The nucleotide sequence of the *Escherichia coli fus* gene, coding for elongation factor G

Janice M.Zengel, Richard H.Archer and Lasse Lindahl

Department of Biology, University of Rochester, Rochester, NY 14627, USA

Received 21 November 1983; Revised and Accepted 24 January 1984

ABSTRACT

We have determined the nucleotide sequence of the Escherichia coli fus gene, which codes for elongation factor G. The protein product of the sequenced gene contains 703 amino acids, with a predicted molecular weight of 77,444. The fus gene shows the nonrandom pattern of codon usage typical of ribosomal proteins and other proteins synthesized at a high level. We have identified several potential promoter sequences within the gene. One of these sequences may correspond to the secondary promoter for expression of the downstream tufA gene (encoding elongation factor Tu) whose activity has been described previously (1,2). A comparison of the nucleotide and amino acid sequences of elongation factors G and Tu reveals a limited but significant homology between the two proteins within the 150 amino acid residues at their amino-terminal ends.

INTRODUCTION

Elongation factors G, Tu, and Ts (EF-G, EF-Tu and EF-Ts, respectively) play essential roles during protein biosynthesis in Escherichia coli. In addition to having a close functional relationship with ribosomes, these proteins are encoded by genes whose expression is tightly coordinated with the expression of ribosomal protein genes (3). We have analyzed the expression of tufA and fus, encoding EF-Tu and EF-G, respectively (1,4). Both of these genes map in the str operon, which also contains genes for two ribosomal proteins, S12 and S7 (5). Interestingly, EF-Tu is also synthesized from the tufB gene, which maps elsewhere on the E. coli chromosome (5). The nucleotide sequences of both tufA and tufB have been published (6,7). As a prerequisite to extending our studies on the expression of the str operon, we have now determined the sequence of the fus gene.

MATERIALS AND METHODS

Bacterial and phage strains

Bacteriophages M13mp8 and M13mp9 and their host strain JM103 were

obtained from BRL. The M13 phages were propagated as described by the BRL manual.

Enzymes and reagents

Restriction endonucleases were obtained from BRL, New England Biolabs, or Boehringer-Mannheim. Mung bean nuclease was obtained from P-L Biochemicals. T4 DNA ligase was purified by E. Lifson in this laboratory. Reagents for DNA sequencing were purchased as a kit from BRL.

Construction of M13 clones

M13 clones for DNA sequencing were constructed by subcloning DNA from plasmid pLL145 (Fig. 1), a pBGP120 (8) derivative carrying a 3.2kb KpnI fragment from fus3 (9) that contains the intact fus gene (1,4). Clones Z11, Z19, Z20, Z21, Z22, and Z24 were constructed by inserting various portions of the 3.2kb KpnI fragment onto M13mp8 or M13mp9 using the restriction enzymes indicated in Fig. 1. [For Z11, the KpnI generated staggered end was converted to a blunt end with mung bean nuclease and ligated to a blunt end HincII site on the M13 vector.] All other clones were derived by deleting, "flipping", and/or subcloning various portions of the original subclones.

DNA sequence determination and analysis

The DNA sequences were obtained from the M13 clones using the dideoxynucleotide chain termination method of Sanger (10) as described by the BRL sequencing manual. Sequencing samples were routinely run on 40 cm long 8% polyacrylamide gels. To extend the sequence of Z21, we also analyzed this sample on a 5% polyacrylamide 87 cm long "King Kong" gel. The sequences of most of the M13 clones were determined from at least two independent sequencing reactions and, in most cases, the sequence was confirmed by sequencing an independently isolated clone containing overlapping DNA sequences or the same sequence in the opposite orientation. The final sequence was constructed as shown in Fig. 1.

The fus sequence was analyzed using programs designed by J. Pustell at the Biological Laboratories at Harvard University (11,12).

RESULTS AND DISCUSSION

Determination of the fus gene sequence

Figure 2 shows the complete 2115 base pair sequence of the fus gene, plus the 5' flanking region up to the UGA termination codon of the gene for r-protein S7, which lies immediately upstream of fus in the str operon. Using the strategy shown in Fig. 1, we determined the sequence of the S7-fus intercistronic region and all of the fus sequence, except for the final 22

