JOURNAL OF BACTERIOLOGY, Jan. 1986, p. 181–192
0021-9193/86/010181-12$02.00/0

Vol. 165, No. 1

Protein Fusions of β-Galactosidase to the Ferrichrome-Iron Receptor of *Escherichia coli* K-12

JAMES W. COULTON,* PATRIZIA MASON, DAVID R. CAMERON, GILLES CARMEL, RICHARD JEAN, AND HAROLD N. RODE

Department of Microbiology and Immunology, McGill University, Montreal, Canada, H3A 2B4

Received 8 August 1985/Accepted 10 October 1985

The fusion-generating phage λ p*lac* Mu1 was used to produce fusions of *lacZ* to *fhuA*, the gene encoding the ferrichrome-iron receptor (FhuA protein) in the outer membrane of *Escherichia coli* K-12. Fusions to the *fhuA* gene in a Δ(*lac*) strain were selected by their resistance to bacteriophage φ80 *vir*. Ten independent (*fhuA'-'lacZ*) fusions were all Lac$^+$ and were resistant to the lethal agents which require the FhuA protein as receptor, i.e., φ80 *vir*, T5, T1, UC-1, and colicin M; none could utilize ferrichrome as the sole iron source. Specialized transducing phages were obtained by illegitimate excision from the chromosome of each of the fusion-bearing strains, and *Eco*RI fragments which encoded the fusions were subcloned into the high-copy plasmid pMLB524. Physical mapping of the fusion-containing plasmids confirmed the presence of three restriction sites which were also located on the chromosomal DNA of sequences near the *fhuA* gene. The direction of transcription of the *fhuA* gene was deduced from the direction of transcription of the (*fhuA'-'lacZ*) gene fusion. Identification of the chimeric proteins was made by both radiolabeling cells and immunoprecipitating the LacZ-containing proteins with antibody to β-galactosidase and by preparing whole cell extracts from Lac$^+$ cells containing the cloned gene fusions. Two sizes of (FhuA'-'LacZ) proteins were detected, 121 kDa and 124 kDa. The DNA sequences at the unique fusion joints were determined. The sequence information allowed us to identify three distinct fusion joints which were grouped as follows, type I fusions, 5'-ACT GCT CAG CCA A-3'; type IIa fusions, 5'-GCG GTT GAA CCG A-3'; and type IIb fusions: 5'-ACC GCT GCA CCT G-3'. To orient these *fhuA* fusion joints, the complete nucleotide sequence of the *fhuA* gene was determined from a 2,902-base-pair fragment of DNA. A single open reading frame was found which translated into a 747-amino acid polypeptide. The signal sequence of 33 amino acids was followed by a mature protein with a molecular weight of 78,992. Alignment of the amino acid sequence of the FhuA protein with the amino acid sequences presented for two other *tonB*-dependent receptor proteins in the outer membrane of *E. coli* showed an area of local homology at the amino terminus of all three proteins.

The uptake of ferrichrome-iron into *Escherichia coli* K-12 is dependent upon the binding of the ligand to a receptor in the outer membrane termed the FhuA protein (25). Transport of iron into the cell is also dependent upon other genetically defined loci, including *fhuC*, *fhuD*, *fhuB* (9, 17), *exbB* (15), and *tonB* (3). The ferrichrome-promoted iron transport system is of interest for a variety of reasons. Bacteriophages T1, T5, φ80, and UC-1, colicin M, and albomycin, which is a structural analogue of ferrichrome, all bind to the FhuA protein. These ligands are structurally unrelated and yet they bind competitively with each other and with ferrichrome (14). In addition, there may be some physical interaction between the FhuA protein and the *tonB* gene product (29) which is responsible for the energy coupling of the transport mechanism (11). Ferric hydroxamate uptake in *E. coli* is a high-affinity transport system and thus differs from the porin-mediated diffusion of a solute across the outer membrane (21, 26) because of its substrate specificity and its requirement for energy. Comparison of the structure of the receptor proteins of the outer membrane with the structure of the porins may identify the common features of these two different systems.

The *fhuA* gene was recently cloned in our laboratory onto a high-copy plasmid containing the pBR322 replicon (7). The clone was isolated by a positive selection, complementation of an *fhuA* mutant to growth in the presence of ferrichrome as the sole iron source. The location of the gene was determined by subcloning. Expression studies in maxicells identified the FhuA protein with a molecular weight of 78,000.

In an attempt to understand the variety of functions of the FhuA protein, β-galactosidase fusions to the *fhuA* gene were constructed by using the transposable fusion-generating phage λ p*lac* Mu1 (4). A selected number of chimeric proteins, (FhuA'-'LacZ), were characterized. The *fhuA* promoter in the fusion-containing strains was found to be regulated by the iron supply. In addition, the nucleotide sequence at the *fhuA* fusion joints was determined, and this sequence information was superimposed upon the complete nucleotide sequence of the *fhuA* gene.

MATERIALS AND METHODS

Bacterial strains and plasmids. *Escherichia coli* K-12 strains and plasmids that were used are listed in Table 1. Plasmid pPM27 was isolated as a subclone of pPM23 (7) and carried the *fhuA* gene on a 4.4-kilobase-pair (kbp) *Sal*I-*Sal*I insert of chromosomal DNA into the *Sal*I site of pBR322.

Media, reagents, and enzymes. All liquid media have been previously described (22, 34). Antibiotics added to media were ampicillin (125 μg/ml), tetracycline (25 μg/ml), and chloramphenicol (25 μg/ml). The indicator added to plates to detect β-galactosidase was 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-Gal). Measurements of the specific activity of β-galactosidase in whole cell extracts were done following

* Corresponding author.

TABLE 1. List of *Escherichia coli* K-12 strains and plasmids

Strain or plasmid	Genotype[a]	Source or reference
Strain		
MC4100	F⁻ *araD* Δ(*argF-lac*)*U169 rspL thi relA flbB deoC pstF rbsR*	T. J. Silhavy
SG303	MC4100 *aroB*	S. Garrett
MC1000	F⁻ *araD* Δ(*araABC-leu*)*7679 galU galK* Δ(*lac*)*X74 rspL thi*	T. J. Silhavy
SG404	F'141/MC4100 *asd* (P1::Tn*9 clr*100)	S. Garrett
JM103	Δ(*lac pro*) *thi str supE endA sbcB hsdR* F'*traD proAB lacI*^{q}ZΔ*M15*	J. Messing
L7	MC4100 (*malK'-'lacZ*)	R. Saunders
L1	MC4100 (*malT'-'lacZ*)	R. Saunders
DRC102 to DRC123	MC4100 (*fhuA'-'lacZ*)	This study
DRC202 to DRC223	SG303 (*fhuA'-'lacZ*)	This study
Plasmids		
pPM18	*bla fhuA* pBR322 replicon, 13.9 kb	7
pPM23	*bla fhuA* pBR322 replicon, 9.6 kb	7
pPM27	*bla fhuA* pBR322 replicon, 8.1 kb	7
pMLB524	*bla 'lacZ* pBR322 replicon, 3251 bp	M. L. Berman
pMLB1104	*bla lacI*^{q} *lacZ*, 7.6 kb	M. L. Berman
pDRC102 to pDRC123	*bla* pMLB524::(*fhuA'-'lacZ*)	This study

[a] Kb, Kilobases; bp, base pairs.

the procedure of Miller (24), using *o*-nitro-phenyl-β-D-galactoside as substrate. Restriction endonucleases and DNA ligase were purchased from Bethesda Research Laboratories and from Boehringer-Mannheim Canada. The Klenow fragment of DNA polymerase, the 17-mer primer, and deoxyadenosine 5'(^{35}S)thiophosphate were obtained from Amersham Canada Ltd.

Construction of β-galactosidase fusions to the *fhuA* gene. The plaque-forming lambda phage, λ *plac* Mu1, was employed to isolate LacZ protein fusions in vivo in a single-step infection (4). MC4100 Δ(*lac*) (1 ml) was coinfected with λ *plac* Mu1 and the helper phage λ pMu507, both at 10^9 PFU/ml. Cells were plated on minimal lactose plates and incubated at 37°C for 48 h. Independent pools of about 3,000 Lac$^+$ transductants were replicated onto L plates that had been seeded with 10^{10} PFU of φ80 *vir*. Selection for resistance to φ80 *vir* was at 37°C for 6 h. Fusion candidates (*fhuA'-'lacZ*) were picked, tested for their sensitivity to φ80 *vir*, λ *vir*, BF23, and λ *c b2*, and for their phenotype on minimal lactose and on MacConkey-lactose plates.

The gene fusions were transduced into a clean genetic background. Lysogens of fusion-containing strains were constructed from a lysate of P1::Tn*9 clr*100 (34). Generalized P1 transduction into MC4100 and into SG303 followed standard genetic protocols (34) with selection for Lac$^+$. Only Lac$^+$ φ80 *vir*-resistant cotransductants were retained.

Specialized lambda transducing phages were isolated by illegitimate excision from fusion-containing strains of *E. coli*. The (*fhuA'-'lacZ*) lysogens were stabbed into a lawn of MC4100 and then exposed to an inducing dose of UV irradiation. After overnight incubation, zones of lysis appeared at the stab and contained lambda phages, some of which had packaged the gene fusion. Single plaques were isolated and purified by plating on MC4100 in the presence of X-Gal. Plaques which stained blue with X-Gal contained the gene fusion and were grown to high titer on the same indicator strain.

Identification of fusion proteins. The following two complementary approaches were employed to identify the fusion proteins and to estimate the molecular size of the hybrids: (i) Sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis of whole cell extracts of Lac$^+$ fusions. Cells were grown to late logarithmic phase, harvested, and suspended in electrophoresis sample buffer (0.0625 M Tris hydrochloride [pH 6.8], 2% [wt/vol] SDS, 5% [vol/vol] 2-mercaptoethanol). Samples were boiled for 5 min, centrifuged in a microfuge to pellet the residue, and loaded onto an 8% polyacrylamide slab gel by using the system of Lugtenberg et al. (20). (ii) Immunoprecipitation of radiolabeled hybrid proteins. Cultures of fusion-containing strains were grown in M63 glycerol, and 1.0 ml of these cells was radiolabeled with (^{35}S)methionine (50 μCi) at 37°C for 2 min. A 0.2-ml amount of a suspension (A_{578} of 100) of nonradiolabeled carrier MC4100 in 10 mM Tris hydrochloride (pH 7.5) at 0°C was then added to these bacteria. Bacteria were collected by centrifugation at 13,000 × *g*, the supernatant was discarded, and the pellet was lysed by suspension in 200 μl of buffer containing 50 mM Tris hydrochloride (pH 7.5), 10 mM EDTA, and 10 μg each of lysozyme, DNase, RNase, and phenylmethyl sulfonyl fluoride per ml (all from Sigma Chemical Co.). The pellets were lysed by three cycles of freeze and thawing in a dry ice-alcohol bath. $MgCl_2$ was added to 10 mM, enabling DNase to reduce the viscosity of the suspension. Triton X-100 to 2% (vol/vol) and EDTA to 10 mM were added to the mixture, followed by an equal volume of immune precipitation buffer (50 mM Tris hydrochloride [pH 7.5], 2.0 M NaCl). After ultracentrifugation of the sample at 150,000 × *g* for 30 min at 4°C, an excess of rabbit anti-β-galactosidase antiserum was added. Immunoprecipitation was overnight at 8°C. The precipitates were collected by adding formalinized, heat-fixed *Staphylococcus aureus* (Cowan strain) and washed with: (i) 50 mM Tris hydrochloride (pH 7.5)–1% Triton X-100–1.0 M NaCl; (ii) 50 mM Tris hydrochloride (pH 7.5)–0.1% SDS–0.5 M LiCl; and (iii) 50 mM Tris hydrochloride (pH 7.5). The pellets were suspended in electrophoresis sample buffer, boiled for 5 min, and centrifuged to pellet the insoluble residue. Samples of the supernatant were electrophoresed on an 8% SDS-polyacrylamide slab gel. The gel was treated with En^{3}Hance before fluorography at −70°C.

