The Cloning, DNA Sequence, and Overexpression of the Gene araE Coding for Arabinose-Proton Symport in Escherichia coli K12*

(Received for publication, October 13, 1987)

Martin C. J. Maiden†, Maurice C. Jones-Mortimer, and Peter J. F. Henderson

From the Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QW, United Kingdom

A XplacMu1 insertion was made into araE, the gene for arabinose-proton symport in Escherichia coli. A phage containing an araE-lacZ fusion was recovered from the lysogen and its restriction map compared with that of the 61-min region of the E. coli genome to establish the gene order thyA araE orf lysR lysA galR; araE was transcribed toward orf. A 2.8-kilobase SalI-EcoRI DNA fragment containing araE was subcloned from the phage λI(yasA+ galR* araE*) into the plasmid vector pBR322. From this plasmid a 2.8-kilobase HindII-PvuII DNA fragment including araE was sequenced and also subcloned into the expression vector pAD284. The araE gene was 1416-base pairs long, encoding a hydrophobic protein of 472 amino acids with a calculated Mr, of 51,883. The amino acid sequence was homologous with the xylose-proton symporter of E. coli and the glucose transporters from a human hepatoma HepG2 cell line, human erythrocytes, and rat brain. The overexpressed araE gene product was identified in Cosmassie-stained sodium dodecyl sulfate-polyacrylamide gel electrophoresis gels of cell membranes as a protein of apparent Mr, 35,000 ± 1,150. Arabinose protected this protein against reaction with N-ethylmaleimide.

The pentose L-arabinose is transported into Escherichia coli by two mechanisms, a binding protein-dependent system and a proton symporter. These can be distinguished by differences in their energetics, genetics, and protein components (1-6). Arabinose-proton symport activity is highly specific for arabinose (and its analogue D-fucose) and is inhibited by N-ethylmaleimide (5,6). It is catalyzed by an integral membrane protein of apparent Mr, 36,000-38,000 (6), coded by the gene araE situated at 61 min on the E. coli linkage map (4, 6-8). A defective phage carrying the araE gene has been isolated (6). The promoter region of araE has been sequenced, but the coding sequence for only 28 N-terminal amino acids was established (9). In the present study the precise location and direction of transcription of the araE gene was established by the comparison of restriction maps of phages inserted near or into araE, enabling the cloning, amplified expression, and sequencing of the gene. The complete amino acid sequence of the arabinose-proton symporter was deduced and a model of its orientation in the membrane proposed.

EXPERIMENTAL PROCEDURES

Genetical Techniques—The E. coli strains, phages, and plasmids are listed in Table I. General procedures were performed as described by Miller (10), and infections with phage XplacMu1 used the method of Bremer et al. (11). Liquid cultures were grown in the minimal medium of Ashworth and Korberberg (12), as modified by Henderson et al. (13), supplemented with 20 mM glycerol, 10 or 1 mM arabinose, and 80 μg/ml amino acids as appropriate. Induction of lysogens of XplacMu1 by ultraviolet light was performed as described by Davis and Henderson (14). The phage excision events occurred by illegitimate recombination, since there was no DNA homology at either end of the phage, so the process was nonrandom and of low frequency (10⁻⁵-10⁻⁶ plaque-forming units/ml).

Lac⁺ phages were identified by plating on a lawn of a lac strain containing X-gal (25 μg/ml) and inducer. Spf phages (15, 16) were selected on a lawn in Tryptone soft agar of the F² lysogen RBS41 on minimal succinate (20 mM) plates. Transformations were performed by the method of Hanahan (17).

Although the araA or B mutation relieved the toxicity of arabinose conferred by the araD mutation (18), growth on certain substrates (including glycerol) was inhibited by arabinose. To overcome this inhibition minimal medium cultures were grown as described (13) with glycerol as the carbon and energy source to a culture density of 0.3-0.45 mg dry mass cells/ml (15-17 h at 30 °C) and induced with arabinose (final concentration 1 mM) for 1 h, when the culture reached a density of 0.5-0.65 mg dry mass/ml. The cells were then harvested and prepared for transport measurements and enzyme assays.

Construction and Characterization of araE-XplacMu Strains—E. coli strain JM2433 (Table I) has a double deficiency in enzymes of arabinose metabolism and a deletion through the lac operon. It is therefore a suitable recipient to select for infection by XplacMu (conferring a Lac⁺ phenotype) and subsequently screen for arabinose-inducibility of growth on lactose, i.e. for in-frame insertions downstream of an arabinose operon.

