JOURNAL OF BACTERIOLOGY, Jan. 1993, p. 148–158
0021-9193/93/010148-11$02.00/0
Copyright © 1993, American Society for Microbiology
Vol. 175, No. 1

Identification, Expression, and DNA Sequence of the GDP-Mannose Biosynthesis Genes Encoded by the O7 *rfb* Gene Cluster of Strain VW187 (*Escherichia coli* O7:K1)

CRISTINA L. MAROLDA AND MIGUEL A. VALVANO*

Department of Microbiology and Immunology, University of Western Ontario, London, Ontario N6A 5C1, Canada

Received 14 August 1992/Accepted 22 October 1992

The O7-specific lipopolysaccharide (LPS) in strains of *Escherichia coli* consists of a repeating unit made of galactose, mannose, rhamnose, 4-acetamido-2,6-dideoxyglucose, and *N*-acetylglucosamine. We have recently cloned and characterized genetically the O7-specific LPS biosynthesis region (rfb_{EcO7}) of the *E. coli* O7:K1 strain VW187 (C. L. Marolda, J. Welsh, L. Dafoe, and M. A. Valvano, J. Bacteriol. 172:3590–3599, 1990). In this study, we localized the *gnd* gene encoding gluconate-6-phosphate dehydrogenase at one end of the rfb_{EcO7} gene cluster and sequenced that end of the cluster. Three open reading frames (ORF) encoding polypeptides of 275, 464, and 453 amino acids were identified upstream of gnd_{EcO7}, all transcribed toward the *gnd* gene. ORF275 had 45% similarity at the protein level with ORF16.5, which occupies a similar position in the *Salmonella enterica* LT2 *rfb* region, and presumably encodes a nucleotide sugar transferase. The polypeptides encoded by ORFs 464 and 453 were expressed under the control of the *ptac* promoter and visualized in Coomassie blue-stained sodium dodecyl sulfate-polyacrylamide gels and by maxicell analysis. ORF464 expressed GDP-mannose pyrophosphorylase and ORF453 encoded a phosphomannomutase, the enzymes for the biosynthesis pathway of GDP-mannose, one of the nucleotide sugar precursors for the formation of the O7 repeating unit. They were designated $rfbM_{EcO7}$ and $rfbK_{EcO7}$, respectively. The $RfbM_{EcO7}$ polypeptide was homologous to the corresponding protein in *S. enterica* LT2, XanB of *Xanthomonas campestris*, and AlgA of *Pseudomonas aeruginosa*, all GDP-mannose pyrophosphorylases. $RfbK_{EcO7}$ was very similar to CpsG of *S. enterica* LT2, an enzyme presumably involved in the biosynthesis of the capsular polysaccharide colanic acid, but quite different from the corresponding RfbK protein of *S. enterica* LT2.

Lipopolysaccharides (LPS) are an integral component of the outer membrane in gram-negative microorganisms (30). Three moieties can be distinguished in the LPS molecule: the O-specific side-chain polysaccharide, the core oligosaccharide, and the lipid A (18). O-specific polysaccharides are composed of repeating oligosaccharide units and display an enormous variation among bacteria from different or similar species in terms of carbohydrate composition and chemical structures (29, 42).

The genes involved in the biosynthesis of the O-specific LPS in *Escherichia coli* and *Salmonella enterica* are localized on the bacterial chromosome in the *rfb* locus, which maps at 44 min on the *E. coli* K-12 linkage map and at 42 min on the *Salmonella* linkage map, in both cases near *gnd* and the *his* operon (1, 47). *gnd* encodes gluconate-6-phosphate dehydrogenase, a housekeeping enzyme involved in the metabolism of carbohydrates, and does not play any direct role in the biosynthesis of LPS (14). The *rfb* genes from several serovars of *S. enterica* have recently been cloned and characterized (6, 19, 24, 27, 58). In contrast, studies on the *rfb* region of *E. coli* K-12 are difficult since this organism does not express a complete LPS with O-specific side chains, presumably due to an unknown mutation(s) in the *rfb* genes.