DNA sequence determination. Plasmid pPM27 was prepared on a large scale from cleared lysates by banding in a cesium chloride-ethidium bromide gradient (22). DNA fragments were isolated by electrophoresis through agarose gels, electroelution, phenol extraction, and ethanol precipitation. Two tandem fragments containing the entire *fhuA* gene were prepared, *Bgl*I-*Sal*I, 1.80 kbp and *Sal*I-*Sma*I, 1.10 kbp. DNA sequence analysis was performed by the enzymatic method of Sanger (30) in conjunction with M13 phage derivatives and the universal primer, a synthetic 17-mer. Deoxyadenosine 5'(^{35}S)thiophosphate and both linear and buffer gradient gels were employed to maximize the data generated by each gel. The sequencing strategy was to isolate *Sau*3A fragments, *Hpa*II fragments, *Taq*I fragments, or *Rsa*I fragments from the DNA containing the *fhuA* gene and to clone selectively these isolated fragments into the unique restriction sites

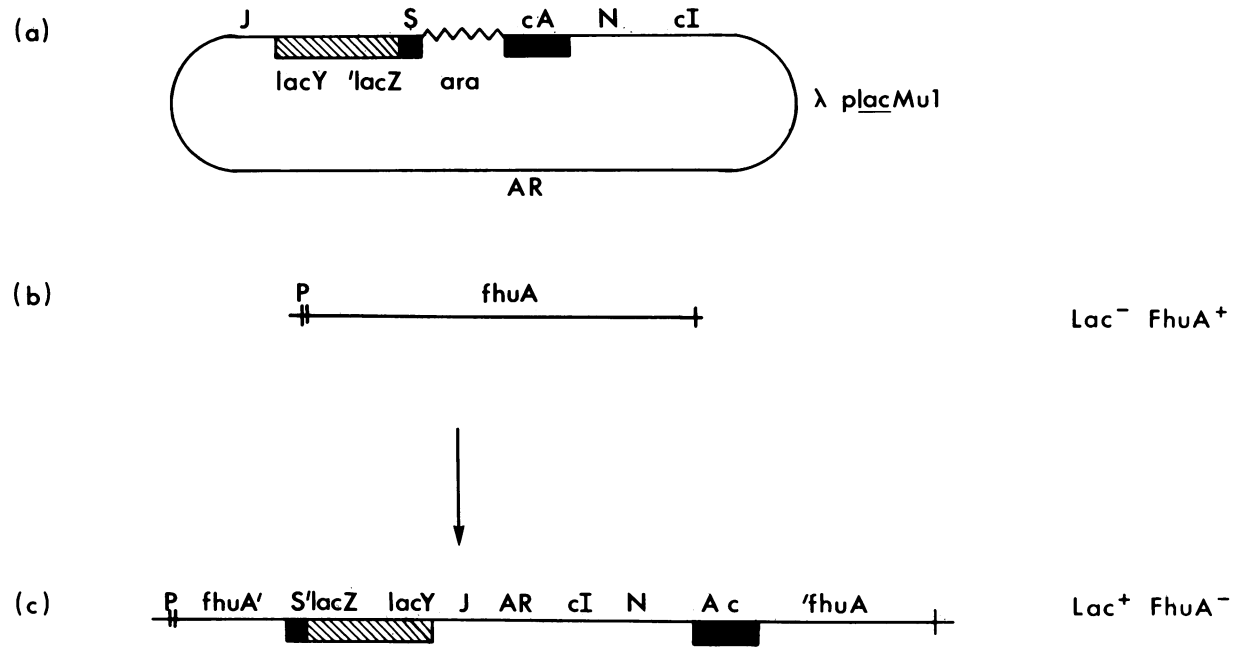


FIG. 1. Construction of β-galactosidase gene fusions to the *fhuA* gene of *E. coli* K-12. The fusion-generating phage λ p*lac* Mu1 carries the *S* and *c* regions of Mu. Upon introduction of phage DNA into the cell, the DNA circularizes (a) and may insert into the target *fhuA* gene of the bacterium (b). If the insertion of the *lac* genes are in the proper orientation and the correct reading frame, a gene fusion is generated which specifies a chimeric protein. The organization of the gene fusion (*fhuA'-'lacZ*) on the bacterial chromosome is shown in (c). Relevant phenotypes are listed at the right.

(*Bam*HI, *Acc*I, *Acc*I, and *Sma*I, respectively) of M13mp8 or M13mp9 (23). Sequence analysis was performed with the aid of the computer program TRANMT which was obtained from the Institut de Recherches Cliniques de Montreal, Canada.

Amino acid sequence analysis. FhuA protein was obtained from *E. coli* PL-6(pPM18) by solubilization of isolated outer membranes at 60°C for 30 min in electrophoresis sample buffer, by electrophoresis on an SDS-polyacrylamide slab gel, and by electroelution of the Coomassie blue-stained protein (6). A preparation of 170 μg of protein was subjected to amino acid sequence determination in an Applied Biosystems gas-phase sequencer. The phenylthiohydantoin derivatives of the amino acids were identified by high-pressure liquid chromatography.

RESULTS

Construction and expression of (*fhuA'-'lacZ*) hybrid genes. The construction of gene fusions to *fhuA* required the use of λ p*lac*Mu, a transposable derivative of bacteriophage lambda (4). This phage carries 117 base pairs from the *S* end of Mu and about 2.8 kbp from the *c* end of Mu. The Mu attachment sites enable the phage to integrate into the chromosome when transposition functions for Mu are provided in *trans*. Since loss of the outer membrane FhuA protein confers resistance to bacteriophage φ80 *vir*, selection for insertion into the *fhuA* gene was made by screening the pool of Lac$^+$ transductants for resistance to this phage. Figure 1 is the construction of (*fhuA'-'lacZ*) gene fusions with λ p*lac* Mu1. To ensure that independent fusions were isolated, only one fusion candidate per plate of 3,000 Lac$^+$ transductants was picked. Each fusion-containing strain was cross-streaked against four other phages: immunity to λ *c b2* verified the presence of the lambda lysogen; λ *vir* sensitivity indicated that the LamB function was unaltered; sensitivity to BF23 demonstrated the presence of the BtuB receptor; and hy2 sensitivity confirmed that the OmpC porin acted as a receptor for this phage. Independently isolated fusions were selected and transduced into MC4100 with P1::Tn*9* *clr*100. From 100,000 Lac$^+$ colonies, 10 transductants, numbered between DRC102 and DRC123, displayed phenotypes that might be expected of a (*fhuA'-'lacZ*) fusion. Each gene fusion was also transduced into SG303; this *aroB* derivative of MC4100 cannot synthesize enterochelin. The *aroB* transductants, numbered between DRC202 and DRC223, could then be tested for their ability to utilize ferrichrome as the sole iron source. Growth promotion by ferrichrome (8) on nutrient broth plates containing 40 μM EDDA [ethylenediamine-di(*o*-hydroxyphenyl acetic acid)] showed that none of the fusion strains could utilize this exogenously added source of iron.

To study the expression of the chromosomally encoded hybrid gene, DRC102 was grown in L broth supplemented with 40 μM EDDA. Although this concentration of the iron-specific chelator caused no apparent limitation of growth compared with cells grown in the absence of EDDA, the level of β-galactosidase increased from a basal level of 105 U to 1,052 U over 3 h. This observation suggested that the activity of the *fhuA* promoter was regulated by the iron supply in the fusion-containing strain. The other nine strains of chromosomal (*fhuA'-'lacZ*) fusions responded similarly to iron deprivation by increased levels of β-galactosidase, although the amount of induction was always lower for DRC109, DRC111, and DRC120 (Table 2). No increase in β-galactosidase activity was found when the iron-specific chelator was added to *E. coli* L7, an (*malK'-'lacZ*) fusion, or to strain L1, an (*malT'-'lacZ*) fusion.

Identification and localization of the hybrid proteins. To identify the (FhuA'-'LacZ) hybrid proteins, the fusion-

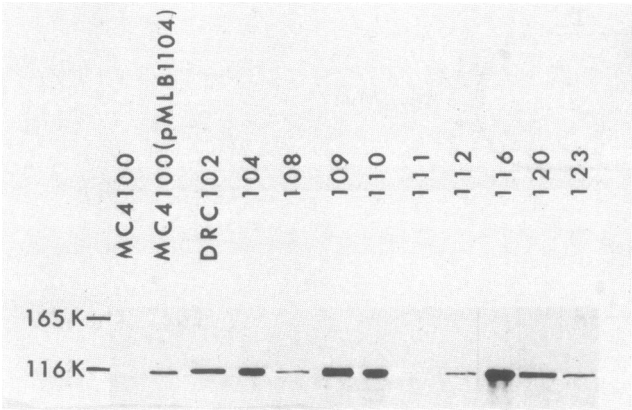


FIG. 2. Identification of chimeric proteins (FhuA'-'LacZ). Chromosomal fusion strains DRC102 to DRC121 were radiolabeled with 50 μCi of (^{35}S)methionine at 37°C for 2 min, chased with an excess of carrier cells, and lysed by repeated freeze-thawing. Rabbit antibody against β-galactosidase was added to immunoprecipitate LacZ-containing fusion proteins. The relevant section of the gel is shown: MC4100 Δ(lac) shows no immunoprecipitate; MC4100 (pMLB1104) identified the monomer of β-galactosidase with an M_r of 116,000; hybrid proteins with an estimated M_r of 121,000 were detected for each of fusion strains DRC102 to DRC121. The size marker of 165 KDa is the position of the β' subunit of RNA polymerase.

containing strains were radiolabeled with (^{35}S)methionine, converted to spheroplasts, and lysed by freeze-thawing. After incubation with anti-β-galactosidase antibody, the immunoprecipitates were analyzed by SDS-polyacrylamide slab gel electrophoresis (Fig. 2). No immunoprecipitate was detected from a lysate of MC4100 Δ(*lac*). MC4100 harboring pMLB1104 (*lacZ*$^+$) was reacted with anti-β-galactosidase antibody; a single immunoprecipitate was observed with molecular weight 116,000. Hybrid proteins of (FhuA'-'LacZ) with a molecular weight greater than 116,000 were detected by fluorography. The amounts of hybrid protein which were immunoprecipitated from a fixed number of bacteria (2×10^{10} cells) varied considerably, as judged by the intensity of the bands on the autoradiogram. The sizes of the 10 hybrid proteins were all estimated as 121,000.