Strain JM2433 was infected with XplacMu1 and the integration helper ApMu507 (11). Colonies selected on arabinose (1 mM) plus lactose (5 mM) as carbon sources (therefore expressing the lac genes) were replicated onto lactose minimal medium and onto arabinose plus lactose. Five colonies out of 167 grew on arabinose plus lactose, but not lactose alone. The XplacMu insertions in these strains could be downstream of three known arabinose promoters, the araD operon located at 1 min (7, 19), the araE transport gene(s) at 61 min (4, 6, 20), or the araF,G operon at 45 min (3, 21). To identify insertions in the araE gene, its known co-transducibility with lysA (3, 6, 7) was tested (Table II). The insertion in strain JM2443 was co-transducible with lysA, from strain CBK140, and was likely to be in araE. The insertions in strains JM2442 and JM2444 were similarly shown to map in the araA,B,C,D region.

Given that the insertion is in the araE gene, the gene order with respect to both adjacent genes (lysA and thyA) was determined as follows. The LyS* Nm⁺ derivatives of strain JM2443 were also tested

---

*This work was supported in part by Science and Engineering Research Council Grant GR/C. 34977 and by the SmithKline Foundation. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) J03752.

†Supported by Studentships from the Science and Engineering Research Council and Sidney Sussex College and by a Fellowship from the Medical Research Council.
for inheritance of the thyA allele of the donor strain. There were 5 Thy" (Ara-+Lac+) recipient-type recombinants, 66 Thy" (Ara-+Lac+) donor type recombinants, 97 Thy" (Ara-+Lac-) and 0 Thy" (Ara-+Lac-) recombinants. If the Thy" (Ara-+Lac+) phenotype represents the four crossovers then the gene order is: lysA araE thyA. This result agrees with that of Kolodrubetz and Schleif (4), but contradicts that of Macpherson et al. (6).

The Preparation of a λ Phage Containing the araE Gene—The lysogen JM1647 contained two phages λ(araE+ lysA+ galR+) and λ(1657857 as integration helper (6). Since the phages were cI857, they were induced by hatching an exponential culture of the lysogen to 42°C. The phages present in the mixed suspension were separated by cesium chloride buoyant density centrifugation (below). The more dense phage, λ(araE+ lysA+ galR+), was about 50 kb in length.

Induction of λ cI857 Lysogens—A 5-ml inoculum was grown overnight with shaking at 35°C and diluted 100-fold into L-medium. This was incubated at 33°C with shaking to a culture density of about 0.5 mg dry mass/ml (A600 = 0.45). The temperature was raised rapidly to 42°C by the addition of an equal volume of L-medium at 52°C. Incubation was continued for 1.5-3 h. The resultant cells were then used to prepare plaque (10) or membranes (see below) as appropriate. For plasmid-containing strains, ampicillin (100 μg/ml) was included in all steps until the heat shock.

DNA Preparations—Phages were propagated in lytic cycle as described by Maniatis et al. (22). Remaining cells were lysed by the addition of chloroform. DNAase and RNase were added (each to about 1 μg/ml) and the suspension incubated for 1 h at room temperature. The phages were concentrated by precipitation with polyethylene glycol 6000 and the phage resuspended in 20 mM NaCl at 22°C and sedimented at 40,000 × g for 3 h. Phages in mixed lysates were separated by a cesium chloride block gradient (3/5 m, 0.45). The more dense phage, λ(araE+ lysA+ galR+), was about 50 kb in length.

Plasmid DNA was extracted by treating cells with lysozyme and Triton as follows. An overnight culture (5 ml) of plasmid-containing organism in L-medium (22) plus ampicillin (100 μg/ml) was harvested and washed once in 0.1 M NaCl (5 ml). The cell pellet was resuspended in 10 mM Tris acetate (pH 8.0), 10% sucrose (200 μl), and the suspension placed in a microcentrifuge tube on ice. The three following additions were made at 5-min intervals on ice: 5 mg/ml lysozyme (10 μl); 100 mM Tris acetate, 10 μl; 1% Triton X-100 (200 μl). The mixture was centrifuged at 12,000 × g for 15 min to sediment chromosomal DNA, and the plasmid DNA was obtained from the supernatant by phenol/chloroform extraction followed by ethanol precipitation (22).

Restriction Analysis—Restriction enzymes were obtained from New England Biolabs, Bethesda Research Laboratories, and Amersham International. Digestions were performed as described by Maniatis et al. (22) and the products separated by electrophoresis on 0.7 or 0.9% agarose gels in 40 mM Tris-CI (pH 8.0). The cell suspension was passed once through a French press at 20,000 p.s.i. and centrifuged at 1,500-3,000 × g to remove unbroken cells and larger cell debris. This cell-free extract was further centrifuged at 145,000 × g to obtain a cytoplasmic fraction (supernatant) and membrane fraction (pellet resuspended to the same volume as the cytoplasmic fraction). All operations were performed at 0-5°C except the β-galactosidase assays on the extract and fractions ("Experimental Procedures"). Strain EJ18 has an operon fusion, i.e. cytoplasmic β-galactosidase (25).

