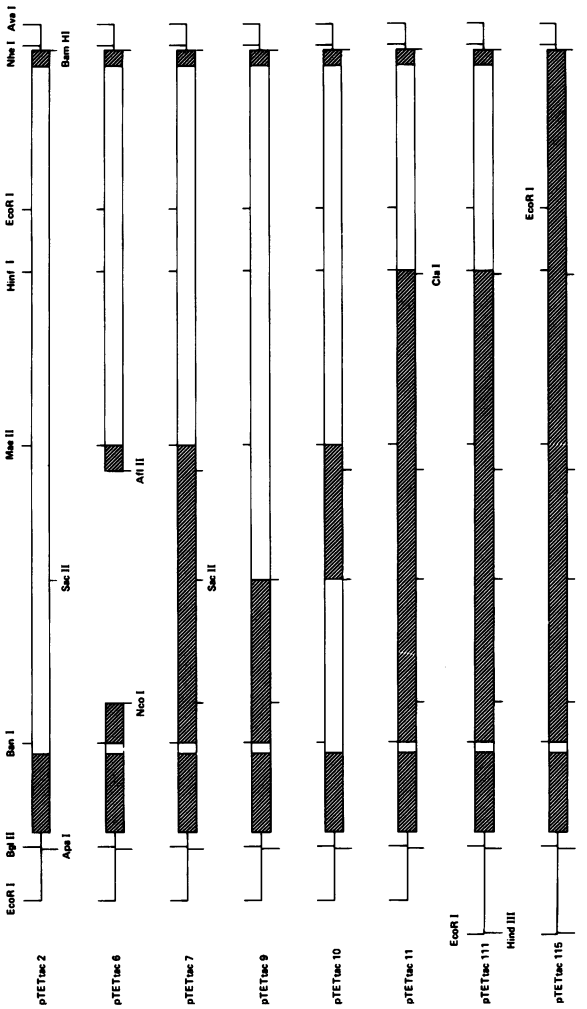
ABSTRACT

Tetanus toxin fragment C had been previously expressed in *Escherichia coli* at 3−4% cell protein. The codon bias for tetanus toxin in *Clostridium tetani* is very different from that of highly expressed homologous genes in *E. coli*, resulting in the presence of many rare *E. coli* codons in the sequence encoding fragment C. We have replaced the coding sequence by sequence optimized for codon usage in *E. coli*, and show that the expression of fragment C is increased. Although the level of mRNA also increased this appeared to be a secondary consequence of more efficient translation. Complete sequence replacement increased expression to approximately 11−14% cell protein but only after the promoter strength had been improved.

INTRODUCTION

It has long been recognized that the choice of codons is not random and varies from organism to organism. The greatest deviation from random codon choice occurs with highly expressed genes (1−3). This is thought to reflect, in part, selection against the use of codons specifying tRNA molecules of low abundance and in part, selection against codons whose binding energy for interactions with the anti-codon loops of the cognate tRNA molecule is suboptimal (4). It is thought that such selection operates against poor rates of expression (5). This is supported by the finding that genes with a high proportion of rare codons are translated at a lower elongation rate than those with a low proportion (6). Insertion of tandem rare codons into homologous genes has also been shown to increase the translation time and to lower expression levels (7, 8).

Because of these considerations it was always expected that when heterologous genes were expressed, a poor fit between the codon bias of the gene and of the host organism would limit expression. However, there are many examples of high level expression in *E. coli* of heterologous genes with an unfavourable codon bias. Table 1 shows some of the published examples of high level expression (greater than 10% cell protein) of heterologous genes in *E. coli* and their usage of 9 of the rare *E. coli* codons as defined by Sharp and Li (5). [The codon GGG was excluded partly because it was less rarely used than the others and partly because it specifies two anticodons (1)]. As can be seen, the total usage of rare codons can be as high as 15% (Influenza NS1 protein), or the individual usage as high as 6.0% (β-interferon) for AGA and AGG by their common tRNA isoacceptor and yet still give high expression levels. Ernst and Kawashima (25) compared constructions of a gene fusion between yeast α-factor and somatomedin C which differed only in the coding sequence. They could find no correlation between codon bias and expression levels in *E. coli* or yeast. By contrast Williams *et al* (26) carried out a similar


Nhe I
Ava I
Bam HI
EcoR I
Hinf I
Mae II
Sac II
Ban I
Nco I
Bgl II
Apa I
EcoR I
Aff II
Cla I
Hind III
pTETtac 2
pTETtac 6
pTETtac 7
pTETtac 9
pTETtac 10
pTETtac 11
pTETtac 111
pTETtac 115

analysis on expression of interleukin 2 and reported a sixteen-fold stimulation by increasing the proportion of preferred codons. However, they obtained much lower levels of expression of interleukin 2 than Devos *et al* (12) who also used the original cDNA, suggesting that rare codons were not limiting. Their data also do not rule out the possibility that the difference in expression level which they ascribed to improving the codon bias is due to changes in mRNA stability.

Fragment C from tetanus toxin shows considerable promise as a sub-unit vaccine against tetanus (27, 28). The coding sequence has been expressed in *E. coli* to give a soluble product at 3−4% cell protein (28). The sequence is very AT-rich (29) and contains a large number of codons which are rarely used in highly expressed *E. coli* genes. In this paper we show that the expression of fragment C is limited by its high demand for rare tRNA molecules.