In addition to the identification of the fusion protein, it was necessary to confirm the absence of wild-type FhuA protein in the fusion-containing strains. Outer membranes were collected from MC4100, MC4100 *fhuA*, MC4100(pPM18), and from the fusion strains DRC102 to DRC123 by a rapid isolation procedure (12, 36), and the proteins were electrophoresed. An FhuA protein with a molecular weight of 78,000 was not observed in the fusion strains (data not shown), although identification was difficult because of the very closely migrating FhuE protein (13). For verification of the absence of FhuA protein, we relied upon the use of a rat monoclonal antibody, 4AA-1, which was raised against FhuA protein and which reacted strongly with FhuA protein upon Western blot analysis (unpublished results). Cell extracts were prepared from the above-listed strains, solubilized at 60°C for 30 min in sample buffer, electrophoresed, blotted onto nitrocellulose paper, probed with rat anti-FhuA monoclonal antibody, and reacted with the developing antibody, and anti-rat immunoglobulin conjugated to peroxidase. Wild-type FhuA protein was identified in MC4100; FhuA protein was not found in the mutant MC4100 *fhuA*; and MC4100(pPM18) containing the cloned *fhuA* gene showed a doublet, a major band with a molecular weight of 78,000 and a minor band of 81,000. When extracts from the fusion-containing strains were electrophoresed and used as antigen for reactivity with monoclonal 4AA-1 antibody, no FhuA protein was identified on the Western blots (gel not shown). Taken together with the results of the immunoprecipitation of the hybrid proteins with anti-β-galactosidase antibody, we concluded that the gene fusions constructed in vivo were transcribed and translated into bona fide (FhuA'-'LacZ) chimeric proteins.

Fractionation of the fusion strains into subcellular compartments was undertaken to localize the hybrid proteins encoded by the chromosomal (*fhuA'-'lacZ*) genes. Total cell envelopes (outer membranes plus inner membranes) were prepared by lysis of spheroplasts and by ultracentrifugation of the lysate at 150,000 × *g*. The specific activities of β-galactosidase in the hybrid proteins were determined in the membrane pellet prepared at 150,000 × *g*, and these measurements were compared to the enzymatic activities for the chimeric proteins found in the high-speed supernatant. The mean values from triplicate experiments are presented in Table 2. The cytoplasmic enzyme β-galactosidase encoded by pMLB1104 was used for reference; its 15% of total activity which was found in the cell envelope fraction could be considered as nonspecific binding to membranes. Whereas seven of the fusion strains consistently showed from 20 to 28% of the total β-galactosidase activity in the pellet of membranes, three strains (DRC109, DRC111, and DRC120) gave between 43 and 58% of the total activity in the cell envelope fraction.

Subcloning of the (*fhuA'-'lacZ*) gene fusions. Identification of some protein fusions to β-galactosidase has been successful with Coomassie blue staining of whole cell extracts (10). This depends upon (i) the paucity of proteins which migrate on an SDS-polyacrylamide slab gel in the molecular weight range 110,000 to 170,000; and (ii) a high level of expression of

TABLE 2. β-galactosidase assays and cellular fractionation of hybrid proteins generated by λ p*lac* Mu1 insertions into the *fhuA* gene of *E. coli* K-12

Strain	β-Galactosidase activity (Miller U)		% Total activity in cell envelope fraction
	Without EDDA	With EDDA	
MC4100	6	6	0
MC4100(pMLB1104)	28	52	15
DRC102	105	1,052	24
MC1000(pDRC102)	11,435		
DRC104	95	966	20
MC1000(pDRC104)	11,696		
DRC108	89	1,141	23
MC1000(pDRC108)	12,385		
DRC109	77	391	48
MC1000(pDRC109)	9,173		
DRC110	123	1,307	22
MC1000(pDRC110)	12,078		
DRC111	81	442	58
MC1000(pDRC111)	6,640		
DRC112	111	1,230	21
MC1000(pDRC112)	10,858		
DRC116	137	1,265	28
MC1000(pDRC116)	12,258		
DRC120	36	211	43
MC1000(pDRC120)	12,941		
DRC123	110	1,274	24
MC1000(pDRC123)	16,133		

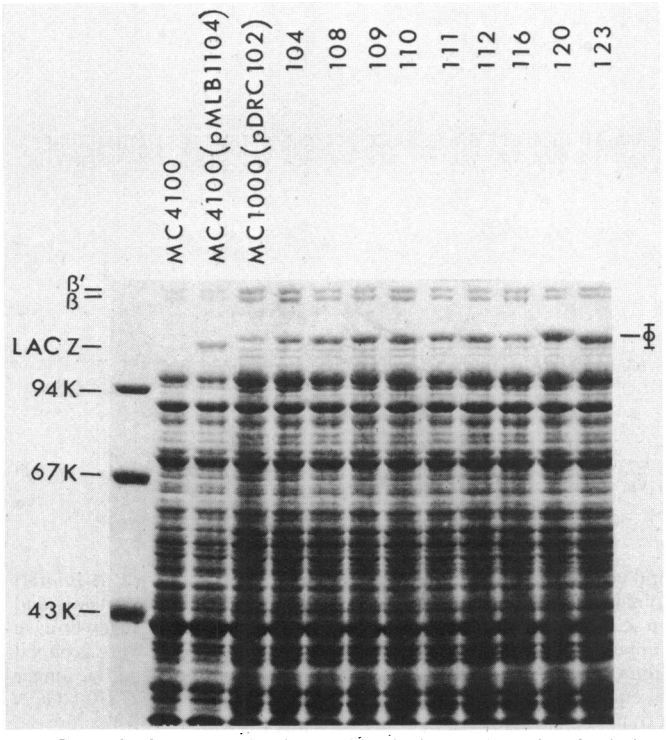


FIG. 3. SDS-polyacrylamide slab gel electrophoresis of whole cell extracts of cells containing the (*fhuA'-'lacZ*) gene cloned into the high-copy plasmid pMLB524. MC1000 was transformed with the Lac$^+$ plasmids pDRC102 to pDRC123. Cell extracts were boiled in electrophoresis sample buffer for 5 min and loaded onto an 8% SDS-polyacrylamide slab gel. The marker proteins were β' and β subunits of RNA polymerase (165 and 155 KDa, respectively), LacZ (116 KDa) from MC4100(pMLB1104), phosphorylase *b* (94 KDa), bovine serum albumin (67 KDa), and ovalbumin (43 KDa).

the chimeric protein from the promoter of the individual gene. Because the expression of the (FhuA'-'LacZ) protein fusion was too low to detect a band by Coomassie staining of an SDS-polyacrylamide slab gel, each of the ten gene fusions was cloned onto the high-copy plasmid pMLB524 (33, 34). β-Galactosidase activities from the plasmid-bearing strains were elevated about 1,000-fold compared with the chromosomal expression (Table 2). These high levels meant that the fusion proteins could be identified by SDS-polyacrylamide slab gel electrophoresis of whole cell extracts and Coomassie blue staining of the gel. Two sizes of the cloned fusion proteins were detected: (i) a 121-KDa protein encoded by pDRC102, pDRC104, pDRC108, pDRC110, pDRC112, pDRC116, and pDRC123; and (ii) a 124-KDa protein from pDRC109, pDRC111, and pDRC120 (Fig. 3). The latter size of protein was not found when the chromosome (FhuA'-'LacZ) fusions were immunoprecipitated (cf. Fig. 2).

Nucleotide sequence of the fusion joints. The subcloning of fusion genes into pMLB524 provided sufficient amounts of DNA for restriction mapping and for sequence determination of the fusion joints. Physical mapping of the pDRC100 series of plasmids showed that the *Eco*RI inserts into pMLB524 were about 10 kbp of DNA and contained both phage DNA and bacterial DNA. The amount of DNA encoding the *lacZ* fragment was 3.0 kbp. In addition, the restriction sites for *Pst*I, *Bgl*II, and *Sal*I were clustered within a 0.2-kbp region in some of these plasmids and could be aligned with the restriction map of these same three enzymes on plasmid pPM23. The *Pst*I, *Bgl*II, and *Sal*I sites were estimated to be 1.4 kbp from the fusion joint (Fig. 4). Moreover, to generate a Lac$^+$ phenotype, the transcriptional orientation of the *fhuA* gene must be the same as the *lacZ* gene in pDRC102. The *Pst*I, *Bgl*II, and *Sal*I sites should therefore be positioned on the 5'-side of the *fhuA* promoter so that transcription proceeds in a rightward direction as shown in Fig. 4.

Determination of the DNA sequence at each of the fusion joints was facilitated by the presence of a *Bam*HI site found within the 117 base pairs of the *S* region of Mu (4). *Pst*I-*Bam*HI fragments from the pDRC100 series of plasmids were cloned into M13 mp9. Sequencing of the single-stranded phage derivatives began from the universal primer, proceeded through all 117 bases from the *S* region of Mu, and then entered the *fhuA* gene. Upon sequencing through the unique fusion joints of the 10 independent (*fhuA'-'lacZ*)

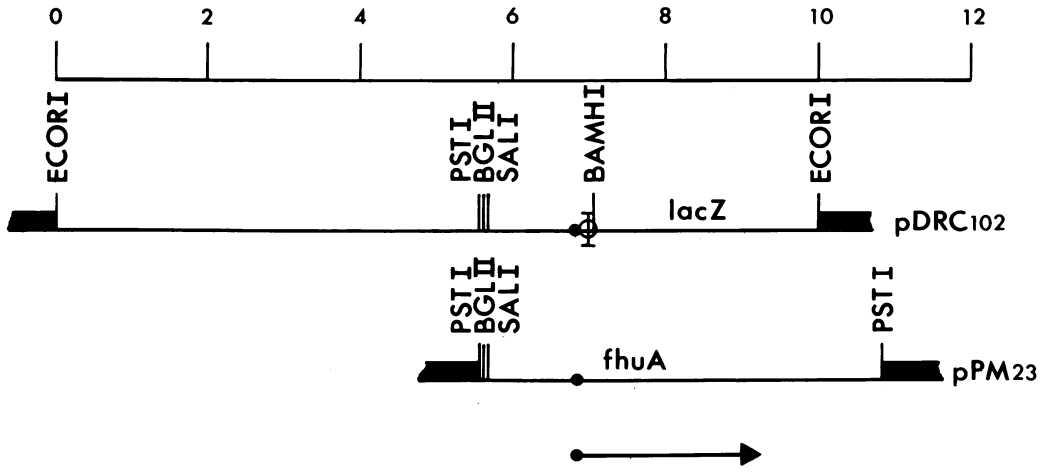


FIG. 4. Comparison of the physical map of the fusion-containing recombinant plasmid pDRC102 with pPM23. Plasmid pDRC102 contains a 10-kb pair *Eco*RI-*Eco*RI insert into pMLB524. The restriction sites for *Pst*I, *Bgl*II, and *Sal*I were mapped to a 0.2-kbp region of the insert, between 4.2- and 4.4-kbp from the right-hand *Eco*RI site. *lacZ* sequences are encoded in 3.0 kbp of DNA. The unique fusion joint between *fhuA* sequences and the *S* region of Mu is indicated (Φ). The restriction map of pPM23 shows the same three sites for *Pst*I, *Bgl*II, and *Sal*I. Approximate positions are indicated for the start of the (*fhuA'-'lacZ*) hybrid gene in pDRC102 (●) and for the start of the *fhuA* gene in pPM23 (●). The direction of transcription of the *fhuA* gene is also shown (→).