To elucidate the genetic organization at the molecular level of O-specific side-chain LPS genes in *E. coli*, we are studying the *rfb* genes from the strain VW187 (36, 55), a clinical isolate of *E. coli* O7:K1 (54, 57). The O7-LPS is a virulence determinant of *E. coli* K1 strains involved in enhancing bacterial survival to serum complement-mediated killing (see reference 53 for a review). The O7 repeating unit consists of a pentasaccharide composed of mannose, galactose, rhamnose, *N*-acetylglucosamine, and 4-acetamido-2,6-dideoxyglucose (31).

We have previously reported the cloning of the rfb_{EcO7} region and its subsequent characterization by transposon mutagenesis in the cloned DNA and site-directed mutagenesis in the chromosome of the wild-type strain VW187 (35, 55). Also, while the rfb_{EcO7} gene cluster proved to be unique with respect to *rfb* genes of other *E. coli* O-types (55), it is highly homologous to the *rfb* gene cluster of *Shigella boydii* type 12 belonging to the SB1 clone (56).

In the present study, we localized the gnd_{EcO7} gene flanking one end of the rfb_{EcO7} gene cluster and sequenced that end of the *rfb* region. We report the genetic and functional characterization of two genes, $rfbK_{EcO7}$ and $rfbM_{EcO7}$, encoding phosphomannomutase and GDP-mannose pyrophosphorylase, respectively. The enzymes are involved in the conversion of mannose-6-phosphate to GDP-mannose, one of the nucleotide sugar precursors required for the synthesis of the O7-specific repeating unit.

MATERIALS AND METHODS

Bacterial strains, plasmids, and bacteriophages. Strains and plasmids used in this study are described in Table 1. CLM4 is a *recA* derivative of strain SØ874 which was constructed by generalized transduction with a P1 lysate (10) obtained from the Δ(*recA-srl*)*306 srl-301*::Tn*10* strain JC10289 (9).

