JOURNAL OF BACTERIOLOGY, Feb. 1995, p. 1059–1068 Vol. 177, No. 4
0021-9193/95/$04.00+0

Molecular, Genetic, and Topological Characterization of O-Antigen Chain Length Regulation in *Shigella flexneri*

RENATO MORONA,* LUISA VAN DEN BOSCH, AND PAUL A. MANNING

Microbial Pathogenesis Unit, Department of Microbiology and Immunology, The University of Adelaide, Adelaide, South Australia, Australia 5005

Received 1 July 1994/Accepted 6 December 1994

The *rfb* region of *Shigella flexneri* encodes the proteins required to synthesize the O-antigen component of its cell surface lipopolysaccharides (LPS). We have previously reported that a region adjacent to *rfb* was involved in regulating the length distribution of the O-antigen polysaccharide chains (D. F. Macpherson et al., Mol. Microbiol. 5:1491–1499, 1991). The gene responsible has been identified in *Escherichia coli* O75 (called *rol* [R. A. Batchelor et al., J. Bacteriol. 173:5699–5704, 1991]) and in *E. coli* O111 and *Salmonella enterica* serovar typhimurium strain LT2 (called *cld* [D. A. Bastin et al., Mol. Microbiol. 5:2223–2231, 1991]). Through a combination of subcloning, deletion, and transposon insertion analysis, we have identified a gene adjacent to the *S. flexneri rfb* region which encodes a protein of 36 kDa responsible for the length distribution of O-antigen chains in LPS as seen on silver-stained sodium dodecyl sulfate-polyacrylamide gels. DNA sequence analysis identified an open reading frame (ORF) corresponding to the *rol* gene. The corresponding protein was almost identical in sequence to the Rol protein of *E. coli* O75 and was highly homologous to the functionally identical Cld proteins of *E. coli* O111 and *S. enterica* serovar typhimurium LT2. These proteins, together with ORF o349 adjacent to *rfe*, had almost identical hydropathy plots which predict membrane-spanning segments at the amino- and carboxy-terminal ends and a hydrophilic central region. We isolated a number of Tn*phoA* insertions which inactivated the *rol* gene, and the fusion end points were determined. The PhoA$^+$ Rol::PhoA fusion proteins had PhoA fused within the large hydrophilic central domain of Rol. These proteins were located in the whole-membrane fraction, and extraction with Triton X-100 indicated a cytoplasmic membrane location. This finding was supported by sucrose density gradient fractionation of the whole-cell membranes and of *E. coli* maxicells expressing L-[^{35}S]methionine-labelled Rol protein. Hence, we interpret these data to indicate that the Rol protein is anchored into the cytoplasmic membrane via its amino- and carboxy-terminal ends but that the majority of the protein is located in the periplasmic space. To confirm that *rol* is responsible for the effects on O-antigen chain length observed with the cloned *rfb* genes in *E. coli* K-12, it was mutated in *S. flexneri* by insertion of a kanamycin resistance cartridge. The resulting strains produced LPS with O antigens of nonmodal chain length, thereby confirming the function of the *rol* gene product. We propose a model for the function of Rol protein in which it acts as a type of molecular chaperone to facilitate the interaction of the O-antigen ligase (RfaL) with the O-antigen polymerase (Rfc) and polymerized, acyl carrier lipid-linked, O-antigen chains. Analysis of the DNA sequence of the region identified a number of ORFs corresponding to the well-known *gnd* and *hisIE* genes. The *rol* gene was located immediately downstream of two ORFs with sequence similarity to the gene encoding UDPglucose dehydrogenase (HasB) of *Streptococcus pyogenes*. The ORFs arise because of a deletion or frameshift mutation within the gene we have termed *udg* (for UDPglucose dehydrogenase).

Shigella flexneri is a causative agent of bacillary dysentery. In common with other members of the family *Enterobacteriaceae*, variations in the O-antigen component of its cell surface lipopolysaccharide (LPS) result in a number of serotypes being recognized (42). The O antigen is encoded by the genes located in the *rfb* locus (42). We have cloned (28) and characterized the genes of the *rfb* region (26, 27, 33, 35). Included within the 12-kb region required for O-antigen biosynthesis are genes encoding the enzymes for dTDP rhamnose synthesis, several rhamnosyl transferases (33), and the *rfc* gene encoding O-antigen polymerase (35). The O antigen has a tetrasaccharide repeat unit composed of the following sugars: →2)-α-L-rhamnose-(1→2)-α-L-rhamnose-(1→3)-α-L-rhamnose-(1→3)-β-D-*N*-acetylglucosamine-(1→. The organization of the *S. flexneri rfb* region is very similar to that of *rfb* regions of *Salmonella enterica* (38).

In our initial study, we observed that *Escherichia coli* K-12 strain DH1 harboring various cosmid clones of the *S. flexneri rfb* region produced either of two types of LPS. One type resembled that observed for the LPS of *S. flexneri*, with the LPS molecules having a nonrandom distribution of O-antigen chain lengths. In this case, sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) revealed that most of the LPS molecules had an O-antigen chain with 10 to 17 O-antigen tetrasaccharide repeat units. This modal chain length distribution contrasted with the second LPS pattern observed, in which the LPS had a nonmodal distribution of O-antigen chain lengths, mostly of high molecular weight (>15 O-antigen repeat units). The plasmids which resulted in this second LPS phenotype had a region adjacent to the *rfb* operon deleted, and a gene(s) which regulated O-antigen length was proposed for this region (28). The effect on the distribution of the length of the O-antigen repeat units by a gene near an *rfb* region has

* Corresponding author. Phone: (61 8) 303 4151. Fax: (61 8) 303 4362. Electronic mail address: RMORONA@MICROB.SCIENCE.ADELAIDE.EDU.AU.

subsequently been described for *E. coli* O75 (6, 7), *S. enterica* serovar typhimurium LT2 (5, 6), *E. coli* O111 (4, 5), and *S. enterica* serovar Muenchen M67 (9).

The gene whose presence is required for wild-type LPS patterns as seen on silver-stained SDS-PAGE gels has been identified and is called either *rol* (regulation of O-antigen length) (6, 7) or *cld* (chain length determination) (4, 5). In this communication, we describe the localization of the *S. flexneri* gene which affects O-antigen chain length distribution and determine its DNA sequence and that of the adjacent genes. This completes the DNA sequence of the entire *rfb* region, including the *gnd* gene and the last gene of the *his* operon. The subcellular location of the *rol* gene product was investigated, and the effect of a mutation in the chromosomal *rol* gene of *S. flexneri* is also reported.

MATERIALS AND METHODS

Bacterial strains. The *E. coli* K-12 and *S. flexneri* strains used are as follows. *E. coli* K-12 strains DH1 and SØ874 (Δ*rfb-his*) have been described previously (27, 37); *phoA* mutant strain CC118 was obtained from C. Manoil (30); strain CSR603, used for maxicell analysis, was also described previously (28); and strain C75a (CGSC 5978) is a constitutive producer of PhoA because of a mutation in *phoR* and was obtained from B. Bachmann, *E. coli* Genetic Stock Center, Yale University, New Haven, Conn. The *S. flexneri* strain used was Sfl1 (obtained from M. Rohde, Braunschweig, Germany) and has the Y serotype. Strains were maintained at −20 or −70°C in a glycerol-peptone mixture (35). Cultures were grown at 37°C with aeration.

Growth media. The growth media used were Luria broth (LB) and agar, and Oxoid nutrient broth (NB) and agar (35). The following antibiotics from either Sigma or Boehringer Mannheim were used at the concentrations indicated in parentheses: ampicillin (50 μg/ml), kanamycin (50 μg/ml), chloramphenicol (25 μg/ml), and tetracycline (10 μg/ml). X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) at 40 μg/ml and IPTG (isopropyl-β-D-thiogalactopyranoside) at 1 mM were used and were purchased from 5-Prime,3-Prime, Inc., Boulder, Colo. 5-Bromo-4-chloro-3-indolyl-phosphate (Xpho; Sigma) was used at 40 μg/ml.

Isolation of Tn*phoA* insertions. Strain CC118 harboring pPM2716 was infected with Tn*phoA* (30), and then kanamycin- and ampicillin-resistant colonies were selected. Plasmid DNA from pools of several thousand colonies was transformed into CC118, and colonies were selected on plates containing Xpho, kanamycin, and ampicillin to detect alkaline phosphatase-positive transformants. Plasmid DNA from each of these isolates was characterized by restriction enzyme mapping.

DNA methods. The preparation of plasmid DNA, restriction enzyme digestion, ligation, and agarose gel electrophoresis were performed as recently described (3, 7). Plasmids were transformed into *E. coli* K-12 strains by the $CaCl_2$ procedure (8) and were transformed into *S. flexneri* by electroporation (43). DNA fragments were purified from agarose gels with the Geneclean II (Bio 101) or Qiaex (Qiagen) kit.