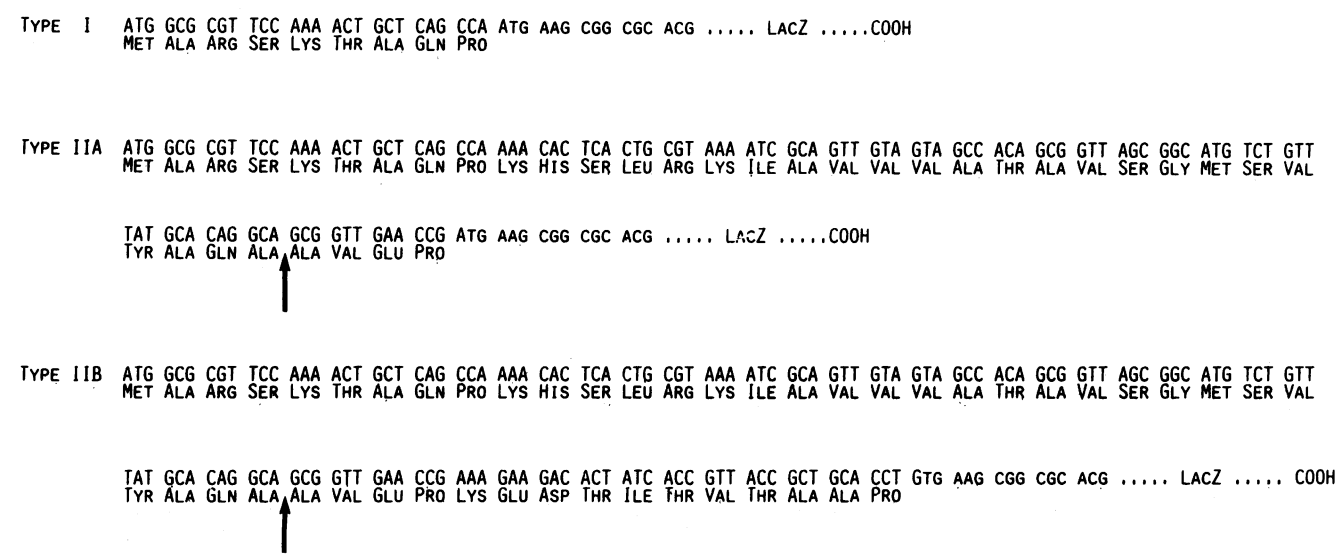


FIG. 5. DNA sequencing of the fusion joints of the (*fhuA'-'lacZ*) hybrid genes. From each of the pDRC series of plasmids, a *Pst*I-*Bam*HI fragment was subcloned into M13mp9. Sequence information was obtained by the method of Sanger, using primer extension and dideoxynucleotide chain termination. After proceeding through the known sequence of 117 bases from the *S* region of Mu shown in lower-case letters, the contiguous *fhuA* sequences were identified and are listed in upper-case letters. The sequences at the fusion joints were grouped as follows: type I, shortest fusions, 5'-ACT GCT CAG CCA A-3' as the *fhuA* sequences before the sequence of the *S* region; type IIa, longer fusions, 5'-GCG GTT GAA CCG A-3' as the *fhuA* sequences before the sequence of *S* region; type IIb, longest fusions, 5'-ACC GCT GCA CCT G-3' as the *fhuA* sequences before the *S* region of Mu. The reading frame for the translation of the chimeric protein is shown below the *fhuA* gene sequence and is taken from the complete nucleotide sequence. The vertical arrow indicates the site of cleavage by signal peptidase and therefore the beginning of the mature FhuA protein.

fusions, we found that the fusions could be grouped as follows: type I fusions contained the nucleotides 5'-ACT GCT CAG CCA A-3' as the *fhuA* sequence before entering the *S* region of Mu (5'-TG AAG CGG-3'). The fusion strains DRC102, DRC104, DRC108, DRC110, DRC112, DRC116, and DRC123 were identical with respect to their sequence and therefore their size of hybrid protein. The other three (*fhuA'-'lacZ*) strains were designated type II fusions and were characterized by longer stretches of *fhuA* sequences; even when an additional 60 nucleotides of the *fhuA* gene were read from the sequencing gels, type II fusions could not be overlapped with type I fusions. The nucleotide sequences of *fhuA* to the 5'-side of the fusion joints were: type IIa, DRC109 and DRC111, 5'-GCG GTT GAA CCG A-3' before the *S* region of Mu; type IIb, DRC120, 5'-ACC GCT GCA CCT G-3' before the *S* region of Mu. A summary of this sequence information is presented in Fig. 5.

Sequencing of the entire *fhuA* gene. Since the sizes of the hybrid (FhuA'-'LacZ) proteins were within a limited range of molecular weights, it was estimated that the gene fusions generated in vivo contained rather short segments at the 5'-end of the *fhuA* gene. To superimpose the nucleotide sequences from the fusion joints onto the nucleotide sequence of the *fhuA* gene, DNA sequencing of the entire gene was undertaken. With the limited number of restriction sites previously identified on pPM23, plasmid pPM27 was constructed for DNA sequence determination (Fig. 6A). The restriction map for pPM27 was expanded to locate several sites with hexanucleotide recognition sequences and to identify a minimum size of DNA fragment that would be suitable for sequencing. Since the distance from the *Pst*I-*Bgl*II-*Sal*I sites to the fusion joint was 1.4 kbp (Fig. 4) and preliminary estimates of the size of the hybrid proteins suggested that the start of the gene was 200 to 400 bases from the fusion joint, the *Bgl*I site was mapped at 500 base pairs to the 5' side of the *fhuA* regulatory region and transcription start.

The nucleotide sequence of the *Bgl*I-*Sma*I fragment carrying the *fhuA* gene was determined by isolating smaller restriction fragments and by cloning them into mp8 or mp9 derivatives of phage M13. Regions which were sequenced are illustrated in Fig. 6B.

The nucleotide sequence and the amino acid sequence which is predicted from its translation are presented in Fig. 7. Sequencing was confirmed on both strands of DNA with the exception of the region of 106 nucleotides, 1881 to 1987. In this region no uncertainty was apparent within the *Rsa*I fragment which was cloned into M13mp9. All other regions and sequences were fully overlapped. Of the six possible reading frames within the 2,902-base-pair fragment, only one maintained an open reading frame: it extended from nucleotides 540 to 2780.

There was only one potential ATG initiation codon which was preceded by the consensus sequence for the Pribnow box of bacterial promoters (TATAAT) (32) beginning at nucleotide 488. A potential Shine-Dalgarno sequence (AGAGAT) (31) from nucleotide 527 was also identified. The sequence ended with the nonsense codon TAA (nucleotide 2781) and was followed by a hairpin loop structure from nucleotides 2795 to 2814. The palindrome of eight nucleotides on either side of the stem, with a four-base loop, gave a structure with a net $\Delta G = -29.8$ Kcal.

It is proposed that the 2,241-base-pair open reading frame encodes the FhuA protein of 747 amino acids, including a signal sequence of 33 amino acids. The molecular weight of the proFhuA was calculated from the amino acid composition to be 82,202. Within the first 33 amino acids of the proFhuA, the consensus sequence for cleavage by leader peptidase (Ala-X-Ala) was predicted from the nucleotide sequence to appear twice: Ala_{21}-Thr_{22}-Ala_{23} and Ala_{31}-Glu_{32}-

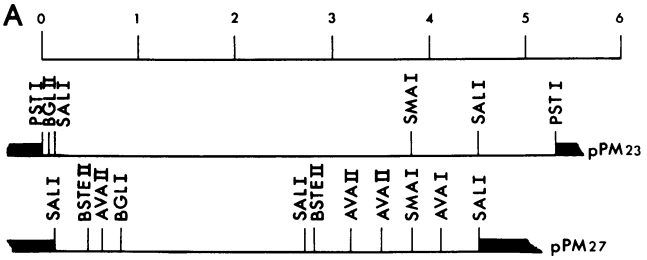
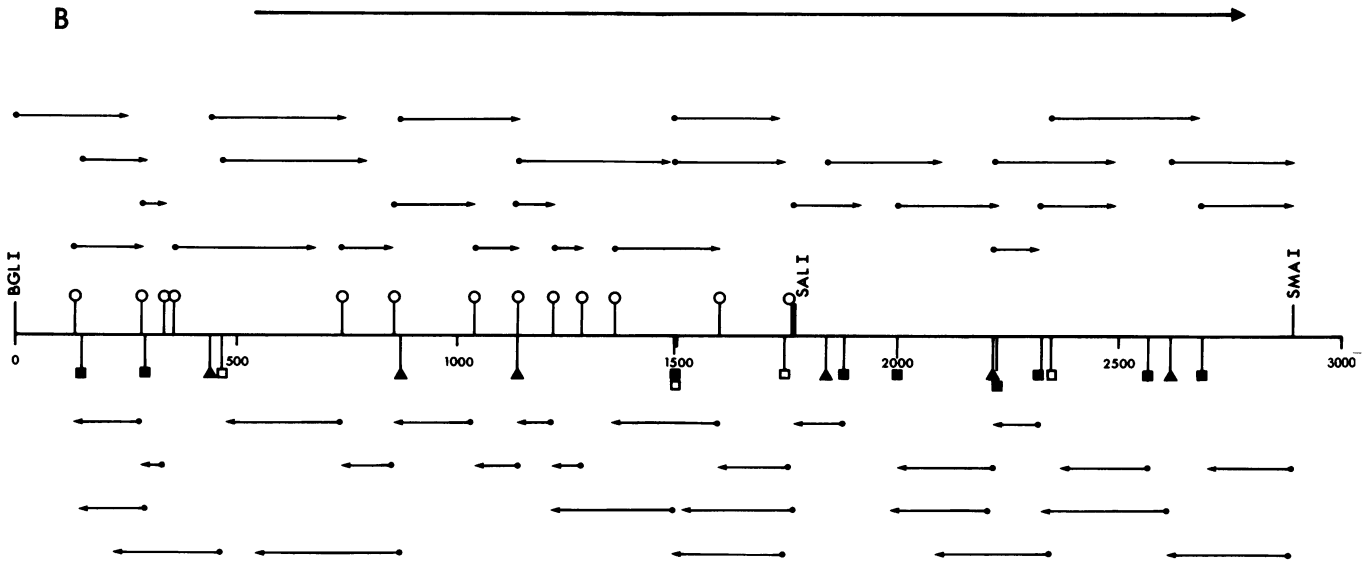