β-Galactosidase activity

| Strain | Inducer | Cell-free extract | Cytoplasmic fraction | Membrane fraction | Membrane-associated | μmol/mg/min | % |
|--------|---------|------------------|---------------------|------------------|--------------------|-------------|----|
| JM2443 | Arabinose (1 mM) | 1623 (318)* | 67 (29) | 997 (1300) | 94 | 60 (0.6) | 19 (6.1) | 76 |
| JM2443 | None | 10 (5.5) | 6 (0.6) | 19 (6.1) | 76 |
| EJ18 | Xylose (10 mM) | 7765 (323) | 7762 (311) | 1900 (534) | 19 |

* Numbers in parentheses are μmol/mg/min.
a host that provided the antiterminator and thus allowed efficient expression from P\textsubscript{ara} (15).

Transport Assays—The accumulation of L- and radiolabeled sugars and sugar-promoted pH changes in bacterial suspensions were measured as described by Henderson et al. (13) and Henderson and Macpherson (8).

β-Galactosidase Assays—Liquid cultures were lysed and assayed for β-galactosidase by the method of Miller (10). Assays on plate-grown colonies were done as described by Davis et al. (25).

Labeling of AraE Protein with Radioactive N-Ethylmaleimide—This was performed by the dual isotope method described by Henderson and Macpherson (8) except that whole cells, rather than vesicles, were used, from which bacterial membranes were prepared (see below). The preferentially labeled proteins were separated by SDS-polyacrylamide gel electrophoresis on 15% gels, which were electroblotted onto cellulose nitrate paper for slicing and determination of radioactivity by scintillation counting (8).

Preparation of Bacterial Membranes and Cytoplast—Spheroplasts were prepared by the method of Witholt et al. (26) and lysed by osmotic shock in ice-cold deionized water (8). The membrane and cytoplasmic fractions of this preparation were separated by centrifugation (40,000 × g for 20 min at 4°C). The membrane pellet was resuspended and the supernatant freeze-dried to concentrate the cytoplasmic proteins.

Separation of Proteins—SDS-polyacrylamide gel electrophoresis was performed as described by Henderson and Macpherson (8).

Fluorography—Amplify (from Amersham International) was used according to the manufacturer’s instructions prior to drying the gel and exposing to x-ray film at −80°C.

DNA Sequencing—The dideoxy chain termination method of Sanger et al. (27) with generation of random clones by sonication (28) was used. The oligonucleotides were separated on wedge-shaped gels cast and electrophoresed in the LKB Macrophor apparatus (14, 29).

The contiguous sequence was assembled using R. Staden’s Database System installed on the University of Cambridge IBM 3080 by M. Bishop and P. Aletounier.

RESULTS

Biochemical Characterization of the Fusion of araE with lacZ—In strain JM2443 (see “Experimental Procedures”) β-galactosidase activity was induced by arabinose, and it was associated with the membrane fraction (Table III); no arabinose-proton symport activity was detected (Fig. 1, Table IV), but there was an arabinose-inducible lactose (or lactose analogue)-proton symport (Fig. 1, Table IV). These were properties expected of an in-frame lacZ fusion in an arabinose-inducible gene coding for a membrane protein, such as araE.

The parent strain (JM2433), in contrast, possessed arabinose-proton symport and lacked methyl-β-D-thiogalactoside-proton (lactose-proton; 30-32) symport (Fig. 1, Table IV). The average arabinose-proton uptake of 331 nmol of H\textsuperscript{+}/mg dry mass cells compared to about 50 nmol of H\textsuperscript{+}/mg dry mass cells for E. coli with the araA mutation (5). This indicated that the mutation in strain JM2433 was in araB (ribulokinase), since an active arabinose isomerase (araA gene product) would convert arabinose to ribulose, explaining the enhanced arabinose influx (Fig. 1, Table IV). It was also consistent with