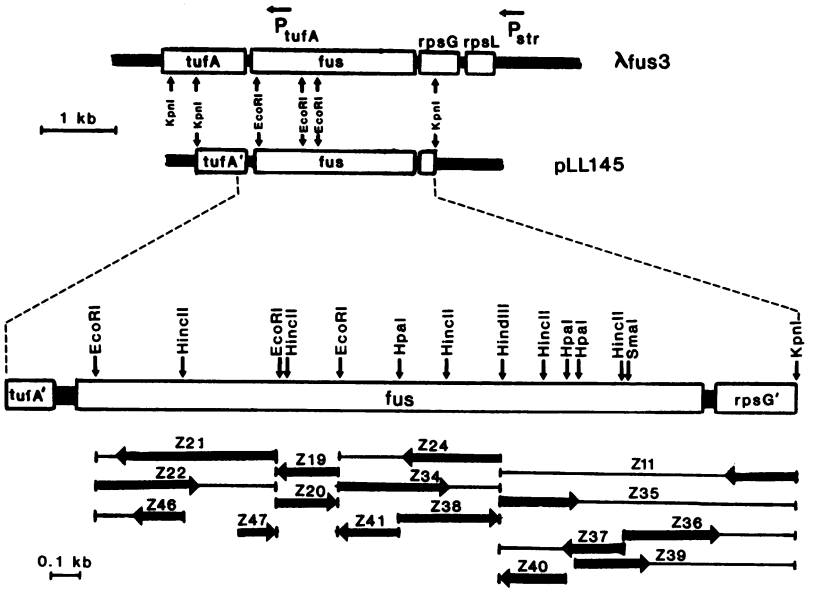


Figure 1.
Strategy for sequencing fus. The top of the figure shows the genetic organization of the str operon and the KpnI fragment from the operon carried on pLL145. P_{str} designates the position of the major promoter of the str operon; P_{tufA}, the approximate position of the secondary promoter for tufA expression (1). Subfragments cloned from pLL145 onto M13 vectors are shown below. Arrows indicate the direction and thick bars indicate the extent of sequencing obtained from each subclone.

codons at the 3' end which were sequenced previously by Yokota et al. (6). The sequence of the S7-fus intercistronic region and the first 279 nucleotides of the fus gene was reported previously by Post and Nomura (13); our sequencing data are consistent with their results.