MATERIALS AND METHODS

Strains

E. coli strain MM294 (30) was used throughout this work.

Plasmids

The construction of pTETtac2 has already been described (28). It was used to construct pTETtac7 via the intermediate plasmid pTETtac6 as shown in Fig. 1. This involved cloning two oligonucleotides between the *Ban*I and *Mae*II sites of pTETtac2 in order to produce the two unique sites *Nco*I and *Afl*II in pTETtac6. Eight more oligonucleotides were then cloned between these two sites to create pTETtac7. Plasmids pTETtac9 and 10 were constructed from pTETtac2 and 7 by ligating DNA from both plasmids from either side of the unique *Sac*II site. Plasmid pTETtac11 was made by cloning six oligonucleotides between the *Afl*II and *Hin*fI sites of pTETtac7 as shown in Fig. 1. The synthetic sequence specified by all the oligonucleotides is given in Fig. 2.

Plasmid pTETtac111 was constructed by cloning four oligonucleotides between the *Eco*RI and *Apa*I sites of pTETtac11 (see Fig. 1). These specified the *tac* promoter from −112 of the *trp* promoter to the 3′ end of the *lac* operator (31). Plasmids pTETtac102 and pTETtac107 were made by ligating DNA from pTETtac111 5′ to the *Bgl*II site with DNA from pTETtac2 or pTETtac7 3′ to the *Bgl*II site. Plasmid pTETtac115 was made by cloning six oligonucleotides between the *Cla*I and *Bam*HI sites of pIC-20H (32) before subcloning the *Cla*I-*Bam*HI fragment into pTETtac111.

All synthetic oligonucleotides were made on a Pharmacia Gene Assembler and purified by electrophoresis on a denaturing polyacrylamide gel. Primers were synthesized and purified in the same way and were used to confirm the sequence of all synthetic oligonucleotide-specified DNA by direct sequencing of the plasmids (33).

Induction of Cultures

Cultures containing various plasmids were induced as described previously (28, 34). The levels of expression of fragment C were quantified either by an ELISA and protein assay (28) or by densitometer scanning of SDS-polyacrylamide gels using a Joyce-Loeble Chromoscan 3.

