Volume 10 Number 19 1982 Nucleic Acids Research

Nucleotide sequence of the *fnr* gene and primary structure of the Fnr protein of *Escherichia coli*

Duncan J.Shaw and John R.Guest

Department of Microbiology, University of Sheffield, Sheffield S10 2TN, UK

Received 28 July 1982; Revised and Accepted 6 August

ABSTRACT

The nucleotide sequence of a 1.64 kb fragment of E. coli DNA containing the fnr gene (regulatory gene for fumarate and nitrate reduction) was determined using the dideoxy chain termination method. The fnr coding region (750 bp) was identified, and the initiation and termination points of fnr transcription were located by RNA:DNA hybridisation with single-stranded M13 probes. The DNA fragment also contained the 5' end of a separately transcribed gene of unknown function. The deduced molecular weight (27947) of the Fnr protein was in agreement with that of the protein identified by the maxicell procedure, and the primary structure contained regions of homology with several transcriptional regulator proteins.

INTRODUCTION

The fnr gene of Escherichia coli K12 is essential for the expression of a variety of anaerobic electron transport systems. It was originally identified by the pleiotropic effects of fnr mutations on fumarate and nitrate reduction (or respiration) (1) and is probably identical to the nirA (2) and nirR (3) genes, all located at 29.3 min in the E. coli linkage map (4). Mutations in fnr lead to deficiencies in fumarate, nitrate and nitrite reductases, hydrogenase, a formate dehydrogenase, formate hydrogenlyase and the anaerobic cytochrome c_{552} (1, 2, 3). These enzyme systems are repressed during aerobic growth, derepressed anaerobically and induced by their substrates, but little is known about the molecular mechanisms that control expression of the corresponding genes (frd, fumarate reductase; chl, nitrate reductase; etc).

The pleiotropic nature of mutations in fnr has prompted the suggestion that the gene product is a positive regulatory protein essential for the transcription of the genes encoding anaerobic functions (5, 6, 7); if so, its action may be controlled by an effector molecule whose concentration reflects the redox or energy status of the cell. This view is supported

© IRL Press Limited, Oxford, England.

0305-1048/82/1019-6119$ 2.00/0

by the observations (i) that a functional *fnr* (*nirR*) gene is required for the expression of a fusion linking the promoter-operator of the nitrate reductase (*chlC*) operon to the *lac* structural genes (6), and (ii) that multiple copies of the *fnr* gene increase anaerobic synthesis and cause partial aerobic derepression of the fumarate and nitrate reductases (7). It is likely that the anaerobic systems are subject also to specific controls because the expression of fumarate and nitrate reductases is differentially regulated by nitrate (8, 9) and multiple copies of the fumarate reductase genes (*frdABCD*) apparently titrate a specific repressor to allow aerobic expression of fumarate reductase but not nitrate or nitrite reductase (10, 11).

A study of the molecular mechanism of action of *fnr* is in progress and previous papers have reported the initial cloning of the *fnr* gene in phage λ (5), and subcloning into pBR322, the effects of gene amplification on anaerobic enzyme activities and tentative identification of the gene product as a protein with M_r=31000 (7). In an attempt to gain further insights into the possible mode of action of *fnr*, the nucleotide sequence of the 1.64 kb DNA fragment containing the *fnr* gene has been determined, the corresponding mRNA identified and the primary structure of the Fnr protein (*fnr* gene product) deduced.

MATERIALS AND METHODS

Bacteria, phages, plasmids and media

The strains of *E. coli* K12 used were: C600 and ED8641 for selection of recombinant plasmids; PL2024 (1) transformed with pGS24 (7) for RNA preparation; JRG861b (*fnr-8*; 1) for testing the *fnr* genotype; JM101 (Δ(*lac-pro*) *supE* *thi*, F' *traD36* *proAB* *lacI*^{q}*Z*Δ*M15*) as host for phage M13 (12). Plasmid pGS24 consists of a 1.64 kb *Bam*HI-*Hin*dIII fragment containing the *fnr* gene (originally derived from a λ*fnr* transducing phage ; 5) cloned in pBR322 (7). The mp8 and mp9 derivatives of phage M13 (J. Messing, unpublished) may be used for cloning specific restriction fragments in either orientation, or for 'shotgun' cloning small fragments. The media used for routine preparations and for testing the Fnr phenotype have been described previously (5).