DNA sequencing. The DNA sequences of the inserts in plasmids pRMA247 and pRMA511 were obtained by obtaining a set of exonuclease III-generated deletions from either end of the inserts, as described by Henikoff (16) and detailed in the Erase-a-Base kit (Promega) protocol. The deleted plasmids were sequenced with the Dye Primer Kits purchased from Applied Biosystems (AB) using M13 forward and reverse primers. Gaps in the sequence were determined by a walking strategy with the use of appropriate primers and the Dye Terminator Kit (AB). DNA sequence data were obtained from the AB model 473A Automated Sequencer, and the program SeqEd (AB) was used to edit the data. The sequence was assembled by using the DNASIS (Hitachi) software package. The determination of Tn*phoA* insertion positions was performed with an oligonucleotide near the end of Tn*phoA* (30) and the Dye Terminator Sequencing Kit (AB) and associated protocols. Plasmid DNA for this purpose was prepared by Triton X-100 lysis (32) followed by RNase treatment and polyethylene glycol 8000 precipitation as recommended by AB.

Sequence analysis. The DNASIS and PROSIS (Hitachi) software packages were used for analysis of DNA and protein sequences. Database searches were conducted by electronic mail with BLAST programs (2) available at the National Center for Biomedical Information, Bethesda, Md. Multiple alignments of protein sequences were achieved with the program CLUSTAL (17, 18), and the program PROFILEGRAPH (20) was used to align protein hydropathy plots by means of the Kyte and Doolittle values (24).

Rapid chromosomal DNA preparation. Chromosomal DNA was prepared from *S. flexneri* strains as follows. Aliquots (1 ml) of LB cultures incubated for 18 h were centrifuged (Biofuge 15; 1 min, 14,000 rpm). The cells were resuspended in MilliQ water and vortex mixed with an equal volume of phenol (Tris saturated [39]). The aqueous phase was removed to a fresh tube, and the phenol treatment was repeated. The aqueous phases were pooled, and the DNA was precipitated by addition of sodium acetate to a final concentration of 0.3 M and ethanol (2 volumes). The DNA was recovered by centrifugation, washed with 70% ethanol, dried in vacuo, and finally resuspended in 10 mM Tris-HCl–1 mM EDTA, pH 7.5.

Southern blotting. DNA fragments were labelled with digoxigenin by random-primer labelling (Boehringer Mannheim). The procedure described by Sambrook et al. (33) was used to transfer DNA from agarose gels to nitrocellulose membranes and for subsequent hybridization to the labelled DNA probe. Anti-digoxigenin antibody (Boehringer Mannheim), goat anti-rabbit horseradish peroxidase secondary antibody, and the electrochemiluminescence detection kit (Amersham) were used to develop the blots.

LPS methods. Small-scale preparations of LPS were made by proteinase K treatment of whole-cell lysates (19). Electrophoresis of LPS on SDS–20% polyacrylamide gels and detection by silver staining were performed as described previously (32).

Maxicell procedure. The maxicell procedure to detect plasmid encoded proteins was performed as described previously (3). Proteins were labelled with L-[^{35}S]methionine (Amersham) or with Tran^{35}S-label (ICN). After solubilization, the samples were electrophoresed on SDS–12% polyacrylamide gels (1.5 mm by 20 cm) which were then stained with Coomasie brilliant blue G250 and dried before the labelled proteins were detected by autoradiography. The protein size markers were obtained from Pharmacia.

Fractionation of *E. coli* cells. The method used to fractionate *E. coli* K-12 cells to identify proteins located in the soluble fractions (cytoplasm and periplasm) and insoluble fractions (inner membrane [IM] and outer membrane [OM]) have been described previously (1, 34). Briefly, the cultures (10 to 50 ml) were grown to mid-exponential phase (optical density at 600 nm of 0.6), and the cells were treated with a buffer containing Tris, EDTA, sucrose, and lysozyme. The supernatant obtained after centrifugation was termed the periplasmic fraction. The spheroplasts were lysed by dilution and sonication, and unlysed cells were then removed by slow-speed centrifugation; the soluble fraction obtained after high-speed centrifugation was termed the cytoplasm, and the pellet was termed the whole-membrane (WM) fraction. The WM fraction was treated with Triton X-100–$MgCl_2$, and after high-speed centrifugation a soluble fraction termed inner membrane (IM) and an insoluble fraction termed outer membrane (OM) were obtained. The separation of IM and OM was also achieved by using lysed spheroplasts and sucrose density gradients (40).

Western blotting (immunoblotting). Detection of PhoA fusion proteins was performed after transfer of SDS-PAGE-separated samples to nitrocellulose membranes (32). Detection was performed with rabbit anti-*E. coli* alkaline phosphatase antiserum (purchased from 5-Prime,3-Prime Inc.) as the primary antibody and with a goat anti-rabbit peroxidase conjugate as the secondary antibody. The staining procedure of Hawkes et al. (15), with 4-chloro-1-naphthol, was used.

Protein SDS-PAGE. Samples were adjusted to represent material from equivalent numbers of cells. Samples were solubilized at 100°C in sample buffer (25) before being applied to SDS–12 or 15% polyacrylamide gels as described previously (1).

Nucleotide sequence accession number. The DNA sequence data obtained in this study are available under GenBank and EMBL accession number X71790 (revised June 1994).

RESULTS

Localization of the *rol* gene. In our earlier study (38), we had found that the gene(s) affecting O-antigen chain length regulation were located in an 8-kb segment adjacent to *rfb* genes. To locate the gene(s) responsible for regulating chain length, we used complementation analysis and Tn*phoA* mutagenesis. A number of deletion derivatives of pPM2213 (pPM2701, pPM2202, pPM2203, pPM2204, and pPM2205) (Fig. 1) were transformed into DH1 harboring pRMA154 (a plasmid having the *rfb* genes but lacking the gene(s) for O-antigen chain length regulation) (Fig. 1). Examination of the LPSs produced by these strains showed that only plasmids pPM2702, pPM2704 and pPM2705 restored wild-type O-antigen length distribution (Fig. 2). This distinction indicated that a *Pst*I fragment contained part or all of the gene. Another plasmid, pRMA247, containing a *Bgl*II fragment from this region failed to complement pRMA154. Hence, the *Pst*I fragment was insufficient. Plasmid pRMA520, containing an *Sph*I fragment encompassing the right-hand end of the O-antigen chain length-determining region, was able to complement the O-antigen chain length regulation defect of pRMA154, and an *Eco*RI deletion derivative of pRMA520 called pRMA532 also effectively complemented pRMA154 (Fig. 2). Thus, the O-antigen chain length-regulating gene (termed *rol*) was located between the

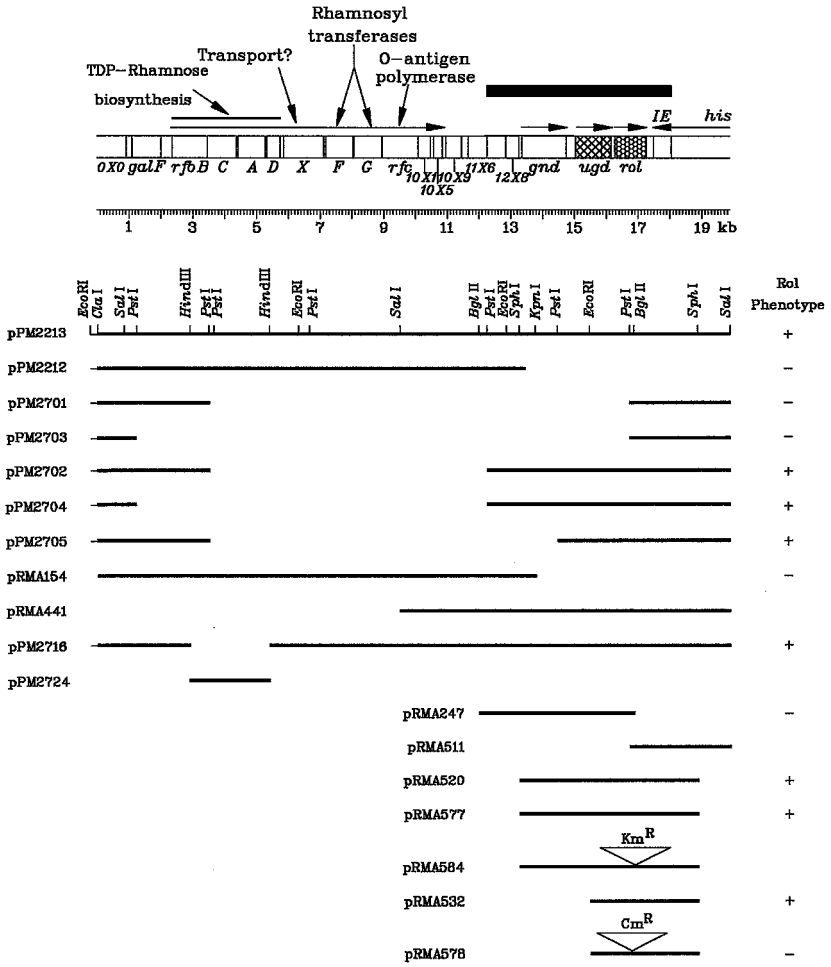

FIG. 1. Organization of the *S. flexneri rfb* region. The regions cloned into the plasmids used in this study are shown. The heavy bar indicates the region whose sequence is presented in this paper (nt 12204 to 18059). The functions encoded by several of the genes are indicated above the map (arrows). The direction of transcription is indicated by horizontal arrows. The construction of plasmids is described in the text. Plasmids pPM2212 and pPM2213 (28), plasmid pRMA154 (35), and plasmids pPM2716 and pPM2724 (26) were described elsewhere. Plasmid pRMA441 has the insert in pACYC184. Plasmids pRMA247 and pRMA511 have their inserts in pBluescript SK. Plasmids pRMA520, pRMA532, and pRMA578 have their inserts in pK184 (22). Plasmids pRMA577 and pRMA584 have their inserts in pCACTUS (10). The Rol phenotype is detected by a complementation test using plasmids pPM2212 and pRMA154. +, complementation; −, no complementation. Km^R, kanamycin resistance cartridge (Pharmacia); Cm^R, chloramphenicol resistance cartridge (12). The *rol* gene spans a *Pst*I site and a *Bgl*II site.