FIG. 6. (A) Restriction maps of the plasmids containing the *fhuA* gene. From plasmid pPM23, the sub-clone pPM27 was constructed as a 4.4-kbp *Sal*I-*Sal*I insert into the *Sal*I site of pBR322. An expanded physical map of pPM27 locates restriction sites for *Ava*I, *Ava*II, *Bgl*I, and *Bst*EII. The scale is in kilobase pairs. (B) Strategy for sequencing of the region encoding the ferrichrome-iron receptor. Plasmid pPM27 was used for determination of the DNA sequence. The *Bgl*I-*Sma*I fragment was sequenced by isolating smaller fragments and cloning them into M13 mp8 or M13 mp9. The restriction sites used for subcloning into M13 were *Sau*3A (■), *Hpa*II (○), *Taq*I (□), and *Rsa*I (▲). The short arrows indicate the point of initiation of the sequence information from an M13 clone, the length of the DNA strand which was sequenced, and the direction of reading. The arrow at the top of the figure shows the location and direction of the FhuA reading frame. The scale (0 to 3,000) is in base pairs.



Ala_{33}. To discriminate between these two possible cleavage sites, amino-terminal sequencing of the mature FhuA protein was undertaken. The protein was isolated from an overproducing strain containing the cloned *fhuA* gene on pPM18 by electroelution into a collodion bag. After exhaustive dialysis against 0.01% SDS, the protein was subjected to amino acid analysis by sequential Edman degradation and by identification of the derivatives on high-pressure liquid chromatography. The sequence of the first 14 amino acids of the mature FhuA protein was determined as Ala-Val-Glu-Pro-Lys-Glu-Asp-Thr-Ile-Thr-Val-Thr-Ala-Ala. These data therefore confirm the length of the signal sequence as 33 amino acids. The signal sequence was followed by a polypeptide with a molecular weight of 78,992. The latter value agrees well with the molecular weight of 78,000 for the size of the ferrichrome-iron receptor as estimated from molecular-weight calibration standards on SDS-polyacrylamide gel electrophoresis (3).

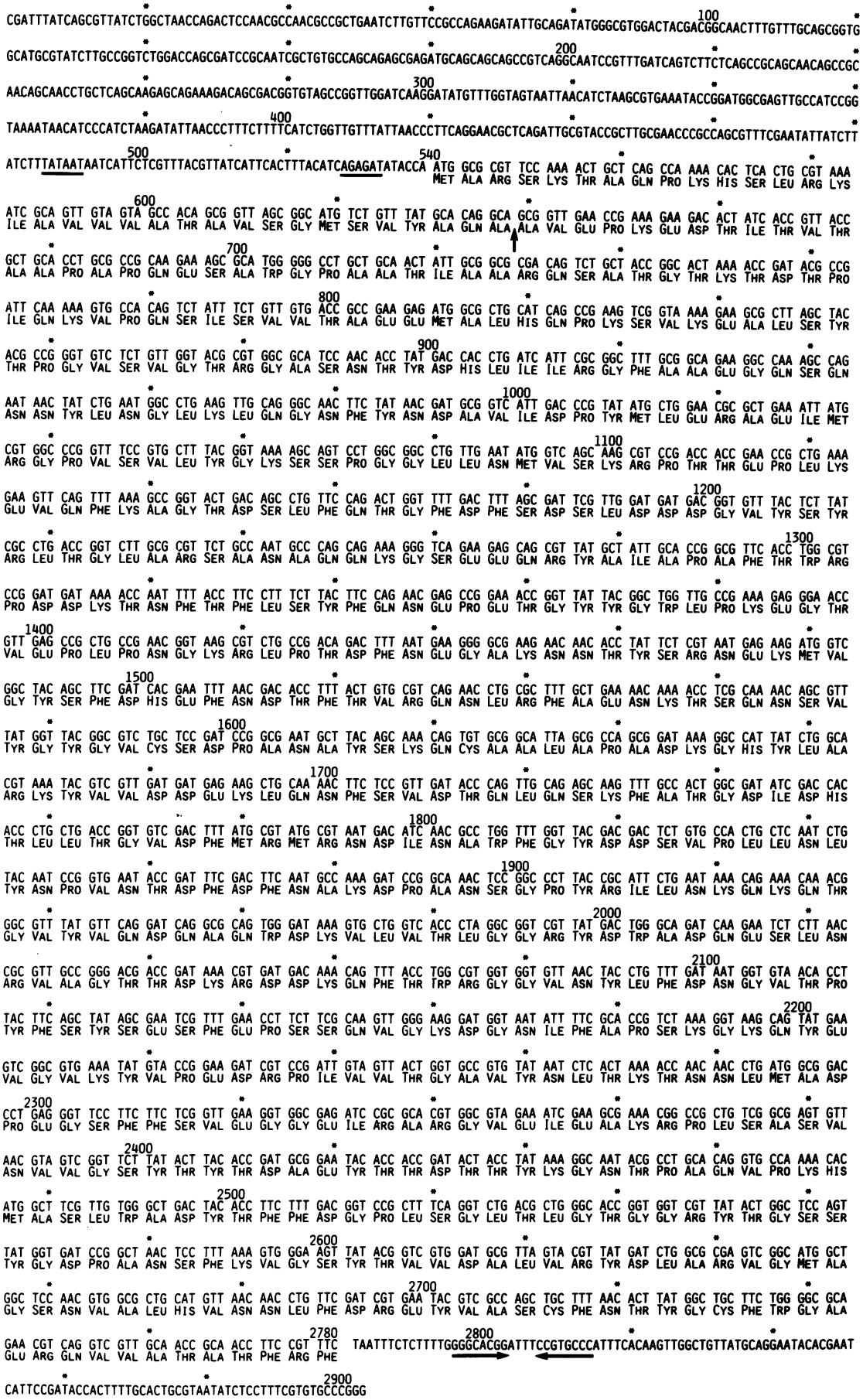
DISCUSSION

We elected to generate protein fusions to the ferrichrome-iron receptor (FhuA protein) of *E. coli* K-12 by using the plaque-forming lambda phage, λ *plac* Mu1. Our objectives were to determine the role of Fe(III) in the regulation of the activity of the *fhuA* gene and to create a range of size classes of chimeric proteins. Such hybrid proteins might be expected to differ in some of the FhuA-associated phenotypes. Domains of the FhuA protein which are responsible for binding of the various ligands have not yet been defined.

A series of 10 independent (*fhuA'-'lacZ*) gene fusions was constructed, all of which were regulated by the iron supply. An increased transcription of the *fhuA* gene was found when the iron-scavenging chelator EDDA was added to the culture, as measured by an increase of up to 12 times the levels of β-galactosidase when compared to the basal level of transcription of the gene. For three of the fusions (DRC109, DRC111, and DRC120), the β-galactosidase activities increased only about fivefold. This result suggested some phenotypic difference in these strains which contributed to lower specific activities of the fusion proteins. Our data confirm the regulatory influence of Fe(III) at the level of transcription of the *fhuA* gene and are similar to the information derived for the operon fusion of (*fhuA'-'lacZ*) (12). The increase in β-galactosidase activity was reported as only two- to threefold under iron-limiting growth conditions.

Perhaps the most striking feature of the (*fhuA'-'lacZ*) fusions which were constructed in vivo is that the sizes of the hybrid proteins were within a very narrow range. Our initial characterization of the chimeric proteins was by immunoprecipitation of (FhuA'-'LacZ), using antibody against β-galactosidase. The molecular weights of all 10 independently isolated strains were estimated to be 121,000 (Fig. 2).

Detailed characterization of each fusion strain was necessary to define the unique fusion joints of the (FhuA'-'LacZ)


