Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 4650–4654, July 1986
Biochemistry

Nucleotide sequence and expression of the selenocysteine-containing polypeptide of formate dehydrogenase (formate-hydrogen-lyase-linked) from *Escherichia coli*

(selenium incorporation/*lacZ* fusion/nonsense suppression)

F. ZINONI*, A. BIRKMANN*, THRESSA C. STADTMAN†, AND A. BÖCK*

*Lehrstuhl für Mikrobiologie der Universität München, Maria-Ward-Strasse 1a, D-8000 München 19, Federal Republic of Germany; and †Laboratory of Biochemistry, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD 20892

Contributed by Thressa C. Stadtman, March 3, 1986

ABSTRACT **The gene (*fdhF*) coding for the selenopolypeptide of the benzylviologen-linked formate dehydrogenase of *Escherichia coli* was cloned and its nucleotide sequence was determined. The *fdhF* gene contains, within an open reading frame coding for a protein of 715 amino acids (calculated molecular weight, 79,087), an opal (UGA) nonsense codon in amino acid position 140. Existence of this nonsense codon was confirmed by physical recloning and resequencing. Internal and terminal deletion clones and *lacZ* fusions of different N-terminal parts of *fdhF* were constructed and analyzed for selenium incorporation. Selenylated truncated polypeptide chains or β-galactosidase fusion proteins were synthesized when the deletion clones or gene fusions, respectively, contained the *fdhF* gene fragment coding for the selenopolypeptide sequence from amino acid residue 129 to amino acid residue 268. Translation of the *lacZ* part of the fusions required the presence of selenium in the medium when the N-terminal *fdhF* part contained the UGA codon and was independent of the presence of selenium when a more upstream part of *fdhF* was fused to *lacZ*. The results are consistent with a co-translational selenocysteine incorporation mechanism.**

Selenium is an essential component of several enzymes from both prokaryotic and eukaryotic organisms; examples are glycine reductase, nicotinic acid hydroxylase, and xanthine dehydrogenase from clostridia, glutathione peroxidase from mammalia and birds, a hydrogenase from *Methanococcus vannielii*, and formate dehydrogenases from certain obligate and facultative anaerobes (for review, see refs. 1 and 2). Many of these enzymes have been shown to contain selenium as a single selenocysteine residue within the polypeptide chain. Although several possibilities have been considered (1), there is no conclusive information yet that explains this high specificity of selenocysteine incorporation.

A resolution of this intriguing question may be approached by genetic analysis. Comparison of the nucleotide sequence of the gene for a selenocysteine-containing polypeptide with the amino acid sequence of this protein can provide information on whether selenocysteine incorporation is directed by an unusual codon or codon context or whether this is accomplished by the posttranslational modification of an amino acid. With this goal in mind we have cloned the gene for the selenopolypeptide of the benzylviologen-linked formate dehydrogenase (FDH_H; a component of the system formerly known as formate hydrogenlyase) from *Escherichia coli*, determined its nucleotide sequence, and studied its expression.

MATERIALS AND METHODS

Strains and Plasmids. Strain FM911 of *E. coli* is a *recA56* derivative of strain FM909 (3) that carries a deletion of the *fdhF* gene. *E. coli* MC4100 (4) and the vector pACYC184 (5) were used for cloning the *fdhF* gene. *fdhF*::*lacZ* gene fusions were constructed with the aid of plasmid pMC1403 (6). In the case of the fusion plasmid pBN2, in-frame fusion was accomplished by fill-in reaction of the sticky ends of the *Kpn* I site at nucleotide position 119 of the *fdhF* gene and of the *Eco*RI site within the multilinker of pMC1403 and consecutive blunt-end ligation. pFM52 was constructed in the same way using the *Eco*RV site (position 807) of *fdhF* and the *Bam*HI site of the pMC1403 multilinker. Direct ligation of the *Bgl* II site (position 1260) of *fdhF* and the *Bam*HI site of pMC1403 resulted in the in-frame fusion carried by plasmid pFM54. The fusion points of all three plasmids were proven by DNA sequencing.

Genetic Techniques. Basic genetic techniques were as given by Miller (7). Recombinant DNA procedures were adopted from Maniatis *et al.* (8). DNA sequencing was carried out as described by Maxam and Gilbert (9) as modified by Gray *et al.* (10). The restriction sites given in Fig. 1 and BAL-31 nuclease fragments (Fig. 3) were used for end-labeling of DNA fragments. Both strands of the DNA fragments were sequenced throughout. The Messing program of an Apple IIe computer was used for sequence analysis.

Microbiological and Biochemical Methods. Screening for $fdhF^+$ transformants was carried out with the benzylviologen agar overlay technique (11). Measurement of the ability to produce gas has been described (3). The assay for selenium incorporation into polypeptide chains was performed according to Cox *et al.* (12). The maxicell expression experiments were done following the standard technique (13) with the modification that all microbiological procedures following UV irradiation were performed under anaerobic conditions.