TABLE IV

Transport activities of bacterial strains

| Strain     | Inducer | Substrate  | Initial rate of proton symport | Extent of proton symport | Energized "C-sugar uptake |
|------------|---------|------------|--------------------------------|--------------------------|---------------------------|
|            |         |            | nmol H\textsuperscript{+}/mg/min | nmol H\textsuperscript{+}/mg | nmol H\textsuperscript{+}/mg/min |
| JM2433     | Arabinose (1 mM) | Arabinose | 2.6 ± 1.2                       | 18.0 ± 3.4               | 6.2 ± 1.3                 |
| JM2443     | Arabinose (1 mM) | TMG       | 0.0                             | 0.0                       | 0.4 ± 0.4                 |
|            | Arabinose (1 mM) | Arabinose | 0.7 ± 0.5                       | 0.4 ± 0.1 | 6.1 ± 2.5 |
|            | Arabinose (1 mM) | TMG       | 1.2 ± 0.5                       | 3.1 ± 1.2 | 4.6 ± 1.7 |
| JM2443     | None    | Arabinose  | 1.5 ± 1.0                       | 0.3 ± 1.0 | 0.1 ± 0.2 |
|            | None    | TMG       | 1.3 ± 1.0                       | 0.2 ± 1.0 | 0.0 ± 0.0 |
| MM23       | Arabinose (1 mM) | Arabinose | 3.8 ± 0.5                       | 5.2 ± 1.1 | 6.1 ± 0.6 |
| MM23 (pMM25) | Arabinose (1 mM) | Arabinose | 16.8 ± 3.2                      | 5.1 ± 2.7 | 6.9 ± 2.4 |
| AD5827 (pMM27) | Arabinose | Arabinose | 0.0                             | 0.0 | 0.3 ± 0.1 |
| AD5827 (pMM27) | None    | Arabinose  | 0.0                             | 0.0 | 0.4 ± 0.1 |
| AD5827 (pMM26) | Arabinose | Arabinose | 0.0                             | 0.0 | 0.4 ± 0.1 |

Some batches of L-arabinose caused a pH change of 7.7 nmol of H\textsuperscript{+} at a rate of 25 nmol/min in the absence of cells. As the incubations contained about 17 mg/dry mass cells, a pH change of 0.45 nmol of H\textsuperscript{+} at a rate of 1.5 nmol/mg/min occurred for every experiment.

* Result of one measurement only.
a relatively low level of induction of β-galactosidase that would result from conversion of the inducer, arabinose, to ribulose. The methyl-β-D-thiogalactoside-proton symport activity in strain JM2443 was similar in extent to measurements on wild-type lactose oerons (31, 32). Uptake of radiolabeled arabinose was present in both strains (Table IV), presumably because the araE,C transport system, present in the parent JM2433, was unaffected by the λlacMu insertion in strain JM2443, an observation consistent with those of Daruwalla et al. (5) for araE-araF' strains.

Procedures. Of eight independently isolated fusion-phages Derived from E. coli strain JM1647 (Ref. 6, "Experimental Procedures"). Comparison of their restriction maps enabled the location of the E. coli DNA insert in the former phage (Fig. 2b) and identified the restriction sites within it.

The Location and Direction of Transcription of the araE Gene—Expanded restriction maps of the E. coli DNA inserts in the λ(araE' lys+ gal+) phages are aligned with the λlac region of the E. coli genome (33) in Fig. 2c. The araE promoter must be located to the left of the point where the insertion in strain JM2443 had occurred, which can be calculated from the BamHI site in lacZ located 1.6 kb away from it (Fig. 2c). The araE gene is transcribed toward the fusion and hence toward lysA. It seemed likely that the

**FIG. 2. Restriction maps of the λ phages and the lysR region of the E. coli chromosome.** DNA preparations were made and restriction digestions performed as described under "Experimental Procedures." The following enzymes were used: BamHI (B), EcoRI (E), SmaI (S), SalI (SI), and HindIII (H). a, location of the E. coli DNA inserts by comparison of phages λp1φ(araE'-'lac) and λp2φ(araE'-'lac) with the original phage λlacMu1 (11). b, location of the E. coli DNA insert by comparison of phage λd(araE' lysA+ galR+) with phage λlac857 S7; the insert is in the opposite orientation to those in λp1φ(araE'-'lac) and λp2φ(araE'-'lac). c, aligned restriction maps of the inserts from phage λp2φ(araE'-'lac), phage λd(araE' lysA+ galR+) and the lysR region of E. coli (kindly given by J. C. Patte, c.f. Ref. 33). The restriction sites within the Xho-EcoRI region were confirmed and extended after its ligation into plasmid pBR322 (text and Fig. 3). The vertical arrows show the location where the araE gene fused with lacZ in the λp2φ(araE'-'lac) phage, and so identify the approximate location of araE in phage λd(araE' lysA+ galR+) and the E. coli DNA. The secondary attachment site where the prophage, from which λd(araE' lysA+ galR+) was derived, originally inserted is shown. d, precise location of the araE, orfX, lysR, and galR genes and the restriction sites derived from the DNA sequences of this work, Stragier and Patte (33), and Stoner and Scalaf (9).