Amino acid sequence of the fus gene product

The fus DNA sequence predicts that the gene product, EF-G, contains 703 amino acids (not including the Met encoded by the initiation codon) and has a molecular weight of 77,444. The amino acid sequence shown in Fig. 2 is in good agreement with the published amino acid sequence of EF-G (14), which contains 701 amino acids. The most serious discrepancy between the published amino acid sequence and the sequence predicted from the nucleotide sequence is in the cysteine content of the protein. According to Ovchinnikov et al. (14), EF-G contains 5 cysteine residues, 2 of which reportedly are involved in a disulfide bond. However, the nucleotide sequence predicts only three

```
-30         -20         -10           1          10          20          30
 *           *           *            *          *           *           *
TGA ACG CCT AAA AGA TAA ACG AGG AAA CAA ATG GCT CGT ACA ACA CCC ATC GCA CGC TAC
                                        Ala Arg Thr Thr Pro Ile Ala Arg Tyr

            40          50          60          70          80          90
             *           *           *          *           *           *
CGT AAC ATC GGT ATC AGT GCG CAC ATC GAC GCC GGT AAA ACC ACT ACT ACC GAA CGT ATT
Arg Asn Ile Gly Ile Ser Ala His Ile Asp Ala Gly Lys Thr Thr Thr Thr Glu Arg Ile

           100         110         120         130         140         150
            *           *           *          *           *           *
CTG TTC TAC ACC GGT GTA AAC CAT AAA ATC GGT GAA GTT CAT GAC GGC GCT GCA ACC ATG
Leu Phe Tyr Thr Gly Val Asn His Lys Ile Gly Glu Val His Asp Gly Ala Ala Thr Met

           160         170         180         190         200         210
            *           *           *          *           *           *
GAC TGG ATG GAG CAG GAG CAG GAA CGT GGT ATT ACC ATC ACT TCC GCT GCG ACT ACT GCA
Asp Trp Met Glu Gln Glu Gln Glu Arg Gly Ile Thr Ile Thr Ser Ala Ala Thr Thr Ala

           220         230         240         250         260         270
            *           *           *          *           *           *
TTC TGG TCT GGT ATG GCT AAG CAG TAT GAG CCG CAT CGC ATC AAC ATC ATC GAC ACC CCG
Phe Trp Ser Gly Met Ala Lys Gln Tyr Glu Pro His Arg Ile Asn Ile Ile Asp Thr Pro

           280         290         300         310         320         330
            *           *           *          *           *           *
GGG CAC GTT GAC TTC ACA ATC GAA GTA GAA CGT TCC ATG CGT GTT CTC GAT GGT GCG GTA
Gly His Val Asp Phe Thr Ile Glu Val Glu Arg Ser Met Arg Val Leu Asp Gly Ala Val

           340         350         360         370         380         390
            *           *           *          *           *           *
ATG GTT TAC TGC GCA GTT GGT GGT GTT CAG CCG CAG TCT GAA ACC GTA TGG CGT CAG GCA
Met Val Tyr Cys Ala Val Gly Gly Val Gln Pro Gln Ser Glu Thr Val Trp Arg Gln Ala

           400         410         420         430         440         450
            *           *           *          *           *           *
AAC AAA TAT AAA GTT CCG CGC ATT GCG TTC GTT AAC AAA ATG GAC CGC ATG GGT GCG AAC
Asn Lys Tyr Lys Val Pro Arg Ile Ala Phe Val Asn Lys Met Asp Arg Met Gly Ala Asn

           460         470         480         490         500         510
            *           *           *          *           *           *
TTC CTG AAA GTT GTT AAC CAG ATC AAA ACC CGT CTG GGC GCG AAC CCG GTT CCG CTG CAG
Phe Leu Lys Val Val Asn Gln Ile Lys Thr Arg Leu Gly Ala Asn Pro Val Pro Leu Gln

           520         530         540         550         560         570
            *           *           *          *           *           *
CTG GCG ATT GGT GCT GAA GAA CAT TTC ACC GGT GTT GTT GAC CTG GTG AAA ATG AAA GCT
Leu Ala Ile Gly Ala Glu Glu His Phe Thr Gly Val Val Asp Leu Val Lys Met Lys Ala

           580         590         600         610         620         630
            *           *           *          *           *           *
ATC AAC TGG AAC GAC GCT GAC CAG GGC GTA ACC TTC GAA TAC GAA GAT ATC CCG GCA GAC
Ile Asn Trp Asn Asp Ala Asp Gln Gly Val Thr Phe Glu Tyr Glu Asp Ile Pro Ala Asp

           640         650         660         670         680         690
            *           *           *          *           *           *
ATG GTT GAA CTG GCT AAC GAA TGG CAC CAG AAC CTG ATC GAA TCC GCA GCT GAA GCT TCT
Met Val Glu Leu Ala Asn Glu Trp His Gln Asn Leu Ile Glu Ser Ala Ala Glu Ala Ser
```

```
          700         710         720         730         740         750
           *           *           *           *           *           *
GAA GAG CTG ATG GAA AAA TAC CTG GGT GGT GAA GAA CTG ACT GAA GCA GAA ATC AAA GGT
Glu Glu Leu Met Glu Lys Tyr Leu Gly Gly Glu Glu Leu Thr Glu Ala Glu Ile Lys Gly

          760         770         780         790         800         810
           *           *           *           *           *           *
GCT CTG CGT CAG CGC GTT CTG AAC AAC GAA ATC ATC CTG GTA ACC TGT GGT TCT GCG TTC
Ala Leu Arg Gln Arg Val Leu Asn Asn Glu Ile Ile Leu Val Thr Cys Gly Ser Ala Phe

          820         830         840         850         860         870
           *           *           *           *           *           *
AAG AAC AAA GGT GTT CAG GCG ATG CTG GAT GCG GTA ATT GAT TAC CTG CCA TCC CCG GTT
Lys Asn Lys Gly Val Gln Ala Met Leu Asp Ala Val Ile Asp Tyr Leu Pro Ser Pro Val

          880         890         900         910         920         930
           *           *           *           *           *           *
GAC GTA CCT GCG ATC AAC GGT ATC CTG GAC GAC GGT AAA GAC ACT CCG GCT GAA CGT CAC
Asp Val Pro Ala Ile Asn Gly Ile Leu Asp Asp Gly Lys Asp Thr Pro Ala Glu Arg His

          940         950         960         970         980         990
           *           *           *           *           *           *
GCA AGT GAT GAC GAG CCG TTC TCT GCA CTG GCG TTC AAA ATC GCT ACC GAC CCG TTT GTT