Materials. Restriction enzymes, polymerase I (Klenow fragment), T4 polynucleotide kinase, and T4 ligase were obtained from Boehringer Mannheim or from Pharmacia. Radionucleotides and ^{75}Se were purchased from Amersham or DuPont.

RESULTS

Cloning and Nucleotide Sequence of the Gene (*fdhF*) for the Selenopolypeptide of Formate Dehydrogenase. For cloning the gene (*fdhF*) for the selenopolypeptide of formate dehydrogenase (FDH_H) from *E. coli*, use was made of a mutant in which integration of phage Mu *d* (Ap *lac*) (4) had led to the formation of a truncated form of the selenopolypeptide (3). A

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: kb, kilobase(s); FDH_H, formate dehydrogenase component of the system formerly known as formate hydrogenlyase.

derivative of this mutant was constructed in which excision of Mu resulted in deletion of the *fdhF* gene (strain FM911); it was used as recipient in transformation experiments with plasmid pACYC184 that contained *Bam*HI restriction fragments of chromosomal DNA from *E. coli* MC4100. Two transformants were obtained that contained a hybrid plasmid with a 6.2-kilobase (kb) *Bam*HI insert. Both of these expressed benzylviologen-linked formate dehydrogenase activity under anaerobic conditions. Strain FM911 containing plasmid pFM3, when grown in the presence of $^{75}SeO_3^{2-}$, synthesized an 80-kDa protein that contained ^{75}Se in the form of selenocysteine‡. Nick-translation of plasmid pFM3 and hybridization with *Bam*HI-digested chromosomal DNA from *E. coli* MC4100 yielded a single hybridization signal of the same size as the *Bam*HI insert of plasmid pFM3. The subcloning experiments shown in Fig. 1 localized the FDH_H complementing activity finally on a 3.3-kb *Sma* I–*Bgl* II fragment (plasmid pFM20).

The nucleotide sequence of this *Sma* I–*Bgl* II fragment was determined. Translation of the sequence in all possible frames showed that there is only one putative open reading frame of respectable size; with some reservation (see below), it codes for a protein of 79,087 daltons, which is in excellent agreement with the molecular weight of about 80,000 determined previously for the selenopolypeptide of FDH_H (3, 12). The nucleotide sequence and the derived amino acid sequence of this open reading frame and of the immediate 5′ and 3′ flanking regions are shown in Fig. 2.

The reservation made above concerns the fact that there is a UGA nonsense codon in amino acid position 140 of the putative open reading frame. Several possibilities exist concerning the relevance of this UGA codon: (*i*) it might have been selected during the cloning procedure counteracting any adverse effect of *fdhF* overexpression; (*ii*) it might have been introduced during *in vitro* manipulation of the DNA; (*iii*) it could constitute a mutation introduced during the long laboratory history of strain MC4100; (*iv*) the *Sma* I–*Bgl* II fragment open reading frame may not code for a protein of M_r 80,000; and (*v*) the UGA codon could be used to direct the incorporation of selenium (or selenocysteine) into the polypeptide chain.

To decide among these possibilities we have recloned the *Kpn* I–*Bam*HI part of the *fdhF* gene by purely physical methods using the internal 1.15-kb *Bgl* II–*Bgl* II fragment as hybridization probe (see Fig. 1). Determination of the nucleotide sequence of the recloned gene fragment gave a sequence identical to that of Fig. 2, which rules out possibilities *i* and *ii* and proves that the UGA codon did not artifactually derive from the cloning procedure (results not shown). A physiological argument against possibility *iii* is that strain MC4100 of *E. coli* actively forms gas and exhibits wild-type-like benzylviologen-linked formate dehydrogenase activity under anaerobic conditions (3).

The Open Reading Frame of pFM20 Codes for the M_r 80,000 Selenopolypeptide of FDH_H. Plasmid pFM20 complements gas formation and FDH_H activity to the *fdhF* mutant FM911. As shown in Fig. 3*B*, pFM20 also enables FM911 to synthesize an 80-kDa selenoprotein. The following experiments were carried out to prove definitively that the open reading frame that contains the UGA nonsense codon directs the synthesis of the 80-kDa selenopolypeptide of FDH_H.

First, pFM20 was transformed into strain CSR603 and expressed in a "maxicell" experiment (13). Using [^{35}S]methionine as label, we detected the anaerobic synthesis of a gene product of the approximate size of the FDH_H selenopeptide. This protein was absent in control experiments in which the vector pACYC184 was used alone (results not shown).

‡Carboxymethyl[^{75}Se]selenocysteine was identified in the acid hydrolysate of a chymotryptic peptide derived from the 80-kDa alkylated selenoprotein. Details will be presented elsewhere.