**FIG. 3. Subcloning of the 4.8-kb XhoI-EcoRI fragment into plasmid pBR322, and of a HincII-PvuII fragment into plasmid pAD284, yielding plasmids pMM25, pMM26, and pMM27.** Plasmid pMM25 was used to confirm that the XhoI-EcoRI fragment contained a functional, intact araE gene (see text and Fig. 4). It was the source of the PvuII-HincII fragment for DNA sequencing. This was also ligated into the expression vector pAD284 to obtain the plasmid pMM27 and the control plasmid pMM26 with the insert in the opposite, unexpressed, orientation.
Arabinose-proton symport carried by plasmid pMM25. Strains MM23 and MM23 transformed with plasmid pMM25 were grown and induced with L-arabinose as described under "Experimental Procedures." pH changes promoted by L-arabinose (10 μmol) were measured as described for Fig. 1.

The comparison of membrane and cytoplasmic proteins from induced and uninduced AD5827 (pMM27) by SDS-polyacrylamide gel electrophoresis. Membrane and cytoplasmic fractions from 100-ml cultures of thermo-induced and uninduced AD5827 (pMM27) (about 40 mg dry mass cells) were prepared ("Experimental Procedures"). The cytoplasmic fractions were freeze-dried and resuspended in about 2 ml of SDS-dissolving buffer (37 °C, 20 min); the membrane pellet was treated in the same way. The proteins in 50-μl samples were separated in a 1.5-mm-thick gel ("Experimental Procedures"). Lane 1, β-galactosidase (M, 116,000), bovine serum albumin (M, 68,000), ovalbumin (M, 43,000), carbonic anhydrase (M, 29,000), and lysozyme (M, 14,500); lane 2, uninduced cytoplasmic fraction; lane 3, uninduced membrane fraction; lane 4, induced membrane fraction; lane 5, induced cytoplasmic fraction.

The identification of overproduced AraE protein by labeling with radioactive N-ethylmaleimide. A dual isotope-labeling strategy was used to compare arabinose (100 mM)-protected with unprotected membrane proteins of heat-induced intact cells of strain AD5827 (pMM27) (details under "Experimental Procedures" and in Ref. 8). a, unprotected cells were reacted with [3H]N-ethylmaleimide and protected cells with [14C]N-ethylmaleimide, and the mixed membrane proteins were separated by SDS-polyacrylamide gel electrophoresis. The gel was cut into slices, the [3H]/[14C] content of each slice determined, and the normalized data (8) were plotted as the percentage difference per slice against M, of the protein(s) in the slice, and the same preparation as in a but with the labels reversed, so that unprotected cells were reacted with [14C]N-ethylmaleimide and protected cells with [3H]N-ethylmaleimide. The calculation of Protected-Unprotected (%) is explained in Ref. 8.

incomplete open reading frame orfY (15 amino acids) was the C-terminal end of the araE coding sequence (33). This was later confirmed by DNA sequencing (below). The position of the secondary attachment site where λd(araE' lysA' galR') originally inserted was also deduced (Fig. 2c).

The Subcloning of the araE Gene and Biochemical Characterization of Its Phenotype—A 4.8-kb XhoI-EcoRI restriction fragment (Fig. 2c) large enough to contain the araE gene had ends convenient for cloning into the high copy number vector pBR322 (34). The λd(araE' lysA' galR') DNA was digested with these two enzymes and ligated into pBR322 digested with EcoRI and SalI (Fig. 3). The ligation mixture was transformed into strain DH1 (Table I) and Ap" recombinants screened by preparation of plasmid DNA and digestion with
The DNA sequence of the 2.8-kb *HindII*-PvuII restriction fragment and the predicted amino acid sequence of the AraE protein. The locations of the PvuII, BamHI, and *HindII* restriction sites are shown for comparison with Figs. 2 and 3. Bases 118-517 are identical to the sequence of Stoner and Schleif (9) except for two small discrepancies (opposite).
restriction enzymes. The DNA of the resultant plasmid, pMM25, was isolated and the positions of its restriction sites confirmed (Fig. 3).

When the araE araF recA strain, MM23 (Table I), was transformed with plasmid pMM25, it acquired both arabinose uptake and arabinose-proton symport (Table IV, Fig. 4). Curiously, the transformants were not Ara" as measured by growth of the strain on minimal arabinose plates, although the specific growth rate of this strain in minimal arabinose liquid cultures was enhanced (0.43 h\(^{-1}\)) compared with the untransformed strain (0.09 h\(^{-1}\)). The absence of enhanced growth on solid media was difficult to explain, but the other phenotype characteristics clearly confirmed the presence of the araE gene on plasmid pMM25.