CGATTTATCAGCGTTATCTGGCTAACCAGACTCCAACGCCAACGCCGCTGAATCTTGTTCCGCCAGAAGATATTGCAGATATGGGCGTGGACTACGACGGCAACTTTGTTTGCAGCGGTG
GCATGCGTATCTTGCCGGTCTGGACCAGCGATCCGCAATCGCTGTGCCAGCAGAGCGAGATGCAGCAGCAGCCGTCAGGCAATCCGTTTGATCAGTCTTCTCAGCCGCAGCAACAGCCGC
AACAGCAACCTGCTCAGCAAGAGCAGAAAGACAGCGACGGTGTAGCCGGTTGGATCAAGGATATGTTTGGTAGTAATTAACATCTAAGCGTGAAATACCGGATGGCGAGTTGCCATCCGG
TAAAATAACATCCCATCTAAGATATTAACCCTTTCTTTTCATCTGGTTGTTTATTAACCCTTCAGGAACGCTCAGATTGCGTACCGCTTGCGAACCCGCCAGCGTTTCGAATATTATCTT
ATCTTTATAATAATCATTCTCGTTTACGTTATCATTCACTTTACATCAGAGATATACCA ATG GCG CGT TCC AAA ACT GCT CAG CCA AAA CAC TCA CTG CGT AAA
MET ALA ARG SER LYS THR ALA GLN PRO LYS HIS SER LEU ARG LYS
ATC GCA GTT GTA GTA GCC ACA GCG GTT AGC GGC ATG TCT GTT TAT GCA CAG GCA GCG GTT GAA CCG AAA GAA GAC ACT ATC ACC GTT ACC
ILE ALA VAL VAL VAL ALA THR ALA VAL SER GLY MET SER VAL TYR ALA GLN ALA ALA VAL GLU PRO LYS GLU ASP THR ILE THR VAL THR
GCT GCA CCT GCG CCG CAA GAA AGC GCA TGG GGG CCT GCT GCA ACT ATT GCG GCG CGA CAG TCT GCT ACC GGC ACT AAA ACC GAT ACG CCG
ALA ALA PRO ALA PRO GLN GLU SER ALA TRP GLY PRO ALA ALA THR ILE ALA ALA ARG GLN SER ALA THR GLY THR LYS THR ASP THR PRO
ATT CAA AAA GTG CCA CAG TCT ATT TCT GTT GTG ACC GCC GAA GAG ATG GCG CTG CAT CAG CCG AAG TCG GTA AAA GAA GCG CTT AGC TAC
ILE GLN LYS VAL PRO GLN SER ILE SER VAL VAL THR ALA GLU GLU MET ALA LEU HIS GLN PRO LYS SER VAL LYS GLU ALA LEU SER TYR
ACG CCG GGT GTC TCT GTT GGT ACG CGT GGC GCA TCC AAC ACC TAT GAC CAC CTG ATC ATT CGC GGC TTT GCG GCA GAA GGC CAA AGC CAG
THR PRO GLY VAL SER VAL GLY THR ARG GLY ALA SER ASN THR TYR ASP HIS LEU ILE ILE ARG GLY PHE ALA ALA GLU GLY GLN SER GLN
AAT AAC TAT CTG AAT GGC CTG AAG TTG CAG GGC AAC TTC TAT AAC GAT GCG GTC ATT GAC CCG TAT ATG CTG GAA CGC GCT GAA ATT ATG
ASN ASN TYR LEU ASN GLY LEU LYS LEU GLN GLY ASN PHE TYR ASN ASP ALA VAL ILE ASP PRO TYR MET LEU GLU ARG ALA GLU ILE MET
CGT GGC CCG GTT TCC GTG CTT TAC GGT AAA AGC AGT CCT GGC GGC CTG TTG AAT ATG GTC AGC AAG CGT CCG ACC ACC GAA CCG CTG AAA
ARG GLY PRO VAL SER VAL LEU TYR GLY LYS SER SER PRO GLY GLY LEU LEU ASN MET VAL SER LYS ARG PRO THR THR GLU PRO LEU LYS
GAA GTT CAG TTT AAA GCC GGT ACT GAC AGC CTG TTC CAG ACT GGT TTT GAC TTT AGC GAT TCG TTG GAT GAT GAC GGT GTT TAC TCT TAT
GLU VAL GLN PHE LYS ALA GLY THR ASP SER LEU PHE GLN THR GLY PHE ASP PHE SER ASP SER LEU ASP ASP ASP GLY VAL TYR SER TYR
CGC CTG ACC GGT CTT GCG CGT TCT GCC AAT GCC CAG CAG AAA GGG TCA GAA GAG CAG CGT TAT GCT ATT GCA CCG GCG TTC ACC TGG CGT
ARG LEU THR GLY LEU ALA ARG SER ALA ASN ALA GLN GLN LYS GLY SER GLU GLU GLN ARG TYR ALA ILE ALA PRO ALA PHE THR TRP ARG
CCG GAT GAT AAA ACC AAT TTT ACC TTC CTT TCT TAC TTC CAG AAC GAG CCG GAA ACC GGT TAT TAC GGC TGG TTG CCG AAA GAG GGA ACC
PRO ASP ASP LYS THR ASN PHE THR PHE LEU SER TYR PHE GLN ASN GLU PRO GLU THR GLY TYR TYR GLY TRP LEU PRO LYS GLU GLY THR
GTT GAG CCG CTG CCG AAC GGT AAG CGT CTG CCG ACA GAC TTT AAT GAA GGG GCG AAG AAC AAC ACC TAT TCT CGT AAT GAG AAG ATG GTC
VAL GLU PRO LEU PRO ASN GLY LYS ARG LEU PRO THR ASP PHE ASN GLU GLY ALA LYS ASN ASN THR TYR SER ARG ASN GLU LYS MET VAL
GGC TAC AGC TTC GAT CAC GAA TTT AAC GAC ACC TTT ACT GTG CGT CAG AAC CTG CGC TTT GCT GAA AAC AAA ACC TCG CAA AAC AGC GTT
GLY TYR SER PHE ASP HIS GLU PHE ASN ASP THR PHE THR VAL ARG GLN ASN LEU ARG PHE ALA GLU ASN LYS THR SER GLN ASN SER VAL
TAT GGT TAC GGC GTC TGC TCC GAT CCG GCG AAT GCT TAC AGC AAA CAG TGT GCG GCA TTA GCG CCA GCG GAT AAA GGC CAT TAT CTG GCA
TYR GLY TYR GLY VAL CYS SER ASP PRO ALA ASN ALA TYR SER LYS GLN CYS ALA ALA LEU ALA PRO ALA ASP LYS GLY HIS TYR LEU ALA
CGT AAA TAC GTC GTT GAT GAT GAG AAG CTG CAA AAC TTC TCC GTT GAT ACC CAG TTG CAG AGC AAG TTT GCC ACT GGC GAT ATC GAC CAC
ARG LYS TYR VAL VAL ASP ASP GLU LYS LEU GLN ASN PHE SER VAL ASP THR GLN LEU GLN SER LYS PHE ALA THR GLY ASP ILE ASP HIS
ACC CTG CTG ACC GGT GTC GAC TTT ATG CGT ATG CGT AAT GAC ATC AAC GCC TGG TTT GGT TAC GAC GAC TCT GTG CCA CTG CTC AAT CTG
THR LEU LEU THR GLY VAL ASP PHE MET ARG MET ARG ASN ASP ILE ASN ALA TRP PHE GLY TYR ASP ASP SER VAL PRO LEU LEU ASN LEU
TAC AAT CCG GTG AAT ACC GAT TTC GAC TTC AAT GCC AAA GAT CCG GCA AAC TCC GGC CCT TAC CGC ATT CTG AAT AAA CAG AAA CAA ACG
TYR ASN PRO VAL ASN THR ASP PHE ASP PHE ASN ALA LYS ASP PRO ALA ASN SER GLY PRO TYR ARG ILE LEU ASN LYS GLN LYS GLN THR
GGC GTT TAT GTT CAG GAT CAG GCG CAG TGG GAT AAA GTG CTG GTC ACC CTA GGC GGT CGT TAT GAC TGG GCA GAT CAA GAA TCT CTT AAC
GLY VAL TYR VAL GLN ASP GLN ALA GLN TRP ASP LYS VAL LEU VAL THR LEU GLY GLY ARG TYR ASP TRP ALA ASP GLN GLU SER LEU ASN
CGC GTT GCC GGG ACG ACC GAT AAA CGT GAT GAC AAA CAG TTT ACC TGG CGT GGT GGT GTT AAC TAC CTG TTT GAT AAT GGT GTA ACA CCT
ARG VAL ALA GLY THR THR ASP LYS ARG ASP ASP LYS GLN PHE THR TRP ARG GLY GLY VAL ASN TYR LEU PHE ASP ASN GLY VAL THR PRO
TAC TTC AGC TAT AGC GAA TCG TTT GAA CCT TCT TCG CAA GTT GGG AAG GAT GGT AAT ATT TTC GCA CCG TCT AAA GGT AAG CAG TAT GAA
TYR PHE SER TYR SER GLU SER PHE GLU PRO SER SER GLN VAL GLY LYS ASP GLY ASN ILE PHE ALA PRO SER LYS GLY LYS GLN TYR GLU
GTC GGC GTG AAA TAT GTA CCG GAA GAT CGT CCG ATT GTA GTT ACT GGT GCC GTG TAT AAT CTC ACT AAA ACC AAC AAC CTG ATG GCG GAC
VAL GLY VAL LYS TYR VAL PRO GLU ASP ARG PRO ILE VAL VAL THR GLY ALA VAL TYR ASN LEU THR LYS THR ASN ASN LEU MET ALA ASP
CCT GAG GGT TCC TTC TTC TCG GTT GAA GGT GGC GAG ATC CGC GCA CGT GGC GTA GAA ATC GAA GCG AAA CGG CCG CTG TCG GCG AGT GTT
PRO GLU GLY SER PHE PHE SER VAL GLU GLY GLY GLU ILE ARG ALA ARG GLY VAL GLU ILE GLU ALA LYS ARG PRO LEU SER ALA SER VAL
AAC GTA GTC GGT TCT TAT ACT TAC ACC GAT GCG GAA TAC ACC ACC GAT ACT ACC TAT AAA GGC AAT ACG CCT GCA CAG GTG CCA AAA CAC
ASN VAL VAL GLY SER TYR THR TYR THR ASP ALA GLU TYR THR THR ASP THR THR TYR LYS GLY ASN THR PRO ALA GLN VAL PRO LYS HIS
ATG GCT TCG TTG TGG GCT GAC TAC ACC TTC TTT GAC GGT CCG CTT TCA GGT CTG ACG CTG GGC ACC GGT GGT CGT TAT ACT GGC TCC AGT
MET ALA SER LEU TRP ALA ASP TYR THR PHE PHE ASP GLY PRO LEU SER GLY LEU THR LEU GLY THR GLY GLY ARG TYR THR GLY SER SER
TAT GGT GAT CCG GCT AAC TCC TTT AAA GTG GGA AGT TAT ACG GTC GTG GAT GCG TTA GTA CGT TAT GAT CTG GCG CGA GTC GGC ATG GCT
TYR GLY ASP PRO ALA ASN SER PHE LYS VAL GLY SER TYR THR VAL VAL ASP ALA LEU VAL ARG TYR ASP LEU ALA ARG VAL GLY MET ALA
GGC TCC AAC GTG GCG CTG CAT GTT AAC AAC CTG TTC GAT CGT GAA TAC GTC GCC AGC TGC TTT AAC ACT TAT GGC TGC TTC TGG GGC GCA
GLY SER ASN VAL ALA LEU HIS VAL ASN ASN LEU PHE ASP ARG GLU TYR VAL ALA SER CYS PHE ASN THR TYR GLY CYS PHE TRP GLY ALA
GAA CGT CAG GTC GTT GCA ACC GCA ACC TTC CGT TTC TAATTTCTCTTTTGGGGCACGGATTTCCGTGCCCATTTCACAAGTTGGCTGTTATGCAGGAATACACGAAT
GLU ARG GLN VAL VAL ALA THR ALA THR PHE ARG PHE
CATTCCGATACCACTTTTGCACTGCGTAATATCTCCTTTCGTGTGCCCGGG