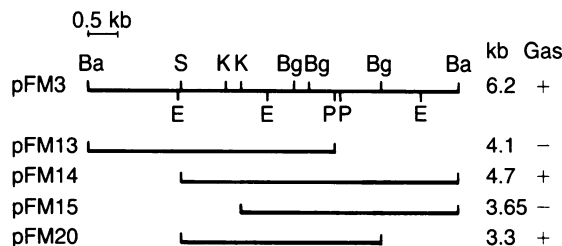

FIG. 1. Restriction map and subcloning of the *fdhF* gene. The sites for restriction endonucleases are abbreviated as follows: Ba, *Bam*HI; Bg, *Bgl* II; E, *Eco*RV; K, *Kpn* I; P, *Pvu* I; S, *Sma* I. The 6.2-kb *Bam*HI insert of pFM3 does not contain sites for *Cla* I, *Eco*RI, *Hin*dIII, *Pst* I, or *Sal* I. The plasmids were transformed into strain FM911 (*fdhF*) and the transformants were tested for their ability for gas formation.

Second, a series of terminal and internal deletions of the open reading frame of pFM20 was constructed by *in vitro* methods and tested for ability to code for the synthesis of selenopeptides in strain FM911. As shown in Fig. 3, shortened selenopeptides were synthesized as a result of these deletions. The physical map of these deletions and a few examples of the migration patterns of the shortened selenopeptides in a $NaDodSO_4$ gel are presented in Fig. 3 *A* and *C*, respectively. The apparent sizes of these truncated polypeptides show a high degree of correlation with the size of the open reading frames from the respective deleted genes (Fig. 3*D*). The slope of the curve extrapolates to zero—i.e., to the start site of the open reading frame.

The UGA Stop Codon Is Translated. The results shown in Fig. 3 demonstrate that the syntheses of truncated selenopolypeptides of the expected sizes are directed by plasmids that contain deletions "downstream" from the UGA codon within the *fdhF* gene. This is strong evidence that the opal codon is translated. Final proof for overreading of the UGA was provided by the study of *in vitro*-constructed *fdhF*::*lacZ* fusions. Three in-frame fusions between different N-terminal parts of the *fdhF* gene and *lacZ* were made using vector pMC1403 (6). All three fusion plasmids contain 700 nucleotides of the 5′ upstream region of the *fdhF* gene; as a consequence, the β-galactosidase fusion proteins are formed only under anaerobic conditions (unpublished results). pBN2 contains the 5′ part of *fdhF* down to the *Kpn* I site at position 119 of the sequence (Fig. 2), pFM52 to the *Eco*RV site at position 807, and pFM54 to the *Bgl* II site at position 1260. Accordingly, the N-terminal parts of these β-galactosidase fusion proteins should contain extra 39, 268, and 420 amino acids, respectively, derived from *fdhF*, resulting in fusion proteins of M_r 115,300, 139,000, and 156,000. Fig. 4 demonstrates that fusion proteins of these sizes, indeed, are synthesized. Radioactive selenium is incorporated into the fusion proteins encoded by pFM52 and pFM54 but not into that encoded by pBN2 (Fig. 4).

DISCUSSION

The results of this study provide the structural basis for the elucidation of the interesting question of the mechanisms of the highly specific incorporation of selenocysteine into a polypeptide chain. Among the more plausible possibilities are co-translational incorporation of selenocysteine directed by some codon or codon context or posttranslational recognition of a specific primary sequence of an apoprotein chain and selenylation of a precursor amino acid.

