JOURNAL OF BACTERIOLOGY, Sept. 1998, p. 4821–4827 Vol. 180, No. 18
0021-9193/98/$04.00+0

Topological Analysis of DcuA, an Anaerobic C_4-Dicarboxylate Transporter of *Escherichia coli*

PAUL GOLBY,[1] DAVID J. KELLY,[2] JOHN R. GUEST,[2] AND SIMON C. ANDREWS[1]*

School of Animal and Microbial Sciences, University of Reading, Reading RG6 6AJ,[1] and Department of Molecular Biology and Biotechnology, The Krebs Institute for Biomolecular Research, University of Sheffield, Sheffield SI0 2TN,[2] United Kingdom

Received 30 April 1998/Accepted 6 July 1998

Escherichia coli **possesses three independent anaerobic C_4-dicarboxylate transport systems encoded by the *dcuA*, *dcuB*, and *dcuC* genes. The *dcuA* and *dcuB* genes encode related integral inner-membrane proteins, DcuA and DcuB (433 and 446 amino acid residues), which have 36% amino acid sequence identity. A previous amino acid sequence-based analysis predicted that DcuA and DcuB contain either 12 or 14 transmembrane helices, with the N and C termini located in the cytoplasm or periplasm (S. Six, S. C. Andrews, G. Unden, and J. R. Guest, J. Bacteriol. 176:6470–6478, 1994). These predictions were tested by constructing and analyzing 66 DcuA-BlaM fusions in which C terminally truncated forms of DcuA are fused to a β-lactamase protein lacking the N-terminal signal peptide. The resulting topological model differs from those previously predicted. It has just 10 transmembrane helices and a central, 80-residue cytoplasmic loop between helices 5 and 6. The N and C termini are located in the periplasm and the predicted orientation is consistent with the "positive-inside rule." Two highly hydrophobic segments are not membrane spanning: one is in the cytoplasmic loop; the other is in the C-terminal periplasmic region. The topological model obtained for DcuA can be applied to DcuA homologues in other bacteria as well as to DcuB. Overproduction of DcuA to 15% of inner-membrane protein was obtained with the *lac*UV5-promoter-based plasmid, pYZ4.**

Escherichia coli can utilize C_4-dicarboxylates as sole carbon and energy sources under both aerobic and anaerobic conditions (7). Aerobically, uptake of C_4-dicarboxylates (fumarate, malate, and succinate) and L-aspartate is mediated by a secondary transporter, designated DctA (14, 18). The corresponding gene, *dctA*, has been sequenced, and the role of its product in the utilization of C_4-dicarboxylates (and the cyclic monocarboxylate orotate) has been established by complementation studies in *Salmonella typhimurium dctA* or *outA* mutants (2, 28).

Uptake, exchange, and efflux of C_4-dicarboxylic acids under anaerobic conditions is mediated by the Dcu systems (K_m for fumarate uptake = 51 μM), which are genetically distinct from the aerobic Dct system (7, 8, 34). Measurements of both C_4-dicarboxylate uptake and exchange have suggested that the Dcu systems are exclusively expressed under anaerobic conditions, activated by the anaerobic activator protein FNR, and repressed in the presence of nitrate. Three independent Dcu systems have been identified, DcuA, DcuB, and DcuC (27, 34). DcuA and DcuB are homologous proteins (36% identical), whereas DcuC is only 22 to 24% identical to DcuA and DcuB. Growth tests and transport studies with *dcuA*, *dcuB*, and *dcuC* single, double, and triple mutants showed that DcuA, DcuB, and DcuC each mediate exchange as well as uptake (27, 34). The triple mutants were almost completely devoid of Dcu activity. The single mutants exhibited no phenotype, but the *dcuA-dcuB* double mutant displayed marked deficiencies in C_4-dicarboxylate transport and growth by fumarate respiration, suggesting that DcuA and DcuB have analogous and mutually complementary transport functions in the anaerobic uptake of C_4-dicarboxylates (27, 34). The affinities of DcuA and DcuB for C_4-dicarboxylates are similar, except for the lower affinity of DcuA for malate (27).

DcuA and DcuB have 433 and 446 amino acid residues, respectively, and their sequences suggest that they are highly hydrophobic and lack N-terminal signal sequences, which together indicate that they are polytopic inner-membrane proteins (27). A combination of SOAP, Helixmem, hydropathy plot, and "von Heijne positive-inside rule" analyses were used to predict that DcuA and DcuB have either 12 or 14 transmembrane spanning helices and that their N and C termini are located in the cytoplasm or periplasm (27).

The reasons for having three independent Dcu systems in *E. coli* and their specific roles in anaerobic C_4-dicarboxylate transport are unknown. In particular, the presence of the homologous and apparently mutually redundant DcuA and DcuB systems remains to be explained. Homologues of DcuA and DcuB are present in *Serratia marcescens*, *Haemophilus influenzae* (two homologues), *Wolinella succinogenes*, *Helicobacter pylori*, and *Salmonella typhimurium*, suggesting that the DcuA-like transporters are widespread among the proteobacteria. However, members of the DcuA family have no significant sequence similarity to members of other transporter families, indicating that the DcuA-like proteins are a distinct group (1). In order to better understand the roles and properties of DcuA-like transporters and to provide a robust topological model that would facilitate more accurate comparisons between this group and other transporter families, the topological organization of DcuA within the inner membrane has been analyzed with in-frame translational fusions between a series of progressively truncated forms of the *dcuA* gene and a downstream "reporter gene" (*blaM*, encoding β-lactamase). The resulting topological model differs from the predicted models and indicates that the DcuA-like transport proteins represent a unique subgroup of the "duo-decimal transporters" (DDTs).

* Corresponding author. Mailing address: School of Animal and Microbial Sciences, University of Reading, Whiteknights, P.O. Box 228, Reading RG6 6AJ, United Kingdom. Phone: 44 118-987-5123, ext. 7045. Fax: 44 118-931-0180. E-mail: S.C.Andrews@reading.ac.uk.

MATERIALS AND METHODS

Media, growth conditions, strains, vectors, and general methods. Standard genetic procedures were performed (25) with *E. coli* DH5α (genotype) as the transformation host. *E. coli* was grown at 37°C in Luria broth (LB; 250 rpm) or Luria agar containing antibiotics where appropriate (ampicillin [Ap], 100 μg ml^{-1}; kanamycin [Kn], 50 μg ml^{-1}). Amino acid sequence similarity searches were performed by using the BLAST program with the nonredundant databases at the Swiss-BLAST interface of the National Center for Biotechnology Information (Bethesda, Md.). Amino acid sequences were analyzed at the Seqnet Computing Facility (Daresbury, Warrington, United Kingdom) with the GCG and EGCG programs (5) (Peter Rice, Sanger Centre, Cambridge, United Kingdom).