proteins. Such an analysis required that the gene fusions be cloned into the high-copy plasmid pMLB524. The cloned (*fhuA'-'-lacZ*) genes on the series of plasmids pDRC102 to pDRC123 served as a source of DNA for sequence analysis through the *fhuA*-Mu fusion joints. From DNA sequencing of the 10 gene fusions, three different sequences were identified (Fig. 5). Type I fusions, of which there were seven, gave the *fhuA* gene sequence at the fusion joint as 5'-ACT GCT CAG CCA A-3'. In the three type II fusions, the sequences were 5'-GCG GTT GAA CCG A-3' (DRC109 and DRC111) and 5'-GCG GTT GAA CCG A-3' (DRC120). Type II fusions were apparently longer than type I fusions because the sequence data could not be overlapped with that of type I. This suggested that the hybrid proteins of the two types differed in size, but this difference was not detected by immunoprecipitation of the fusion proteins. An indication of some difference in molecular weights of the chimeric proteins came from the electrophoresis of whole cell extracts of MC1000 containing the Lac^+ plasmids. Upon Coomassie blue staining of the SDS-polyacrylamide slab gel, strains containing pDRC109, pDRC111, and pDRC120 encoded hybrid proteins which were estimated to be of slightly lower mobility compared with other proteins of the pDRC series of plasmids. From the determination of the complete nucleotide sequence of the *fhuA* gene, the above sequences were found within the first 145 nucleotides of the coding region of the *fhuA* gene. It was then possible to calculate the molecular weight of each of the fusion proteins based upon (i) the amino acid sequence derived from the truncated FhuA protein; plus (ii) the amino acids translated from the open reading frame of the 117-base-pair region of Mu; plus (iii) the amino acid sequence of β-galactosidase, shortened by eight codons at the amino terminus. In summary (Fig. 5), the FhuA-derived regions of the chimeric proteins and their calculated molecular weights were as follows: type I, 9 amino acids of the signal sequence, M_r = 120,984; type IIa, the signal sequence of 33 amino acids plus 4 amino acids of the mature protein, M_r = 123,846; type IIb, the signal sequence of 33 amino acids plus 15 amino acids of the mature protein, M_r = 124,974. If the type II fusions were to be cleaved by signal peptidase, the signal sequence (M_r = 3,210) would reduce the molecular weight of a type II protein to about 121,000. This may explain the apparent similarity of the molecular weights of the immunoprecipitated proteins. On the other hand, if the amount of signal peptidase were limiting or if processing of the type II fusion proteins from the high-copy plasmids of the pDRC series were inefficient, it might account for the discrimination of molecular weights which was observed upon Coomassie staining of the proteins from the overproducing strains.

Although the use of λ p*lac* Mul represents an attractive vector for the production of some Lac protein fusions, it may be of limited value when constructing fusions to outer membrane proteins. The idea of overproduction lethality has been addressed in attempting to generate (OmpA'-'LacZ) fusions (27). Even in the case of (LamB'-'LacZ) fusions, it was suggested that because the sizes of the hybrid proteins fell into four distinct size classes, the technique to generate fusions might be the result of nonrandom events (10). We have also employed the "superhopper" phage λ p*lac* Mu2 which was constructed to transpose into cloned genes (E. Bremer, personal communication). Insertion of λ p*lac* Mu2 into the *fhuA* gene which was cloned in pPM25 and which contained the pBR322 replicon (7) did not generate chimeric proteins which appeared larger than those of type II fusions (unpublished results). This implies that for (FhuA'-'LacZ) fusions, there is an inherent limitation in constructing fusions of greater length and at the same time maintaining the viability of the cells.

Two additional possibilities might be entertained to explain why larger fusion proteins were not obtained. First, such (FhuA'-'LacZ) fusions might not confer a Lac^+ phenotype; they would therefore not have grown on minimal lactose. Alternatively, longer fusions might have conferred sensitivity to the selective agent. We are now constructing longer (*fhuA'-'lacZ*) fusions in vitro with pMLB1034 (34) to address both of these suggestions.

The complete nucleotide sequencing of the *fhuA* gene was undertaken to obtain detailed molecular information concerning the FhuA protein and the (FhuA'-'LacZ) hybrid proteins. Sequencing was completed for 2,902 base pairs on a *BglI-SmaI* fragment from pPM27 which contained the *fhuA* gene. A strong stem-loop structure ($\Delta G = -33.4$ Kcal) was detected between positions 339 and 359 and was preceeded by an open reading frame at the beginning of the sequence given in Fig. 7. This inverted repeat likely functions as the transcription terminator for the upstream *ponB* gene which codes for penicillin-binding protein 1B (5). Indeed, we find for a stretch of 430 nucleotides, there is complete identity between the published sequence of the carboxy terminus of the *ponB* gene product plus the 3'-tandem noncoding sequences (bases 2328 to 2758) (5) and our sequence data in the intragenic region on the 5'-side of the *fhuA* gene.

The nucleotide sequence of the *fhuA* gene allowed prediction of the amino acid sequence of the ferrichrome-iron receptor. The FhuA protein was previously shown to be synthesized as a preprotein (28). Our data identified a signal peptide of 33 amino acids, which may be the longest signal peptide thus far identified for an exported protein of *E. coli*. This signal sequence is characterized by two positively charged residues within the first five amino acids, by a net positive charge of plus six, by an extended region of hydrophobic residues beginning at position 16, and by a consensus sequence for recognition by leader peptidase of Ala_{31}-X_{32}-Ala_{33}. Amino-terminal-sequencing of the isolated mature FhuA protein identified the first 14 residues, beginning with Ala_1-Val_2-Glu_3-Pro_4-Lys_5 and provided both a verification of the fidelity of the nucleotide sequence and of the reading frame.

The overall hydropathy along the amino acid sequence of the FhuA protein was evaluated by the method of Kyte and Doolittle (19) with an average hydropathy of segments of 15 residues being plotted at the mid-point of the segment. Hydrophobic regions gave positive scores; hydrophilic seg-

FIG. 7. Nucleotide sequence of the DNA region of pPM27 containing the gene encoding the ferrichrome-iron receptor protein. The amino acid sequence of the predicted translation product is shown below the nucleotide sequence. The numbers refer to the nucleotide sequence and an asterisk is indicated every 20 nucleotides. The consensus sequence of the Pribnow box (TATAAT) is underlined beginning at nucleotide 485, as is a potential Shine-Dalgarno ribosome-binding site (AGAGAT) from nucleotide 527. The initiation codon ATG is at position 540 and an open reading frame continues until the TAA stop codon at position 2781. The vertical arrow indicates the start of the mature protein as determined by amino-terminal sequencing. A region which shows dyad symmetry (nucleotides 2795 to 2814) is shown by the paired arrows beneath the sequence (→ ←).

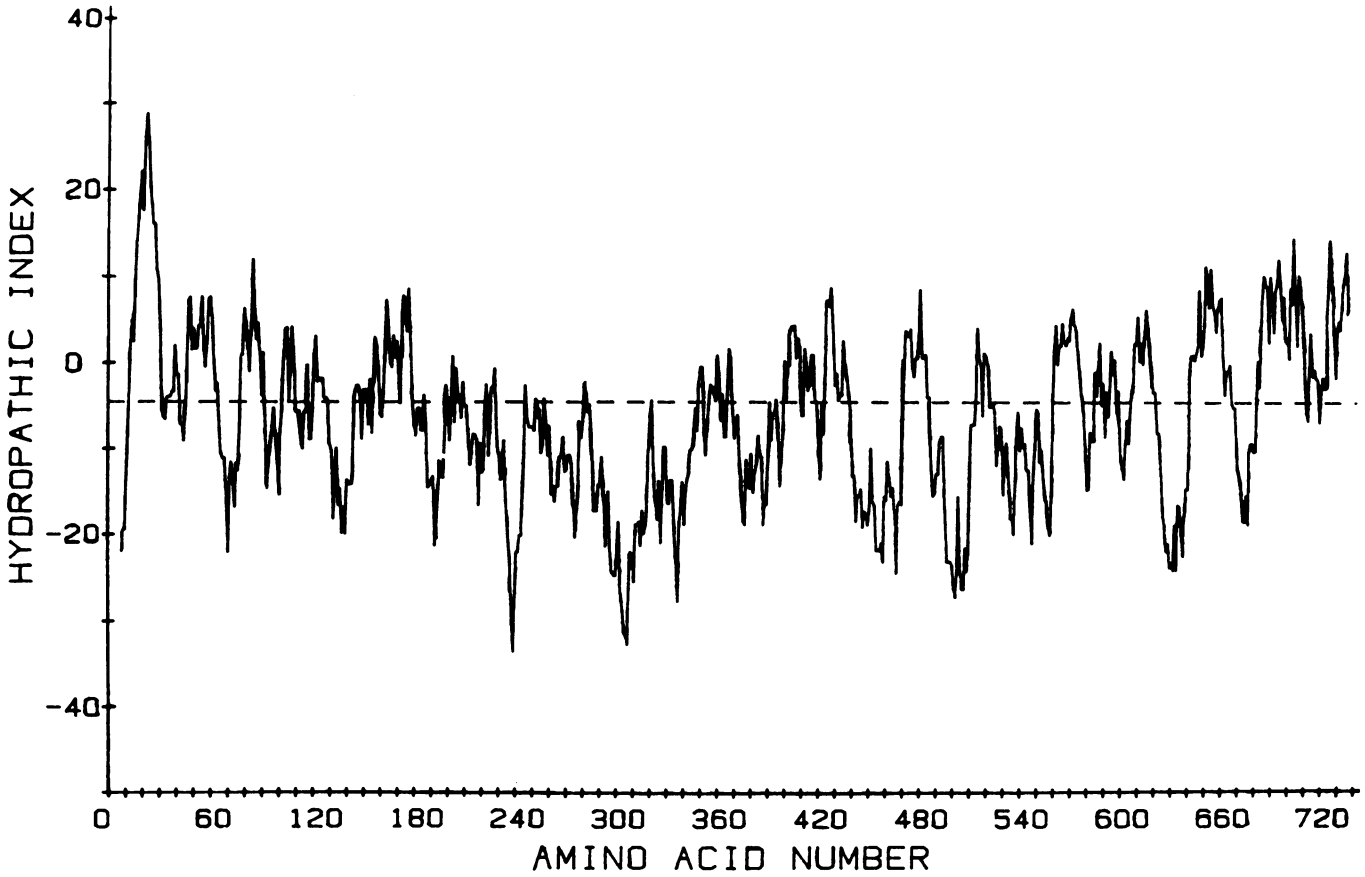


FIG. 8. Hydropathy profile of the FhuA protein. The hydropathy index was evaluated according to the parameters presented by Kyte and Doolittle, using a span setting of 15. Hydrophobic regions of the protein extend above the dotted line.

ments gave negative scores. Figure 8 displays the hydropathy profile for the FhuA protein and indicates that it is not appreciably hydrophobic in overall character. The most hydrophobic segments of the protein are the signal sequence and two domains at the carboxy terminus. Alternating regions of hydrophobic and hydrophilic character are indicated throughout the protein.

The data presented on the localization of the (FhuA'-'LacZ) hybrid proteins (Table 2) should be interpreted with some caution. Whereas all seven of the type I fusions appeared to be found in the cytoplasm, our fractionation data indicated that type II fusions showed up to 58% of the β-galactosidase activity in the total membranes (outer membranes plus inner membranes). This apparent association of type II fusion proteins with the cell envelope may be due to the inherently high hydrophobicity of the first 48 amino acids of the proFhuA covalently linked to the amino-terminal regions of β-galactosidase. Such "localization" may be only nonspecific binding to the cell envelope preparation. Further information on the cellular location of the type II fusions as obtained by gold labeling of thin sections for immunoelectron microscopy would provide an independent confirmation of these fractionation data. In the case of a (PhoE'-'LacZ) hybrid protein in which 300 out of 351 amino acids of PhoE were fused to LacZ, cell fractionation experiments suggested that the hybrid gene was transported to the outer membrane; immunocytochemical labeling identified the hybrid protein in the cytoplasm. Tomassen et al. (35) therefore concluded that data on the localization of hybrid proteins based merely on cell fractionation experiments were not reliable. Current models for the translocation of outer membrane proteins are partly derived from studies of (LamB'-'LacZ) hybrid proteins (1, 2). Since the fractionation data in the case of the (PhoE'-'LacZ) hybrid protein indicated some anomaly with respect to envelope localization, it may be necessary to reevaluate such information by using another technique.