*Eco*RI and *Sph*I sites of pRMA520 (Fig. 1) and it was likely to contain a *Pst*I site and a *Bgl*II site.

The above analysis was supported by the results of Tn*phoA* mutagenesis. Plasmid pPM2716 (Fig. 1) is a *Hin*dIII deletion derivative of pPM2213, which inactivates all four dTDP-rhamnose biosynthesis genes (*rfbBCAD*), and can be complemented with pPM2724 in *E. coli* DH1 (26). As a result of a large number of independent Tn*phoA* mutagenesis experiments, eight $PhoA^+$ insertions (blue colonies on Xpho-containing LB agar plates) were isolated in pPM2716. These insertions were all mapped in the *rol* region identified above. Complementation of the pPM2716::Tn*phoA*$^+$ insertions with pPM2724 revealed that the LPSs had a random O-antigen chain length distribution, indicating that Tn*phoA* was inserted in the O-antigen chain length regulation gene(s) (Fig. 2). The cause of the difference between the LPS banding pattern observed for these plasmids and that obtained with pRMA154 or pPM2212 is not known but may be differences in *rfb* gene expression. In addition to these Tn*phoA* insertions, others which affected O-antigen chain length were detected by SDS-PAGE screening of the LPSs of DH1 containing pPM2724 and randomly selected pPM2716::Tn*phoA* insertions which were O antigen positive but were $PhoA^-$. This method detected a number of pPM2716::Tn*phoA* insertions in which *rol* was inactivated. All were mapped within the same region as the pPM2716::Tn*phoA* $PhoA^+$ fusions. The precise locations of the fusion junctions were determined and are reported below. These observations are consistent with those previously (28) reported and indicate that the strain DH1 in our collection, like the Δ*rfb* strain SØ874, is unable to complement the *rol* mutations that we have generated.

DNA sequence of the O-antigen regulator gene. We have determined the DNA sequence of a 5,856-bp region, starting from a *Pst*I site at nucleotide (nt) 12204 of our previously published sequence (35), within which the gene(s) regulating O-antigen length was localized (Fig. 1). A number of open reading frames (ORFs) were identified. At the 5′ end of the sequence, a short ORF was detected and this completes the sequences of the previously identified *orf11×6* (35). The ORF commencing at nt 13363 corresponds to *gnd*, encoding gluconate-6-phosphate dehydrogenase, and the gene is almost identical to that of *E. coli* K-12 and other *E. coli* strains (data

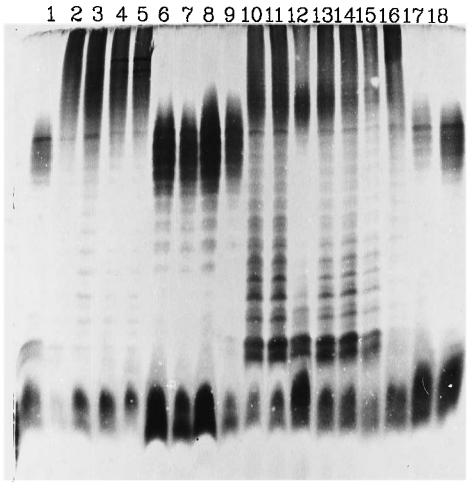

FIG. 2. Analysis of LPS to assess *rol* activity by complementation and effect of *rol*::Tn*phoA* insertions. LPSs were prepared by proteinase K treatment of whole-cell lysates of *E. coli* K-12 DH1 derivatives with various plasmids. Samples (equivalent to 2×10^8 cells) were electrophoresed on an SDS–20% polyacrylamide gel, and LPSs were detected by silver staining. The strains contained the following plasmids: pPM2213 (lane 1); pPM2212 (lane 2); pRMA154 and pHC79 (lane 3); pRMA154 and pPM2701 (lane 4); pRMA154 and pPM2703 (lane 5); pRMA154 and pPM2702 (lane 6); pRMA154 and pPM2704 (lane 7); pRMA154 and pPM2705 (lane 8); pPM2724 and pPM2716 (lane 9); pPM2724 and pRMA278 (lane 10); pPM2724 and pRMA290 (lane 11); pPM2724 and pRMA465 (lane 12); pPM2724 and pRMA467 (lane 13); pPM2724 and pRMA469 (lane 14); pPM2724 and pRMA471 (lane 15); pPM2212 and pK184 (lane 16); pPM2212 and pRMA520 (lane 17); and pPM2212 and pRMA532 (lane 18).

not shown). The DNA sequence between the end of *orf11×6* and *gnd* contains only a short ORF, termed *orf12×8*, and no other ORF in any reading frame. The entire region preceding the *gnd* gene has a low G+C content of 31%, and this is a continuation of the low-percent G+C which starts with the *rfbX* gene (Fig. 1).

The region between *rfc* and the *gnd* gene contains a number of ORFs whose products have no sequence similarity to any protein in the databases. As mentioned above, one ORF completes the previously described *orf11×6* and terminates at nt 12446. A second ORF (unnamed) commences at nt 12188 and stops at nt 12898. This ORF overlaps at a third ORF (*orf12×8*) which starts at nt 12715 and ends at nt 13254. A potential start codon for *orf12×8* is a GTG codon at nt 12832.

The sequence immediately after the *gnd* gene contains two slightly overlapping ORFs which have a high degree of identity (Table 1) with the ORF between *gnd* and *cld* of *S. enterica* serovar typhimurium LT2, with an ORF adjacent to *E. coli* O111 *gnd* (5), and with an incomplete ORF from *Streptococcus pneumoniae* type 3 (14). The products of these ORFs all have a high degree of similarity to the *Streptococcus pyogenes hasB* gene product, UDPglucose dehydrogenase (13), and it seems likely that the ORFs encode this enzyme (Table 1). We propose that the gene in *Shigella* and *Salmonella* species and *E. coli* be termed *ugd* (UDPglucose dehydrogenase). The *S. flexneri* gene, designated *ugd**, apparently has a mutation as a result of a deletion of 38 or 40 bp, resulting in a frame shift. Figure 3 shows an alignment of the *S. pyogenes* HasB protein with the products of *S. enterica* LT2 ORF1, *E. coli* O111 ORF1, *S. pneumoniae* type 3 Cps1-3, and the two *S. flexneri* ORFs. The product of each ORF shows identity with HasB along its entire length. The deletion within *ugd** may have arisen in the cloned *S. flexneri* genes; we have not investigated whether it is present in the *S. flexneri* chromosomal *udg* gene.

The *rol* gene. The ORF following the *udg* gene corresponds to the *rol* gene, and its position correlates with the complementation and Tn*phoA* insertion data described above. The *S. flexneri rol* gene and its product are almost identical to those of *E. coli* O75 (Table 2). This identity is remarkable in that the *E. coli* O75 gene and product have 65% identity with counterparts from *E. coli* O111. An alignment of the Rol and Cld proteins is presented in Fig. 4. As noted by Bastin et al. (5), sequence similarity extends along the length of the proteins. Several small regions of sequence difference between the *S. flexneri* and *E. coli* O75 proteins and the other Cld proteins can be seen. We have also included the ORF product adjacent to the *E. coli rfe* gene, termed o349 (11, 31), which is also a member of this family of proteins. The *S. flexneri rol* gene includes *Pst*I and *Bgl*II sites, explaining why deletions, or subclones, using these sites inactivate the *rol* gene.

Located immediately after the *rol* gene is an inverted repeat sequence typical of a transcriptional terminator (6). On the 3′ side of this terminator and on the opposite strand, the *hisIE* gene is located (Fig. 1). The *S. flexneri hisIE* gene is almost identical to that of *E. coli* K-12 (data not shown). All the features of the *col-his* sequences reported for *E. coli* O75 and *S. enterica* LT2 (6) can be found in the *S. flexneri rol-hisIE* sequences and are not described further in this report.

Characterization of *rol*::Tn*phoA* insertions. The positions of the Tn*phoA* insertions within the *rol* gene were determined with an oligonucleotide primer reading out of the end of Tn*phoA* (30). The positions of the insertions are shown in Fig. 5. One insertion is located in the −10 region of the *rol* gene (pRMA475), and the other two PhoA$^-$ insertions are out of frame and located near the 5′ end of *rol* (pRM473 and pRMA479). The remaining Tn*phoA* insertions are in frame

TABLE 1. Comparison of *S. pyogenes* HasB with proteins encoded by *S. flexneri udg** ORFs and other ORFs

ORF[a]	% Identity[b]					
	STRHASB	EO111O1	LT2O1	SFUDGHN	SFUDGHC	SPCPS13
STRHASB	100	53 (402)	53.2 (402)	56 (294)	52.7 (93)	47.9 (140)
EO111O1		100	81.2 (388)	81.5 (287)	80.9 (94)	51.6 (128)
LT2O1			100	87.5 (287)	90.4 (94)	40.7 (140)
SFUDGHN				100	ND	42.9 (140)
SFUDGHC					100	ND
SPCPS13						100

[a] STRHASB, *S. pyogenes* HasB, UDPglucose dehydrogenase (13); EO111O1, *E. coli* O111 ORF1 (5); LT2O1, *S. enterica* serovar typhimurium strain LT2 ORF1 (5); SFUDGHN, *S. flexneri udg** ORF (nt 15019 to 15861); SFUDGHC, *S. flexneri udg** ORF (nt 15855 to 16136); SPCPS13, *S. pneumoniae* Cps1-3 ORF (14).