The plasmid pGS745 was constructed by cloning the 6-kb *Sph*I-*Sal*I *dcuA*-containing fragment of plasmid pGS73 (11) into the corresponding sites of pBR322.

Subcloning the *dcuA* gene into pYZ4. A 1.3-kb DNA fragment containing the *dcuA* coding region was PCR amplified by using the *dcuA*-containing plasmid pGS73 as a template, *Pfu* DNA polymerase (Stratagene), and two primers—DCUA2, 5′-CCGAATTCC^{2129}*ATG*GTAGTTGTAGAACTC^{2146}-3′; and DCUA4, 5′-CCGGATCC^{3436}TGATCATTACAGCATGAAG^{3418}-3′ (start codon is underlined, mismatches are in small capitals, *Eco*RI and *Bam*HI sites are in boldface, the *Nco*I site is in italics, and the coordinates are from Six et al. [27])—designed to introduce flanking *Eco*RI-*Nco*I and *Bam*HI sites. The *Eco*RI- and *Bam*HI-digested PCR product was subcloned initially into plasmid pUC118 (31), generating pGS748. Plasmid pGS767 was then constructed by subcloning the 1.3-kb *Nco*I-*Bam*HI *dcuA* fragment of pGS748 into the corresponding sites of the Kn^r plasmid pYZ4 (33) such that the *dcuA* gene is appropriately positioned downstream of the IPTG (isopropyl-β-D-thiogalactopyranoside)-inducible *lac*UV5 promoter.

Construction of *dcuA-blaM* fusions. Nested 3′ deletions of *dcuA* were generated by digesting *Bam*HI- and *Kpn*I-treated pGS767 with exonuclease III using the Double-Stranded Nested Deletion Kit (Pharmacia) according to the manufacturer's instructions. The products were digested with *Sac*I and ligated with the 0.85-kb *Sma*I-*Sac*I *blaM* cassette from the plasmid pLH21 (supplied by J. K. Broome-Smith). The ligation products were used to transform competent *E. coli* TG1, and Kn^r transformants were selected. Kn^r transformants were tested for growth when inoculated at low or high density on Luria agar containing Ap (35 μg ml^{-1}) plus Kn. The MICs of Ap for transformants carrying *dcuA-blaM* fusions were determined as previously described (33). Plasmids were isolated from Ap^r Kn^r transformants and analyzed by restriction enzyme mapping to determine the approximate positions of *dcuA-blaM* fusion sites.

Seventeen codon-specific fusions were made by amplifying the desired *dcuA* fragment by PCR as before, except that codon-specific primers (26-mers, with the *Eco*RV site-containing sequence CCGATATC at the 5′ termini, and 18 homologous bases at the 3′ termini) were used in place of DCUA4. The 0.4- to 1.3-kb *Nco*I- and *Eco*RV-treated PCR fragments and the 0.85-kb *SmaI-SacI blaM* cassette of pLH21 were coligated into the corresponding sites of pYZ4, and the desired transformants were selected and treated as described above. It should be noted that the PCR-generated fusions contain an Asp-Gly linker derived from codons at the *Eco*RV-*Sma*I hybrid site, whereas the exonuclease III-generated fusions contain a linking Gly residue specified by the *Sma*I half site.

DNA sequencing. DNA was sequenced by the chain termination method with the Sequenase kit (U.S. Biochemicals, Ltd.). The *dcuA-blaM* junctions were sequenced with a primer (BLAM1, 5′-CTCGTGCACCCAACTGA-3′) identical to codons 14 to 18 of *blaM*.

Immunodetection of DcuA-BlaM fusion proteins. TG1 transformants expressing *dcuA-blaM* fusions were grown on 0.4% glucose minimal medium plates plus Kn (50 μg ml^{-1}) for ~36 h. Samples of bacteria were then taken from the plates, resuspended in 1 ml of ice-cold saline, centrifuged (10,000 × *g*, 10 min), and resuspended in buffer A (100 mM Tris-HCl, pH 6.8; 5% [vol/vol] β-mercaptoethanol; 1% bromophenol blue; 7% [vol/vol] glycerol) to give an optical density at 650 nm of 5. Samples (10 μl) were incubated at 37°C for 1 h (or boiled for 10 min), centrifuged as described above, and fractionated by electrophoresis in a sodium dodecyl sulfate (SDS)–12.5% polyacrylamide gel. The proteins were then transferred to a nitrocellulose membrane (Schleicher and Schuell-Protan BA83) with a Trans-Blot electrophoretic transfer cell according to the manufacturer's instructions (Bio-Rad). The β-lactamase fusions were detected by using the Bio-Rad protocol (Immuno-Blot Assay Kit) except that a 1/500 dilution of anti-ampicillinase polyclonal antiserum (5 Prime→3 Prime, Inc.) was used as the primary antibody and a 1/30,000 dilution of alkaline-phosphatase-conjugated anti-rabbit antibody (Sigma) was the secondary antibody.

Preparation of membrane fractions. TG1 transformants were grown at 250 rpm for 16 h in ca. 300 ml of LB plus Kn at 37°C. The cultures were harvested by centrifugation (3,000 × *g*, 30 min), and membrane fractions were prepared by sucrose density-gradient centrifugation as previously described (21). The membrane fractions were analyzed by SDS-polyacrylamide gel electrophoresis (PAGE). Protein concentrations were determined by a modification of the Lowry method (22).