DNA manipulations

Plasmid and phage M13 replicative form DNAs were prepared on a large scale from cleared lysates by banding in CsCl gradients (13, 14) and on a small scale by the method of Birnboim and Doly (15). Restriction endonuclease digestions, agarose gel electrophoresis and *in vitro* ligation of

DNA fragments with phage T4 DNA ligase were as previously described (5, 7). DNA fragments were purified after agarose gel electrophoresis by dissolving gel pieces in saturated KI solution and recovering the DNA by chromatography on hydroxylapatite (16).

DNA sequencing

Derivatives of phage M13mp8 and M13mp9 carrying inserted DNA fragments were identified by plaque morphology (colourless) on plates containing the chromogenic indicator 5-bromo-4-chloro-3-indolyl-β galactoside. Single-stranded DNA was prepared from the plaques as described (17), and used as a template for the dideoxy chain termination sequencing reactions devised by Sanger et al. (18). Labelled poly-nucleotide fragments were identified by autoradiography following electrophoresis on polyacrylamide gels (6% w/v) containing 8M urea. To avoid generation of redundant data by sequencing identical fragments, all clones were first characterised by 'tracking' (i.e. sequenced using only 1 of the 4 reactions). Data from the sequencing gels were stored and manipulated by computer using the programs described by Staden (19, 20).

Maxicell labelling of polypeptides

Plasmid-coded proteins were labelled with ^{35}S-methionine and identified by polyacrylamide gel electrophoresis and autoradiography using the 'maxicell' procedure (21) as described previously (7).

RNA:DNA hybridisation

RNA was prepared from early stationary phase cultures of E. coli and hybridised with single-stranded M13 derivatives containing cloned DNA fragments, according to the method of Squires et al. (22). Double-stranded, nuclease S1-resistant hybrids were identified by polyacrylamide gel electrophoresis and staining with ethidium bromide.

Enzymes and radioisotopes

Restriction endonucleases and phage T4 DNA ligase were purchased from Bethesda Research Ltd. and New England Biolabs, and DNA polymerase (Klenow fragment) and S1 nuclease from Boehringer Corporation Ltd. All enzymes were used as directed by the suppliers. L-[^{35}S]-methionine and deoxyadenosine-5'[α^{32}P]-triphosphate were supplied by the Radiochemical Centre, Amersham.

RESULTS AND DISCUSSION

Sequencing strategy

The 1.64 kb BamHI-HindIII fragment containing the fnr gene was purified

from a restriction digest of pGS24 DNA. The fragment was digested with EcoRI yielding 3 subfragments of similar size (7) which were cloned in both orientations in M13mp8 and M13mp9. The complete sequence of each subfragment was obtained by sequencing in from both ends until an overlap was reached (Fig. 1). The 1.64 kb fragment was also digested with Sau3A and shotgun-cloned in the BamHI site of M13mp9. Most of the fragment was recovered and sequenced in this way (Fig. 1) but no Sau3A clone containing the fragment S_1-S_2 and overlapping the EcoRI target R_1 was found. In order to overlap the sequences on either side of target R_1 the 1.64 kb fragment was cloned intact in M13mp9 and sequenced from the HindIII target (H) to beyond R_1. Most of the fragment (94%) was sequenced from at least two independent clones, and 81% was sequenced on both strands. There were no ambiguities in the final sequence, presented in Fig. 2.