The amino acid sequence of the FhuA protein represents the third published sequence of a *tonB*-dependent receptor protein. The other two sequences are the vitamin B12 receptor (BtuB protein) (16) and the aerobactin receptor (18). Heller and Kadner identified the *btuB451* mutation as being a leucine-to-proline transition at the eighth amino acid of the putative mature protein. This mutation blocked vitamin B12 uptake but did not affect other receptor functions. Their suggestion was that this region might be involved in some interaction with the *tonB* gene product (29). In searching for regions of homology within the FhuA protein, the BtuB protein, and the aerobactin receptor, it was appropriate to direct some attention to the amino acid sequences of the three amino-terminal regions. An alignment to provide best fit within the first 15 amino acids is presented in Fig. 9. The correspondence of the amino acid sequences in this region was remarkable. Even without allowing for equivalent substitutions (for example, Asp to Glu), it was found that Thr_8, Val_{11}, and Ala_{13} from the FhuA protein could be aligned with the identical amino acids in the other two proteins; correspondence of amino acids in six other posi-

FHUA PROTEIN	ALA$_1$	VAL	GLU	PRO	LYS	GLU	ASP	THR	ILE	THR	VAL$_{11}$	THR	ALA	ALA	PRO
BTUB PROTEIN		GLN$_1$	ASP	THR	SER	PRO	ASP	THR	(LEU)	VAL	VAL$_{10}$	THR	ALA	ASN	ARG
AEROBACTIN RECEPTOR		GLN$_1$	GLN	THR	ASP	ASP	GLU	THR	PHE	VAL	VAL$_{10}$	SER	ALA	ASN	ARG

FIG. 9. Amino acid sequence homology among TonB-dependent proteins. The regions of homology in the amino-terminal sequences of the three receptor proteins FhuA, BtuB, and aerobactin receptor are indicated by a vertical alignment which provides the best fit. Single underlining indicates the correspondence of two amino acids; double underlining indicates correspondence of all three amino acids. A *btuB* mutation which gave a leucine-to-proline change in the eighth amino acid (indicated by the circled Leu) of the mature protein is apparently blocked in vitamin B12 uptake and may be involved in some interaction with the *tonB* gene product.

tions was also noted when the proteins were paired. From the hydropathy profile, it would appear that this limited region is a weakly hydrophobic segment of the protein. Because of the strong amino acid sequence homology which we have now identified for three *tonB*-dependent receptors and in anticipation of ascribing some activity to the *tonB* gene product, continuing studies on the topology and functional behavior of this domain of the receptor should prove fruitful.

ACKNOWLEDGMENTS

Construction of (*fhuA'-'lacZ*) gene fusions was begun when J.W.C. participated in the Advanced Bacterial Genetics course at Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., July, 1983. Tom Silhavy, Mike Berman, Lynn Enquist, and Erhard Bremer provided sagacious teaching. We thank Michel van der Rest for conducting the amino acid sequencing, Stephen Garrett for constructing appropriate bacterial strains, Judith Kashul for contributing the figures, and Anna Campana for providing photographic services. The Institut de Recherches Cliniques de Montreal made available programs for assembly and analysis of nucleotide sequences.

This research was supported by grant A7289 from the Natural Sciences and Engineering Research Council and by grant MT-6911 from the Medical Research Council of Canada.

LITERATURE CITED

1. **Benson, S. A., E. Bremer, and T. J. Silhavy.** 1984. The use of gene fusions to define intragenic information required for LamB export to the outer membrane. Proc. Natl. Acad. Sci. U.S.A. **81**:3830–3834.
2. **Benson, S. A., M. N. Hall, and T. J. Silhavy.** 1985. Genetic analysis of protein export in *Escherichia coli* K12. Annu. Rev. Biochem. **54**:101–134.
3. **Braun, V., R. E. W. Hancock, K. Hantke, and A. Hartmann.** 1976. Functional organization of the outer membrane of *Escherichia coli*: phage and colicin receptors as components of iron uptake systems. J. Supramol. Struct. **5**:37–58.
4. **Bremer, E., T. J. Silhavy, J. M. Weisemann, and G. M. Weinstock.** 1984. λ p*lac*Mu: a transposable derivative of bacteriophage lambda for creating *lacZ* protein fusions in a single step. J. Bacteriol. **158**:1084–1093.
5. **Broome-Smith, J. K., A. Edelman, S. Yousif, and B. G. Spratt.** 1985. The nucleotide sequences of the *ponA* and *ponB* genes of *Escherichia coli* K12. Eur. J. Biochem. **147**:437–466.
6. **Coulton, J. W.** 1982. The ferrichrome-iron receptor of *Escherichia coli* K-12. Antigenicity of the *fhuA* protein. Biochim. Biophys. Acta **717**:154–162.
7. **Coulton, J. W., P. Mason, and M. S. DuBow.** 1983. Molecular cloning of the ferrichrome-iron receptor of *Escherichia coli* K-12. J. Bacteriol. **156**:1315–1321.
8. **Coulton, J. W., H.-U. Naegeli, and V. Braun.** 1979. Iron supply of *Escherichia coli* with polymer-bound ferricrocin. Eur. J. Biochem. **99**:39–47.
9. **Fecker, L., and V. Braun.** 1983. Cloning and expression of the *fhu* genes involved in iron(III)-hydroxamate uptake by *Escherichia coli*. J. Bacteriol. **156**:1301–1314.
10. **Hall, M. N., M. Schwartz, and T. J. Silhavy.** 1982. Sequence information within the *lamB* gene is required for proper routing of the bacteriophage λ receptor protein to the outer membrane of *Escherichia coli* K-12. J. Mol. Biol. **156**:93–112.
11. **Hancock, R. E. W., and V. Braun.** 1976. Nature of the energy requirement for the irreversible adsorption of bacteriophages T1 and φ80 to *Escherichia coli*. J. Bacteriol. **125**:409–415.
12. **Hantke, K.** 1981. Regulation of ferric iron transport in *Escherichia coli* K-12: isolation of a constitutive mutant. Mol. Gen. Genet. **182**:288–292.
13. **Hantke, K.** 1983. Identification of an iron uptake system specific for coprogen and rhodotorulic acid in *Escherichia coli* K-12. Mol. Gen. Genet. **191**:301-306.
14. **Hantke, K., and V. Braun.** 1978. Functional interaction of the *tonA*/*tonB* receptor system in *Escherichia coli*. J. Bacteriol. **135**:190–197.
15. **Hantke, K., and L. Zimmermann.** 1981. The importance of the *exbB* gene for vitamin B12 and ferric iron transport. FEMS Microbiol. Lett. **12**:31–35.
16. **Heller, K., and R. J. Kadner.** 1985. Nucleotide sequence of the gene for the vitamin B12 receptor protein in the outer membrane of *Escherichia coli*. J. Bacteriol. **161**:904–908.
17. **Kadner, R. J., K. Heller, J. W. Coulton, and V. Braun.** 1980. Genetic control of hydroxamate-mediated iron uptake in *Escherichia coli*. J. Bacteriol. **143**:256–264.
18. **Krone, W. J. A., F. Stegehuis, G. Koningstein, C. van Doorn, B. Roosendaal, F. K. de Graaf, and B. Oudega.** 1985. Characterization of the pColV-K30 encoded cloacin DF13/aerobactin outer membrane receptor protein of *Escherichia coli*; isolation and purification of the protein and analysis of its nucleotide sequence and primary structure. FEMS Microbiol. Lett. **26**:153–161.
19. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.
20. **Lugtenberg, B., J. Meijers, R. Peters, and P. van der Hock.** 1975. Electrophoretic resolution of the "major outer membrane protein" of *Escherichia coli* K-12 into four bands. FEBS Lett. **58**:254–258.
21. **Lugtenberg, B., and L. van Alphen.** 1983. Molecular architecture and functioning of the outer membrane of *Escherichia coli* and other gram-negative bacteria. Biochim. Biophys. Acta **737**:51–115.
22. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning. A laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
23. **Messing, J., and J. Vieira.** 1982. A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. Gene **19**:269–276.
24. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
25. **Neilands, J. B.** 1982. Microbial envelope proteins related to iron. Annu. Rev. Microbiol. **36**:285–309.
26. **Nikaido, H., and M. Vaara.** 1985. Molecular basis of bacterial outer membrane permeability. Microbiol. Rev. **49**:1–32.
27. **Palva, E. T., and T. J. Silhavy.** 1984. *lacZ* fusions to genes that specify exported proteins: a general technique. Mol. Gen.

Genet. **194**:388–394.
28. **Plastow, G. S., J. M. Pratt, and I. B. Holland.** 1981. The ferrichrome receptor protein (*tonA*) of *Escherichia coli* is synthesized as a precursor *in vitro*. FEBS Lett. **131**:262–264.
29. **Postle, K., and R. F. Good.** 1983. DNA sequence of the *Escherichia coli tonB* gene. Proc. Natl. Acad. Sci. U.S.A. **80**:5235–5239.
30. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. **74**:5463–5467.
31. **Shine, J., and L. Dalgarno.** 1974. The 3′ terminal sequences of *Escherichia coli* 16S rRNA: complementary to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. U.S.A. **71**:1342–1346.
32. **Siebenlist, U., R. B. Simpson, and W. Gilbert.** 1980. *E. coli* RNA polymerase interacts homologously with two different promoters. Cell **20**:269–281.
33. **Silhavy, T. J., and J. Beckwith.** 1983. Isolation and characterization of mutants of *Escherichia coli* K12 affected in protein localization. Methods Enzymol. **97**:11–40.
34. **Silhavy, T. J., M. L. Berman, and L. W. Enquist.** 1984. Experiments with gene fusions. Cold Spring Harbor Laboratory, Cold Spring Harbor, N. Y.
35. **Tommassen, J., J. Leunissen, J. van Damme-Jongsten, and P. Overduin.** 1985. Failure of *E. coli* K-12 to transport PhoE-LacZ hybrid proteins out of the cytoplasm. EMBO J. **4**:1041–1047.
36. **Witholt, B., M. Boekhout, J. Kingma, H. Herrikhuizen, and L. Leij.** 1976. An efficient and reproducible procedure for the formation of spheroplasts from variously grown *Escherichia coli*. Anal. Biochem. **74**:160–170.