From the expected position of the fusion with lacZ (Fig. 2) and our suggestion (above) that orfY (33) was the C-terminal end of araE, it was deduced that the intact araE gene was located within a 2.8-kb HindII-PvuII restriction fragment (Figs. 2 and 3). This was isolated and ligated into the expression vector pAD284 in strain AD5827 ("Experimental Procedures," Fig. 3). Two plasmids were obtained, pMM27 in which the araE gene was in the correct orientation for expression from Pl, and pMM26 with the reverse, incorrect, orientation (Fig. 5). Plasmid pMM26 was important for control experiments as it enabled the same host with the same DNA (vector and insert) to be subjected to the induction conditions.

Transport of [\(^{14}\)C]arabinose and arabinose-proton symport activities of strain AD5827 (pMM27), but not of strain AD5827 (pMM26), were increased after heat shock (Table IV). This confirmed the successful expression of araE from Pl.

Analysis of the Membrane Proteins of Induced Overexpressing Cells—Membranes were prepared from samples of the following cultures: thermo-induced AD5827 (pMM27), uninduced AD5827 (pMM27), and thermo-induced AD5827 (pMM26). These were solubilized (37 °C, 20 min) and the proteins separated by SDS-polyacrylamide gel electrophoresis (Fig. 5). In several experiments a protein of apparent M, 35,000 ± 1,150 was present in induced AD5827 (pMM27) but not in either control (see, e.g., Fig. 5), consistent with the apparent M, found by Macpherson et al. (6). Furthermore, this protein (M, 35,000, 35,000, 37,000, and 38,000 in four experiments) was protected against reaction with N-ethylmaleimide by arabinose (Fig. 6), confirming that it was the araE gene product (6). AraE was about 7% of the total membrane proteins (estimated by scanning the absorbance of the Coomassie-stained gel at 560 nm), an amplification factor of about 10-50-fold. The protein could not be detected when solubilization was performed at 100 °C for 1.5 min. Whole cell preparations showed an additional protein of M, 25,000 (30), which may correspond to the product of orfX (calculated M, 25,207 (33)).

DNA Sequencing—The sequence of the 2.8-kb DNA fragment was determined (Fig. 7). At the position of the gene anticipated by the restriction analyses (above) was an open reading frame of 1416 base pairs corresponding to a protein of 472 amino acids. There was total agreement with the partial sequence of orfY determined by Stragier and Patte (33) and two discrepancies with the N-terminal bases determined by Stoner and Schleif (Ref. 9, Fig. 7). The reading frame was confirmed by sequencing the araE-lacZ fusion (Fig. 7) and by its correspondence with the promoter and N terminus determined by Stoner and Schleif (9). At the C-terminal end of araE (orfY, above) was a sequence typical of a p-independent terminator (Fig. 7; Refs. 33 and 35). The other open reading frame in the sequence corresponded precisely to orfX, identified by Stragier and Patte (33), which starts immediately after the terminator. The frequency of use of optimal codons calculated by the method of Ikemura (36) was 0.74, similar to XylE (0.65), LacY (0.62), and MelB (0.57), but lower than values for OmpA (0.92) and Lpp (0.98), which are expressed at much higher levels.

Analysis of the Amino Acid Sequence—As expected for an integral membrane protein AraE was highly hydrophobic (hydropathic index of 0.63; Ref. 37), with 12 hydrophobic domains separated by hydrophilic segments (37). The M, of the protein calculated from the sequence was 51,683, larger than the apparent M, of 36,000 (Figs. 5 and 6). Such discrepancies are shown by other integral membrane proteins (38-40), which appear smaller on SDS-polyacrylamide gels than their sequences predict.

DISCUSSION

The structure and function of membrane proteins is poorly understood compared with soluble proteins. This is largely due to their hydrophobic character, which makes them difficult to examine by standard techniques (41, 42). A further complication with proteins such as AraE is their low abundance in the membrane when fully induced in wild-type strains.

The first discrepancy was:

\[ \text{Stoner and Schleif (9)} \ldots \text{ATGCAGGTTTTTATA} \ldots \]

\[ \text{Present work} \ldots \text{ATGCAGGTTTTTATA} \ldots \]

168 (Number in 2866-base pair sequence)

In this case Stoner and Schleif (9) sequenced only one strand (the sense strand in terms of araE) where there was a compression of three superimposed G-ending nucleotides. The sequence read correctly on the other strand. This did not affect the coding sequence or the control regions of araE.