Location of open reading frames

With the aid of the computer program FRAMESCAN (23) both strands of the DNA sequence were examined for potential coding regions (60 or more codons starting with ATG or GTG). The most likely candidate for the *fnr* coding region was found at positions 520-1272 in the DNA strand shown in Fig. 2 (CR1 in Fig. 1). A second potential coding region, truncated by the terminal BamHI target, was found at positions 1424-1641 in the same strand (Fig. 2: CR2 in Fig. 1), and two other open reading frames were located in the complementary strand (CR3 and CR4, Fig. 1). The length of CR1 is

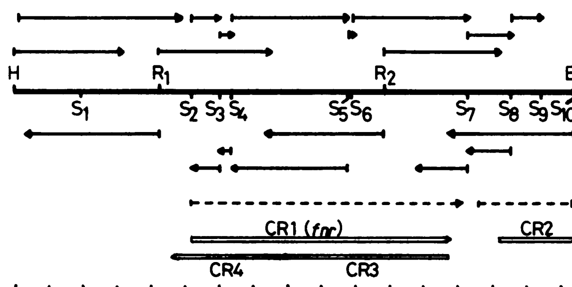

Figure 1. Restriction map of the 1.64 kb DNA fragment showing sequencing strategy, transcribed regions and potential coding regions. Heavy single line denotes the DNA fragment, with restriction targets for BamHI (B), EcoRI (R), HindIII (H) and Sau3A (S). Light single lines with arrowheads indicate position and extent of sequencing. Broken lines show position and orientation (5' → 3') of mRNA transcripts. Open boxes show the positions of potential coding regions, CR1 and CR2 in the DNA strand of Fig. 2, CR3 and CR4 in the complementary strand. Left to right on the map corresponds to anticlockwise on the *E. coli* linkage map, and each scale division equals 100 bp.

consistent with the size of the protein tentatively identified as Fnr and its position agrees with the location of individual fnr mutations between targets R_1 and R_2 (7). The codon usage in CR1 and CR2 and the amino acid compositions of the products are typical for E. coli (Tables 1 and 2), whereas those for CR3 and CR4 are highly atypical. Thus it seems reasonable to conclude that CR1 is the fnr structural gene and that neither CR3 nor CR4 corresponds to an active gene.

Detection of messenger RNAs

In order to determine which of the potential coding regions were transcriptionally active, RNA was prepared and hybridised to single-stranded M13 DNA probes containing all or part of either strand of the 1.64 kb fragment. After digestion with nuclease S1, double-stranded hybrid molecules were identified on polyacrylamide gels. The source of the RNA was the wild-type E. coli strain PL2024 transformed with pGS24 to provide higher levels of the relevant mRNAs.

Of the two DNA probes containing each strand of the entire 1.64 kb fragment only one (mp8HB6) hybridised to RNA (Fig. 3). The negative

Table 1.

Codon usage in the fnr gene

TTT	3	TCT	1	TAT	3	TGT	2
TTC	8	TCC	4	TAC	2	TGC	2
TTA	1	TCA	0	TAA	0	TGA	1
TTG	2	TCG	4	TAG	0	TGG	0
CTT	6	CCT	2	CAT	5	CGT	9
CTC	1	CCC	0	CAC	1	CGC	5
CTA	0	CCA	0	CAA	3	CGA	1
CTG	17	CCG	5	CAG	12	CGG	2
ATT	5	ACT	3	AAT	4	AGT	1
ATC	17	ACC	4	AAC	5	AGC	7
ATA	1	ACA	1	AAA	9	AGA	0
ATG	9	ACG	6	AAG	5	AGG	0
GTT	1	GCT	4	GAT	8	GGT	10
GTC	1	GCC	6	GAC	5	GGC	11
GTA	2	GCA	3	GAA	11	GGA	1
GTG	1	GCG	3	GAG	5	GGG	0

Table 2.