Ala Ser Asp Asp Glu Pro Phe Ser Ala Leu Ala Phe Lys Ile Ala Thr Asp Pro Phe Val

         1000        1010        1020        1030        1040        1050
           *           *           *           *           *           *
GGT AAC CTG ACC TTC TTC CGT GTT TAC TCC GGT GTG GTT AAC TCT GGT GAT ACC GTA CTG
Gly Asn Leu Thr Phe Phe Arg Val Tyr Ser Gly Val Val Asn Ser Gly Asp Thr Val Leu

         1060        1070        1080        1090        1100        1110
           *           *           *           *           *           *
AAC TCC GTG AAA GCT GCA CGT GAG CGT TTC GGT CGT ATC GTT CAG ATG CAC GCT AAC AAA
Asn Ser Val Lys Ala Ala Arg Glu Arg Phe Gly Arg Ile Val Gln Met His Ala Asn Lys

         1120        1130        1140        1150        1160        1170
           *           *           *           *           *           *
CGT GAA GAG ATC AAA GAA GTT CGC GCG GGC GAC ATC GCT GCT GCT ATC GGT CTG AAA GAC
Arg Glu Glu Ile Lys Glu Val Arg Ala Gly Asp Ile Ala Ala Ala Ile Gly Leu Lys Asp

         1180        1190        1200        1210        1220        1230
           *           *           *           *           *           *
GTA ACC ACT GGT GAC ACC CTG TGT GAC CCG GAT GCG CCG ATC ATT CTG GAA CGT ATG GAA
Val Thr Thr Gly Asp Thr Leu Cys Asp Pro Asp Ala Pro Ile Ile Leu Glu Arg Met Glu

         1240        1250        1260        1270        1280        1290
           *           *           *           *           *           *
TTC CCT GAG CCG GTA ATC TCC ATC GCA GTT GAA CCG AAA ACC AAA GCT GAC CAG GAA AAA
Phe Pro Glu Pro Val Ile Ser Ile Ala Val Glu Pro Lys Thr Lys Ala Asp Gln Glu Lys

         1300        1310        1320        1330        1340        1350
           *           *           *           *           *           *
ATG GGT CTG GCT CTG GGC CGT CTG GCT AAA GAA GAC CCG TCT TTC CGT GTA TGG ACT GAC
Met Gly Leu Ala Leu Gly Arg Leu Ala Lys Glu Asp Pro Ser Phe Arg Val Trp Thr Asp

         1360        1370        1380        1390        1400        1410
           *           *           *           *           *           *
GAA GAA TCT AAC CAG ACC ATC ATC GCG GGT ATG GGC GAA CTG CAC CTC GAC ATC ATC GTT
Glu Glu Ser Asn Gln Thr Ile Ile Ala Gly Met Gly Glu Leu His Leu Asp Ile Ile Val
```

```
         1420        1430        1440        1450        1460        1470
           *           *           *           *           *           *
GAC CGT ATG AAG CGT GAA TTC AAC GTT GAA GCG AAC GTA GGT AAA CCG CAG GTT GCT TAC
Asp Arg Met Lys Arg Glu Phe Asn Val Glu Ala Asn Val Gly Lys Pro Gln Val Ala Tyr

         1480        1490        1500        1510        1520        1530
           *           *           *           *           *           *
CGT GAA ACT ATC CGC CAG AAA GTT ACC GAT GTT GAA GGT AAA CAC GCG AAA CAG TCT GGT
Arg Glu Thr Ile Arg Gln Lys Val Thr Asp Val Glu Gly Lys His Ala Lys Gln Ser Gly

         1540        1550        1560        1570        1580        1590
           *           *           *           *           *           *
GGT CGT GGT CAG TAT GGT CAT GTT GTT ATC GAC ATG TAC CCG CTG GAG CCG GGT TCA AAC
Gly Arg Gly Gln Tyr Gly His Val Val Ile Asp Met Tyr Pro Leu Glu Pro Gly Ser Asn

         1600        1610        1620        1630        1640        1650
           *           *           *           *           *           *
CCG AAA GGC TAC GAG TTC ATC AAC GAC ATT AAA GGT GGT GTA ATC CCT GGC GAA TAC ATC
Pro Lys Gly Tyr Glu Phe Ile Asn Asp Ile Lys Gly Gly Val Ile Pro Gly Glu Tyr Ile

         1660        1670        1680        1690        1700        1710
           *           *           *           *           *           *
CCG GCC GTT GAT AAA GGT ATC CAG GAA CAG CTG AAA GCA GGT CCG CTG GCA GGC TAC CCG
Pro Ala Val Asp Lys Gly Ile Gln Glu Gln Leu Lys Ala Gly Pro Leu Ala Gly Tyr Pro

         1720        1730        1740        1750        1760        1770
           *           *           *           *           *           *
GTA GTA GAC ATG GGT ATT CGT CTG CAC TTC GGT TCT TAC CAT GAC GTT GAC TCC TCT GAA
Val Val Asp Met Gly Ile Arg Leu His Phe Gly Ser Tyr His Asp Val Asp Ser Ser Glu

         1780        1790        1800        1810        1820        1830
           *           *           *           *           *           *
CTG GCG TTT AAA CTG GCT GCT TCT ATC GCC TTT AAA GAA GGC TTT AAG AAA GCG AAA CCA
Leu Ala Phe Lys Leu Ala Ala Ser Ile Ala Phe Lys Glu Gly Phe Lys Lys Ala Lys Pro

         1840        1850        1860        1870        1880        1890
           *           *           *           *           *           *
GTT CTG CTT GAG CCG ATC ATG AAG GTT GAA GTA GAA ACT CCG GAA GAG AAC ACC GGT GAC
Val Leu Leu Glu Pro Ile Met Lys Val Glu Val Glu Thr Pro Glu Glu Asn Thr Gly Asp

         1900        1910        1920        1930        1940        1950
           *           *           *           *           *           *
GTT ATC GGT GAC TTG AGC CGT CGT CGT GGT ATG CTC AAA GGT CAG GAA TCT GAA GTT ACT
Val Ile Gly Asp Leu Ser Arg Arg Arg Gly Met Leu Lys Gly Gln Glu Ser Glu Val Thr

         1960        1970        1980        1990        2000        2010
           *           *           *           *           *           *
GGC GTT AAG ATC CAC GCT GAA GTA CCG CTG TCT GAA ATG TTC GGA TAC GCA ACT CAG CTG
Gly Val Lys Ile His Ala Glu Val Pro Leu Ser Glu Met Phe Gly Tyr Ala Thr Gln Leu

         2020        2030        2040        2050        2060        2070
           *           *           *           *           *           *
CGT TCT CTG ACC AAA GGT CGT GCA TCA TAC ACT ATG GAA TTC CTG AAG TAT GAT GAA GCG
Arg Ser Leu Thr Lys Gly Arg Ala Ser Tyr Thr Met Glu Phe Leu Lys Tyr Asp Glu Ala

         2080        2090        2100        2110
           *           *           *           *
CCG AGT AAC GTT GCT CAG GCC GTA ATT GAA GCC CGT GGT AAA TAA
Pro Ser Asn Val Ala Gln Ala Val Ile Glu Ala Arg Gly Lys ---
```