The second discrepancy was:

\[ \text{Stoner and Schleif (9)} \ldots \text{TTGTTCAGATAAGG} \]

\[ \text{Present work} \ldots \text{TTGTTCAGATAAGGCTGCT} \ldots \]

520

This was located in the structural region of araE at the extreme end of Stoner and Schleif's (9) sequence and was covered only once on one strand by them. In the present work it was covered with complete agreement seven times (six times on the sense and once on the non-sense strand). Stoner and Schleif (9) identified the binding sites for CAP protein (-----), the AraR protein (-----), and RNA polymerase (-----), the mRNA start (****) and a potential ribosome binding site (+++++). A potential p-independent terminator is indicated by boxes around the palindromic sequences (see also Ref. 33). Residues 1803 onward were identical to the sequences of Stragier and Patte (33), whose identification of the orfX and lysR reading frames is shown. Residues 1-1270 were also sequenced in a BamHI fragment derived from phage Ap2 of araE-lac and the point of fusion is shown by a vertical arrow (D. C. M. Moore, unpublished data).
We circumvented these problems by cloning the araE gene, by sequencing its DNA, and by overexpressing its product in a λPL vector. The cloning strategy illustrated the utility of λpacMu phage (11) for locating and isolating an individual gene. Furthermore, comparison of the derived physical maps and DNA sequence with the published sequences (9, 35) established the precise position and direction of transcription of araE on the E. coli linkage map (Fig. 2d).

Knowledge of the primary sequence of the AraE protein (Fig. 7) enabled us to construct a model of the folding of the protein with 12 α-helices through the membrane. This was based on the structural algorithms of Eisenberg (41) and Robinson (43) and on comparisons with the homologous glucose transporters from mammals (38, 44, 45) and the XylE xylose-proton symporter from E. coli (14, 38). Hence, these membrane transport proteins of E. coli provide convenient model systems for investigating the molecular mechanism of sugar transport in man. They also show similarities with the E. coli citrate-proton symporter (38) and the plasmid- or transposon-encoded tetracycline-proton antiporters (30, 46, 47).

Such comparative data permit the prediction of some functionally important residues, the roles of which might be investigated by site-specific mutagenesis (see e.g. Ref. 48). Pure protein obtained in large amounts from the overexpressing plasmid will facilitate parallel structural, biochemical, and immunological studies to validate models based on the sequences (14, 44).

Acknowledgments—We thank Drs. E. O. Davis and D. C. M. Moore for discussion and assistance during this work. Drs. P. Stragier and J. C. Patte kindly provided a restriction map of the lysA region prior to its publication.