The finding of a UGA stop codon in the open reading frame that is translated suggests that selenylation may occur at the level of an aminoacyl-tRNA and that the resulting seleno-

```
-50                                                      ACTGTAGTCG AGAGCGCGTA TGCGTGATTT GATTAACTGG AGCGAGACCG

   1  ATG AAA AAA GTC GTC ACG GTT TGC CCC TAT TGC GCA TCA GGT TGC AAA ATC AAC CTG GTC GTC GAT AAC GGC AAA ATC GTC CGG
   1  Met Lys Lys Val Val Thr Val Cys Pro Tyr Cys Ala Ser Gly Cys Lys Ile Asn Leu Val Val Asp Asn Gly Lys Ile Val Arg
                                                 pBN2 ↓
  85  GCG GAG GCA GCG CAG GGG AAA ACC AAC CAG GGT ACC CTG TGT CTG AAG GGT TAT TAT GGC TGG GAC TTC ATT AAC GAT ACC CAG
  29  Ala Glu Ala Ala Gln Gly Lys Thr Asn Gln Gly Thr Leu Cys Leu Lys Gly Tyr Tyr Gly Trp Asp Phe Ile Asn Asp Thr Gln

 169  ATC CTG ACC CCG CGC CTG AAA ACC CCC ATG ATC CGT CGC CAG CGT GGC GGC AAA CTC GAA CCT GTT TCC TGG GAT GAG GCA CTG
  57  Ile Leu Thr Pro Arg Leu Lys Thr Pro Met Ile Arg Arg Gln Arg Gly Gly Lys Leu Glu Pro Val Ser Trp Asp Glu Ala Leu

 253  AAT TAC GTT GCC GAG CGC CTG AGC GCC ATC AAA GAG AAG TAC GGT CCG GAT GCC ATC CAG ACG ACC GGC TCC TCG CGT GGT ACG
  85  Asn Tyr Val Ala Glu Arg Leu Ser Ala Ile Lys Glu Lys Tyr Gly Pro Asp Ala Ile Gln Thr Thr Gly Ser Ser Arg Gly Thr

 337  GGT AAC GAA ACC AAC TAT GTA ATG CAA AAA TTT GCG CGC GCC GTT ATT GGT ACC AAT AAC GTT GAC TGC TGC GCT CGT GTC[TGA]
 113  Gly Asn Glu Thr Asn Tyr Val Met Gln Lys Phe Ala Arg Ala Val Ile Gly Thr Asn Asn Val Asp Cys Cys Ala Arg Val[XXX]

 421  CAC GGC CCA TCG GTT GCA GGT CTG CAC CAA TCG GTC GGT AAT GGC GCA ATG AGC AAT GCT ATT AAC GAA ATT GAT AAT ACC GAT
 141  His Gly Pro Ser Val Ala Gly Leu His Gln Ser Val Gly Asn Gly Ala Met Ser Asn Ala Ile Asn Glu Ile Asp Asn Thr Asp

 505  TTA GTG TTC GTT TTC GGG TAC AAC CCG GCG GAT TCC CAC CCA ATC GTG GCG AAT CAC GTA ATT AAC GCT AAA CGT AAC GGG GCG
 169  Leu Val Phe Val Phe Gly Tyr Asn Pro Ala Asp Ser His Pro Ile Val Ala Asn His Val Ile Asn Ala Lys Arg Asn Gly Ala

 589  AAA ATT ATC GTC TGC GAT CCG CGC AAA ATT GAA ACC GCG CGC ATT GCT GAC ATG CAC ATT GCA CTG AAA AAC GGC TCG AAC ATC
 197  Lys Ile Ile Val Cys Asp Pro Arg Lys Ile Glu Thr Ala Arg Ile Ala Asp Met His Ile Ala Leu Lys Asn Gly Ser Asn Ile

 673  GCG CTG TTG AAT GCG ATG GGC CAT GTC ATT ATT GAA GAA AAT CTG TAC GAC AAA GCG TTC GTC GCT TCA CGT ACA GAA GGC TTT
 225  Ala Leu Leu Asn Ala Met Gly His Val Ile Ile Glu Glu Asn Leu Tyr Asp Lys Ala Phe Val Ala Ser Arg Thr Glu Gly Phe
                                                                 pFM52 ↓
 757  GAA GAG TAT CGT AAA ATC GTT GAA GGC TAC ACG CCG GAG TCG GTT GAA GAT ATC ACC GGC GTC AGC GCC AGT GAG ATT CGT CAG
 253  Glu Glu Tyr Arg Lys Ile Val Glu Gly Tyr Thr Pro Glu Ser Val Glu Asp Ile Thr Gly Val Ser Ala Ser Glu Ile Arg Gln

 841  GCG GCA CGG ATG TAT GCC CAG GCG AAA AGC GCC GCC ATC CTG TGG GGC ATG GGT GTA ACC CAG TTC TAC CAG GGC GTG GAA ACC
 281  Ala Ala Arg Met Tyr Ala Gln Ala Lys Ser Ala Ala Ile Leu Trp Gly Met Gly Val Thr Gln Phe Tyr Gln Gly Val Glu Thr

 925  GTG CGT TCT CTG ACC AGC CTC GCG ATG CTG ACC GGT AAC CTC GGT AAG CCG CAT GCG GGT GTT AAC CCG GTT CGT GGT CAG AAC
 309  Val Arg Ser Leu Thr Ser Leu Ala Met Leu Thr Gly Asn Leu Gly Lys Pro His Ala Gly Val Asn Pro Val Arg Gly Gln Asn

1009  AAC GTT CAG GGT GCC TGC GAT ATG GGC GCG CTG CCG GAT ACG TAT CCG GGA TAC CAG TAC GTG AAA GAT CCG GCT AAC CGC GAG