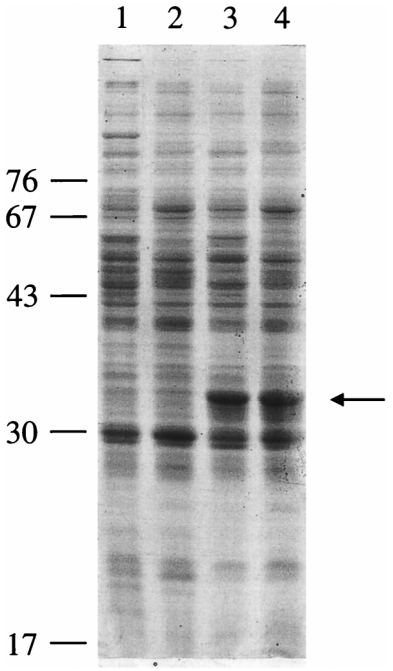

FIG. 1. SDS-PAGE analysis of membrane proteins produced by TG1(pGS767). Samples (50 μg of protein) containing total-membrane and inner-membrane fractions (see Materials and Methods) were fractionated by SDS-PAGE (12.5%) and stained with Coomassie brilliant blue R250. Prior to electrophoresis, samples were dissolved in SDS-containing loading buffer (see Materials and Methods) and incubated for 1 h at 37°C. Lanes 1 and 2 show total-membrane and inner membrane proteins, respectively, of TG1(pYZ4); lanes 3 and 4 show total-membrane and inner-membrane proteins, respectively, of TG1(pGS767). The positions of molecular mass markers and DcuA (arrow) are indicated.

RESULTS AND DISCUSSION

Overproduction of DcuA. A 1.3 kb *Nco*I-*Bam*HI PCR-generated DNA fragment containing the *dcuA* coding region was cloned immediately downstream of, and with the same polarity as, the IPTG-inducible *lac*UV5 promoter of the Kn^r expression vector pYZ4 (33), generating the plasmid pGS767 (see Materials and Methods). Transformants of *E. coli* TG1 carrying pGS767 were analyzed for DcuA overproduction after overnight growth in LB at 37°C in the presence (data not shown) or absence of the inducer IPTG (Fig. 1). A major component (apparent molecular mass, 34 kDa) was present at 10% of total membrane protein (15% of inner-membrane protein). No equivalent component was observed with the control strain, TG1(pYZ4) (Fig. 1), or in whole-cell extracts (data not shown). The "34-kDa" protein probably corresponds to DcuA, which has a predicted molecular mass of 45.8 kDa. The 11.8-kDa (26%) underestimation is consistent with previous reports that integral membrane proteins have higher electrophoretic mobilities than nonmembrane proteins during SDS-PAGE (12, 23, 29). The apparent presence of overproduced DcuA in the inner-membrane fraction suggests that pGS767 can be used for functional overexpression of *dcuA*. However, measurements of [^{14}C]fumarate (Sigma) uptake failed to demonstrate any enhanced Dcu activity for strain TG1(pGS767) when compared to the control strain TG1 (data not shown). This is presumably because DcuA is inactive under aerobic conditions (27).

Construction and analysis of *dcuA′-blaM* fusions. To test the predicted topology of DcuA (27), fusions (*dcuA′-blaM*) were created between a nested series of exonuclease III-truncated *dcuA* derivatives and a *blaM* gene encoding a leaderless β-

lactamase (see Materials and Methods). The MIC of Ap was determined for all Kn^r Ap^r transformants obtained (Fig. 2). Transformants for which the Ap MIC was 5 to 10 μg ml^{-1} were presumed to encode fusion proteins in which the BlaM region remains in the cytoplasm or is located in the membrane, whereas those for which the Ap MIC was ≥10 μg ml^{-1} were considered to encode fusion proteins in which the BlaM segment is exported to the periplasm (2). The junctions of the *dcuA′-blaM* fusions were defined by sequence analysis (see Materials and Methods), and a total of 49 unique in-frame *dcuA′-blaM* fusions were characterized. Although the fusion sites were fairly evenly distributed within the *dcuA* coding region, attempts to isolate fusions in some regions, in particular between codons 140 and 200, were not very successful. The occurrence of such "cold spots" has been noted previously (30). To overcome this problem, 17 directed fusions (Fig. 3) were generated by cloning specific PCR-amplified *dcuA* fragments (together with *blaM*) into pYZ4, and the MICs of the resulting constructs were determined as described above.

Analysis of DcuA′-BlaM fusions. The DcuA hydropathy profile (Fig. 2A) contains 12 highly hydrophobic regions (regions 1 to 9, 5a, 10a, and 10b) separated by 11 hydrophilic regions (regions A to K). The shaded-solid and shaded-broken bars superimposed upon the hydropathy plot denote the 12 well-predicted membrane spanning α-helices (spanners) and the two (spanners 4a and 9a) tentatively predicted spanners, respectively (27). These observations suggest that DcuA consists of either 12 strongly predicted spanners with the N and C termini in the cytoplasm or 14 spanners with the N and C termini located in the periplasm (Fig. 2C) (27). The distribution of positively charged residues in both models shows good agreement with the positive-inside rule (32), although the charge bias is twice as high for the 14-spanner model (Fig. 2C).

The MICs of strains expressing in-frame *dcuA′-blaM* fusions identified five periplasmic regions for DcuA (labeled I to V in Fig. 2B), three of which (I, III, and IV) correspond to hydrophilic regions B, G, and I, respectively, in the hydropathy profile (Fig. 2A). Periplasmic loop II includes the hydrophilic region D and the weakly hydrophobic region 4a (Fig. 2A). This implies that region 4a is not a spanner but instead forms a hydrophobic domain within periplasmic loop II. This is not surprising since 4a is not recognized as a potential spanner by the Helixmem and SOAP programs. Periplasmic segment V extends from residue 384 to residue 433 (the last residue of DcuA) and includes hydrophilic region K, the hydrophobic region 10b, and part of the hydrophobic region 10a (Fig. 2C). This leads to several conclusions: (i) the C terminus is oriented towards the periplasm; (ii) region 10b does not contain a spanner; and (iii) there is only one (not two) spanner (designated spanner 10) in the 9a-to-10a region. The finding that hydrophobic region 10b is nonspanning was unexpected and indicates that this region could form a hydrophobic domain within the C-terminal periplasmic segment that may interact with the membrane or another periplasmic portion of DcuA. Alternatively, it is possible that region 10b is indeed membrane spanning in the "natural" DcuA protein and that the C-terminally linked BlaM domain interferes with the folding or membrane insertion of region 10b, resulting in its false location in the periplasm. To test this possibility, a DcuA-Asp-Gly-$(Gly)_4$-BlaM fusion was generated comprising residues 1 to 433 of DcuA, the BlaM domain, and an Asp-Gly-$(Gly)_4$ flexible linker. The four additional Gly residues (note that all PCR-derived fusions contain the Asp-Gly dipeptide at the fusion site; see Materials and Methods) were included to provide a spacer region that would tether the BlaM domain and region 10b less tightly, thereby minimizing the potential for an interfering interaction. The MIC for the resulting fusion was high (100 μg ml^{-1} with the Gly_4 linker, cf. 50 μg ml^{-1} without the linker; see Fig. 3), indicating that the C terminus is periplasmically located in the presence or absence of the flexible linker. This supports the notion that the hydrophobic 10b region is indeed periplasmic in the natural DcuA protein.