Fig. 1 Restriction maps of the fragment C-encoding part of expression plasmids. For complete map of pTETtac2 see ref. 28. Box depicts fragment C sequence: open box native sequence; hatched box synthetic sequence. *Tac* promoter is between the first *Eco*RI and *Apa*I sites.

```
                                        30                                                  60                                                  90
ATG AAA AAT CTG GAT TGT TGG GTT GAT AAT GAA GAA GAT ATA GAT GTT ATA TTA AAA AAG AGT ACA ATT TTA AAT TTA GAT ATT AAT AAT
      C   T             C   C   C                 C   C           C C G           TC    C       C G   C   G   C   C   C
Met Lys Asn Leu Asp Cys Trp Val Asp Asn Glu Glu Asp Ile Asp Val Ile Leu Lys Lys Ser Thr Ile Leu Asn Leu Asp Ile Asn Asn

                                       120                                                 150      BanI                                       180
GAT ATT ATA TCA GAT ATA TCT GGG TTT AAT TCA TCT GTA ATA ACA TAT CCA GAT GCT CAA TTG GTG CCC GGA ATA AAT GGC AAA GCA ATA
        C   C   C   C       T   C   C   C       T   C                                                   G   C   C   C           T   C
Asp Ile Ile Ser Asp Ile Ser Gly Phe Asn Ser Ser Val Ile Thr Tyr Pro Asp Ala Gln Leu Val Pro Gly Ile Asn Gly Lys Ala Ile

                                       210                                                 240                                                 270
CAT TTA GTA AAC AAT GAA TCT TCT GAA GTT ATA GTG CAT AAA GCT ATG GAT ATT GAA TAT AAT GAT ATG TTT AAT AAT TTT ACC GTT AGC
  C C G   T       C                       C       C   G   C       C   C       C   C   C       C   C   C   C
His Leu Val Asn Asn Glu Ser Ser Glu Val Ile Val His Lys Ala Met Asp Ile Glu Tyr Asn Asp Met Phe Asn Asn Phe Thr Val Ser

                                       300                                                 330                                                 360
TTT TGG TTG AGG GTT CCT AAA GTA TCT GCT AGT CAT TTA GAA CAA TAT GGC ACA AAT GAG TAT TCA ATA ATT AGC TCT ATG AAA AAA CAT
  C     C   C C       G       T           TCC   C C G       G   C       T   C       C   C   C   C                   G       C
Phe Trp Leu Arg Val Pro Lys Val Ser Ala Ser His Leu Glu Gln Tyr Gly Thr Asn Glu Tyr Ser Ile Ile Ser Ser Met Lys Lys His

                                       390                                                  20      SacII                                      450
AGT CTA TCA ATA GGA TCT GGT TGG AGT GTA TCA CTT AAA GGT AAT AAC TTA ATA TGG ACT TTA AAA GAT TCC GCG GGA GAA GTT AGA CAA
TCC   G   C   C   C             TC    T   C   G   G       C     C G   C         C G       C             C         C T   G
Ser Leu Ser Ile Gly Ser Gly Trp Ser Val Ser Leu Lys Gly Asn Asn Leu Ile Trp Thr Leu Lys Asp Ser Ala Gly Glu Val Arg Gln

                                       480                                                 510                                                 540
ATA ACT TTT AGG GAT TTA CCT GAT AAA TTT AAT GCT TAT TTA GCA AAT AAA TGG GTT TTT ATA ACT ATT ACT AAT GAT AGA TTA TCT TCT
  C       C C C   C C G   G   C   G   C   C   G   C C G   T   C                   C   C       C       C     C T C G
Ile Thr Phe Arg Asp Leu Pro Asp Lys Phe Asn Ala Tyr Leu Ala Asn Lys Trp Val Phe Ile Thr Ile Thr Asn Asp Arg Leu Ser Ser

                                       570                                                 600                                                 630
GCT AAT TTG TAT ATA AAT GGA GTA CTT ATG GGA AGT GCA GAA ATT ACT GGT TTA GGA GCT ATT AGA GAG GAT AAT AAT ATA ACA TTA AAA
      C C G   C   C   C   C   T   G       C TCC   T       C       C G   C       C C T       C   C   C   C   T C T   G
Ala Asn Leu Tyr Ile Asn Gly Val Leu Met Gly Ser Ala Glu Ile Thr Gly Leu Gly Ala Ile Arg Glu Asp Asn Asn Ile Thr Leu Lys

                                    MaeII                                                  690                                                 720
CTA GAT AGA TGT AAT AAT AAT AAT CAA TAC GTT TCT ATT GAT AAA TTT AGG ATA TTT TGC AAA GCA TTA AAT CCA AAA GAG ATT GAA AAA
  G   C C T   C   C   C   C   C   G       A   C   C   C   G   C C T   C   C               C G   C   G           C
Leu Asp Arg Cys Asn Asn Asn Asn Gln Tyr Val Ser Ile Asp Lys Phe Arg Ile Phe Cys Lys Ala Leu Asn Pro Lys Glu Ile Glu Lys

                                       750                                                 780                                                 810
TTA TAC ACA AGT TAT TTA TCT ATA ACC TTT TTA AGA GAC TTC TGG GGA AAC CCT TTA CGA TAT GAT ACA GAA TAT TAT TTA ATA CCA GTA
C G   T   C   C   C C G       C       C C G C T               T       G C G   T   C   C   C           C C G   C   G
Leu Tyr Thr Ser Tyr Leu Ser Ile Thr Phe Leu Arg Asp Phe Trp Gly Asn Pro Leu Arg Tyr Asp Thr Glu Tyr Tyr Leu Ile Pro Val

                                       840                                                 870                                                 900
GCT TCT AGT TCT AAA GAT GTT CAA TTG AAA AAT ATA ACA GAT TAT ATG TAT TTG ACA AAT GCG CCA TCG TAT ACT AAC GGA AAA TTG AAT
          C           C       G C         C   C   T   C   C       C C     C   C       G   C   C           T       C   C
Ala Ser Ser Ser Lys Asp Val Gln Leu Lys Asn Ile Thr Asp Tyr Met Tyr Leu Thr Asn Ala Pro Ser Tyr Thr Asn Gly Lys Leu Asn

                                       930                                                 960      HinfI                                      990
ATA TAT TAT AGA AGG TTA TAT AAT GGA CTA AAA TTT ATT ATA AAA AGA TAT ACA CCT AAT AAT GAA ATA GAT TCT TTT GTT AAA TCA GGT
  C   C   C C   C T C G   C   C   C   G       C   C   C     C C   C   T   G   C   C       C           C           T
Ile Tyr Tyr Arg Arg Leu Tyr Asn Gly Leu Lys Phe Ile Ile Lys Arg Tyr Thr Pro Asn Asn Glu Ile Asp Ser Phe Val Lys Ser Gly

                                      1020                                                1050                                                1080
GAT TTT ATT AAA TTA TAT GTA TCA TAT AAC AAT AAT GAG CAC ATT GTA GGT TAT CCG AAA GAT GGA AAT GCC TTT AAT AAT CTT GAT AGA
  C   C   C G   C   T   T   C       C   C   A       C   T       C           C   T   C   T   C   C   C   G   C
Asp Phe Ile Lys Leu Tyr Val Ser Tyr Asn Asn Asn Glu His Ile Val Gly Tyr Pro Lys Asp Gly Asn Ala Phe Asn Asn Leu Asp Arg

                                      1110                                                1140                                                1170
ATT CTA AGA GTA GGT TAT AAT GCC CCA GGT ATC CCT CTT TAT AAA AAA ATG GAA GCA GTA AAA TTG CGT GAT TTA AAA ACC TAT TCT GTA
      G C T   T       C   C   T   G           G   G   C                   T   T       C         C C G           C       T
Ile Leu Arg Val Gly Tyr Asn Ala Pro Gly Ile Pro Leu Tyr Lys Lys Met Glu Ala Val Lys Leu Arg Asp Leu Lys Thr Tyr Ser Val

                                      1200                                                1230                                                1260
CAA CTT AAA TTA TAT GAT GAT AAA AAT GCA TCT TTA GGA CTA GTA GGT ACC CAT AAT GGT CAA ATA GGC AAC GAT CCA AAT AGG GAT ATA
  G   G     C G   C   C   C       C   T     C G   T   G   T           C   C       G   C   T       C   G   C C T   C   C
Gln Leu Lys Leu Tyr Asp Asp Lys Asn Ala Ser Leu Gly Leu Val Gly Thr His Asn Gly Gln Ile Gly Asn Asp Pro Asn Arg Asp Ile

                                      1290                                                1320                                                1350
TTA ATT GCA AGC AAC TGG TAC TTT AAT CAT TTA AAA GAT AAA ATT TTA GGA TGT GAT TGG TAC TTT GTA CCT ACA GAT GAA GGA TGG ACA
C G   C   T TCT             C   C   C C G       C       C C G   T   C   C           C   T   G   C                   T       C
Leu Ile Ala Ser Asn Trp Tyr Phe Asn His Leu Lys Asp Lys Ile Leu Gly Cys Asp Trp Tyr Phe Val Pro Thr Asp Glu Gly Trp Thr

AAT GAT TAA
  C   C
Asn Asp STOP
```

Fig. 2 Complete sequence of fragment C. Top line shows ATG followed by the *C. tetani* sequence (29) with the changes in the synthetic sequence given below. Restriction sites used in the construction are also shown.