Amino acid compositions of the Fnr and 'average' E.coli proteins (28)

Amino acid	Mol % in: Fnr	Average
Asp	5.2	5.2
Asn	3.9	3.5
Thr	5.6	5.8
Ser	6.8	6.3
Glu	6.4	5.5
Gln	6.0	4.2
Pro	2.8	3.5
Gly	8.8	7.2
Ala	6.4	11.1
Val	2.0	7.5
Met	3.6	2.2
Ile	9.2	6.1
Leu	10.8	7.9
Tyr	2.0	2.6
Phe	4.4	3.6
Lys	5.6	6.4
His	2.4	2.5
Arg	6.8	6.6
Cys	1.6	1.1
Trp	0.0	1.2

```
AAGCTTCGTGAATATTTTGCCGGTAATCTTAGCATTATTGATACGCTTCCCACTGCTACG
        10        20        30        40        50        60

GGGGGACGCCATTTCAGCGCGAAGTCTGGAAAACACTACGCACTATCCCCTGCGGCAGG
        70        80        90       100       110       120

TAATGCATTACGGCCAACTGGCTGAGCAATTGGGCCGTCCTGGCGCGGCGCGTCGCCGTT
       130       140       150       160       170       180

GGTGCGGCAAACGGATCGAATCCCATCAGCATCGTCGTACCTTGCCATCGGGTTATTGGC
       190       200       210       220       230       240

CGAAACGGCACCATGACCGGATATGCAGGCGGAGTTCAGCGAAAAGAGTGGTTATTGCGC
       250       260       270       280       290       300

CATGAAGGTTATCTTTTGCTGTAAACATTAAACAATTTGTCGCCAGCTTGTTCACACTTT
       310       320       330       340       350       360

TATGTAAAGTTACCCTTAACAACTTAAGGGTTTTCAAATAGATAGACATATATTTACATC
       370       380       390       400       410       420

TAATATCGGAATTCTCTGCTGTTAAGGTTTGCTTAGACTTACTTGCTCCCTAAAAAGATG
       430       440       450       460       470       480

                                          MetIleProGluLysArgIle
TTAAAATTGACAAATATCAATTACGGCTTGAGCAGACCTATGATCCCGGAAAAGCGAATT
       490       500       510       520       530       540

IleArgArgIleGlnSerGlyGlyCysAlaIleHisCysGlnAspCysSerIleSerGln
ATACGGCGCATTCAGTCTGGCGGTTGTGCTATCCATTGCCAGGATTGCAGCATCAGCCAG
       550       560       570       580       590       600

LeuSerIleProPheThrLeuAsnGluHisGluLeuAspGlnLeuAspAsnIleIleGlu
CTTTCGATCCCGTTCACACTCAACGAACATGAGCTTGATCAGCTTGATAATATCATTGAG
       610       620       630       640       650       660

ArgLysLysProIleGlnLysGlyGlnThrLeuPheLysAlaGlyAspGluLeuLysSer
CGGAAGAAGCCTATTCAGAAAGGCCAGACGCTGTTTAAGGCTGGTGATGAACTTAAATCG
       670       680       690       700       710       720

LeuTyrAlaIleArgSerGlyThrIleLysSerTyrThrIleThrGluGlnGlyAspGlu
CTTTATGCCATCCGCTCCGGTACGATTAAAAGTTATACCATCACTGAGCAAGGCGACGAG
       730       740       750       760       770       780

GlnIleThrGlyPheHisLeuAlaGlyAspLeuValGlyPheAspAlaIleGlySerGly
CAAATCACTGGTTTCCATTTAGCAGGCGACCTGGTGGGATTTGACGCCATCGGCAGCGGC