The fusion data were further interpreted by identifying regions where periplasmic loops are separated by a pair of predicted spanners which are themselves separated by a hydrophilic cytoplasmic loop. This revealed three cytoplasmic loops flanked by spanners: 3-C-4, 7-H-8, and 9-J-10 (Fig. 2A). Hydrophobic regions 5 and 6 are likely to be spanners since fusions in these regions gave low MICs suggestive of a cytoplasic location, yet the adjacent hydrophilic regions D and G correspond to periplasmic loops II and III (Fig. 2 and 3). Hydrophobic region 5a must therefore form a hydrophobic domain (that is possibly membrane associated) within the large central cytoplasmic loop that includes hydrophilic regions E and F (Fig. 3). This is surprising given the highly hydrophobic nature of region 5a (Fig. 2A). However, in order to accommodate a spanner in region 5a, a second adjacent spanner would also be required and the fusion data, hydropathy plot, and computer predictions do not support this.

Fusions in the region preceding periplasmic loop I gave MICs of 5 to 10 μg ml^{-1}. These data are difficult to interpret because fusions close to the N terminus are likely to contain insufficient truncated polypeptide to allow meaningful insertion into the membrane (4). However, since the N-terminal region of DcuA contains two strongly predicted and highly hydrophobic spanners, it is likely that hydrophobic regions 1 and 2 are membrane spanning even though spanner 1 is predicted to be of insufficient length to completely span the membrane. Also, since hydrophilic region A is not apparently in the periplasm (and is therefore likely to be cytoplasmic), whereas hydrophilic region B is periplasmic, it is most probable that the N terminus is orientated towards the periplasm as shown in the new model (Fig. 3).

Immunodetection of DcuA′-BlaM fusion proteins. The β-lactamase approach for investigating membrane protein topology is based on the assumption that levels of Ap resistance exhibited by *blaM* fusion strains are directly related to the subcellular locations of the BlaM component of corresponding fusion proteins. However, it is possible that Ap resistance reflects the cellular amounts of the fusion protein, rather than the topology. In order to address this possibility, the amounts of DcuA′-BlaM fusion protein present in the 66 *dcuA′-blaM* strains were compared by anti-β-lactamase Western blot analysis (see Materials and Methods and Fig. 4). Most of the DcuA′-BlaM fusion strains had similar quantities of multiple immunoreactive proteins (data not shown). It was assumed that the proteins with the greatest molecular mass correspond to the full-length fusions, whereas the others represent breakdown products. The instability of topology probe fusions has been documented previously (26, 30). A Western blot analysis of nine representative fusions and two controls is shown in Fig. 4. The observed sizes of the full-length DcuA′-BlaM fusions are similar to the predicted sizes when allowances are made for the abnormal mobility of the DcuA protein. Although the DcuA′-BlaM fusions containing residues 1 to 46, 1 to 110, and 1 to 129 of DcuA conferred high MICs (90 to 115 μg of Ap ml^{-1}), they were present at levels similar to those obtained for the fusions conferring MICs of just 5 to 10 μg of Ap ml^{-1} (Fig. 4). These examples clearly illustrate that there is no correlation between the level of Ap resistance and the amount of DcuA′-BlaM fusion protein. Furthermore, they strongly suggest that the observed level of resistance is related to the subcellular