[b] Numbers are percent identity as determined by the FASTA program and implemented in PROSIS. The numbers in parentheses indicate the number of amino acids over which the percent identity occurs. ND, not done.

```
STRHASB   MKIAVAGSGYVGLSLGVLLSLQNEVTIVDILPSKVDKINNGLSPIQDEYIEYYLKSKQLS 60
E111O1    MKITISGTGYVGLSNGILIAQHHEVVALDIVPTKVEMLNQKKSPIVDKEIEEYLATKQLN 60
LT2O1     MKITISGTGYVGLSNGLLIAQHHDVVALDIVPSRVELLNDRISPIVDKEIQQFLKEDNIR 60
SFUDGH    MKITISGTGYVGLSNGLLIAQNHEVVALDILPSRVAMLNDRISPIVDKEIQQFLQSDKIH 60
SPCPS13   MKIAIAGSGYVGLSLAVLLAQHHEVKVIDVIKDKVESINNRKSPIKDEAIEKYLVEKELN 60
          ***...*.****** ..*.. ...*  .*.. ..*. .*.  *** *. *. .* ....

STRHASB   IKATLDSKAAYKEAELVIIATPTNYNSRINYFDTQHVETVIKEVLSVNSHATLIIKSTIP 120
E111O1    FRATTDKYDAYRDGTYVIIATPTDYDPKTNYFNTSSVESVIRDVVDINPNAVMVIKSTIP 120
LT2O1     FRATLDKFDAYQNADYVIIATPTDYDPKTNYFNTSSVESVMQDVISINPAAVMIIKSTVP 120
SFUDGH    FNATLDKNEAYRDADYVIIATPTDYDPKTNYFNTSSVESVIKDVVEINPYAVMVIKSTVP 120
SPCPS13   LEASLDPAHVYKDVEYAIIATPTNYDVDLNQFDTSSVEAAIKTCMEYNDTCTIVIKS-LP 119
          . *. *   .*.. . .******.*.   * *.*  **..... .. *   ...*** .*

STRHASB   IGFITEMRQKFQTDRIIFS-PEFLRESKALYDNLYPSRIIVSCEENDSPKVKADAEKFAL 179
E111O1    VGFTNLLKERLGIDNIFFS-PEFLREGRALYDNLHPSRIVIG-ERSER------AGRFAA 172
LT2O1     VGFTAAMRQKFATENIIFS-PEFLREGKALYDNLYPSRIVIG-EQSER------AREFAA 172
SFUDGH    VGFTAAMHKKYRTENIIFS-PEFLREGKALYDNLHPSRIVIG-ERSER------AERFAA 172
SPCPS13   PKFRN-LSKTFSQALLLYSS                                        138
            * . . .    . ...* ******..******.****... * ...      *  **

STRHASB   LLKSAAKKNNVPVLIMGASEAEAVKLFANTYLALRVAYFNELDTYAESRKLNSHMIIQGI 239
E111O1    LLQEGAVKKDIPTLFTDSTEAEAIKLFANTYLALRVAYFNELDSYAESLGLNSRQIIEGV 232
LT2O1     LLQEGAIKQEIPTLFTDSTEAEAIKLFANTYLAMRVAYFNELDSYAETLGLNTRQIIEGV 232
SFUDGH    LLQEGAIKQNIPTLFTDSTEAEAIKLFANTYLAMRVAYFNELDSYAESLGLNTRQIIEGV 232
          **...* *...*.*. ...****.*********.*********.***.  **.. **.*.

STRHASB   SYDDRIGMHYNNPSFGYGGYCLPKDTKQLLANYNNIPQTLIEAIVSSNNVRKSYIAKQII 299
E111O1    CLDPRIGNHYNNPSFGYGGYCLPKDTKQLLANYASVPNNIIGAIVDANRTRKDFIADSIL 292
LT2O1     CLDPRIGNHYNNPSFGYGGYCLPKDTKQLLANYQSVPNNIISAIVEANRTRKDFIADAIL 292
SFUDGH    CLDPRIGNHYNNPSFGYGGYCLPKDTKQLLANYQSVPNNLISAIVDANR----------- 281
          . * *** ************************...*...*.***..*.

STRHASB   NVLKEQESPVKVVGVYRLIMKSNSDNFRESAIKDVIDILKSKDIKIIIYEPMLNKLESED 359
E111O1    A------RKPKVVGVYRLIMKSGSDNFRASSIQGIMKRIKAKGVPVIIYEPVMVE-DEFF 345
LT2O1     A------RKPKVVGIYRLIMKSGSDNFRASSIQGIMKRIKAKGVEVIIYEPVMEE-DTFF 345
SFUDGH    --------KPQVVGIYRLIMKSGSDNFRASSIQGIMKRIKAKGVEVIIYEPVMKG-DSFF
                  .***.*******.*****.*.*..... .*.*.. .*****..   ..

STRHASB   QSVLVNDLENFKKQANIIVTNRYDNELQDVKNKVYSRDIFGRD 402
E111O1    HSRVVRDLTAFKNEADIIISNRMTSELADVADKVYTRGLFGSD 388
LT2O1     NSRLERDLHCFKQQADVIISNRMAAELLDVAEKVYTRDLFGSD 388
SFUDGH    NSRLERDLATFKQQADVIISNRMAEELKDVADKVYTRDLFGSD
          .* . .**  **..*..*..** ..** ** .***.*..**.*
```

FIG. 3. Alignment of *S. pyogenes hasB* gene product with similar ORFS. The program CLUSTAL (using default settings) was used to align the amino acid sequences of the following ORFS: *S. pyogenes hasB* gene (STRHASB) (13), *E. coli* O111 ORF1 (E111O1) (4), *S. enterica* serovar typhimurium strain LT2 ORF1 (LT2O1) (4), *S. flexneri udg*∗ ORFs (nt 15019 to 15861 and nt 15855 to 16136) (SFUDGH), and *S. pneumoniae* Cps1-3 ORF (SPCPS13) (14). Symbols: ∗, identical amino acids; ⊙, similar amino acids. The amino acids in boldface type indicate a potential nucleotide binding motif. The amino acid sequence is numbered on the right.

and are located in a region between the middle and near the end of the gene. The Rol::PhoA fusion proteins could be detected in Western blots of whole-cell lysates (data not shown; see Fig. 6), and their apparent molecular size (Fig. 5) correlated with the fusion point within Rol.

Examination of the hydropathy plot of the SfRol protein indicated that it has two hydrophobic regions which are potential membrane-spanning regions (Fig. 5) and that all members of the Rol family have almost identical plots (data not shown). The Rol::PhoA$^+$ proteins have fusion junctions between these two hydrophobic domains. These data suggest that the region of the SfRol protein between these domains is located in the periplasmic space and that the Rol protein is anchored in the IM near its amino and carboxy termini (Fig. 4 and 5).

TABLE 2. Comparison of *S. flexneri* Rol with Rol and Cld proteins

ORF[a]	% Identity[b]				
	SFROL	E75ROL	LT2CLD	E111CLD	RFEORF2
SFROL	100	94.5 (327)	72 (325)	66.5 (325)	22.3 (345)
E75ROL		100	73.4 (327)	65.1 (327)	23.5 (345)
LT2CLD			100	64.7 (327)	24.3 (345)
E111CLD				100	24.2 (345)
RFEORF2					100

[a] SFROL, *S. flexneri* Rol protein; E75ROL, *E. coli* O75 Rol protein (6); LT2CLD, *S. enterica* serovar typhimurium strain LT2 Cld protein (5); E111CLD, *E. coli* O111 Cld protein (5); RFEORF2, *E. coli* K-12 ORF o349 adjacent to *rfe* gene (35).

[b] Numbers are percent identity as determined by the FASTA program and implemented in PROSIS. The numbers in parentheses indicate the number of amino acids over which the percent identity occurs.

Location of the Rol protein. Initially, the Rol::PhoA fusion proteins were localized to the WM by fractionation of *E. coli* DH1 encoding the Rol::PhoA fusion proteins. We subsequently separated the IM and OM by using either Triton X-100 extraction or sucrose density gradients. *E. coli* DH1 producing the Rol::PhoA467 protein was fractionated to give soluble (cytoplasm or periplasm) and Triton X-100-soluble (IM or cytoplasmic membrane) and insoluble (OM) fractions. Figure 6 shows that of Rol::PhoA which was located in the WM fraction, most could be solubilized by a single Triton X-100 extraction. Hence, this result suggested an IM location. A similar result was obtained for Rol::PhoA fusion 471 (data not shown). To confirm this observation, we separated the WM containing Rol::PhoA471 on sucrose gradients. The fractions from the gradients were subjected to SDS-PAGE and either stained with Coomassie brilliant blue or subjected to Western blotting to detect PhoA. The experiment showed that the fractions (Fig. 7, lanes 8 to 12) which were enriched with the OM proteins (OmpF, OmpC, and OmpA) contained little or no Rol::PhoA471 fusion protein. The Rol::PhoA471 fusion protein was found in fractions of lower density (Fig. 7, lanes 13 to 18). Hence, the Rol::PhoA471 fusion protein is associated with the IM and our data provide direct evidence for the location of the Rol protein within the cytoplasmic membrane.