       790       800       810       820       830       840

HisHisProSerPheAlaGlnAlaLeuGluThrSerMetValCysGluIleProPheGlu
CATCACCCGAGCTTCGCGCAGGCGCTGGAAACCTCGATGGTATGTGAAATCCCGTTCGAA
       850       860       870       880       890       900

ThrLeuAspAspLeuSerGlyLysMetProAsnLeuArgGlnGlnMetMetArgLeuMet
ACGCTGGACGATTTGTCCGGTAAAATGCCGAATCTGCGTCAGCAGATGATGCGTCTGATG
       910       920       930       940       950       960
```

```
SerGlyGluIleLysGlyAspGlnAspMetIleLeuLeuLeuSerLysLysAsnAlaGlu
AGCGGTGAAATCAAAGGCGATCAGGACATGATCCTGCTGTTGTCGAAGAAAAATGCCGAG
       970       980       990      1000      1010      1020

GluArgLeuAlaAlaPheIleTyrAsnLeuSerArgArgPheAlaGlnArgGlyPheSer
GAACGTCTGGCTGCATTCATCTACAACCTGTCCCGTCGTTTTGCCCAACGCGGCTTCTCC
      1030      1040      1050      1060      1070      1080

ProArgGluPheArgLeuThrMetThrArgGlyAspIleGlyAsnTyrLeuGlyLeuThr
CCTCGTGAATTCCGCCTGACGATGACTCGTGGCGATATCGGTAACTATCTGGGCCTGACG
      1090      1100      1110      1120      1130      1140

ValGluThrIleSerArgLeuLeuGlyArgPheGlnLysSerGlyMetLeuAlaValLys
GTAGAAACCATCAGCCGTCTGCTGGGTCGCTTCCAGAAAAGCGGCATGCTGGCAGTCAAA
      1150      1160      1170      1180      1190      1200

GlyLysTyrIleThrIleGluAsnAsnAspAlaLeuAlaGlnLeuAlaGlyHisThrArg
GGTAAATACATCACCATCGAAAATAACGATGCGCTGGCCCAGCTTGCTGGTCATACGCGT
      1210      1220      1230      1240      1250      1260

AsnValAla***
AACGTTGCCTGATTTTTCCGCATAACTCACTATCCTTCTGTCATATCATTAAATTTTTCT
      1270      1280      1290      1300      1310      1320

GATTTATTGATCTGGCAGAAGGTTCATCACTGTTTCATTCACCAGATATGGGTTAATCTT
      1330      1340      1350      1360      1370      1380

                                          MetAlaMetTyrGlnAsn
TTAATTACAAACTGCGTTGACAGTTGTTGTAAGGAGACCCTGTATGGCTATGTATCAGAA
      1390      1400      1410      1420      1430      1440

 MetLeuValValIleAspProAsnGlnAspAspGlnProAlaLeuArgArgAlaValTyr
CATGCTCGTTGTTATCGATCCTAACCAGGACGACCAACCAGCATTGCGGCGAGCTGTTTA
      1450      1460      1470      1480      1490      1500

 LeuHisGlnArgIleGlyGlyLysIleLysAlaPheLeuProIleTyrAspPheSerTyr
TTTACATCAACGGATTGGTGGCAAAATTAAAGCCTTTTTGCCGATCTATGACTTCTCATA
      1510      1520      1530      1540      1550      1560

 GluMetThrThrLeuLeuSerProGluArgThrTyrArgTyrAlaSerGlyArgHisGln
CGAAATGACCACCCTGCTCTCCCCGGAGCGAACGTACCGCTATGCGTCAGGGCGTCATCA
      1570      1580      1590      1600      1610      1620

 ProAlaTyrSerLeuAspPro
GCCAGCGTACAGCCTGGATCC
      1630      1640
```