Table 1 Percentage utilization of each rare codon by highly expressed heterologous genes in *E. coli*

Gene	CUA	AUA	UCG	CCC	CGA	CGG	AGA AGG	GGA	Total	Expression*	Ref.
FluNS1	2.6	1.7	0.4	0.4	1.7	2.6	2.6	3.0	15.2	20–25	9
bGH	0	0	0.5	1.6	0	1.6	1.6	0.5	5.8	NE	10
pGH	0	0	2.1	1.6	0	2.1	1.1	1.1	7.9	NE	10
hIL-2	3.0	0.8	0	2.2	0	0	3.0	1.5	9.8	>10	11
hγ-IF	0	1.4	1.4	0	2.8	0	2.1	0.7	9.0	25	12
hβ-IF	1.2	1.2	0	0	0.6	0	6.0	1.2	10.2	15	13
bPC	0.3	0.3	1.1	0.8	0.3	0	1.6	0.3	4.7	20	14
mTNF	0	0.6	0.6	3.8	0.6	0	0	1.9	7.7	24	15
sGH	1.1	1.1	0	1.1	0	1.1	2.7	1.1	8.0	15	16
PP	1.4	1.4	0	0.5	0.5	0	2.9	3.8	10.6	15	17
hR	0.3	0.3	0.9	2.4	0.3	0.3	0.9	2.6	7.9	>10	18
ras21	0.5	0.5	0	1.0	0	3.7	1.0	0.5	7.3	>20	19
Cel	0	0	2.7	2.1	0	1.2	0.2	0	6.2	>20	20
cL	0.8	0.8	0	0	0.8	1.6	2.3	4.7	10.9	25	21
HIVRT	1.4	4.1	0.2	0.9	0	0	3.2	3.6	13.4	10	22
cAK	0	0	1.0	2.6	0	3.6	0	1.0	8.2	10	23
hSOD	0.5	0.5	0	1.5	0	0.5	1.0	1.0	5.1	25	24

Abbreviations: FluNS1=influenza non-structural protein 1; bGH, pGH and sGH=bovine, porcine and salmon growth hormones; hIL-2=human interleukin 2; hγ-IF, hβ-IF=human γ and β interferons; bPC=calf (bovine) prochymosin; mTNF=murine tumour necrosis factor; PP=poliovirus protease; hR-human renin; ras21=p21 of Ha-ras; Cel=*C. fimi* cellulose; cL=chicken lysozyme; HIVRT=human immunodeficiency virus reverse transcriptase; cAK=chicken muscle adenylate kinase; hSOD=human superoxide dismutase. NE = not estimated. *Expression is given as a percentage of total cell protein.

Northern blots

RNA was prepared from frozen pellets from induced cultures as described by von Gabain *et al* (35). All apparatus was pre-treated with diethylpyrocarbonate. The RNA was used immediately or stored at −70°C before use. When required, it was glyoxylated and 5μg was applied to each track of a 1% agarose gel in 10mM phosphate buffer pH7.0 as previously described (36). The RNA was blotted overnight onto a nylon membrane (Gene Screen), which was then baked for 2 hours. It was prehybridized for several hours at 42°C and hybridized overnight at 42°C with an oligonucleotide or DNA probe labelled with ^{32}P by random priming (37). For hybridizing to fragment C-specific transcripts, oligo-3 (28) which covers sequence from the *BglII* site to the end of the first hatched area of Fig.1, was used to generate the probe. The membrane was then washed with 2×SSC at 25°C, 2×SSC + 0.1% SDS at 65°C followed by 0.1×SSC at 25°C. The hybridizing mRNA was visualized by autoradiography. The strength of the signal detected by autoradiography was quantified by densitometer scanning of the autoradiographs.

RESULTS

Limitation of Expression by Rare Codons

We have previously shown that induced cultures of *E. coli* containing pTETtac2 express fragment C at approximately 3–4% of total cell protein (28). The plasmid pIFGtac124A (38) which differs from pTETtac2 only in having the coding sequence for γ-interferon (γ-IF) in place of tetanus toxin fragment C, expresses γ-IF at 30–35% of total cell protein

Table 2 Frequency of *E. coli* rare codons in fragment C sequence

Amino Acid	Codon	pTETtac2 pTETtac102	pTETtac7 pTETtac107	pTETtac11 pTETtac111	pTETtac115
Leu	CUA	5(1.1)	3	2	0
Ile	AUA	19(4.2)	9	2	0
Ser	UCG	1(0.2)	1	0	0
Pro	CCC	1(0.2)	0	0	0
Arg	CGA	1(0.2)	1	1	1
	CGG	0	0	0	0
	AGA AGG	14(3.1)	8	3	1
Gly	GGA	12(2.7)	6	3	0

Percentage use shown in brackets.

(unpublished observations). On a molar basis this corresponds to a 30-fold higher level of expression. Northern blotting showed that the γ-IF transcript was only two to three-fold more abundant than the fragment C transcript (data not shown) and therefore the difference in expression was mainly at or after the level of translation.

There are a large number of rare codons in the fragment C coding sequence and in particular those encoded by the minor tRNA molecules for isoleucine (AUA), arginine (AGA, AGG) and glycine (GGA) (Table 2). Because the rare codons were fairly evenly spread, in order to reduce their number, almost the entire coding sequence was replaced by a synthetic sequence (Fig. 2). This was done in several stages to provide a series of plasmids which contained an increasing proportion of synthetic sequence (Fig. 1). The effect of these changes on the number of rare codons in each of the plasmids is shown in Table 2. Extracts of induced cultures containing these plasmids were analysed by SDS-polyacrylamide gel electrophoresis. As shown in Fig. 3 and Table 3, pTETtac7 which has approximately half the number of rare codons as pTETtac2 shows a significant increase in expression.