* Corresponding author. Electronic mail address: 37_1510@uwovax.ca

TABLE 1. *E. coli* strains, plasmids, and bacteriophages used in this study

Strains, plasmids, and phages	Relevant properties[a]	Source or reference
E. coli		
DH5α	F⁻ φ80*lac*dZM15 *endA1 recA1 hsdR17*(r_K^- m_K^-) *supE44 thi-1 gyrA96 relA1*? Δ(*lacZYA-argF*)*U169*	Laboratory stocks
SØ874	*lacZ2286 trp-49* Δ(*sbcB-rfb*)*86 upp-12 relA1 rps1150* λ⁻	39
CLM3	SØ874 *recA* Tc^r	This work
CLM4	CLM3 Tc^s	This work
VW187	Prototrophic, O7:K1:H?	54, 57
HB101	*hsdS serA ara proA lacY galK rspL xyl mtl supE*	Laboratory stocks
DF710	*fhuA22*? *edd-1 gnd-1 tyrA2 relA1*? *rpsL125 pit-10*? *spoT1*? *thi-1* λ⁻?	44
GMS343	*aroD6 argE3 lacY1 galK2 man-4 mtl-1 rpsL700 tsx-29*? *supE44*? λ⁻	41
JC10289	*thr-1 ara-14 leuB6* Δ(*gpt-proA*)*62 lacY1 tsx-33 supE44 galK2* λ⁻ *rac hisG4*(Oc) *rfbD1 mgl-51* Δ(*recA-srl*)*306 srl-301*::Tn*10-84 rpsL31 kdgK51 xyl-5 mtl-1 argE3 thi-1*	9
Plasmids		
pACYC184	Cloning vector, Cm^r Tc^r	7
pGEM3	Cloning and sequencing vector, Ap^r	Promega Biotech
pGEM4	Cloning and sequencing vector, Ap^r	Promega Biotech
pBR322	Cloning vector, Ap^r Tc^r	Laboratory stocks
pEX1	p*tac lacI*^{q} *rrnBt* Ap^r	43
pEXZ	Promoterless *lacZ* (*Eco*RI-*Dra*I from pTL38) Ap^r	T. Linn (25)
pMN1	3.7-kb *Hin*dIII fragment from (λ *h*80 *dgnd his*) cloned into pBR322, Ap^r	38
pJHCV31	O7⁺ cosmid, Tc^r	55
pJHCV32	O7⁺ cosmid, Tc^r	55
pJHCV64	O7⁺ 14-kb *Bst*EII fragment from pJHCV32 cloned into pHP45Ω	55
pMAV57	10.2-kb *Hin*dIII partial from pJHCV32 cloned into pBR322, Ap^r	This work
pCM10	8.1-kb *Hin*dIII fragment of pJHCV31 cloned into pACYC184, Cm^r	This work
pCM29	5.1-kb *Hin*dIII fragment from pJHCV32 cloned into pGEM3, Ap^r	This work
pCM39	2.2-kb *Eco*RI fragment from pJHCV32 cloned into pGEM3, Ap^r	This work
pCM60	1.4-kb *Eco*RV fragment from pJHCV32 cloned into pGEM3, Ap^r	This work
pCM70	0.5-kb *Eco*RV-*Kpn*I fragment from pMAV57 cloned into pGEM3, Ap^r	This work
pCM101	3.7-kb *Eco*RI from pJHCV64 cloned into pEX1, Ap^r	This work
pCM103	3.6-kb *Eco*RI-*Sma*I from pCM101 cloned into pEX1, Ap^r	This work
pCM104	2.3-kb *Spe*I-*Sma*I fragment from pCM103 cloned into pEX1, Ap^r	This work
pCM108	3.0-kb *Pst*I fragment from pJHCV32 cloned into pEX1, Ap^r	This work
pJHCV31::TnHoHo1-315L	*rfbM*$_{EcO7}$::Tn*3*HoHo1-α315, O7⁻	36
pJHCV31::TnHoHo1-312R	*rfbM*$_{EcO7}$::Tn*3*HoHo1-α312, O7⁻	36
pJHCV31::TnHoHo1-310R	*rfbM*$_{EcO7}$::Tn*3*HoHo1-α310, O7⁻	36
pJHCV32::TnHoHo1-37L	*rfbM*$_{EcO7}$::Tn*3*HoHo1-ω37, O7⁻	36
pJHCV32::TnHoHo1-31L	*rfbM*$_{EcO7}$::Tn*3*HoHo1-ω31, O7⁻	36
pJHCV31::TnHoHo1-406R	*rfbM*$_{EcO7}$::Tn*3*HoHo1-α406, O7⁺	36
Phage P1 vir		C. Schnaitman

[a] Abbreviations: Ap, ampicillin; Tc, tetracycline; Cm, chloramphenicol.

Tetracycline-resistant transductants unable to grow in minimal medium without histidine were screened for the *recA* phenotype by UV sensitivity. Cells were streaked onto LB plates, and different sectors of the plates were exposed to UV light for 0, 15, 30, and 45 s and incubated overnight at 37°C. *recA* mutants were not able to grow when exposed to UV light for more than 15 s; one of these mutants was designated CLM3. To obtain a tetracycline-sensitive derivative, we plated a late-logarithmic-phase culture of CLM3 onto Luria plates containing chlortetracycline and fusaric acid as described by Maloy and Nunn (33). This method selects for tetracycline-sensitive colonies which are resistant to fusaric acid. One of the colonies, designated CLM4, was kept and used in our studies. For the determination of the *gnd* phenotype, the *gnd edd* strain DF710 containing various plasmids was grown on MacConkey gluconate agar plates. *gnd edd* strains give white colonies in this medium, whereas *gnd*⁺ *edd* strains give rise to pink colonies (3). Experiments with strain GMS343 (*manA*) were done in M9 minimal medium containing either glucose or glucose and D-mannose as carbon sources.