cysteines. The two missing residues are those reported to be in the S-S bond: the DNA sequence predicts a Gly instead of Cys at nucleotides 889-891, and a Thr instead of Cys at nucleotides 1186-1188 (Fig. 2).

Other differences predicted by the nucleotide sequence include an Asn instead of Asp at nucleotides 886-888, a Glu instead of Gln at nucleotides 1876-1878, 1936-1938, and 1969-1971, an Ile instead of Val at nucleotides 1726-1728, a Lys instead of His at nucleotides 1780-1782, and a His instead of Lys at nucleotides 1750-1752. In addition, the DNA sequence predicts an additional three amino acids, Asp-Asp-Gly, at nucleotides 898-906, not found in the published amino acid sequence, and does not predict a Gln between nucleotides 1986 and 1987 that is present in the previously published sequence.

We have double-checked our nucleotide sequence in regions where we have detected a discrepancy between our results and those reported by Ovchinnikov et al. (14), and are confident that our data are correct. Some of the differences may be due to errors in the original amino acid sequence or, possibly, to strain specific differences in the fus sequence between the cells used in the amino acid analysis and the cells from which the DNA for cloning was derived. Alternatively, the discrepancies could have resulted from mutations which arose during the construction of the fus clones. We would like to point out in this connection, however, that the source of the sequenced DNA, pLL145, contains a fus gene from which a functional EF-G protein is synthesized, as defined by its ability to convert a fusidic acid resistant host to fusidic acid sensitivity (4). Also, the EF-G derived from pLL145 comigrates with the chromosomally derived EF-G on a two-dimensional O'Farrell gel (4). In any event, given the size of the fus gene and its protein product, there are surprisingly few discrepancies between the published amino acid sequence and the sequence predicted by our DNA sequence analysis.

Codon frequencies

Table 1 shows the frequency of codon usage in fus. The pattern of codon

Figure 2.
Nucleotide sequence of E. coli fus. The DNA strand with the fus mRNA polarity is given, along with the deduced amino acid sequence. The nucleotide positions are numbered beginning with the initiation codon for fus. The sequence through nucleotide 279 was reported previously by Post and Nomura (13); the nucleotide sequence from 2047 through 2115 was reported previously by Yokota et al. (6).

Table 1. Comparison of codon usage in the *E. coli fus*, *tufA/tufB* and ribosomal protein genes.