Identification of *S. flexneri* Rol protein. We identified the SfRol protein by comparing the proteins expressed from several plasmids in maxicells. Figure 8 shows that an L-[^{35}S]me-

```
SFROL     MRVENNNVSGQNHDPEQIDLIDLLVQLWRGKMTIIISVIVAIALAIGYLAVAKEKWTSTA  60
E75ROL    MRVENNNVSGQNHDPEQIDLIDLLVQLWRGKMTIIISVIVAIALAIGY--VAKEKWTSTA  58
LT2CLD    MTVDSNTSSGRGNDPEQIDLIELLLQLWRGKMTIIVAVIIAILLAVGYLMIAKEKWTSTA  60
E111CLD   MSISDNNMSGRSHDPEQIDLIDLLMQLWRGKVTIIICIVVAIALAVGYLAVAKEKWTSIA  60
RFEORF2   MTQPMPGKPAEDAE-NELDIRGLFRTLWAGKLWIIGMGLAFALIALAYTFFARQEWSSTA  59
          *      . .. . . ...*. .*.  ** **. **    .  .*..*  *...*.*.*

SFROL     IITQPDVGQIA------GYNNAMNVIYGQAAPKVSDLQETLIGRFSSAFSA--------- 105
E75ROL    IITQPDVGQIAV-----GSYNNAMNVIYGQAAPKVSDLQETLIGRFSSAFSA--------- 105
LT2CLD    IITQPDAAQVA------TYTNALNVLYGGNAPKISEVQANFISRFSSAFSA--------- 105
E111CLD   IVTQPDAGQLA------SYNNAMDVLFGANAPEMTDVQAGFIGRFSSAFSA--------- 105
RFEORF2   ITDRPTVNMLGGYYSQQQFLRNLDVRSNMASADQPSVMDEAYKEFVMQLASWDTRREFWL 119
          *...*...  .. .  . ....*   . ....  ...  ..     *   ....

SFROL     ----LAETLDNQEEPEKLTIEPSVKN---------QQLPLTVSYVGQTAEGAQMKLAQYI 152
E75ROL    ----LSEVLDNQKERKKLTIEPSVKN---------QQLPLTVSYVGQTAEGAQMKLAQYI 152
LT2CLD    ----LSEVLDNQKERKKLTIEQSVKG---------QALPLSVSYVSTTAEGAQRRLAEYI 152
E111CLD   ----LAETLDNQQEPEKLSIDAAVKG---------QTLPLKVSYTGKTSEEAQKTLAQYI 152
RFEORF2   QTDYYKQRMVGNSKADAALLDEMINNIQFIPGDFTRAVNDSVKLIAETAPDANNLLRQYV 179
             . . .. . .   ..  ...        . .  .*. .. *. .*.  * .*.

SFROL     QQVDDKVNQELEKDLKDNIALGRKNLQDSLRTQEVVAQEQKDLRIRQIQEALQYANQAQV 212
E75ROL    QQVDDKVNQELEKDLKDNIALGRKNLQDSLRTQEVVAQEQKDLRIRQIQEALQYANQAQV 212
LT2CLD    QQVDEEVAKELEVDLKDNITLQTKTLQESLETQEVVAQEQKDLRIKQIEEALRYADEAKI 212
E111CLD   QQVDEGVAKELNADLSMSMETRLADLQKSLAAQEAVAKEQKTLRIAQMTQALKVAQQSNI 212
RFEORF2   AFASQRAASHLNDELKGAWAARTIQMKAQVKRQEEVAKAIYDRRMNSIEQALKIAEQHNI 239
          . ... .. .*. .*. . .     ..  .  ** **..  . *.  . .**. *.. ..

SFROL     TKPQVQ-QTEDVTQDTLFLLGSEALESMIKHEATRPLVFSPNYYQTRQNLLDIEKLKFDD 271
E75ROL    TKPQIQ-QTQDVTQDTMFLLGSEALESMIKHEATRPLVFSSNYYQTRQNLLDIDNLDVDK 271
LT2CLD    TQPQIQ-QTQDVTQDTMFLLGSDALKSMIQNEATRPLVFSPAYYQTKQTLLDIKNLKVTA 271
E111CLD   KLPQVQ-QVDQVSQDSMFMLGSEALSSMVENEATRPLTFSDQYYQTRQNLLEVQALEVAP 271
RFEORF2   SRSATDVPAEELPDSEMFLLGRPMLQARLENLQAVGPAFDLDYDQNRAMLNTLNVGPTLD 299
          . .... .........*.**.  *  ..... ..   .*. .* *... *  ...

SFROL     LDIHAYRYVMKPTLPIRRDSPKKAITLILAVLLGGMVGAGIVLGRNALRNYNAK--      325
E75ROL    LDIHAYRYVMKPTLPIRRDSPKKAITLILAVLLGGMVGAGIVLGRNALRNYNAK--      325
LT2CLD    DTVHVYRYVMKPTLPVRRDSPKTAITLVLAVLLGGMIGAGIVLGRNALRSYKPKAL      327
E111CLD   DSVHAYRYVMKPTLPIRRDSPKKAITLVLAVLIGGMIGAGVVLGRNALRGYKAKAE      327
RFEORF2   PRFQTYRYLRTPEEPVKRDSPRRAFLMIMWGIVGGLIGAGVALTRRCSK-------      348
             ..***...*. *..****. *. ...  ..**..***..*.*.  .
```

FIG. 4. Alignment of Rol and Cld protein sequences. The amino acid sequences of the following proteins were aligned by using the program CLUSTAL at the default settings. The protein sequences aligned were *S. flexneri* Rol (SFROL), *E. coli* O75 Rol (E75ROL [5]), *S. enterica* serovar typhimurium strain LT2 Cld (LT2CLD [5]), *E. coli* O111 (E111CLD [6]), and *E. coli* K-12 ORF o349 adjacent to the *rfe* gene (RFEORF2 [31]). Symbols: *, identical amino acids at this position in all the ORFs; ⊙, similar amino acids at this position in all the ORFs. Amino acids identical to those in SFROL at a particular position are in boldface type. The boxed regions correspond to hydrophobic segments as determined from hydrophobicity plots.

thionine-labelled protein of the size predicted for SfRol (36 kDa) can be detected in strain E1664 harboring pRMA520 and pRMA532. When plasmid pRMA527 having a Cm^r (chloramphenicol resistance) cartridge (12) inserted into the *Pst*I site within the *rol* gene is used, the 36-kDa protein is no longer detected. The gene products for the *S. flexneri gnd* and mutant *udg** genes were not detected but could be detected in autoradiographs from other experiments (data not shown).

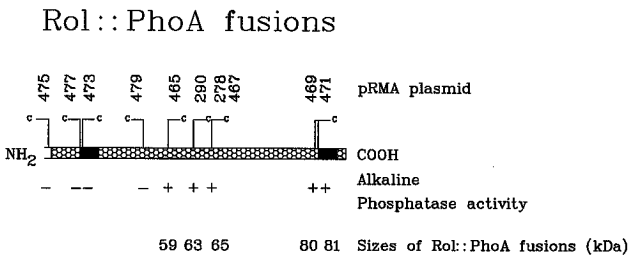

FIG. 5. Location of *rol*::Tn*phoA* insertions and Rol::PhoA fusion proteins. The locations of Tn*phoA* insertions inactivating the *rol* gene are shown. The numbers above the map correspond to the pRMA plasmids in which the insertions are located. The orientation of each insertion is shown, with "c" indicating the carboxy-terminal end of the PhoA protein. Insertion 475 is located outside of the coding region. The insertions are situated after the nucleotide positions as follows: 475, nt 16242; 477, nt 16343; 473, nt 16348; 479, nt 16564; 465, nt 16711; 290, nt 16801; 278 and 467, nt 16843; 469, nt 17104; and 471, nt 17125. Insertions 475, 477, 473, and 479 are either out of frame or in the wrong orientation. Fusions yielding a strong blue color on indicator plates containing Xpho are scored as positive (+). The apparent molecular sizes of the Rol::PhoA fusion proteins are indicated. The honeycomb areas correspond to the ORF, and the filled areas correspond to predicted membrane-spanning regions of the polypeptide chain.