Materials. Chemicals and antibiotics were obtained from Sigma Chemical Co., St. Louis, Mo. Restriction enzymes and T4 DNA ligase were purchased from Boehringer Mannheim, Dorval, Quebec, Canada. T7 Sequencing Kit was purchased from Pharmacia LKB Biotechnology, Baie d'Urfe, Quebec, Canada. Kits and enzymes were used according to the conditions suggested by the suppliers.

Recombinant DNA methods. Small-, intermediate-, and large-scale purifications of plasmid DNA (from 5-, 100-, and 1,000-ml cultures, respectively) were done as described elsewhere (36). Electrophoresis of plasmid DNA and DNA fragments was performed as previously described (54). Transformations were done by the calcium chloride method (8). pCM104 (Fig. 1) was constructed as follows. (i) pCM101 resulted from the cloning into the expression vector pEX1 of a 3.7-kb *Eco*RI fragment from pJHCV64 containing *rfbM*$_{EcO7}$ and part of *rfbK*$_{EcO7}$; (ii) a *Sma*I deletion was made in pCM101 removing the downstream *Eco*RI site, giving rise to pCM103; and (iii) an *Eco*RI-*Spe*I deletion was made in pCM103, giving rise to pCM104. pCM108 is a 3-kb *Pst*I fragment from pJHCV32 cloned into pEX1 (Fig. 1).

DNA sequencing and analysis. DNA was sequenced by the dideoxy method of Sanger et al. (48) modified to use with the

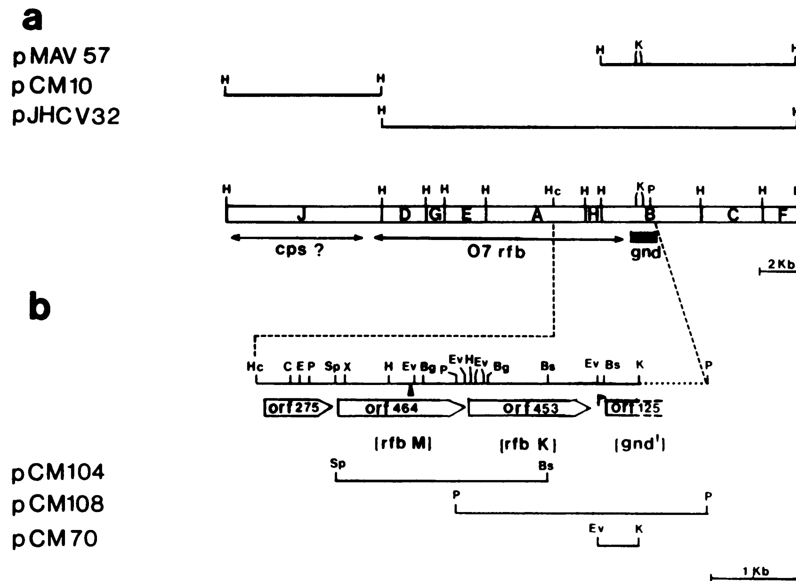

FIG. 1. Genetic analysis of the rfb_{EcO7} region. (a) J, D, G, E, A, H, B, C, and F denote the various *Hin*dIII fragments spanning the O7-LPS region and flanking sequences (36). Arrows beneath the restriction map indicate the boundaries of rfb_{EcO7}, and *cps* ? denotes a region upstream of rfb_{EcO7} suspected to encode part of the colanic acid biosynthesis genes. Box indicates the boundaries of the *gnd* gene. (b) Restriction endonuclease map of the expanded 4,274-bp *Hin*cII-*Kpn*I fragment containing the GDP-mannose biosynthesis genes and part of *gnd*. orf275, orf464 (*rfbM*), orf453 (*rfbK*), and orf125 (*gnd'*) were identified by DNA sequencing (see Fig. 3). Flag upstream of *gnd'* indicates the location of the *gnd* promoter region. ▲ indicates the position of insertion 37 (see also Fig. 3). H, *Hin*dIII; K, *Kpn*I; Hc, *Hin*cII; P, *Pst*I; C, *Cla*I; E, *Eco*RI; Sp, *Spe*I; X, *Xba*I; Ev, *Eco*RV; Bg, *Bgl*II; Bs, *Bst*EII.