Codons	EF-G	EF-Tu	r-proteins	Codons	EF-G	EF-Tu	r-proteins
TTT Phe	4	2	17	TAT Tyr	4	3	7
TTC Phe	20	26	38	TAC Tyr	15	17	23
TTA Leu	0	0	4	TAA ---	1	2	13
TTG Leu	1	0	7	TAG ---	0	0	1
CTT Leu	1	2	7	CAT His	6	3	9
CTC Leu	3	1	9	CAC His	9	19	11
CTA Leu	0	0	1	CAA Gln	0	0	18
CTG Leu	38	53	108	CAG Gln	24	16	44
ATT Ile	9	6	37	AAT Asn	0	0	15
ATC Ile	41	52	80	AAC Asn	27	14	52
ATA Ile	0	0	0	AAA Lys	37	35	134
ATG Met	22	20	54	AAG Lys	7	11	50
GTT Val	38	46	88	GAT Asp	10	8	35
GTC Val	0	1	25	GAC Asp	32	41	51
GTA Val	19	21	59	GAA Glu	50	60	95
GTG Val	3	6	21	GAG Glu	12	13	28
TCT Ser	15	14	39	TGT Cys	2	2	4
TCC Ser	8	6	26	TGC Cys	1	4	4
TCA Ser	2	0	2	TGA ---	0	0	0
TCG Ser	0	0	2	TGG Trp	6	2	3
CCT Pro	3	0	13	CGT Arg	29	41	84
CCC Pro	1	0	2	CGC Arg	7	5	47
CCA Pro	2	2	7	CGA Arg	0	0	3
CCG Pro	27	38	45	CGG Arg	0	0	1
ACT Thr	14	25	48	AGT Ser	3	0	3
ACC Thr	21	31	41	AGC Ser	1	1	13
ACA Thr	3	3	4	AGA Arg	0	0	1
ACG Thr	0	1	6	AGG Arg	0	0	0
GCT Ala	26	24	113	GGT Gly	46	38	107
GCC Ala	5	2	27	GGC Gly	11	41	65
GCA Ala	16	11	68	GGA Gly	1	0	1
GCG Ala	20	17	45	GGG Gly	1	2	7

The values for EF-Tu are the combined frequencies for *tufA* (6) and *tufB* (7). The values for r-proteins include all 10 r-protein genes from the *spc* operon (16) plus the genes for L11, L1, L10, and L7/L12 (15). The initiation codons are not included.

usage shows the same, highly non-random, pattern observed for r-proteins and EF-Tu (see, e.g., 6,7,15,16). Since EF-G is synthesized in equimolar amounts with ribosomal proteins, this result is consistent with the hypothesis that