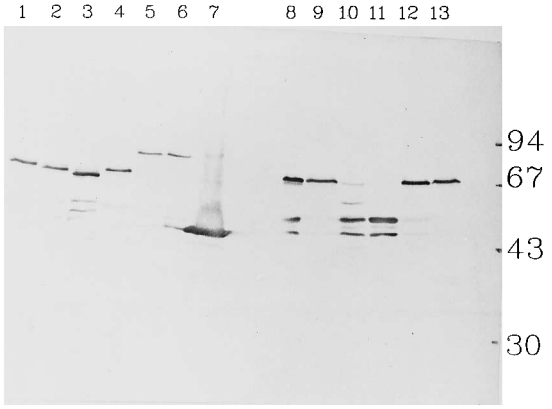

FIG. 6. Western immunoblot of *E. coli* K-12 strains producing Rol::PhoA fusion proteins. Strain CC118 harboring plasmids with *rol*::Tn*phoA*$^+$ insertions was grown to mid-exponential phase (optical density at 600 nm of 0.5) in nutrient broth. WM were prepared as detailed in Materials and Methods. In a separate experiment, strain CC118 with pRMA467 was similarly grown and treated to obtain subcellular fractions as detailed in Materials and Methods. The samples were solubilized in sample buffer and electrophoresed on an SDS–15% polyacrylamide gel. The proteins were transferred to nitrocellulose membranes and detected with rabbit antibacterial alkaline phosphatase as detailed in Materials and Methods. Lanes: 1, pRMA278 WM; 2, pRMA290 WM; 3, pRMA465 WM; 4, pRMA467 WM; 5, pRMA469 WM; 6, pRMA471 WM; 7, whole-cell lysate (WCL) of *E. coli* K-12 C75a, a PhoA overproducer. Lanes 8 to 13 (from CC118[pRMA467]): lane 8, Triton X-100 insoluble fraction; lane 9, Triton X-100 soluble fraction; lane 10, periplasmic fraction; lane 11, cytoplasmic fraction, lane 12, WM fraction; lane 13, WCL. The equivalent of 2×10^8 cells was loaded into each lane. Coomassie brilliant blue staining of an identically loaded gel showed that lanes 8 to 11 had similar amounts of protein. The migration positions of the following molecular mass standards (Pharmacia) are indicated on the right (in kilodaltons): carbonic anhydrase (30), ovalbumin (43), bovine serum albumin (67), and phosphorylase *b* (94).

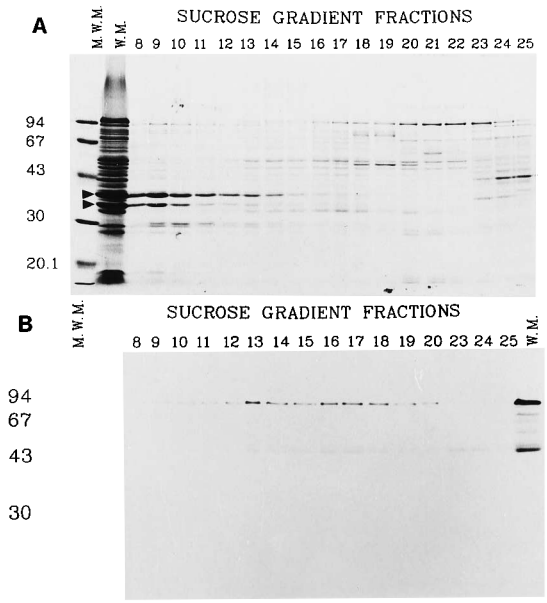

FIG. 7. Subcellular localization of Rol::PhoA471 fusion protein. WM were prepared by French press lysis of *E. coli* K-12 strain CC118 harboring pRMA471. The WMs were centrifuged in a 30 to 50% sucrose density gradient as detailed in Materials and Methods. Fractions (0.5 ml; numbers 8 to 25) were collected from the bottom of the centrifuge tube, and samples of each were electrophoresed on two SDS–12% polyacrylamide gels; fractions 1 to 7 did not contain protein and were omitted. One gel was stained with Coomassie brilliant blue (A), and the other was used for a Western immunoblot with rabbit antibacterial alkaline phosphatase to detect the Rol::PhoA471 fusion protein (B). The migration positions of the following molecular mass standards (M.W.M.) (Pharmacia) are indicated on the left in kilodaltons: soybean trypsin inhibitor, 20.1; carbonic anhydrase, 30; ovalbumin, 43; bovine serum albumin, 67; phosphorylase *b*, 94. In panel A, the major OM proteins OmpF+OmpC and OmpA are indicated by the two arrowheads in the M.W.M. lane. In panel B, lane W.M., of the two dark bands, the upper band shows Rol::PhoA471 protein and the band at approximately 47 kDa shows a stable degradation product of a size similar to that of alkaline phosphatase.

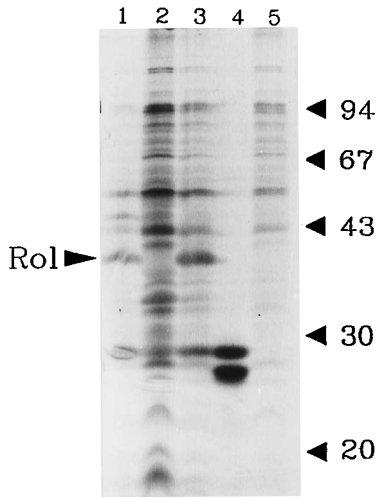

FIG. 8. Identification of the *rol* gene product in *E. coli* maxicells. Maxicell strains containing various plasmids were labelled with L-[^{35}S]methionine. The samples were electrophoresed on an SDS–12% polyacrylamide gel, and the labelled proteins were detected by autoradiography of the dried gel. The plasmids used were pRMA520 (lane 1), pK184 (lane 2), pRMA532 (lane 3), pRMA578 (lane 4) and no plasmid (lane 5). A protein band corresponding to the Rol protein is indicated on the left by an arrowhead. The migration positions of the following molecular mass standards (Pharmacia) are indicated (arrowheads on the right; in kilodaltons): soybean trypsin inhibitor, 20; carbonic anhydrase, 30; ovalbumin, 43; bovine serum albumin, 67; and phosphorylase *b*, 94.

Since Rol appeared to be well expressed and easily detectable in maxicells, we decided to fractionate maxicells containing L-[^{35}S]methionine-labelled Rol. Figure 9 shows that the Rol protein was located in the WM fraction (WM) and that about half of it could be solubilized by a single extraction with Triton X-100. These observations correlate well with the behavior of the Rol::PhoA fusion protein and provide strong evidence for its localization in the IM.

Construction of an *S. flexneri rol*::Km mutant. Since all work on the *rol* or *cld* gene undertaken thus far involved use of multicopy *rfb* and *rol* gene, we decided to mutate the chromosomal *rol* gene of *S. flexneri* and determine if this had the expected effect on O-antigen chain length regulation.

The *Sph*I-*Sph*I fragment containing the *rol* gene (Fig. 1) was cloned from pRMA520 into a suicide vector with the *sacB* gene as a contraselective marker (pCACTUS [10]) to yield pRMA577, and a kanamycin resistance (Km^r) cartridge was inserted into the *Bgl*II site to inactivate *rol* (Fig. 1). The resulting plasmid, pRMA584, was transformed into *S. flexneri* Sfl1 by electroporation prior to selection for resistance to kanamycin and chloramphenicol and growth at 30°C. Growth at 42°C and selection for chloramphenicol and kanamycin resistance resulted in chromosomal integration of the plasmid. Growth at 37°C without selection for the plasmid resulted in resolution of the cointegrate. Finally, selection for sucrose and kanamycin resistance resulted in isolation of strains having the *rol*::Km^r insertion and being cured of the plasmid. The resulting strains were confirmed to have an insertion of the appropriate size by Southern hybridization with the *Sph*I-*Sph*I fragment from pRMA520 as a probe (data not shown).

The LPS produced by the *rol*::Km^r strains (e.g., RMA585) showed that, as expected, the O-antigen chain length control was deregulated (Fig. 10, lane 3), although the O-antigen chains were shorter than those observed for the cloned *rfb* genes in *E. coli* K-12 (Fig. 2 and reference 28). This presumably reflects differences in either LPS core structure, interactions with other O-antigen biosynthesis proteins, or gene expression levels (as discussed below). The wild-type chain length distribution pattern was restored when pRMA577 having the *rol*$^+$ gene was introduced into RMA585 but not when pCAC TUS was electroporated into this strain (Fig. 10).

DISCUSSION

In a previous study, we observed that *E. coli* K-12 DH1 harboring cosmid clones with the genes for *S. flexneri* O-antigen biosynthesis produced LPS with two different banding patterns as detected by silver staining and Western immunoblotting of LPSs separated by SDS-PAGE (28). A large region which contained the gene(s) whose product regulated O-antigen chain length was identified. A subsequent communication (7) described a gene, termed *rol*, whose product regulated the chain length of the *E. coli* O75 O antigen. It has been further reported that the DNA sequence of the *E. coli* O75 and the *S. enterica* serovar typhimurium LT2 *rol* genes (6) and their products were approximately 70% identical. Bastin et al. (4, 5) reported the DNA sequence of genes termed *cld*, from *S. enterica* serovar typhimurium LT2 and *E. coli* O111, whose products regulate O-antigen chain length, and showed these sequences to be almost the same as that of the *rol* gene. A further *cld* function associated with the biosynthesis of the *S. enterica* serovar Muenchen strain M67 O antigen has also been