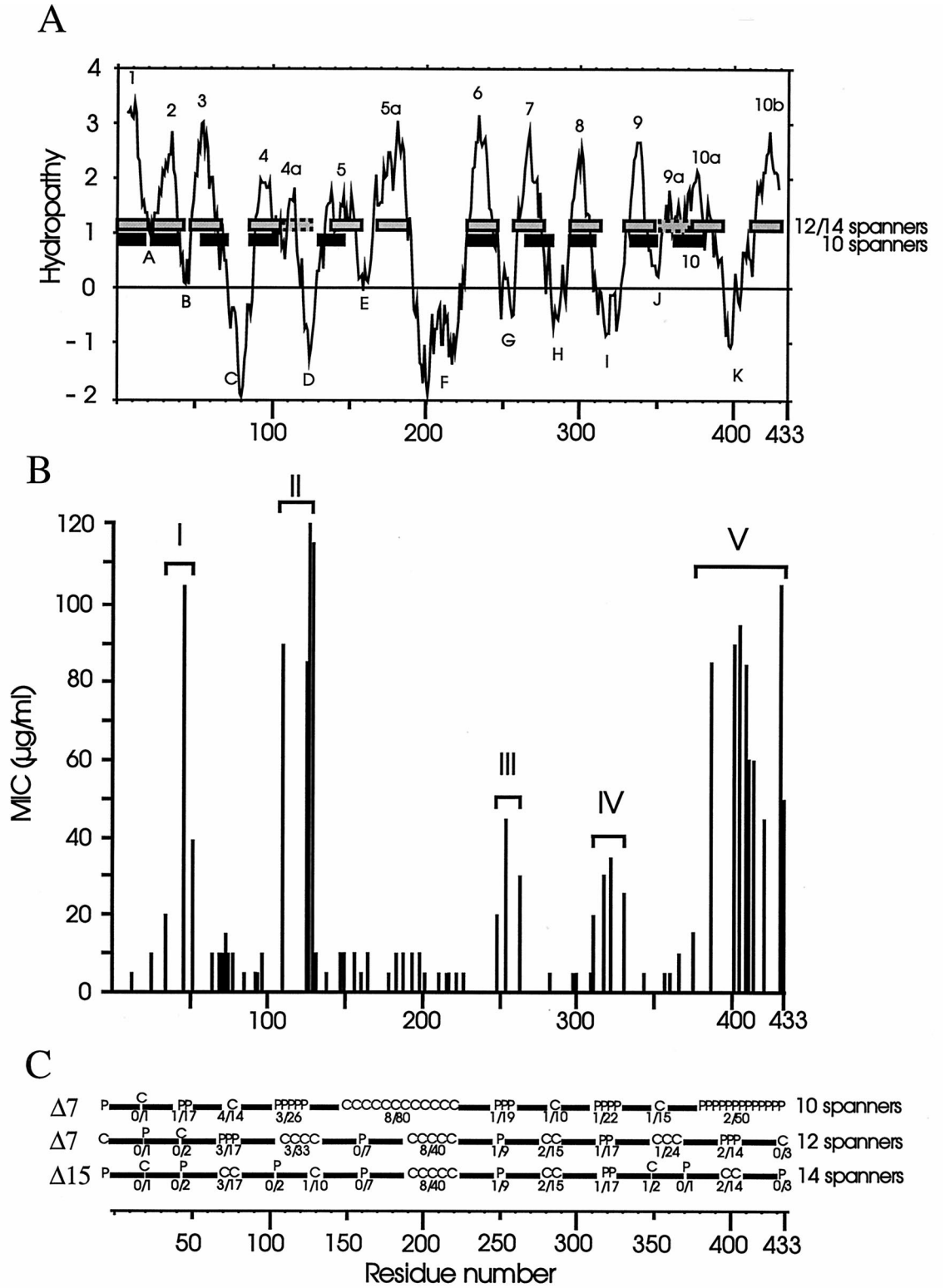

FIG. 2. Hydropathy plot of the DcuA protein and properties of DcuA′-BlaM fusion proteins. (A) Hydropathy was calculated by using the algorithm of Kyte and Doolittle (16) with a window of 11 residues (27). Regions of hydrophobicity (labeled 1 through 12) and hydrophilicity (labeled A through K) are indicated. The 12 shaded solid bars show the putative membrane-spanning helices of the 12-spanner model which were all strongly predicted by the SOAP and Helixmem programs (6, 15). The shaded broken bars denote the two additional helices of the 14-spanner model that are weakly predicted by the hydropathy plot. The 10 solid bars show the helices predicted from the DcuA′-BlaM fusion data. (B) Ap resistance of strains expressing DcuA′-BlaM fusion proteins. Vertical lines represent the Ap MICs conferred by 65 unique DcuA′-BlaM fusion proteins. The BlaM fusion point for each of the DcuA′-BlaM proteins is shown on the horizontal axis. Five regions predicted to be in the periplasm are indicated by the numerals I through IV. (C) Linear representations of the topologies of the computer-derived 12- and 14-spanner models for DcuA, together with the 10-spanner model generated by using the fusion data in Fig. 2B. Solid bars show the predicted helices, "C" indicates the putative cytosolic loops, and "P" shows the potential periplasmic loops. The numbers of Arg+Lys residues, the lengths of the extramembranous segments, and the charge biases (Δ) are indicated.

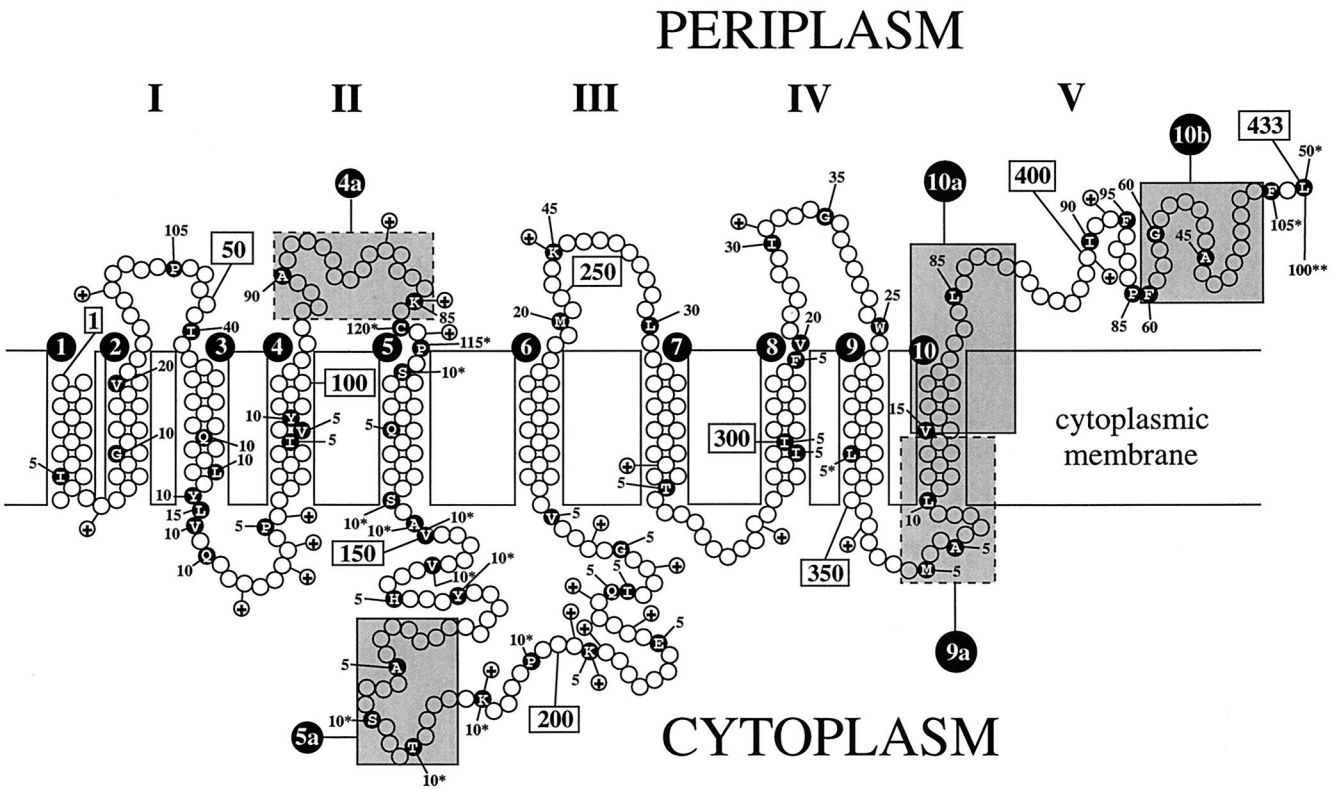

FIG. 3. Proposed model of the membrane topology of the DcuA protein. Ten transmembrane helices (labeled 1 through 10) are connected by four periplasmic loops (labeled I through IV) and five cytoplasmic domains. The N and C termini are located in the periplasm. The nonspanning, highly hydrophobic regions (5a, 10b, and part of 10a) or weakly hydrophobic regions (4a and part of 9a) are enclosed in shaded solid or shaded broken boxes, respectively. Residue numbers are boxed. DcuA′-BlaM fusion positions are indicated by black circles with numbers denoting the MICs of Ap ($\mu g\ ml^{-1}$). Fusions generated by PCR amplification are indicated by a single asterisk, and the fusion containing the Asp-Gly-$(Gly)_4$ linker is indicated by two asterisks. Encircled plus signs denote positively charged residues.

location of BlaM rather than to the amounts of BlaM in the cell. The Ap^s control strain, TG1(pYZ4), produced a weakly staining immunoreactive unknown protein that appears to be present in all the samples. The Ap^r control strain, TG1(pUC118), produced two closely spaced, strongly staining immunoreactive proteins (molecular masses of ca. 29 kDa) corresponding to the immature and mature forms of β-lactamase encoded by the *bla* gene of pUC118.