The increase in expression in pTETtac7 over pTETtac2 is consistent with the limiting factor being the dependence on rare tRNA molecules during translation. However, it could also be due to the removal of a specific sequence element which decreases expression, for example, a transcriptional termination signal or an endonucleolytic cleavage site in the mRNA molecule. To investigate this, the unique *Sac*II site present in both pTETtac2 and 7 was utilized to create two new plasmids pTETtac9 and 10 (Fig. 1). If such an element had been removed in pTETtac7, then one of the new plasmids should give levels of expression like pTETtac7 while the other would behave like pTETtac2. Extracts of induced cultures containing all 4 plasmids were analysed both by SDS-polyacrylamide gel electrophoresis (not shown) and by an ELISA (see Materials and Methods). As shown in Table 3 the two new plasmids produced fragment C at levels intermediate between those of pTETtac2 and 7. (However, the differences were only significant in the case of pTETtac10.) It is therefore very unlikely that the native region of the sequence, altered in pTETtac7, contains specific elements for RNA degradation or termination.

Reducing the number of rare codons still further, in pTETtac11 (Table 2 and Fig. 1), made no detectable improvement in the level of expression (Fig. 3 and Table 3). The level of expression of approximately 6–7% total cell protein is still much lower than that achieved for γ-IF by pIFGtac124A, and yet rare codons are presumably no longer limiting.

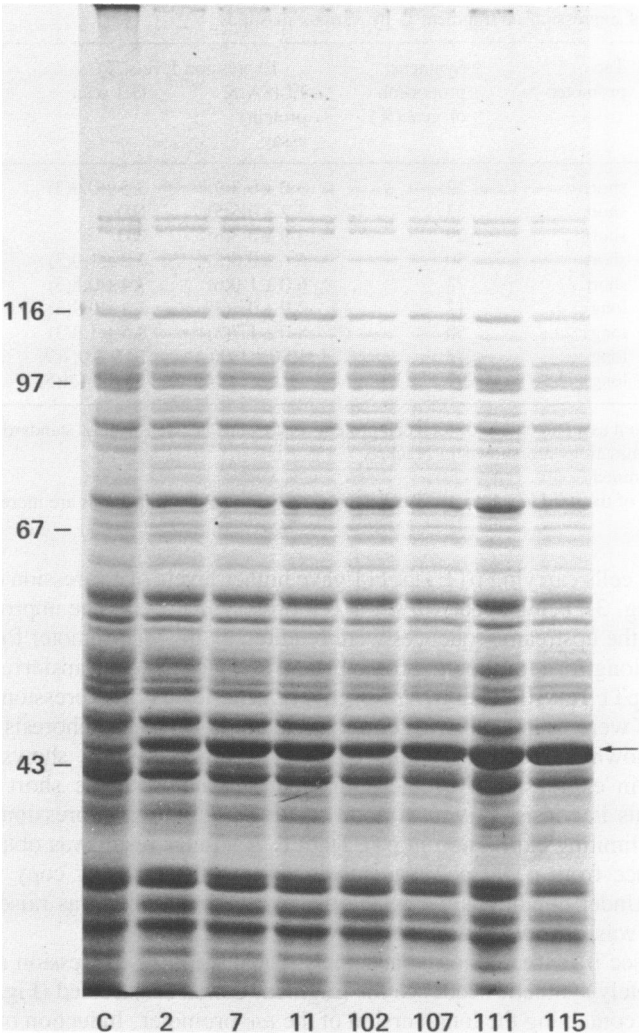


Fig. 3 Coomassie blue stained SDS-polyacrylamide gel analysis of induced extracts containing various plasmids. The arrow indicates the position of fragment C. The numbers of each track refer to the plasmid present in the culture; C refers to the control track. The size markers are given in Kdaltons.

Limitation of expression by transcription

The *tac* promoter used in the plasmids described so far lacks *trp* promoter sequence upstream of the *Ssp*I site (AATATT) near position −50 (31). However, work by Nishi and Itoh (39) has shown that part of this upstream sequence (particularly two AT rich blocks) is required for full *trp* promoter function. In order to see if this sequence was required for full *tac* promoter function, oligonucleotides were cloned upstream of the *tac* promoter in pTETtac11 to replace some of the missing sequence, including the two AT-rich blocks, to give pTETtac111 (Fig. 2).

Table 3 Levels of expression of fragment C by various plasmids

Plasmid	Tac promoter	Synthetic proportion of gene(%)	Expression levels(%)* ELISA & protein assay	Gel scan	Relative levels (combined)
pTETtac2	short	12	3.0±0.3(9)	3.5±0.9(3)	1.0
9	short	33	3.2±0.5(5)	ND	–
10	short	29	3.9±0.3(6)	ND	–
7	short	50	5.6±0.6(8)	7.2±1.0(3)	2.0
11	short	73	6.0±1.0(6)	7.4±0.8(3)	2.1
102	long	12	2.9±0.4(3)	4.2±0.9(3)	1.1
107	long	50	8.0±1.7(3)	8.5±1.3(3)	2.5
111	long	73	10.0±1.0(6)	10.9±0.7(9)	3.2
115	long	99	11.4±1.3(4)	14.2±1.4(5)	3.9

*Expression is given as a percentage of total cell protein. Each value is given with ± standard error of the mean. Number of determinations are shown in brackets
ND = not determined
Note that because of the need to subtract a control value, the gel scan determinations are more accurate at higher expression levels.