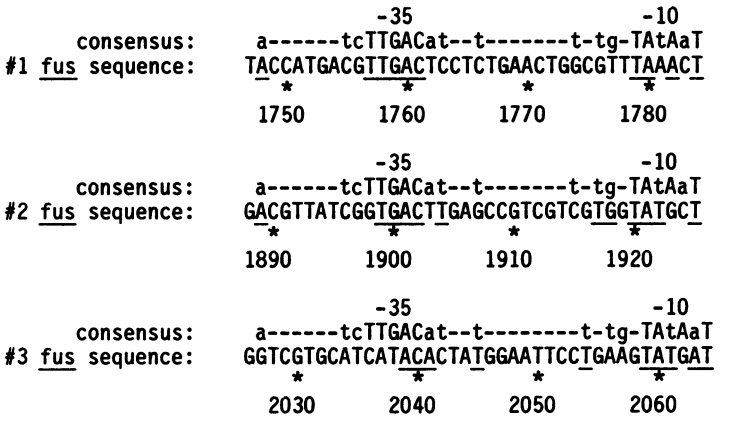


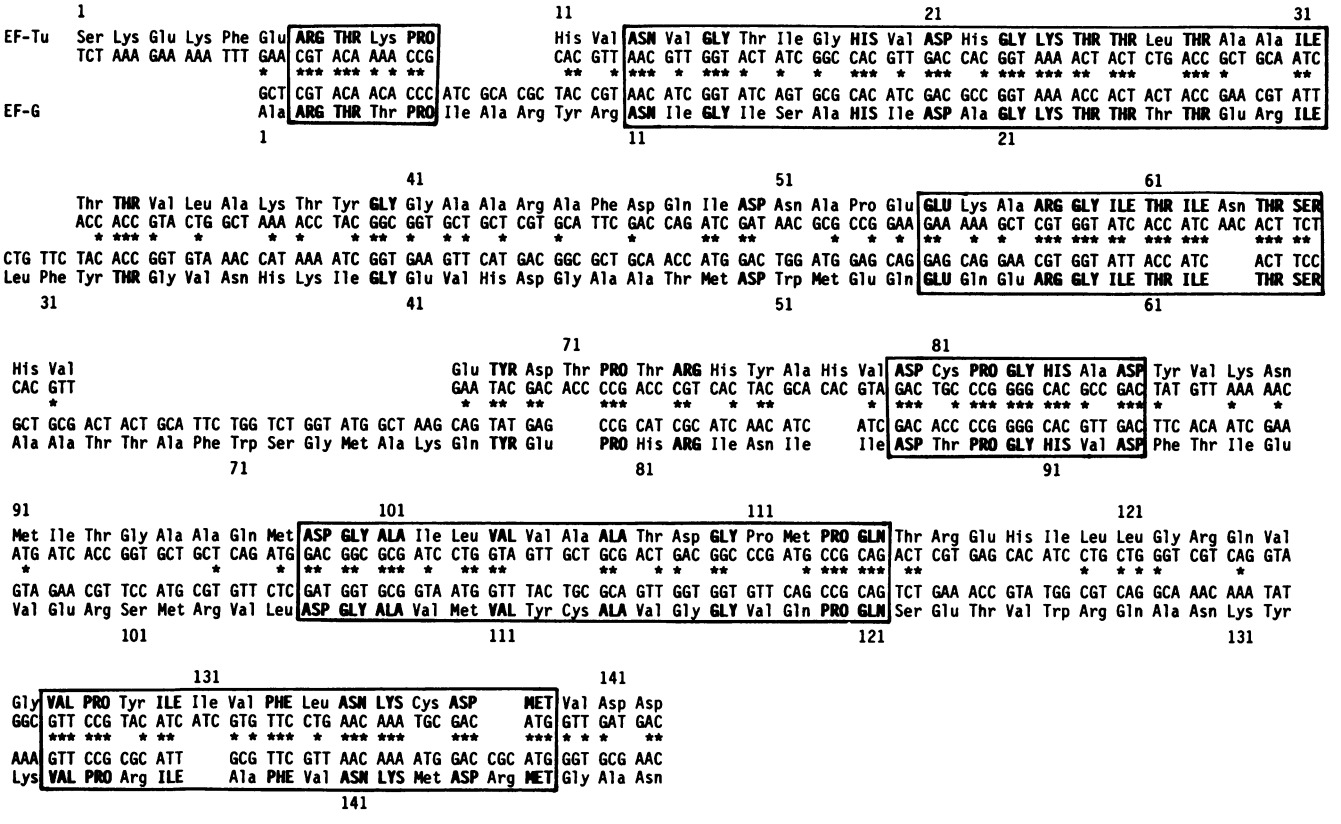
Figure 3.
Potential promoter sequences within the fus gene. Shown are three sequences within the structural gene for EF-G which resemble the consensus promoter sequence (19). The bases with exact homology are underlined. The numbering system for fus sequences corresponds to the system used in Fig. 2.

the genes for proteins synthesized at very high rates, such as ribosomal proteins and translation elongation factors, preferentially use codons recognized by the most abundant of the isoaccepting species of tRNA (15,17).

Possible secondary promoter sequences

One strong incentive for sequencing the fus gene was to facilitate our analysis of a secondary promoter in the str operon, apparently mapping within the structural gene for EF-G (1,2). Analysis of the nucleotide sequence has turned up several sequences (Fig. 3) which contain "-35" and "-10" regions typical of E. coli promoters (18,19). Two of these sequences, #1 and #2, are located within a 600 base pair EcoRI fragment internal to the fus gene. Hybridization experiments measuring the transcription rates from various regions of the str operon in cells carrying multicopy plasmid pLL145 (Fig. 1) indicate that the secondary promoter is either within the 600 base pair EcoRI sequence or just upstream of this region (J. M. Zengel, unpublished experiments). Therefore, both of these sequences, particularly #1 because of its strong homology with the consensus -35 sequence, are reasonable candidates for the secondary promoter.

Yokota et al. (6) pointed out a possible Pribnow box at the very end of the fus gene, just downstream from the EcoRI site at nucleotide 2047. At the time, the upstream sequence containing the putative -35 region was not known.


EF-Tu
EF-G

Now that we have completed the fus sequence, the -35 region can be analyzed. As shown in Fig. 3, this potential secondary promoter, #3, has a relatively good Pribnow box, but a relatively weak -35 region. Since our hybridization results (described above) suggest that the promoter is further upstream, it seems unlikely that this promoter contributes significantly to the expression of the tufA gene. In any event, experiments are in progress to map more precisely the secondary promoter.