Figure 2. Nucleotide sequence of the 1.64 kb DNA fragment containing the fnr gene. The deduced amino acid sequences of the products of the fnr gene (520-1272) and 'gene X' (1424-1641) are shown. Regions of dyad symmetry in the DNA are underlined with arrows. Potential Shine-Dalgarno sequences (ribosome binding sites) are boxed. Lines above the nucleotide sequence indicate possible Pribnow (-10) and -35 promoter sequences.

result obtained with the other probe (mp9HB3; Fig. 3) indicates that there is no detectable hybridisation between RNA and M13 vector DNA sequences. The results obtained with mp8HB6 and mp9HB3 suggest that of all the four potential coding regions, only the proposed fnr coding region and 'gene X' (CR1 and CR2, Fig. 1) are transcribed. Two hybrid bands (sizes 790 ± 40 and 280 ± 20) were obtained with mp8HB6 showing that the two genes are transcribed separately.

To position the two mRNA transcripts on the 1.64 kb fragment, probes containing the coding strand (i.e. the same strand as in mp8HB6) of the three EcoRI subfragments were used (mp8HR5, subfragment H-R_1; mp8R9, R_1-R_2; mp9BR10, R_2-B). No hybrids were formed with mp8HR5, but mp8R9 produced one hybrid band (570 ± 30 bp), which presumably represents hybridisation to part of the 790 bp transcript detected with mp8HB6. This indicates that transcription of the fnr gene starts at position 520 ± 30 and continues rightward beyond R_2, terminating at position 1310 ± 40. It should have been possible to detect a hybrid of size 220 bp with mp9BR10 (resulting from hybridisation of this DNA to the right-hand end of the 790 bp transcript) but the band may have been obscured by undigested low molecular weight RNA (Fig. 3). Probe mp9BR10 gave rise to one hybrid of size 280 ± 20 bp, apparently identical to that obtained with mp8HB6, confirming that this mRNA is

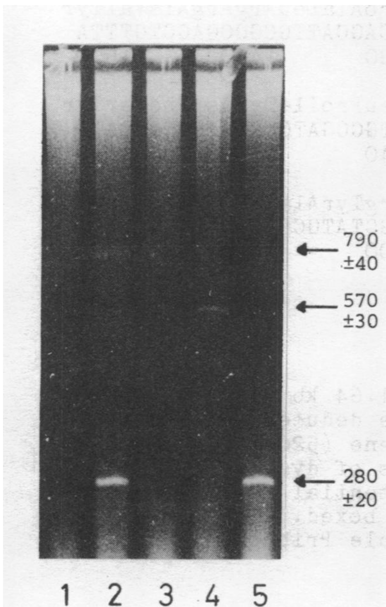

Figure 3. Hybrids formed between RNA from strain PL2024(pGS24), and single-stranded M13 DNA probes. Track 1, probe mp9HB3; track 2, mp8HB6; track 3, mp8HR5; track 4, mp8R9; track 5, mp9BR10. Size markers (not shown) were Sau3A-digested pBR322, HindIII-digested and HindIII plus EcoRI-digested λ. Arrows indicate positions and sizes of hybrid molecules.

generated by transcription of gene X starting at position 1360 ± 20.

Gene-polypeptide relationships

Using the maxicell system it has been shown that pGS24 directs the synthesis of a protein with M_r=31000, tentatively identified as the fnr gene product, and three minor polypeptides (M_r=28000, 27000, 19000) that may be processed or degraded forms of the 31000 protein (7). The proposed fnr gene coding region (520-1272, Fig. 2) would express a protein of molecular weight 27947, in good agreement with the estimate of 31000.

The estimated start-point of fnr transcription is position 520 ± 30, suggesting that the fnr promoter is located rightwards of target R_1 (429 in Fig. 2). A plasmid lacking the DNA between this target and the HindIII target (position 1, Fig. 2) was constructed by partial digestion of pGS24 with EcoRI and religation. This derivative (pGS70), which has also lost the residue of the vector tetracycline resistance promoter (24), was indistinguishable from pGS24 in terms of both complementation of fnr mutations and synthesis of polypeptides in maxicells (results not shown), confirming the view that the fnr promoter is located between 429 and 519 (Fig. 2).