Induction of cells carrying pTETtac111 gave higher levels of expression compared with pTETtac11 (Fig. 3), being about 10−11% cell protein (Table 3). The improved expression suggested that the upstream sequences were required by the *tac* promoter for full promoter function. The long form of the *tac* promoter of pTETtac111 was transferred to pTETtac2 and 7 to give pTETtac102 and 107 respectively. The levels of expression obtained with these plasmids were estimated by SDS-polyacrylamide gel electrophoresis and ELISA as before. As shown in Table 3, unlike pTETtac111, pTETtac102 shows no detectable improvement in expression over the equivalent plasmid with the short *tac* promoter, pTETtac2. This is consistent with a poor codon bias limiting expression in pTETtac2, but no longer limiting expression in pTETtac11. A similar result was obtained when the coding sequence from pTETtac2 was subcloned into the runaway copy number vector pMM1 (40). Under conditions where the plasmid copy number was raised, no increase in expression was observed (data not shown).

In order to see if more codon replacements would improve expression still further, an almost completely synthetic sequence for fragment C was constructed (Fig. 2) and cloned into a plasmid containing the long version of the *tac* promoter. Induction of cells carrying the resulting plasmid pTETtac115 (Fig. 1 and Table 2) gave slightly higher expression at 11−14% cell protein (Table 3). We reported previously that fragment C produced by pTETtac2 was soluble (28). We investigated induced cultures containing pTETtac115 in the same way as before (28) and found that even when expressed to a much higher level, fragment C was still fully soluble (data not shown).

Increase in mRNA levels by sequence replacements

The levels of mRNA specific for fragment C were estimated by Northern blotting of RNA isolated from three induced independent cultures of cells each containing pTETtac2, 11, 102 or 111. Two samples of each RNA extract were separated on an agarose gel, and blotted onto a nylon membrane. One blot was hybridized with an oligonucleotide probe specific for a fragment C sequence common to all four plasmids (see Materials and Methods) and the other as a control was hybridized to a probe specific for β-lactamase sequence.

Table 4 Relative levels of transcripts from pTETtac2, 11, 102 and 111

Plasmid	Fragment C probe	B-lactamase probe
pTETtac2	1.0 ± 0.4	1.0 ± 0.3
11	2.9 ± 0.3	0.5 ± 0.03
102	2.0 ± 0.2	0.8 ± 0.2
111	7.0 ± 1.0	0.6 ± 0.1

Levels given (± standard error of the mean) are relative to the level for pTETtac2. See text for details.

A set of DNA fragments specific to the fragment C probe was run as markers (1389bp, 819bp and 253bp). In each track of the Northern blot shown in Fig. 4, the probe hybridized to a transcript slightly larger than the 1389bp marker, consistent with a full length fragment C transcript. Each track also contained a number of smaller transcripts in approximately the same amount relative to the full length transcript. The relative abundance of each full length transcript was estimated by densitometer scanning and is shown in Table 4. This shows that the plasmids with the long form of the *tac* promoter (pTETtac102 and 111) both produce approximately twice as much specific mRNA as the comparable plasmid with the short form of the promoter (pTETtac2 and 11). It also shows that the plasmids with more synthetic fragment C-specific DNA (pTETtac11 and 111) both produce approximately three times more transcript than their counterparts containing more native *C.tetani* DNA (pTETtac2 and 102). The control blot using a β-lactamase-specific probe was quantified in a similar way. The relative amounts of this transcript are also given in Table 4. For reasons that are unclear, there appear to be some differences in β-lactamase mRNA levels, but in no case do they parallel the differences in fragment C mRNA levels. Therefore, the differences in levels of fragment C-specific transcripts are unlikely to be due to different amounts of mRNA being loaded onto the gel.

DISCUSSION

The coding sequence for tetanus toxin fragment C has many codons which are rarely used by *E. coli* in highly expressed genes. We have synthesized almost the entire coding sequence for fragment C in order to remove most of these rare codons. On induction, this resulted in high level expression of fragment C in *E. coli* at 11−14% of total cell protein. By comparison, the mainly native sequence was only expressed at approximately 3% of cell protein (28, this study). We have demonstrated that this four-fold improvement in expression is almost certainly due to relief from a requirement for rare tRNA molecules during translation.

By replacing a large sequence, we have potentially altered many factors which affect expression. Besides codon bias these include (i) plasmid copy number, (ii) promoter strength, (iii) premature termination of transcription at internal sites, (iv) mRNA turnover, and (v) efficiency of the ribosome binding site. Factors (i) and (v) are most unlikely to be involved since no sequence changes were made within 221 nucleotides of the presumed start of transcription or within 158 nucleotides of the AUG initiation codon. Any long range interactions at the DNA or mRNA level between the promoter region or the ribosome binding site and a distant site within the coding sequence cannot be affecting expression, since a substantial region of the sequence had to be altered in order for the maximum improvement in expression to be observed. Alterations in smaller regions led to smaller improvements in expression.

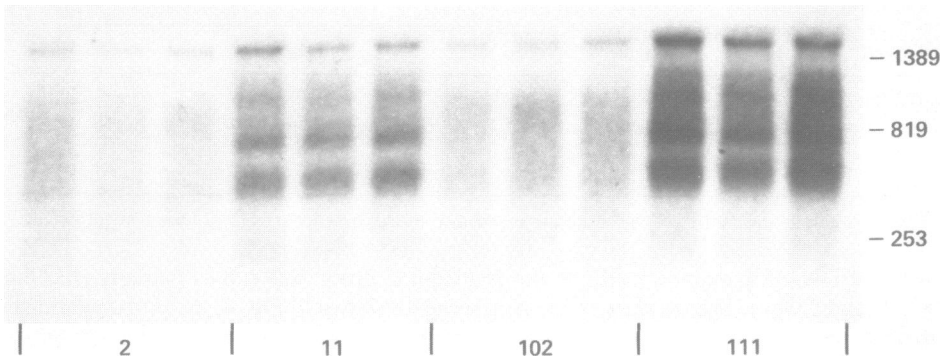


Fig. 4 Northern blot of RNA isolated from induced cultures containing various plasmids. Each track refers to an individual culture. The number on each group of three tracks refer to the plasmid in the three cultures. The size markers are DNA fragments and are in nucleotides.