The DcuA topological model. The DcuA topological model derived from the hydropathic profile, spanner prediction (Fig. 2A), BlaM fusion data (Fig. 2B) and multiple alignment (Fig. 5) contains 10 spanners arranged as two clusters of five helices separated by an 80-residue cytoplasmic loop (Fig. 3). The N terminus is located in the membrane and orientated towards the periplasm, and the C-terminal region consists of a relatively hydrophobic 50-residue periplasmic tail. The 433 residues of DcuA are fairly evenly distributed among the three subcellular compartments: 120 (28%) residues in the cytoplasm, 134 (31%) residues in the periplasm, and 179 (41%) residues in the membrane. This is a typical feature of bacterial secondary transporters (10). In accordance with the positive-inside rule (32), the distribution of positively charged residues shows a strong bias towards the cytoplasm: 8 Arg and Lys residues in the periplasm and 15 in the cytoplasm, giving a charge bias of 7; Arg and Lys residues represent 12.5% of cytosolic residues but only 6% of periplasmic residues (Fig. 2C). This distribution is better than for the 12-spanner model (12% of cytoplasmic and 11% of periplasmic residues are Arg or Lys), but less favorable than for the 14-spanner model, where 17% of cytoplasmic and 5% of periplasmic residues are Arg or Lys res-

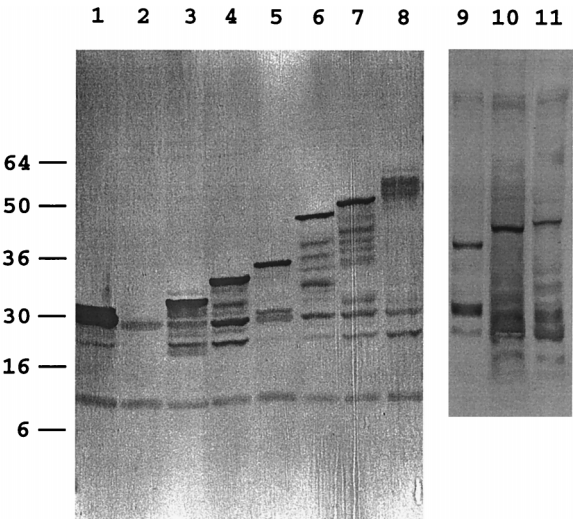

FIG. 4. Immunodetection of DcuA′-BlaM fusion proteins. Extracts of *E. coli* TG1 transformants were prepared as described in Materials and Methods. Lanes 1 to 11 contained whole-cell extracts of strains carrying the following plasmids: 1, pUC118 (Ap^r control); 2, pYZ4 (control); 3 to 11, plasmids encoding DcuA′-BlaM fusions containing the following DcuA residues, with the apparent and expected molecular masses of the fusion proteins (in kilodaltons): 1 to 46 (32 and 35), 1 to 84 (33 and 38), 1 to 110 (35 and 40), 1 to 129 (39 and 44), 1 to 160 (42 and 48), 1 to 183 (44 and 50), 1 to 227 (47 and 53), 1 to 282 (52 and 59), and 1 to 433 (60 and 75). The positions of molecular mass markers are indicated.

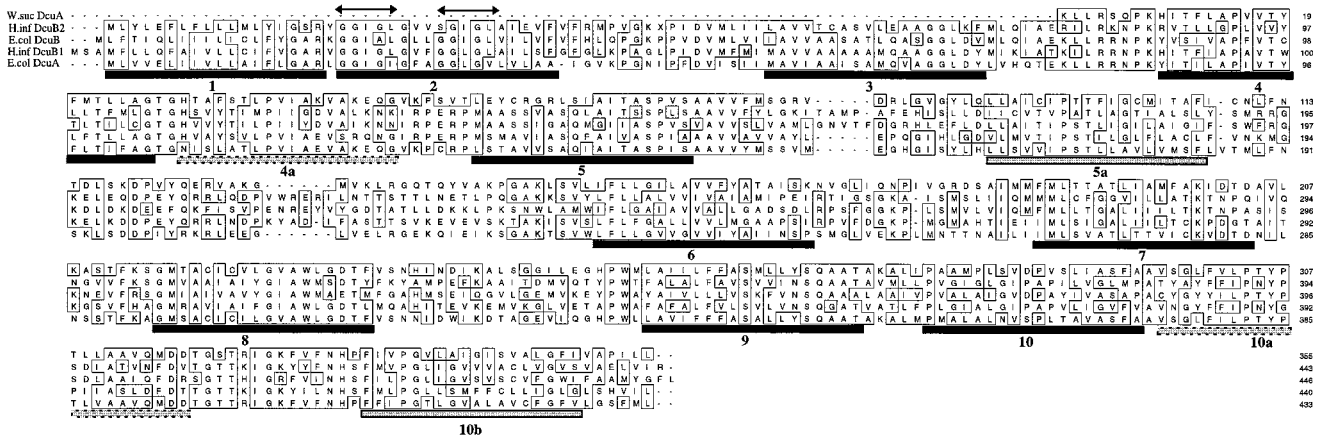

FIG. 5. Multiple alignment and positions of predicted membrane-spanning helices for the DcuA family. The amino acid sequences (and database reference numbers) are as follows: DcuA from *W. succinogenes* (e1186590); DcuB1 and DcuB2 from *H. influenzae* (P44855 and gi1573110); DcuA from *H. pylori* (gi2313848); and DcuA and DcuB from *E. coli* (P04539 and P14409). Sequences were aligned by using PILEUP and displayed by using Prettyplot (5). The *S. typhimurium oxy-d* (U84267) and *S. marcescens* DcuA (P40684) sequences are not included because they are incomplete (36 and 132 residues, respectively) and are very similar (100 and 91% identity) to *E. coli* DcuA and DcuB, respectively. The codon corresponding to residue X (position 46) in the *H. influenzae* DcuB2 sequence contains an extra base which was inserted to generate an in-frame translation in the encoding gene. Regions of sequence similarity are boxed and a conserved, repeated motif (GGIGL) is marked by double-headed arrows. The 10 transmembrane helices of *E. coli* DcuA (1 through 10) are denoted by solid bars, and the highly hydrophobic (5a, 10b, and 10a) and weakly hydrophobic (4a and 9a) regions are indicated by gray solid bars and gray broken bars, respectively.