Homology between EF-G and EF-Tu

A comparison of the amino acid sequence of the amino-terminal region of EF-G with the amino acid sequence of elongation factor Tu (EF-Tu) revealed a region of homology between the two elongation factors (20,21). Jones et al. (21) have suggested that at least the regions comprising amino acid residues 25 to 150 of EF-G and EF-Tu have evolved from a common ancestral protein by gene duplication. We have searched for homology between EF-Tu and EF-G, both at the nucleotide level and at the amino acid level, using the homology matrix program developed by Pustell and Kafatos (12). Our analysis revealed no significant regions of homology between EF-Tu and EF-G, with the exception of the region of weak homology already pointed out (20,21). However, by introducing two additional "gaps" in the alignment of the amino acid sequences, we were able to extend this region of homology to include the first 25 amino acids at the amino-terminal ends of the two proteins. Figure 4 shows the comparison of the amino acid and nucleotide sequences in the regions of homology, aligned for maximum amino acid homology. Note that short stretches of strong homology are flanked by regions of essentially no homology and that a number of gaps have been introduced in the EF-Tu and EF-G sequences to maximize the homology (Fig. 4). These factors plus our failure to identify any significant homology elsewhere in the two proteins makes it difficult to evaluate the evolutionary relationship between the two elongation factors.

Figure 4.
Comparison of the amino acid sequences of EF-G and EF-Tu. The corresponding nucleotide sequences of fus and tufA are also shown. The numbers refer to the amino acid sequences. Amino acids with exact homology are shown in boldface, upper case letters. Exact nucleotide homologies are indicated with asterisks. Regions of extensive sequence homology are enclosed in blocks. The amino acid alignment beginning with residue 32 of the two proteins is essentially the same as that pointed out previously by Jones et al. (21).

ACKNOWLEDGEMENTS

We thank Mr. J. Pustell for providing us with the microcomputer programs for DNA sequence analysis. This work was supported by Public Health Service grant AI 15286 from the National Institute of Allergy and Infectious Diseases.

REFERENCES

1. Zengel, J. M. and Lindahl, L. (1982) Mol. Gen. Genet. 185: 487-492.
2. An, G., Lee, J. S., and Friesen, J. D. (1982) J. Bacteriol. 149: 548-563.
3. Lindahl, L. and Zengel, J. M. (1982) Adv. in Genetics 21: 53-121.
4. Zengel, J. M. and Lindahl, L. (1982) J. Bacteriol. 149: 793-797.
5. Jaskunas, S. R., Lindahl, L., Nomura, M., and Burgess, R. R. (1975) Nature 257: 458-462.
6. Yokota, T., Sugisaki, H., Takanami, M., and Kaziro, Y. (1980) Gene 12: 25-31.
7. An, G. and Friesen, J. D. (1980) Gene 12: 33-39.
8. Polisky, B., Bishop, R. J., and Gelfand, D. H. (1976) Proc. Natl. Acad. Sci. USA 73: 3900-3904.
9. Jaskunas, S. R. and Nomura, M. (1977) J. Biol. Chem. 252: 7337-7343.
10. Sanger, F., Wicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA 74: 5463-5467.
11. Pustell, J. and Kafatos, F. C. (1982) Nucleic Acids Res. 10: 51-59.
12. Pustell, J. and Kafatos, F. C. (1982) Nucleic Acids Res. 10: 4765-4782.
13. Post, L. E. and Nomura, M. (1980) J. Biol. Chem. 255: 4660-4666.
14. Ovchinnikov, Yu. A., Alakhov, Yu. B., Bundulis, Yu. P., Bundule, M. A., Dovgas, N. V., Kozlov, V. P., Motuz, L. P., and Vinokurov, L. M. (1982) FEBS Lett. 139: 130-135.
15. Post, L. E., Strycharz, G. D., Nomura, M., Lewis, H., and Dennis, P. P. (1979) Proc. Natl. Acad. Sci. USA 76: 1697-1701.
16. Cerretti, D. P., Dean, D., Davis, G. R., Bedwell, D. M., and Nomura, M. (1983) Nucleic Acids Res. 11: 2599-2616.
17. Ikemura, T. (1981) J. Mol. Biol. 151: 389-409.
18. Rosenberg, M. and Court, D. (1979) Ann. Rev. Genet. 13: 319-353.
19. Hawley, D. K. and McClure, W. R. (1983) Nucleic Acids Res. 11: 2237-2255.
20. Laursen, R. A. and Duffy, L. (1978) FEBS Lett. 92: 200-202.
21. Jones, M. D., Petersen, T. E., Nielsen, K. M., Magnusson, S., Sottrup-Jensen, L., Gausing, K., and Clark, B. F. C. (1980) Eur. J. Biochem. 108: 507-526.