Changes in factors (i)−(iv) which lead to an increase in expression, would all be expected to raise the steady state level of fragment C-specific mRNA. Clearly the specific mRNA levels did increase when only the coding sequence was changed (compare pTETtac11 with pTETtac2, and pTETtac111 with pTETtac102 in Fig. 4 and Table 4). However, there are two lines of evidence which are consistent with the increase in expression being due to a limitation of rare codons and not due to an increase in the level of mRNA. Firstly, our experiments which deliberately increased the fragment C-specific mRNA by increasing the promoter strength failed to increase expression (compare pTETtac102 with pTETtac2 in Tables 3 and 4). This is supported by a similar failure to increase expression by increasing the plasmid copy number. Both experiments clearly show that the mRNA level in inductions with pTETtac2 is not the limiting factor. Secondly, with the weaker *tac* promoter, successive changes in the coding sequence all increased expression until a limit was reached. With the stronger *tac* promoter, the same trend was observed but a limit was not reached until more sequence had been replaced. It is difficult to see how this could happen if the sequence replacements were acting by increasing mRNA levels. Yet, this is exactly as would be predicted if limiting rare codons were successively being replaced until they were no longer the limiting factor. The point at which they ceased to be limiting would depend on mRNA levels. Robinson et *al* (8) reported a similar effect when they inserted a block of rare codons into the chloramphenicol transacetylase gene. Expression decreased if the cells were fully induced, but the rare codons had no effect if the cells were only partially induced.

For these reasons, the changes in mRNA levels following the coding sequence changes appear to be a consequence of their effect on translation via rare codons. It is known that rare codons are translated more slowly than more commonly used ones (6, 7, 41, 42). Since in prokaryotes, transcription and translation are coupled, pausing of ribosomes will result in an increase in the length of 'naked' RNA between the transcribing RNA polymerase molecule and the first ribosome. Such stretches of mRNA are thought to be particularly susceptible to the action of endonucleases (43, 44) and to rho-dependent termination (45). In either event, the presence of a large number of rare codons might be expected to decrease the population of fragment C-specific mRNA molecules. A similar phenomenon has been described with mRNA molecules which are poorly translated due to a weak ribosome binding site (46).

If so many other heterologous genes containing a high proportion of rare codons are expressed to high levels (Table 1) why is the expression of fragment C limited by rare codons? It has been suggested that rare codons are more likely to limit expression if they are clustered (47). However, that cannot be relevant to fragment C since its rare codons are fairly evenly distributed throughout the sequence and only in one instance (AGA, AGG at positions 910–915 in Fig. 2) are there adjacent codons which specify the same rare tRNA molecule. Since it is clear from our experiments that when rare codons are limiting expression, raising the level of the specific mRNA will not raise the expression level, any differences in mRNA levels between the various expression vectors are also unlikely to be relevant. One possibility might be that fragment C is more rapidly turned over in the cell than are all the other heterologous proteins quoted in Table 1. This is plausible, especially since in almost all of these other examples, the heterologous protein is expressed in an insoluble form where it is presumably relatively inaccessible to the action of host proteases. Fragment C, by contrast is soluble (28).

There is an intriguing possibility that a single rare codon might be responsible for the limitation of expression of fragment C in *E. coli*. As shown in Table 2, the rare codon most frequently used by the fragment C gene is AUA which is used at 4.2%. With one exception (HIV RT), this is at a much higher frequency than any of the other highly expressed heterologous genes (see Table 1). Ikemura (1) showed by direct measurement that the tRNA specific for codon AUA is among the least abundant. Subsequently, Gouy and Cautier (3) have extrapolated from codon usage analyses to conclude that the tRNA for AUA is actually the rarest. It is therefore possible that a high dependence on the AUA codon is particularly likely to limit expression but, as discussed above, the effect may be masked by a very stable protein.

ACKNOWLEDGEMENTS

We would like to thank A. Smallwood for her help in carrying out preliminary experiments, C.H. Reynolds for advice on the statistical analysis and T. Silva for typing this paper.