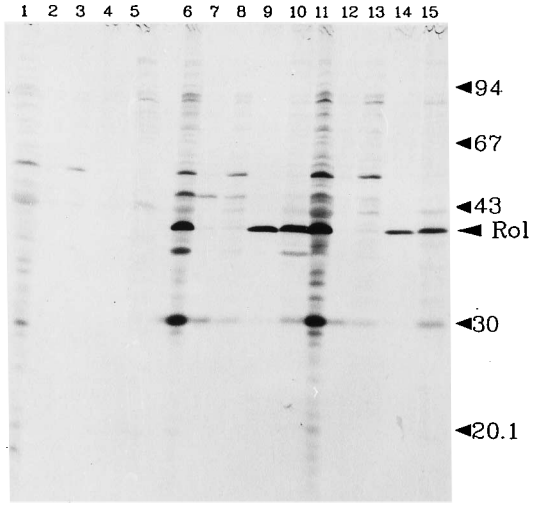

FIG. 9. Subcellular localization of the Rol protein in *E. coli* maxicells. Maxicell strain CSR603, containing no plasmid (lanes 1 to 5) or containing plasmid pRMA520 (lanes 6 to 10) or plasmid pRMA532 (lanes 10 to 15), was grown and UV irradiated, and the encoded Rol protein was labelled with L-[^{35}S]methionine. The labelled maxicells were then treated as detailed in Materials and Methods to yield the following fractions: whole cells (lanes 1, 6, and 11), cytoplasm (lanes 2, 7, and 12), periplasm (lanes 3, 8, and 13), Triton X-100-soluble fraction or IM (lanes 4, 9, and 14), and Triton X-100-insoluble fraction or OM (lanes 5, 10, and 15). The samples (equivalent to 2×10^8 cells) were electrophoresed on an SDS–15% polyacrylamide gel, and the labelled proteins were detected by autoradiography of the dried gel. The position of the Rol protein is indicated by the large arrowhead on the right. Controls in which the maxicell strain harboring the cloning vector (pK184) or inactivated *rol* gene (plasmid pRMA578) were also labelled and fractionated, but these data are not shown. The migration positions of the following molecular mass standards (Pharmacia) are indicated (arrowheads on the right; in kilodaltons): soybean trypsin inhibitor, 20.1; carbonic anhydrase, 30; ovalbumin, 43; bovine serum albumin, 67; phosphorylase *b*, 94.

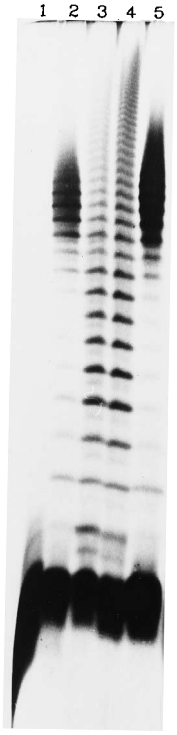

FIG. 10. LPSs of a *S. flexneri rol* mutant. LPSs were prepared by proteinase K treatment of whole-cell lysates of the strains indicated below. Samples (equivalent to 2×10^8 cells) were electrophoresed on an SDS–20% polyacrylamide gel, and LPSs were detected by silver staining. Lanes: 1, *E. coli* K-12 DH1; 2, *S. flexneri* Sfl1 (Y serotype); 3, *S. flexneri* RMA585 (Sfl1 *rol*::Km^r); 4, *S. flexneri* RMA588 (RMA585 with pCACTUS); 5, *S. flexneri* RMA587 (RMA585 with pRMA577, which is pCACTUS *rol*$^+$).

detected (9). The available evidence suggests that the Rol and Cld proteins are functionally interchangeable (5, 6). We believe that our original observations were made possible by the fortuitous occurrence of a mutation in the *rol* gene of the *E. coli* K-12 DH1 strain that we used.

DNA sequencing of the region adjacent to that whose sequence is already published (26, 33, 35) located the *rol* gene. An interesting feature of the 3,264 bp of DNA sequence of the region between the end of the *rfc* gene (35) and the start of the *gnd* gene is the uniformly low percent G+C content of the sequence. Although a number of ORFs have been identified, none have significant DNA or protein sequence similarity to any sequence in the databases. We know that several of these ORFs are not needed for O-antigen biosynthesis (28, 35).

The arrangement and DNA sequence of the *S. flexneri rol* and *hisIE* genes are almost identical to those in *E. coli* O75 (6). The *udg*, *rol*, and *hisIE* genes are organized as in *S. enterica* serovar typhimurium LT2 (5). We chose to call the gene for chain length *rol* instead of *cld* as it is nearly identical to the *rol* gene of *E. coli* O75 (Table 2). This identity is remarkable, since the *E. coli* O75 Rol and *E. coli* O111 Cld proteins have only 70% sequence identity. The possibility that *E. coli* O75 may be more closely related to *S. flexneri* is also reflected in the O-antigen sugar composition: the *E. coli* O75 O antigen contains the two sugars (*N*-acetylglucosamine and rhamnose) which are the components of the *S. flexneri* O-antigen tetrasaccharide repeat unit (21).

The hydropathy profile of the *S. flexneri* Rol protein (and those of similar proteins) indicates that it has two potential membrane-spanning regions, located near the amino and carboxy termini of the protein. We investigated the location and topology of Rol in the cell. A number of Tn*phoA* insertions inactivating the *rol* gene were isolated. All of the PhoA$^+$ fusions were in the hydrophilic segment of the Rol protein, between the putative membrane-spanning domains. The fusion proteins were all associated with the membrane fraction. Hence, our data suggest that the membrane-spanning segment at the amino terminus anchors the Rol::PhoA fusion proteins to the IM and that the hydrophilic region of the Rol protein is located in the periplasmic space. We assume, but have no direct evidence, that the predicted membrane-spanning segment near the carboxy terminus anchors this end of the Rol protein to the IM. We localized L-[^{35}S]methionine-labelled Rol protein (expressed in maxicells) by cell fractionation and extraction with Triton X-100 to the membrane, and a large proportion of it was solubilized with a single extraction, suggesting an IM location. Cells producing an Rol::PhoA fusion protein were fractionated, and it was found to be membrane associated and could be largely solubilized with Triton X-100. This was further confirmed by locating the Rol::PhoA fusion protein to the cytoplasmic membrane fraction by sucrose density gradient centrifugation (Fig. 7). In summary, our data suggest that the Rol protein is located in the IM and has a large hydrophilic domain located in the periplasmic spaces. This provides experimental evidence for the model proposed by Bastin et al. (5) for the Cld protein which was based only on the protein hydropathy plot and a model for its mode of action.

The model for Cld action proposed by Bastin et al. (5) has the Cld protein acting on the O-antigen polymerase (Rfc protein) in a time-dependent manner to switch this enzyme be-

tween cycles of O-antigen polymerization and transfer of the growing, acyl carrier lipid-linked, O-antigen chain to the O antigen: lipid A ligase (RfaL protein [41]) for ligation to lipid A-core sugar molecules. The molecular basis of the timing mechanism is unknown. We suggest that the Rol protein acts as a type of molecular chaperone which interacts with the RfaL protein to assemble a complex. This complex would give a specific ratio of RfaL to Rfc, thereby altering the overall kinetics of the ligation reaction (enzyme-substrate ratio) to give the observed nonrandom O-antigen chain length distribution. In the absence of Rol protein, the ratio of RfaL to Rfc is not fixed and the proteins are free to diffuse in the 2-dimensional plane of the membrane. Random interaction (diffusion-limited collision) results in nonmodal O-antigen chain length distribution. The available data (5, 6, 43) indicate that while Rol and Cld proteins from different sources are functionally interchangeable, the LPS banding patterns observed are slightly different when heterologous Rol and Cld proteins are used in complementation tests. Indeed, the LPS banding pattern for the *S. flexneri* rol::Km^r mutant was different from that observed for *E. coli* K-12 harboring cloned *rfb* genes. This difference may arise because of one or a combination of the following reasons: some specificity of the Rol protein for the O-antigen polysaccharide, for Rfc, or for RfaL; gene dosage effects due to the presence of *rfb* genes on plasmids; and/or differences in the structure of the core sugars attached to lipid A. Interestingly, the RfaL protein of *E. coli* K-12 has little specificity for the O antigen, since it efficiently forms LPSs with O antigens as diverse as those from *Vibrio cholerae* O1 (3, 29) and *S. enterica* serovar typhimurium (36). Hence, RfaL recognizes the acyl carrier lipid and either a specific lipid A-core sugar structure as modified by RfaK or the RfaK protein itself (41). Like Rfc proteins (35), the only two characterized RfaL proteins (from *E. coli* K-12 and *S. enterica* serovar typhimurium) have little identity at the amino acid sequence level but have nearly identical hydrophobicity plots, suggesting that they are structurally similar (23). The Rol protein may act by recognizing a structural feature of the RfaL and Rfc proteins. Hence Rol and Cld proteins from different sources may function in heterologous systems because the protein(s) they interact with is structurally similar. Small differences between these interaction systems may be reflected in the differences in O-antigen chain length distribution patterns. The identification, localization, and topology of all the proteins involved in O-antigen synthesis are required for a full understanding of this process.

We have constructed the first chromosomal mutation affecting a *rol* gene. The LPS produced by the *S. flexneri rol* mutants has the expected deregulated O-antigen chain length distribution pattern. These and similar strains are being used to explore the biological significance of regulating the length distribution of O-antigen polysaccharide chains.

ACKNOWLEDGMENTS

We thank the National Health and Medical Research Council of Australia and the Diarrhoeal Diseases Vaccines Programme of the World Health Organization for support. R.M. is an R. Douglas Wright Fellow.