 337  Asn Val Gln Gly Ala Cys Asp Met Gly Ala Leu Pro Asp Thr Tyr Pro Gly Tyr Gln Tyr Val Lys Asp Pro Ala Asn Arg Glu

1093  AAA TTC GCC AAA GCC TGG GGC GTG GAA AGC CTG CCA GCG CAT ACC GGC TAT CGC ATC AGC GAG CTG CCG CAC CGC GCA GCG CAT
 365  Lys Phe Ala Lys Ala Trp Gly Val Glu Ser Leu Pro Ala His Thr Gly Tyr Arg Ile Ser Glu Leu Pro His Arg Ala Ala His
                                                                                                         pFM54 ↓
1177  GGC GAA GTG CGT GCC GCG TAC ATT ATG GGC GAA GAT CCG CTA CAA ACT GAC GCG GAG CTG TCG GCA GTA CGT AAA GCC TTT GAA
 393  Gly Glu Val Arg Ala Ala Tyr Ile Met Gly Glu Asp Pro Leu Gln Thr Asp Ala Glu Leu Ser Ala Val Arg Lys Ala Phe Glu

1261  GAT CTG GAA CTG GTT ATC GTT CAG GAC ATC TTT ATG ACC AAA ACC GCG TCG GCG GCG GAT GTT ATT TTA CCG TCA ACG TCG TGG
 421  Asp Leu Glu Leu Val Ile Val Gln Asp Ile Phe Met Thr Lys Thr Ala Ser Ala Ala Asp Val Ile Leu Pro Ser Thr Ser Trp

1345  GGC GAG CAT GAA GGC GTG TTT ACT GCG GCT GAC CGT GGC TTC CAG CGT TTC TTC AAG GCG GTT GAA CCG AAA TGG GAT CTG AAA
 449  Gly Glu His Glu Gly Val Phe Thr Ala Ala Asp Arg Gly Phe Gln Arg Phe Phe Lys Ala Val Glu Pro Lys Trp Asp Leu Lys

1429  ACG GAC TGG CAA ATC ATC AGT GAA ATC GCC ACC CGT ATG GGT TAT CCG ATG CAC TAC AAC AAC ACC CAG GAG ATC TGG GAT GAG
 477  Thr Asp Trp Gln Ile Ile Ser Glu Ile Ala Thr Arg Met Gly Tyr Pro Met His Tyr Asn Asn Thr Gln Glu Ile Trp Asp Glu

1513  TTG CGT CAT CTG TGC CCG GAT TTC TAC GGT GCG ACT TAC GAG AAA ATG GGC GAA CTG GGC TTC ATT CAG TGG CCT TGC CGC GAT
 505  Leu Arg His Leu Cys Pro Asp Phe Tyr Gly Ala Thr Tyr Glu Lys Met Gly Glu Leu Gly Phe Ile Gln Trp Pro Cys Arg Asp

1597  ACT TCA GAT GCC GAT CAG GGG ACT TCT TAT CTG TTT AAA GAG AAG TTT GAT ACC CCG AAC GGT CTG GCG CAG TTC TTC ACC TGC
 533  Thr Ser Asp Ala Asp Gln Gly Thr Ser Tyr Leu Phe Lys Glu Lys Phe Asp Thr Pro Asn Gly Leu Ala Gln Phe Phe Thr Cys

1681  GAC TGG GTA GCG CCA ATC GAC AAA CTC ACC GAC GAG TAC CCG ATG GTA CTG TCA ACG GTG CGT GAA GTT GGT CAC TAC TCT TGC
 561  Asp Trp Val Ala Pro Ile Asp Lys Leu Thr Asp Glu Tyr Pro Met Val Leu Ser Thr Val Arg Glu Val Gly His Tyr Ser Cys

1765  CGT TCG ATG ACC GGT AAC TGT GCG GCA CTG GCG GCG CTG GCT GAT GAA CCT GGC TAC GCA CAA ATC AAT ACC GAA GAC GCC AAA
 589  Arg Ser Met Thr Gly Asn Cys Ala Ala Leu Ala Ala Leu Ala Asp Glu Pro Gly Tyr Ala Gln Ile Asn Thr Glu Asp Ala Lys

1849  CGT CTG GGT ATT GAA GAT GAG GCA TTG GTT TGG GTG CAC TCG CGT AAA GGC AAA ATT ATC ACC CGT GCG CAG GTC AGC GAT CGT
 617  Arg Leu Gly Ile Glu Asp Glu Ala Leu Val Trp Val His Ser Arg Lys Gly Lys Ile Ile Thr Arg Ala Gln Val Ser Asp Arg

1933  CCG AAC AAA GGG GCG ATT TAC ATG ACC TAC CAG TGG TGG ATT GGT GCC TGT AAC GAG CTG GTT ACC GAA AAC TTA AGC CCG ATT
 645  Pro Asn Lys Gly Ala Ile Tyr Met Thr Tyr Gln Trp Trp Ile Gly Ala Cys Asn Glu Leu Val Thr Glu Asn Leu Ser Pro Ile

2017  ACG AAA ACG CCG GAG TAC AAA TAC TGC GCC GTT CGC GTC GAG CCG ATC GCC GAT CAG CGC GCC GCC GAG CAG TAC GTG ATT GAC
 673  Thr Lys Thr Pro Glu Tyr Lys Tyr Cys Ala Val Arg Val Glu Pro Ile Ala Asp Gln Arg Ala Ala Glu Gln Tyr Val Ile Asp

2101  GAG TAC AAC AAG TTG AAA ACT CGC CTG CGC GAA GCG GCA CTG GCG TAA TACCGTCCTT TCTACAGCCT CCTTTCGGAG GCTGTTTTTT TAT
 701  Glu Tyr Asn Lys Leu Lys Thr Arg Leu Arg Glu Ala Ala Leu Ala ***

2192  CCATTCG AACTCTTTAT ACTGGTTACT TCCCG
```