*To whom correspondence should be addressed

REFERENCES

1. Ikemura, T. (1981) J. Mol. Biol. 146, 1–21.
2. Ikemura, T. (1981) J. Mol. Biol. 151, 389–409.
3. Gouy, M. and Gautier, C. (1982) Nucleic Acids Res. 10, 7055–7074.
4. Grosjean, H. and Fiers, W. (1982) Gene 18, 199–209.
5. Sharp, P.M. and Li, W-H. (1986) Nucleic Acids Res. 14, 7737–7749.
6. Pedersen, S. (1984) EMBO J. 3, 2895–2898.
7. Sorensen, M.A., Kurland, C.G. and Pedersen, S. (1989) J. Mol. Biol. 207, 365–377.
8. Robinson, M., Lilley, R., Little, S., Emtage, J.S., Yarranton, G., Stephens, P., Millican, A., Eaton, M., and Humphreys, G. (1984) Nucleic Acids Res. 12, 6663–6671.
9. Young, J.F., Desselberger, U., Palese, P., Ferguson, B., Shatzman, A.R. and Rosenberg, M. (1983) Proc. Natl. Acad. Sci. USA. 6105–6109.
10. Seeburg, P.H., Sias, S., Adelman, J., de Boer, H.A., Hayflick, J., Jhurani, P., Goeddel, D.V. and Heyneker, H.L. (1983) DNA 2, 37–45.
11. Devos, R., Plaetinck, G., Cheroutre, H., Simons, G., Degrave, W., Tavernier, J., Remaut, E., and Fiers, W. (1983) Nucleic Acids Res. 11, 4307–4322.
12. Simons, G., Remaut, E., Allet, B., Devos, R. and Fiers, W. (1984) Gene 28, 55–64.
13. Whitehorn, E.A., Livak, K.J. and Petteway Jr, S.R. (1985) Gene 36, 375–379.
14. Schoemaker, J.M., Brasnett, A.H. and Marston, F.A.O. (1985) EMBO J. 4, 775–780.

15. Pennica, D., Hayflick, J.S., Bringman, T.S., Palladino, M.A. and Goeddel, D.V. (1985) Proc. Natl. Acad. Sci. USA. 82, 6060–6064.
16. Sekine, S., Mizukami, T., Nishi, T., Kuwana, Y., Saito, A., Sato, M., Itoh, S. and Kawauchi, H. (1985) Proc. Natl. Acad. Sci. USA. 82, 4306–4310.
17. Ivanoff, L.A., Towatari, T., Ray, J., Korant, B.D. and Petteway Jr, S.R. (1986) Proc. Natl. Acad. Sci. USA. 83, 5392–5396.
18. Kaytes, P.S., Theriault, N.Y., Poorman, R.A., Murakami, K. and Tomich C-S.C. (1986) J. Biotechnol. 4, 205–218.
19. Tucker, J., Sczakiel, G., Feuerstein, J., John, J., Goody, R.S. and Wittinghofer, A. (1986) EMBO J. 5, 1351–1358.
20. O'Neill, G.P., Kilburn, D.G., Warren, R.A.J. and Miller Jr, R.C. (1986) Appl. Environmental Microbiol. 52, 737–743.
21. Miki, T., Yasukochi, T., Nagatani, H., Furuno, M., Orita, T., Yamada, H., Imoto, T. and Horiuchi, T. (1987) Protein Eng. 1, 327–332.
22. Larder, B., Purifoy, D., Powell, K. and Darby, G. (1987) EMBO J. 6, 3133–3137.
23. Tanizawa, Y., Kishi, F., Kaneko, T. and Nakazawa, A. (1987) J. Biochem. 101, 1289–1296.
24. Beck, Y., Bartfeld, D., Yavin, Z., Levanon, A., Gorecki, M. and Hartman, J.R. (1988) Bio/Technology 6.
25. Ernst, J.F. and Kawashima, E. (1988) J. Biotechnol. 7, 1–10.
26. Williams, D.P., Regier, D., Akiyoshi, D., Genbauffe, F. and Murphy, J.R. (1988) Nucleic Acids Res. 16, 10453–10467.
27. Helting, T.B. and Nau, H.H. (1984) Act. Pathol. Microbiol. Scan. Sect. C, 92, 59–63.
28. Makoff, A.J., Ballantine, S.P., Smallwood, A.E. and Fairweather, N.F. (1989) Bio/Technology 7, 1043–1046.
29. Fairweather, N.F. and Lyness, V.A. (1986) Nucleic Acids Res. 14, 7809–7812.
30. Meselson, M. and Yuan, R. (1968) Nature 217, 1110–1114.
31. Amann, E., Brosius, J. and Ptashne, M. (1983) Gene 25, 167–178.
32. Marsh, J.L., Erfle, M. and Wykes, E.J. (1984) Gene 32, 481–485.
33. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
34. Makoff, A.J., Parry, N. and Dicken, L.P. (1989) J. Gen. Microbiol. 135, 11–24.
35. von Gabain, A., Belasco, J.G., Schottel, J.L., Chang, C.Y., and Cohen, S.N. (1983) Proc. Natl. Acad. Sci. USA 80, 653–657.
36. Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Laboratory, New York.
37. Feinberg, A. and Vogelstein, B. (1983) Anal. Biochem. 132, 6–13.
38. Makoff, A.J. and Smallwood, A.E. (1988) Biochem. Soc. Trans. 16, 48–49.
39. Nishi, T. and Itoh, S. (1986) Gene 44, 29–36.
40. Wong, E.M., Muesing, M.A. and Polisky, B. (1982) Proc. Natl. Acad. Sci. USA 79, 3570–3574.
41. Varenne, S., Buc, J., Lloubes, R. and Lazdunski, C. (1984) J. Mol. Biol. 180, 549–576.
42. Bonekamp, F., Andersen, H.D., Christensen, T. and Jensen, K.F. (1985) Nucleic Acids. Res. 13, 4113–4123.
43. Cannistraro, V.J., Subbarao, M.N. and Kennell, D. (1986) J. Mol. Biol. 192, 257–274.
44. Higgins, C.F. and Smith, N.H. (1986) in Regulation of Gene Expression 25 Years On. 39th Symposium of Soc. for Gen. Microbiol. (Cambridge University Press).
45. Platt, T. (1986) Ann. Rev. Biochem. 55, 339–72.
46. Stanssens, P., Remaut, E. and Fiers, W. (1986) Cell 44, 711–718.
47. Varenne, S. and Lazdunski (1986) J. Theor. Biol. 120, 99–110.

This article, submitted on disc, has been automatically converted into this typeset format by the publisher.