REFERENCES

1. Schleif, R. (1989) J. Mol. Biol. 46, 185–196
2. Clark, A. F., and Hogg, R. W. (1981) J. Bacteriol. 147, 920–924
3. Kolodrubetz, D., and Schleif, R. (1981) J. Mol. Biol. 151, 215–227
4. Kolodrubetz, D., and Schleif, R. (1981) J. Bacteriol. 148, 472–479
5. Darwallas, K. R., Paxton, A. T., and Henderson, P. J. F. (1981) Biochem. J. 200, 611–627
6. Macpherson, A. J. S., Jones-Mortimer, M. C., and Henderson, P. J. F. (1981) Biochem. J. 196, 269–283
7. Bachmann, B. J. (1983) Microbiol. Rev. 47, 180–230
8. Henderson, P. J. F., and Macpherson, A. J. S. (1986) Methods Enzymol. 125, 387–429
9. Stoner, C., and Schleif, R. (1983) J. Mol. Biol. 171, 369–381
10. Miller, J. H. (1972) Experiments in Molecular Genetics, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
11. Bremer, E., Silhavy, T. J., Weisemann, J. M., and Weinstock, G. M. (1984) J. Bacteriol. 158, 1084–1093
12. Ashworth, J. M., and Kornberg, H. L. (1966) Proc. R. Soc. Lond. B Biol. Sci. 165, 179–188
13. Henderson, P. J. F., Giddens, R. A., and Jones-Mortimer, M. C. (1977) Biochem. J. 162, 309–320
14. Davis, E. O., and Henderson, P. J. F. (1987) J. Biol. Chem. 262, 13926–13932
15. Murray N. E. (1983) in Lambda II (Hendrix, R. W., Roberts, J. W., Stahl, F. W., and Weisberg, R. A., eds) pp. 395–432, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
16. Silhavy, T. J., Berman, M. L., and Enquist, L. W. (1984) in Experiments with Gene Fusions, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
17. Hanshan, D. (1983) J. Mol. Biol. 166, 557–580
18. Englesberg, E., Anderson, R. L., Weinberg, R., Lee, N., Hoffee, R., Huttenhauzer, G., and Boyer, H. (1983) J. Bacteriol. 154, 157–146
19. Englesberg, E., Irr, J., Power, J., and Lee, N. (1965) J. Bacteriol. 90, 946–957
20. Brown, C. E., and Hogg, R. W. (1972) J. Bacteriol. 111, 606–613
21. Heffner, L., Bass, R., and Englesberg, E. (1976) J. Bacteriol. 126, 1119–1131
22. Matiasis, T., Fritsch, E. F., and Sambrook, J. (1982) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
23. Davis, R. W., Botstein, D., and Roth, J. R. (1980) Advanced Bacterial Genetics, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
24. Dretzen, G., Ballard, M., Sassone-Corsi, P., and Chambon, P. (1981) Anal. Biochem. 112, 290–299
25. Davis, E. O., Jones-Mortimer, M. C., and Henderson, P. J. F. (1984) J. Biol. Chem. 259, 1520–1525
26. Walthot, B., Boekhout, M., Brock, M., Kingma, J., van Heerikhuizen, H., and de Leij, L. (1976) Anal. Biochem. 74, 160–170
27. Sanger, F., Couchon, A. R., Barrell, B. G., Smith, A. J. H., and Roe, B. A. (1983) J. Mol. Biol. 143, 161–178
28. Bankier, A. T., and Barrell, B. G. (1983) in Techniques in the Life Sciences, B6, Nucleic Acid Biochemistry B608 (Flavell, R. A., ed) pp. 1–34, Elsevier Scientific Publishers, Ireland
29. Olesen, A., Moks, T., Uhlen, M., and Gaal, A. B. (1984) J. Biochem. Biophys. Methods 10, 83–90
30. Maiden, M. C. J. (1986) Arabinose-Proton Symport in Escherichia coli. Ph.D. Thesis, University of Cambridge
31. West, I. C. (1970) Biochem. Biophys. Res. Commun. 41, 655–661
32. West, I. C., and Mitchell, P. (1972) J. Bioenerg. 3, 445–462
33. Stragier, P., and Patte, J.-C. (1983) J. Mol. Biol. 168, 333–350
34. Bolivar, F., Rodrigues, R. L., Greene, P. J., Betlock, M. C., Heyneker, H. L., Cross, J. H., and Falkow, S. (1977) Gene (Amst.) 3, 95–100
35. Rosenberg, M. and Court, D. (1979) Annu. Rev. Genet. 13, 319–383
36. Ikemura, T. (1981) J. Mol. Biol. 151, 389–409
37. Kyte, J., and Doolittle, R. F. (1982) J. Mol. Biol. 157, 105–132
38. Maiden, M. C. J., Davis, E. O., Baldwin, S. A., Moore, D. C. M., and Henderson, P. J. F. (1987) Nature 325, 641–643
39. Ehren, R., Beyreuther, K., Wright, J. K., and Overath, P. (1980) Nature 283, 537–540
40. Rizolari, L. L., le Maire, M., Reynolds, J. A., and Tanford, C. (1976) Biochemistry 15, 3433–3437
41. Eisenberg, D., Schwarz, E., Komaromy, M., and Wall, R. (1984) J. Mol. Biol. 179, 125–142
42. Walker, J. E., and Fearnley, I. M. (1986) in Techniques for the Analysis of Membrane Proteins (Ragan, C. I., and Cherry, R., eds) pp. 235–273, Chapman and Hall, London
43. Garnier, J., Osguthorpe, D. J., and Robson, B. (1978) J. Mol. Biol. 120, 97–120
44. Mueckler, M., Caruso, C., Baldwin, S. A., Panico, M., Blench, I., Morris, H. R., Allard, W. J., Lienhard, E., and Lobish, H. F. (1985) Science 229, 941–945
45. Birnbaums, M. J., Haspel, H. C., and Rosen, O. M. (1986) Proc. Natl. Acad. Sci. U. S. A. 83, 5784–5788
46. Foster, T. J. (1983) Microbiol. Rev. 47, 361–409
47. Hillen, W., and Schollmeier, K. (1983) Nucleic Acids Res. 11, 595–539
48. Sarkar, H. K., Viitanen, P. V., Padan, E., Trumble, W. R., Poonian, M. S., McComas, W., and Kaback, H. R. (1983) Methods Enzymol. 125, 214–230
49. Mott, J. E., Grant, R. A., Ho, Y.-S., and Platt, T. (1985) Proc. Natl. Acad. Sci. U. S. A. 82, 88–92
50. Davis, E. O. (1986) Xylose Transport in Escherichia coli. Ph.D. Thesis, University of Cambridge