idues (Fig. 2C). The failure of the previously published analysis to correctly predict the topology of DcuA was due to the false identification of two (or four) membrane-spanning helices (Fig. 2A, 2C, and 3). It is not clear how the two strongly predicted spanners could have been correctly identified as being extramembranous given their highly hydrophobic nature and high scores in the Helixmem analysis. However, their incorrect identification highlights the potential unreliability of current computer-based predictions, showing that such predictions are not satisfactory replacements for experimental data and should be treated with caution.

Many bacterial transporters have 12 membrane-spanning helices, and such proteins are designated "duo-decimal transporters" or DDTs (13). The topological and sequence similarities of DDTs (10, 19), suggest that they possess similar structural and mechanistic properties (19, 24). A common feature of many DDTs (e.g., AraE, UhpT, and XylE, but not DctA, GltP, LacY, and MelB) is the presence of a large, central cytoplasmic loop between spanners 6 and 7 which divides the 12 spanners into two six-helix clusters. The topology of DcuA conforms to this consensus, except that there are only five helices in each half of the protein (i.e., helices 1 and 12 are absent). This suggests that DcuA could have structural and functional properties similar to those of the DDTs.

The presence of N and C termini in the periplasm is unusual for a membrane transport protein. Other membrane proteins possessing periplasmic termini include leader peptidase (20), F_1F_0 ATP synthase (17), and the rhamnose sugar transport protein RhaT (30). Interestingly, topological analysis revealed that RhaT, like DcuA, possesses 10 spanners. However, it lacks a C-terminal tail, its central loop contains only 25 amino acid residues, and there is no sequence similarity between DcuA and RhaT. Therefore, DcuA and RhaT appear to be structurally distinct.

DcuA family. Database searches revealed that there are six *dcuA* homologues (two completely sequenced and two partially sequenced) in other bacteria. Most of these genes, like the *E. coli dcuA* and *dcuB* genes, are adjacent to genes encoding enzymes with roles in C_4 metabolism: *dcuA* of *S. marcesens* is next to an aspartase (*aspA*) gene; and one (*dcuB1*) of the two *dcuB* genes of *H. influenzae* and the *dcuA* genes of *W. succinogenes* and *H. pylori* are next to genes (*asnB* or *asnA*) that apparently encode a periplasmic L-asparaginase and cytosolic L-asparaginases, respectively. The other *dcuB* gene (*dcuB2*) of *H. influenzae* is adjacent to genes with no known role in C_4 metabolism, and the location of the *dcuB* homologue (*oxd*-5) of *S. typhimurium* has not been reported, but it is likely to be next to the *fumB* gene, as in *E. coli*. Thus, of the eight genes encoding DcuA-like proteins, it appears that seven are linked to genes involved in the interconversion of asparagine, aspartate, fumarate, and malate. This is fully consistent with the deduced roles of the *E. coli* DcuA and DcuB proteins in the transport of C_4-dicarboxylates and supports the notion that DcuA and DcuB homologues from other bacteria have roles in C_4-dicarboxylate transport. It also indicates that those *dcuA* genes adjacent to asparaginase genes could encode transporters with specificities for asparagine as well as, or instead of, C_4-dicarboxylates. This would be consistent with many previous examples of close linkage between genes encoding transport systems and enzymes that utilize the same substrates.

The six DcuA and DcuB sequences aligned in Fig. 5 have a high degree of sequence similarity (32 to 57% identity), evenly distributed throughout the membrane-spanning and extramembranous sections. There are only eight regions where padding characters are required to optimize the alignment, and these lie in loop regions: four in the central cytoplasmic loop, two between spanners 9 and 10 in the cytoplasmic loop, and one each between spanners 2 and 3 and spanners 6 and 7 in periplasmic loops I and III. The alignment indicates that all six proteins have a similar topological organization. The hydropathy profiles and predicted membrane-spanning helices are likewise very similar, indicating that the topological model produced for *E. coli* DcuA can be used to predict the topologies of the other family members (Fig. 5). Interestingly, spanner 2 contains a highly conserved and duplicated motif, GGIGL, of unknown function (Fig. 5).

Sequence searches of available databases using either single amino acid sequences or a "profile" (9) generated from the alignment in Fig. 5 failed to detect any other proteins with significant sequence similarity to those in the DcuA family. This contrasts with a previous report suggesting that the DcuA, DcuC, and DctA proteins (34) are homologous. Furthermore,

since the protein fusion-derived topology model of the DctA protein of *Rhizobium meliloti* contains 12 spanners, with C and N termini in the cytosol, no cytoplasmic central loop, and no periplasmic C-terminal tail, it is unlikely that the DcuA and DctA families are closely related.

Conclusion. Members of the DcuA family of bacterial polytopic transporters are apparently associated with the metabolism of C_4 compounds and are presumed to mediate the exchange, uptake, and export of C_4-dicarboxylates. Their topological organization suggests that they have structural similarities to the bacterial DDT proteins, but there are sufficient differences to warrant classification as a novel subgroup of DTT-like proteins currently containing eight members from six proteobacterial species.

ACKNOWLEDGMENTS

This work was supported by a BBSRC project grant (S.C.A. and J.R.G.) and a BBSRC Advanced Fellowship (S.C.A.).

We thank J. Broome-Smith for the supply of plasmids pYZ4 and pLH21 and R. Seabrook and R. Trewinnard for technical assistance.