The discovery of gene X raised the possibility that one or more of the maxicell proteins may have resulted from fusion between this gene and the pBR322 tetracycline resistance gene. A plasmid lacking fragment R_1-R_2 (429-1087, Fig. 2) was constructed by partial digestion of pGS24 with EcoRI and religation, so as to retain gene X but not fnr. The plasmid (pGS69) gave rise to no unique polypeptides in the maxicell system, indicating that the fusion products (if any) are not abundant or stable enough to be detected, and confirming that the 31000 protein is the product of the proposed fnr gene.

Features of the DNA sequence

The proposed fnr gene is preceded by a correctly placed ribosome binding site (Shine-Dalgarno sequence; 25), containing GAG at positions 510-512 (Fig. 2) which is complementary to part of the 3' end of the 16s ribosomal RNA. Gene X is also preceded by a ribosome binding site containing the sequence TAAGGAG (1410-1416, Fig. 2). Neither of the potential coding regions CR3 and CR4 (Fig. 1), which were not found to be transcribed, is preceded by a recognisable ribosome binding site.

The fnr promoter is probably located between 429 and 519, and inspection of this region revealed a likely Pribnow (-10) sequence at

482-487. The sequence (TAAAAT) is identical in 5 out of 6 positions (including the highly conserved 1st, 2nd and 6th bases) to the 'consensus' Pribnow sequence TATAAT (26). Homologies to the sequence TTGACA are found in the -35 region of many E. coli promoters and the fnr gene has homologous sequences at 449-454 (TTGCTT), 453-458 (TTAGAC) and 459-464 (TTACTT), the third of which is most favourably positioned relative to the Pribnow sequence (26). The proposed fnr Pribnow sequence forms part of a short region of dyad symmetry (Fig. 2).

The RNA:DNA hybridisation studies showed that transcription of fnr terminates at position 1310 ± 40. Examination of this region of the sequence revealed a large potential stem-loop structure centred at 1317, involving 19 base-pairings with a free energy of formation of -81.1 kJ mol^{-1} (Fig. 2). Such structures are associated with transcription termination and the absence of a run of T residues following the stem suggests that termination is rho-dependent (26, 27).

The Pribnow and -35 sequences of two possible promoters for gene X, consistent with the estimate of its transcriptional start-point (1360 ± 20) are indicated in Fig. 2.

The codon usage of the proposed fnr gene (Table 1) is highly non-random and very similar to the average codon usage in E. coli (28). The pattern of degenerate codon choice has been used to distinguish between 'weakly' and 'strongly' expressed mRNAs (28) but no definite assignment to one of these classes could be made for fnr.

The orientation of the DNA containing the fnr gene relative to the E. coli linkage map has been established (Fig. 1 ; 7) and combined with the above results suggests that fnr and gene X are both transcribed with anti-clockwise polarity (4).

Primary structure of the Fnr protein

The primary structure of the protein, deduced from the nucleotide sequence of the putative fnr gene, is shown in Fig. 2. Due to the absence of relevant information it was not possible to confirm the amino acid sequence directly, but application of the FRAMESCAN program (23) to the nucleotide sequence indicated that there were no frameshift errors (FRAMESCAN determines the probability that any region of a sequence, in any of the three reading frames, codes for protein). The amino acid composition of the Fnr protein is given in Table 2: it is not strikingly different from that of an 'average' E. coli protein (28) with the possible exception of the rather low valine and alanine, and high leucine and isoleucine, contents.

Comparisons between the amino acid sequences of Fnr and other proteins showed that there are significant homologies with the catabolite activator protein CAP (29) and other transcriptional regulator proteins of E. coli, e.g. the lac and gal repressors (30, 31). The detailed analysis of the comparisons, which will be published separately (32), provides strong evidence that Fnr is a positively-acting regulatory protein controlling diverse functions associated with anaerobic respiration.