FIG. 2. Nucleotide sequence and derived amino acid sequence of the *fdhF* gene from *E. coli* MC4100. The sequence of the noncoding, RNA-like strand is given in the 5′ to 3′ direction, the numbering starting with the adenosine of the putative ATG initiation codon. The putative Shine–Dalgarno sequence is underlined, the in-frame UGA codon is boxed (positions 418–420), and the termination codon TAA (positions 2146–2148) is indicated by asterisks. Palindromes or inverted repeat sequences are indicated by arrows above the sequence. Vertical arrows denote the fusion points of the N-terminal fragments of the *fdhF* gene with *lacZ* carried by vector pMC1403 resulting in the *fdhF*::*lacZ* gene fusions of plasmids pBN2, pFM52, and pFM54.

cysteinyl-tRNA acts as a suppressor of UGA. Although this is an attractive possibility, it provokes a number of questions such as how the level of selenylation of this aminoacyl-tRNA is quantitatively controlled or why selenocysteine is not incorporated into other protein chains containing a translated stop codon (14). The occurrence of UGA in the middle of a short inverted repeat sequence (Fig. 2) might account for the specificity with which this putative suppression takes place.

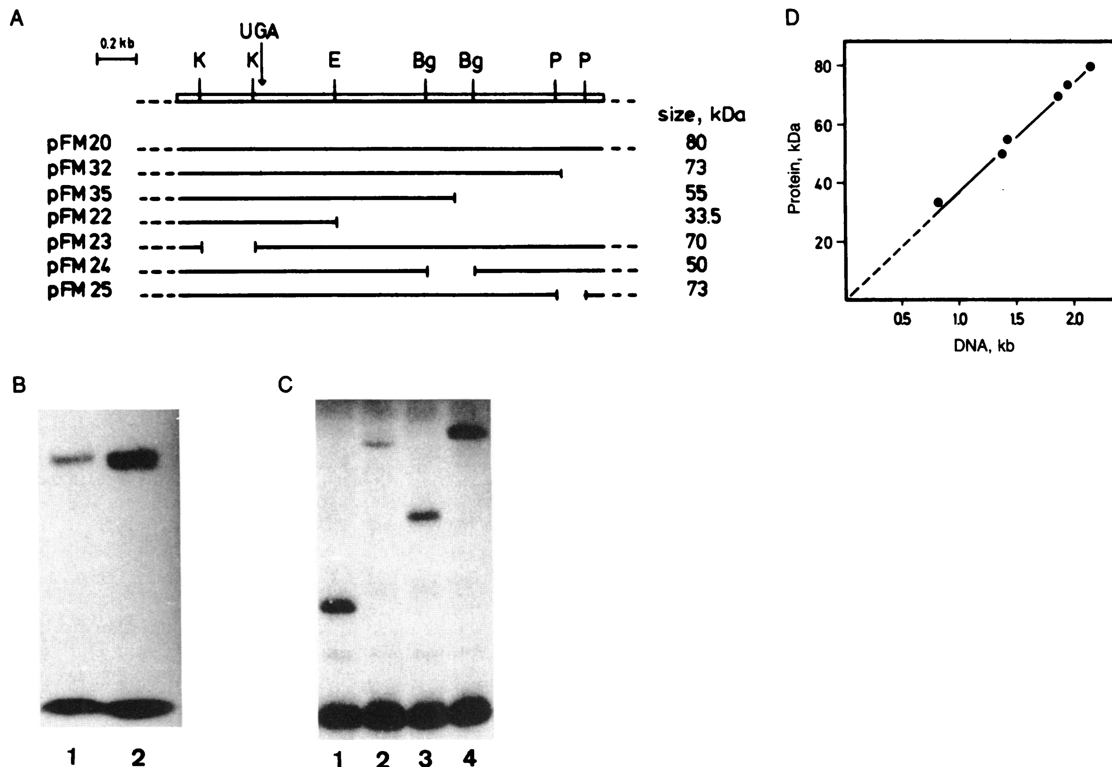

FIG. 3. Plasmid-encoded expression of wild-type and truncated selenopolypeptides. (*A*) Restriction maps of *in vitro*-constructed *fdhF* deletions. Horizontal lines give the lengths of the DNA inserts. Plasmids pFM32 and pFM35 contain nuclease BAL-31-generated terminal deletions of *fdhF*; the endpoints of the deletions were determined by DNA sequence analysis. Plasmids pFM22, pFM23, pFM24, and pFM25 contain *fdhF* genes in which removal of the internal *Eco*RV-*Xba* I, *Kpn* I, *Bgl* II, and *Pvu* I fragments, respectively, has generated internal deletions that are in-frame in the case of the *Kpn* I-deleted fragment and cause frameshifts and premature termination in the cases of the *Eco*RV-*Xba* I, *Bgl* II, and *Pvu* I deletions. The position of the UGA is indicated. The sizes of the truncated selenopolypeptides are given on the right-hand side. (*B*) Autoradiograph of ^{75}Se-labeled NaDodSO$_4$-cell lysates separated by NaDodSO$_4$/polyacrylamide gel electrohoresis. Lane 1: MC4100 *recA56*; lane 2: FM911/pFM20. (*C*) Autoradiograph of NaDodSO$_4$ lysates from ^{75}Se-labeled cells of strain FM911 with plasmids pFM22 (lane 1), pFM23 (lane 2), pFM24 (lane 3), and pFM25 (lane 4), respectively. Strain FM911 does not deliver any Se-labeled band (not shown). (*D*) Relation between the size of the open reading frame of the *fdhF* deletions and the size of the selenopolypeptides synthesized.