T7 Sequencing Kit. Various overlapping fragments spanning a 4,271-bp *Hin*cII-*Kpn*I fragment (Fig. 1) were cloned into either pGEM3 or pGEM4. These recombinant plasmids were sequenced by using the T7 or SP6 promoter primers. Nested deletions by the method described by Henikoff (17) were made in pCM29, pCM39, and pCM60 (Table 1) to complete and confirm the DNA sequence of some regions. The endpoints of Tn*3*HoHo1 transposon insertions in plasmids pJHCV31 and pJHCV32 were sequenced by using the synthetic primer 5'-CGGTCATCTGAGACCATTA-3', which anneals to the left end of the transposon element proximal to the promoterless *lacZ* gene (51).

DNA sequences were analyzed with the GCG package version 7 (University of Wisconsin [11]). The FASTA program (26) was used for searching protein and DNA sequences data bases (GenBank, EMBL and Swissprot). The program MOTIFS was used to search the PROSITE data base (2). Paired sequencing alignments and multiple alignments were done with the GCG programs BESTFIT and PILEUP, respectively. Protein was analyzed with the program PROFILEGRAPH version 1.3 obtained from K. O. Hofmann, Institüt für Biochemie, Universitätzu Köln, Cologne, Germany. Hydropath profiles were calculated by the method of Kyte and Doolitle (22).

Hybridizations. Hybridization experiments were done to investigate the presence of *gnd* in the vicinity of the O7-LPS region. DNA fragments, obtained from pMN1 (Table 1), were recovered by electrophoresis into a strip of DE-81 filter paper (Whatman, Inc., Clifton, N.J.) as previously described (36) and labelled with [^{32}P]ATP (Amersham Canada Ltd., Oakville, Ontario, Canada) by oligonucleotide synthesis (13). The *gnd* probe was a 0.5-kb *Kpn*I fragment internal to the *gnd* gene, and the IS*5* probe was a 2.0-kb *Hin*dIII-*Bcl*I fragment containing the IS*5* insertion element. Southern blot hybridizations were done as described previously (54, 55).

Preparation of cell extracts. Cell extracts for the determination of GDP-mannose pyrophosphorylase and phosphomannomutase were prepared from 500-ml cultures. Cells were grown for 5 h prior to induction with 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG). At 2 h postinduction, cells were centrifuged at 5,930 × *g* and washed once with 150 mM NaCl. Sedimented cells were resuspended in either (i) 50 mM Tris-HCl buffer (pH 7.0)–10 mM $MgCl_2$–1 mM EDTA for the determination of GDP-mannose pyrophosphorylase or (ii) 10 mM 4-morpholinepropanesulfonic acid (MOPS) (pH 7.0)–2 mM 1,4-dithiothreitol for the determination of phosphomannomutase activity. Cell suspensions were sonicated (four 80-s pulses) with a Sonifier Cell Disruptor 350 (Branson Ultrasonics Corp., Danbury, Conn.). Cell lysates were centrifuged for 30 min at 41,000 × *g*, and the supernatants were used for enzyme assays. Protein concentration was determined by the Lowry method (28).