ACKNOWLEDGEMENTS

We would like to thank P. Stephens and M. Darlison for advice with cloning and sequencing, F. Sanger and A. Coulson for advice and for a generous gift of primer, and I. Duckenfield for assistance with computing. The work was supported by the U.K. Science and Engineering Research Council.

REFERENCES

1. Lambden, P. R. and Guest, J. R. (1976) J. Gen. Microbiol. 97, 145-160.
2. Newman, B. M. and Cole, J. A. (1978) J. Gen. Microbiol. 106, 1-12.
3. Chippaux, M., Giudici, D., Abou-Jaoudé, A., Casse, F. and Pascal, M. C. (1978) Mol. Gen. Genet. 160, 225-229.
4. Bachmann, B. J. and Low, K. B. (1980) Microbiol. Rev. 44, 1-56.
5. Shaw, D. J. and Guest, J. R. (1981) Mol. Gen. Genet. 181, 95-100.
6. Chippaux, M., Bonnefoy-Orth, V., Ratouchniak, J. and Pascal, M. C. (1981) Mol. Gen. Genet. 182, 477-479.
7. Shaw, D. J. and Guest, J. R. (1982) J. Gen. Microbiol.128, 2221-2228.
8. Wimpenny, J. W. T. and Cole, J. A. (1967) Biochim. Biophys. Acta 148, 233-242.
9. Cole, J. A. and Wimpenny, J. W. T. (1968) Biochim. Biophys. Acta 162, 39-48.
10. Cole, S. T. and Guest, J. R. (1979) FEMS Microbiol. Letts. 5, 65-67.
11. Cole, S. T. and Guest, J. R. (1980) Mol. Gen. Genet. 179, 377-385.
12. Messing, J. (1979) Recombinant DNA Tech. Bull. 2, 43-48.
13. Clewell, D. B. and Helinski, D. R. (1970) Biochemistry 9, 4428-4440.
14. Clewell, D. B. (1972) J. Bacteriol. 110, 667-676.
15. Birnboim, H. C. and Doly, J. (1979) Nucleic Acids Res. 7, 1513-1523.
16. Smith, H. O. (1980) Methods in Enzymol. 65, 371-380.
17. Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. and Roe, B. A. (1980) J. Mol. Biol. 143, 161-178.
18. Sanger, F., Nicklen, S. and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
19. Staden, R. (1979) Nucleic Acids Res. 6, 2601-2610.
20. Staden, R. (1980) Nucleic Acids Res. 8, 3673-3694.

21. Sancar, A., Hack, A. M. and Rupp, W. D. (1979) J. Bacteriol. 137, 692-693.
22. Squires, C., Krainer, A., Barry, G., Shen, W. F. and Squires, C. L. (1981) Nucleic Acids Res. 9, 6827-6840.
23. Staden, R. and McLachlan, A. D. (1982) Nucleic Acids Res. 10, 141-156.
24. Sutcliffe, J. G. (1979) Cold Spring Harbor Symp. Quant. Biol. 43, 77-90.
25. Gold, L., Pribnow, D., Schneider, T., Shinedling, S., Singer, B. S. and Stormo, G. (1981) Ann. Rev. Microbiol. 35, 365-403.
26. Rosenberg, M. and Court, D. (1979) Ann. Rev. Genet. 13, 319-353.
27. Platt, T. (1981) Cell 24, 10-23.
28. Grantham, R., Gautier, C., Gouy, M., Jacobzone, M. and Mercier, R. (1981) Nucleic Acids Res. 9, r43-r74.
29. Cossart, P. and Gicquel-Sanzey, B. (1982) Nucleic Acids Res. 10, 1363-1378.
30. Farabaugh, P. J. (1978) Nature 274, 765-769.
31. Von Wilcken-Bergmann, B. and Müller-Hill, B. (1982) Proc. Natl. Acad. Sci. USA 79, 2427-2431.
32. Shaw, D. J. and Guest, J. R. (Submitted to Nature).