It is important to stress in this context that any shift of the reading frame at the site of the UGA codon would result in the creation of frequent termination codons in downstream positions. Frameshift suppression that allows translation of the termination codon would not allow completion of *fdhF* mRNA translation as it does in the translation of release factor 2 (14).

Evidence supporting the notion that selenocysteine is inserted co-translationally by suppression of the opal nonsense codon comes from recent experiments in which the effect of selenium deficiency in the medium was investigated on expression of the β-galactosidase fusion proteins. Plasmid pBN2, in which the N-terminal 39 amino acids of FDH$_H$ are fused to β-galactosidase, expresses identical levels of β-galactosidase activity under anaerobic conditions in the presence or absence of selenium. This demonstrates that transcription of the *fdhF* gene is not controlled by the availability of selenium. However, when plasmids pFM52 or pFM54 were analyzed (which code for *lacZ* fusions with N-terminal parts of *fdhF* including the UGA codon; Fig. 2), the expression of the fusion protein (determined as β-galactosidase activity) in the absence of selenium was only 10% of the value measured for cells grown with added selenite present (A.B. and F.Z., unpublished results). Selenylation of the *fdhF* part of the fusion protein, therefore, may be required for translation of the downstream *lacZ* part of the gene fusion. Thus, in the absence of the appropriate selenocysteinyl-tRNA, translation of the *fdhF* gene would cease at the in-frame nonsense codon at amino acid position 140.

The nature of the putative selenocysteinyl-tRNA involved in the postulated mechanism is not yet known. There could be a specific UGA-reading tRNA species accepting selenocysteine or a precursor thereof. Another possibility, in view of the fact that UGA is recognized by tRNA^{Trp} in other biological systems (15, 16), is the formation of a selenocysteine aminoacyl-tRNA^{Trp} by modification of a tryptophane residue esterified to tRNA^{Trp}. In such a reaction replacement of the indole group of the tryptophanyl-tRNA^{Trp} with -SeH would result in the formation of selenocysteinyl-tRNA^{Trp}. A mechanism of this type would afford an explanation for the failure to detect a tRNA species in *E. coli* that can be specifically aminoacylated with selenocysteine as donor (17). Although nonspecific esterification of cysteine tRNA with selenocysteine has been shown to occur in *E. coli* (17), under normal selenium/sulfur ratios this pathway would appear to be inadequate for essential selenoprotein biosynthesis since the sulfur amino acid is the preferred substrate for cysteine tRNA synthetase. In other studies the tRNA population of *E. coli* cells, labeled with ^{75}Se by growth in the presence of 1 μM $^{75}SeO_3^{2-}$ or $^{75}SeO_4^{2-}$, was examined for the presence of esterified [^{75}Se]selenocysteine (18). No labeled amino acid, rigorously identifiable as selenocysteine, was detected. Proof of the tRNA^{Trp} hypothesis requires the *in vitro* demonstration of the postulated selenylation mechanism together with identification of a selenocysteine residue in the polypeptide chain at the location specified by the UGA sequence position.

The *fdhF* sequence can be translated into a protein of M_r 79,087 that exactly matches the size previously determined for the selenoprotein of FDH$_H$ (3, 12). The N terminus does not bear significant resemblance to the consensus features of a leader peptide that might have explained the fact that part of FDH$_H$ is membrane associated (19). Lack of a signal