REFERENCES

1. **Andrews, S. C.** Unpublished data.
2. **Baker, K. E., K. P. Ditullio, J. Neuhard, and R. A. Kelln.** 1996. Utilization of orotate as a pyrimidine source by *Salmonella typhimurium* and *Escherichia coli* requires the dicarboxylate transport protein encoded by *dctA*. J. Bacteriol. **178:**7099–7105.
3. **Broome-Smith, J. K., and B. G. Spratt.** 1986. A vector for the construction of translational fusions to TEM β-lactamase and the analysis of protein export signals and membrane protein topology. Gene **49:**341–349.
4. **Calamia, J., and C. Manoil.** 1990. *lac* permease of *Escherichia coli*: topology and sequence elements promoting membrane insertion. Proc. Natl. Acad. Sci. USA **87:**4937–4941.
5. **Devereux, J.** 1989. The GCG sequence analysis software package, version 6.0. Genetics Computer Group, Inc., Madison, Wis.
6. **Eisenberg, D., E. Schwarz, M. Komaromy, and R. Wall.** 1984. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. J. Mol. Biol. **179:**125–142.
7. **Engel, P., R. Krämer, and G. Unden.** 1992. Anaerobic fumarate transport in *Escherichia coli* by an *fnr*-dependent dicarboxylate uptake system which is different from aerobic dicarboxylate uptake. J. Bacteriol. **174:**5533–5539.
8. **Engel, P., R. Krämer, and G. Unden.** 1994. Transport of C_4-dicarboxylates by anaerobically grown *Escherichia coli*: energetics and mechanism of exchange, uptake and efflux. Eur. J. Biochem. **222:**605–614.
9. **Gribskov, M., R. Luthy, and D. Eisenberg.** 1990. Profile analysis. Methods Enzymol. **183:**146–159.
10. **Griffith, J. K., M. E. Baker, D. A. Rouch, M. G. P. Page, R. A. Skurray, L. Paulsen, K. F. Chater, S. A. Baldwin, and P. J. F. Henderson.** 1992. Membrane transport proteins: implications of sequence comparisons. Curr. Opin. Cell Biol. **4:**684–695.
11. **Guest, J. R., R. E. Roberts, and R. J. Wilde.** 1984. Cloning of the aspartase gene (*aspA*) of *Escherichia coli*. J. Gen. Microbiol. **130:**1271–1278.
12. **Gunn, F. J., C. G. Tate, and P. J. F. Henderson.** 1994. Identification of a novel sugar-H^+ symport protein, FucP, for transport of L-fucose into *Escherichia coli*. Mol. Microbiol. **12:**799–809.
13. **Henderson, P. J. F.** 1993. The 12-transmembrane helix transporters. Curr. Opin. Cell Biol. **5:**708–721.
14. **Kay, W. W., and H. L. Kornberg.** 1971. The uptake of C_4 dicarboxylic acids by *Escherichia coli*. Eur. J. Biochem. **18:**274–281.
15. **Klein, P., M. Kanehisa, and C. DeLisi.** 1985. The detection and classification of membrane spanning proteins. Biochim. Biophys. Acta **815:**468–476.
16. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157:**105–132.
17. **Lewis, M. J., and R. D. Simoni.** 1992. Deletions in hydrophilic domains of subunit-A from the *Escherichia coli* F_1F_0-ATP synthetase interfere with membrane insertion or F_0 assembly. J. Biol. Chem. **267:**3482–3489.
18. **Lo, T. C. Y.** 1977. The molecular mechanism of dicarboxylic acid transport in *Escherichia coli* K12. J. Supramol. Struct. **7:**463–480.
19. **Maloney, P. C.** 1990. A consensus structure for membrane proteins. Res. Microbiol. **141:**374–383.
20. **Moore, K. E., and S. Miura.** 1987. A small hydrophobic domain anchors leader peptidase to the cytoplasmic membrane of *Escherichia coli*. J. Biol. Chem. **262:**8806–8813.
21. **Osborn, M. J., and R. Munson.** 1974. Separation of the inner (cytoplasmic) and outer membranes of gram-negative bacteria. Methods Enzymol. **31:**642–653.
22. **Peterson, G. L.** 1977. A simplification of the protein assay method of Lowry et al. which is more generally acceptable. Anal. Biochem. **83:**346–356.
23. **Pourcher, T., M. Bassilana, H. K. Sarkar, R. Kaback, and G. Leblanc.** 1990. Melibiose permease and α-galactosidase of *Escherichia coli*: identification by selective labelling using a T7 RNA polymerase/promoter expression system. Biochemistry **29:**690–696.
24. **Saier, M. H.** 1990. Evolution of permease diversity and energy-coupling mechanisms—an introduction. Res. Microbiol. **141:**282–286.
25. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
26. **Sarsaro, J. P., and A. J. Pittard.** 1995. Membrane topology analysis of *Escherichia coli* K-12 Mtr permease by alkaline phosphatase and β-galactosidase fusions. J. Bacteriol. **177:**297–306.
27. **Six, S., S. C. Andrews, G. Unden, and J. R. Guest.** 1994. *Escherichia coli* possesses two homologous anaerobic C_4-dicarboxylate membrane transporters (DcuA and DcuB) distinct from the anaerobic dicarboxylate transport system (Dct). J. Bacteriol. **176:**6470–6478.
28. **Sofia, H. J., V. Burland, D. L. Daniels, G. Plunkett III, and F. R. Blattner.** 1994. Analysis of the *Escherichia coli* genome. V. DNA sequence of the region from 76.0 to 81.5 minutes. Nucleic Acids Res. **22:**2576–2586.
29. **Tate, C. G., J. A. R. Muiry, and P. J. F. Henderson.** 1992. Mapping, cloning, expression, and sequencing of the *rhaT* gene, which encodes a novel L-rhamnose-H^+ transport protein in *Salmonella typhimurium* and *Escherichia coli*. J. Biol. Chem. **267:**6923–6932.
30. **Tate, C. G., and P. J. F. Henderson.** 1993. Membrane topology of the L-rhamnose-H^+ transport protein (RhaT) from enterobacteria. J. Biol. Chem. **268:**26850–26857.
31. **Vieira, J., and J. Messing.** 1987. Production of single-stranded plasmid DNA. Methods Enzymol. **153:**3–11.
32. **von Heijne, G.** 1992. Membrane protein structure prediction: hydrophobicity analysis and the positive inside rule. J. Mol. Biol. **225:**487–494.
33. **Zhang, Y., and J. K. Broome-Smith.** 1990. Correct insertion of a simple eukaryotic plasma-membrane protein into the cytoplasmic membrane of *Escherichia coli*. Gene **96:**51–57.
34. **Zientz, E., S. Six, and G. Unden.** 1996. Identification of a third secondary carrier (DcuC) for anaerobic C_4-dicarboxylate transport in *Escherichia coli*: roles of the three Dcu carriers in uptake and exchange. J. Bacteriol. **178:**7241–7247.