U. Stroeher and S. Williams are thanked for their stimulating discussions on the function of Rol.

REFERENCES

1. **Achtman, M., P. A. Manning, C. Edelbluth, and P. Herrlich.** 1979. Export without proteolytic processing of inner and outer membrane proteins encoded by F sex factor *tra* cistrons in *Escherichia coli* minicells. Proc. Natl. Acad. Sci. USA **76:**4837–4841.
2. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**404–410.
3. **Attridge, S. R., D. Daniels, J. K. Morona, and R. Morona.** 1990. Surface coexpression of *Vibrio cholerae* and *Salmonella typhi* O-antigens on Ty21a clone EX210. Microb. Pathog. **8:**177–188.
4. **Bastin, D. A., L. K. Romana, and P. R. Reeves.** 1991. Molecular cloning and expression in *Escherichia coli* K-12 of the *rfb* gene cluster determining the O antigen of an *E. coli* O111 strain. Mol. Microbiol. **5:**2223–2231.
5. **Bastin, D. A., G. Stevenson, P. K. Brown, A. Haase, and P. R. Reeves.** 1993. Repeat unit polysaccharides of bacteria: a model for polymerisation resembling that of ribosomes and fatty acid synthetase, with a novel mechanism for determining chain length. Mol. Microbiol. **7:**725–734.
6. **Batchelor, R. A., P. Alifano, E. Biffali, S. I. Hull, and R. A. Hull.** 1992. Nucleotide sequence of the genes regulating O-polysaccharide antigen chain length (*rol*) from *Escherichia coli* and *Salmonella typhimurium*: protein homology and functional complementation. J. Bacteriol. **174:**5228–5236.
7. **Batchelor, R. A., G. E. Haraguchi, R. A. Hull, and S. I. Hull.** 1991. Regulation by a novel protein of the bimodal distribution of lipopolysaccharide in the outer membrane of *Escherichia coli*. J. Bacteriol. **173:**5699–5704.
8. **Brown, M. C. M., A. Weston, J. R. Saunders, and G. O. Humphreys.** 1979. Transformation of *E. coli* C600 by plasmid DNA at different phases of growth. FEMS Microbiol. Lett. **5:**219–222.
9. **Brown, P. K., L. K. Romana, and P. R. Reeves.** 1991. Cloning of the *rfb* gene cluster of group C2 *Salmonella*: comparison with the *rfb* regions of groups B and D. Mol. Microbiol. **5:**1873–1881.
10. **Clark, C. A., and P. A. Manning.** Unpublished data.
11. **Daniels, D. L., G. Plunket III, V. D. Burland, and F. R. Blattner.** 1992. Analysis of the *Escherichia coli* genome: DNA sequence of the region from 84.5 to 86.5 minutes. Science **257:**771–778.
12. **Dobrowski, P.** 1991. Plasmids with easily excisable *cat* gene cartridges. Gene **122:**139–140.
13. **Dougherty, B. A., and I. van de Rijn.** 1993. Molecular characterization of *hasB* from an operon required for hyaluronic acid synthesis in group A streptococci: demonstration of UDP-glucose dehydrogenase activity. J. Biol. Chem. **268:**7118–7124.
14. **Garcia, E., P. Garcia, and R. Lopez.** 1993. Cloning and sequencing of a gene involved in the synthesis of the capsular polysaccharide of *Streptococcus pneumoniae* type 3. Mol. Gen. Genet. **239:**188–195.
15. **Hawkes, R. E., E. Niday, and J. Gordon.** 1982. A dot blot-immunobinding assay for monoclonal and other antibodies. Anal. Biochem. **119:**142–147.
16. **Henikoff, S.** 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene **28:**351–359.
17. **Higgins, D. G., and P. M. Sharp.** 1988. CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene **73:**237–244.
18. **Higgins, D. G., and P. M. Sharp.** 1989. Fast and sensitive multiple sequence alignments on a microcomputer. Comput. Appl. Biosci. **5:**151–153.
19. **Hitchcock, P. J., and T. M. Brown.** 1983. Morphological heterogeneity among *Salmonella* lipopolysaccharide chemotypes in silver-stained polyacrylamide gels. J. Bacteriol. **154:**269–277.
20. **Hofmann, K., and W. Stoffel.** 1992. An interactive graphical tool for protein sequence analysis. Comput. Appl. Biosci. **8:**331–337.
21. **Jann, K., and B. Jann.** 1985. Cell surface components and virulence: *Escherichia coli* O and K antigens in relation to virulence and pathogenicity, p. 157–176. *In* M. Sussman (ed.), The virulence of *Escherichia coli*: reviews and methods. Society for General Microbiology, Academic Press, London.
22. **Jobling, M. G., and R. K. Holmes.** 1990. Construction of vectors with the p15a replicon, kanamycin resistance, inducible *lacZ* and pUC18 or pUC19 multiple cloning sites. Nucleic Acids Res. **18:**5315–5316.
23. **Klena, J. D., E. Pradel, and C. A. Schnaitman.** 1992. Comparison of lipopolysaccharide biosynthesis genes *rfaK*, *rfaL*, *rfaY*, and *rfaZ* of *Escherichia coli* K-12 and *Salmonella typhimurium*. J. Bacteriol. **174:**4746–4752.
24. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157:**105–132.
25. **Lugtenberg, B., J. Meijers, R. Peters, P. van der Hoek, and L. van Alphen.** 1975. Electrophoretic resolution of the major membrane proteins of *Escherichia coli* K-12 into four major bands. FEBS Lett. **58:**254–258.
26. **Macpherson, D. F., P. A. Manning, and R. Morona.** 1994. Characterisation of the TDP-rhamnose biosynthetic genes encoded in the *rfb* locus of *Shigella flexneri*. Mol. Microbiol. **11:**281–292.
27. **Macpherson, D. F., P. A. Manning, and R. Morona.** Submitted for publication.
28. **Macpherson, D. F., R. Morona, D. W. Beger, K.-C. Cheah, and P. A. Manning.** 1991. Genetic analysis of the *rfb* region of *Shigella flexneri* encoding the Y serotype O-antigen specificity. Mol. Microbiol. **5:**1491–1499.
29. **Manning, P. A., M. W. Heuzenroeder, J. Yeadon, D. I. Leavesley, P. R. Reeves, and D. Rowley.** 1986. Molecular cloning and expression in *Escherichia coli* K-12 of the O antigens of the Inaba and Ogawa serotypes of the *Vibrio cholerae* O1 lipopolysaccharides and their potential for vaccine development. Infect. Immun. **53:**272–277.
30. **Manoil, C., and J. Beckwith.** 1985. Tn*phoA*: a transposon probe for protein export signals. Proc. Natl. Acad. Sci. USA **82:**8129–8133.
31. **Meier-Dieter, U., K. Barr, R. Starman, L. Hatch, and P. D. Rick.** 1992. Nucleotide sequence of the *Escherichia coli rfe* gene involved in the synthesis of enterobacterial common antigen. J. Biol. Chem. **267:**746–753.

32. **Morona, R., M. H. Brown, J. Yeadon, M. W. Heuzenroeder, and P. A. Manning.** 1991. Effect of lipopolysaccharide core synthesis mutations on the production of *Vibrio cholerae* O-antigen in *Escherichia coli* K-12. FEMS Microbiol. Lett. **82:**279–286.
33. **Morona, R., D. F. Macpherson, L. Van Den Bosch, N. I. A. Carlin, and P. A. Manning.** Submitted for publication.
34. **Morona, R., P. A. Manning, and P. R. Reeves.** 1983. Identification and characterization of the TolC protein, an outer membrane protein from *Escherichia coli*. J. Bacteriol. **153:**693–699.
35. **Morona, R., M. Mavris, A. Fallarino, and P. A. Manning.** 1994. Characterization of the *rfc* region of *Shigella flexneri*. J. Bacteriol. **176:**733–747.
36. **Neal, B. L., P. K. Brown, and P. R. Reeves.** 1993. Use of *Salmonella* phage P22 for transduction in *Escherichia coli*. J. Bacteriol. **175:**7115–7118.
37. **Neuhard, J., and E. Thomassen.** 1976. Altered deoxyribonucleotide pools in P2 eductants of *Escherichia coli* K-12 due to deletion of the *dcd* gene. J. Bacteriol. **126:**999–1001.
38. **Reeves, P. R.** 1993. Evolution of *Salmonella* O-antigen variation by interspecific gene transfer on a large scale. Trends Genet. **9:**17–22.
39. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed., vol. 1, 2, and 3. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
40. **Sarvas, M.** 1985. Membrane fractionation methods, p. 111–122. *In* T. K. Korhonen, E. A. Dawes, and P. H. Mäkelä (ed.), Enterobacterial surface antigens: methods for molecular characterisation. Elsevier Science Publishing, Amsterdam.
41. **Schnaitman, C. A., and J. D. Klena.** 1993. Genetics of lipopolysaccharide biosynthesis in enteric bacteria. Microbiol. Rev. **57:**655–682.
42. **Simmons, D. A. R., and E. Romanowska.** 1987. Structure and biology of *Shigella flexneri* O-antigens. J. Med. Microbiol. **23:**289–302.
43. **Stoebner, J. A., and S. M. Payne.** 1988. Iron-regulated hemolysin production and utilization of heme and hemoglobin by *Vibrio cholerae*. Infect. Immun. **56:**2891–2895.