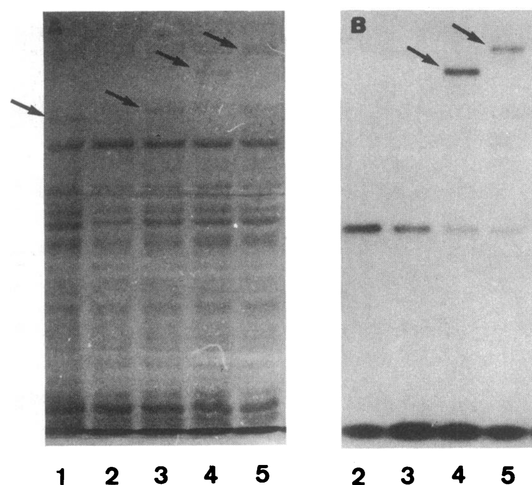

FIG. 4. Expression of *fdhF*::*lacZ* fusion proteins. Plasmids pMC1403 (vector) and pBN2, pFM52, and pFM54 were transformed into strain MC4100 *recA56*. Transformants were grown anaerobically in the presence of $Na_2{}^{75}SeO_3$, the cells were lysed in $NaDodSO_4$, and the proteins were separated by electrophoresis through 6% $NaDodSO_4$/polyacrylamide gels. (*A*) Electropherogram stained with Coomassie fast blue R250. (*B*) Autoradiograph of *A*. Lane 1, *E. coli* Hfr 3300 (*lacI22, lacZ*$^+$); lanes 2–5, transformants of MC4100 *recA56*-carrying plasmids pMC1403, pBN2, pFM52, and pFM54, respectively. The arrows in *A* denote the migration of β-galactosidase (lane 1) and of the respective fusion proteins (lanes 3–5). Note that MC4100 *recA56* carries an intact chromosomal *fdhF* gene that is expressed in all strains under anaerobic conditions. The radioactive bands at the front are seleno-tRNAs that contain 5-methylaminomethyl-2-selenouridine.

sequence is also indicated by the fact that the fusion protein encoded by pBN2 (Fig. 3) shows the size increment of β-galactosidase expected from the additional 39 amino acids. The open reading frame codes for a conspicuously high content of hydrophobic amino acids with more than 10% being alanine. It contains 15 cysteine residues, some of them (especially the N-terminal sequence *Cys*-Pro-Tyr-*Cys*-Ala-Ser-Gly-*Cys*) in a context resembling those of iron–sulfur proteins (20). No statistically significant homology to iron–sulfur proteins could be detected in a computer search, however.

The *fdhF* sequence is preceded by a consensus Shine–Dalgarno motif and followed by a classical ρ-independent transcription termination structure (see Fig. 2).

Altogether, the results of this study show that the nucleotide sequence coding for a selenocysteine-containing polypeptide contains an in-frame opal nonsense codon that is translated when selenium is present in the medium. Selenium is incorporated into truncated polypeptide chains or into fusion proteins that contain that part of the FDH_H polypeptide primary structure coded for by the gene sequence around the UGA. The results, therefore, suggest that selenium enters the protein chain co-translationally. A translational pathway for incorporation of selenocysteine has been postulated previously for glutathione peroxidase from rat liver solely on the basis of *in vivo* and *in vitro* isotope studies (21).

We thank Friedhelm Pfeiffer for the homology analysis. This work was supported by a grant from the Deutsche Forschungsgemeinschaft (SFB 145) to A.B.

1. Stadtman, T. C. (1980) *Annu. Rev. Biochem.* **49**, 93–110.
2. Stadtman, T. C. (1985) *Phosphorus Sulfur* **24**, 199–216.
3. Pecher, A., Zinoni, F. & Böck, A. (1985) *Arch. Microbiol.* **141**, 359–363.
4. Casadaban, M. J. & Cohen, S. N. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 4530–4533.
5. Chang, A. C. Y. & Cohen, S. N. (1978) *J. Bacteriol.* **134**, 1141–1156.
6. Casadaban, M. J., Chou, J. & Cohen, S. N. (1980) *J. Bacteriol.* **143**, 971–980.
7. Miller, J. H. (1972) *Experiments in Molecular Genetics* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
8. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
9. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
10. Gray, C. P., Sommer, R., Polke, C., Beck, E. & Schaller, H. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 50–53.
11. Mandrand-Berthelot, M.-A., Wee, M. Y. K. & Haddock, B. A. (1978) *FEMS Microbiol. Lett.* **4**, 37–40.
12. Cox, J. C., Edwards, E. S. & DeMoss, J. A. (1981) *J. Bacteriol.* **145**, 1317–1324.
13. Sancar, A., Hack, A. M. & Rupp, W. D. (1979) *J. Bacteriol.* **137**, 692–693.
14. Craigen, W. J., Cook, R. G., Tate, W. P. & Caskey, C. T. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 3616–3620.
15. Barrell, B. G., Bankier, A. T. & Drouin, J. (1979) *Nature (London)* **282**, 189–194.
16. Yamao, F., Muto, A., Kawauchi, Y., Iwami, M., Iwagami, S., Azumi, Y. & Osawa, S. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 2306–2309.
17. Young, P. A. & Kaiser, I. I. (1975) *Arch. Biochem. Biophys.* **171**, 483–489.
18. Wittwer, A. J. (1983) *J. Biol. Chem.* **258**, 8637–8641.
19. Gray, C. T., Wimpenny, J. W. T., Hughes, D. E. & Mossman, M. R. (1966) *Biochim. Biophys. Acta* **117**, 22–32.
20. Adam, E. T., Sieker, L. C. & Jensen, L. H. (1973) *J. Biol. Chem.* **248**, 3987–3996.
21. Hawkes, W. C. & Tappel, A. L. (1983) *Biochim. Biophys. Acta* **739**, 225–234.