GDP-mannose pyrophosphorylase assay. GDP-mannose pyrophosphorylase (EC 2.7.7.13) was assayed by the procedure of Munch-Petersen (37) with some modifications. The final 1-ml volume of reaction mixture contained cell extract (50 μg of protein), 10 mM sodium fluoride, 1 mM $MgCl_2$ 50 mM Tris-HCl buffer (pH 7.0), 0.4 mM glucose, 0.1 mM ADP, 0.1 mM GDP-mannose, 12 U of hexokinase, and 3 U of glucose-6-phosphate dehydrogenase. The reaction was started by the addition of Na PP_i (freshly prepared) to give a final concentration of 1 mM. The enzyme activity was monitored by measuring the A_{340} with an Ultrospec Plus spectrophotometer (Pharmacia LKB Biotechnology, Baie d'Urfe, Quebec, Canada). Background levels were subtracted from enzyme assays in which PP_i was omitted. One unit of enzyme activity was defined as that which reduced 1 μM of NADP min^{-1} at 25°C.

Phosphomannomutase assay. Phosphomannomutase (EC 5.4.2.8) was assayed as described by Sá-Correia et al. (46)

with some modifications. A 1-ml reaction mixture contained 200 μg of protein, 50 mM Tris-HCl, 0.25 μM α-D-glucose-1-6-diphosphate, 0.5 U of phosphoglucose isomerase, 0.5 U of glucose-6-phosphate dehydrogenase, 0.5 U of phosphomannose isomerase (PMI), 1 μM NADP, 1 μM mannose-1-phosphate, and 1 μM cysteine (40). The enzyme activity was monitored as described above for GDP-mannose pyrophosphorylase. Enzyme units were calculated after subtracting background levels from parallel assays in which mannose-1-phosphate was omitted. One unit of enzyme activity was defined as that which reduced 1 μM of NADP min^{-1}.

LPS and protein analysis. The presence of O7-specific LPS in whole cells was examined by coagglutination (55). Immunoblot analysis was done on LPS samples prepared and processed as described elsewhere (36). To investigate the expression of *rfbM*$_{EcO7}$ and *rfbK*$_{EcO7}$ gene products encoded by pCM104 and pCM108, respectively, we induced early-logarithmic-phase cultures with IPTG (1 mM) for 30 min. Aliquots (1.5 ml) were centrifuged, and cell pellets were suspended in 25 μl of distilled H_2O–25 μl of 2× tracking dye (0.125 M Tris-HCl [pH 7], 4% sodium dodecyl sulfate [SDS], 20% glycerol, 0.02% bromophenol blue, 10% 2-mercaptoethanol). The suspensions were boiled for 10 min, and aliquots (10 μl) were loaded into SDS 10% polyacrylamide gels. Samples from parallel cultures not induced with IPTG were used as controls. After electrophoresis (23), gels were stained with Coomassie blue. Expression of polypeptides in vivo was examined by maxicells labelled with [^{35}S]methionine (5).

Nucleotide sequence accession number. The nucleotide sequence presented in this report has been deposited with GenBank under accession number L04596.

RESULTS AND DISCUSSION

Localization of *gnd* flanking one end of the *rfb*$_{EcO7}$ region. The *gnd* gene of *E. coli* K-12 is located at approximately 44 min on the linkage map between the *rfb* region and the *his* operon (1). We have previously shown that the *rfb*$_{EcO7}$ gene cluster of strain VW187 is linked to the *his* operon (55) as in other strains of *E. coli* and *S. enterica* (42, 47). In *S. enterica* LT2, *gnd* is located immediately adjacent to the downstream end of the *rfb* gene cluster (19). Since *E. coli* and *Salmonella* species are genetically related, this may apply to *E. coli* VW187. The *gnd*$_{EcO7}$ gene was located with the purpose of defining precisely one of the endpoints of the *rfb*$_{EcO7}$ region. Plasmids pCM10 and pMAV57 containing sequences flanking the *rfb*$_{EcO7}$ gene cluster (Fig. 1a) were transformed into strain DF710. Only pMAV57 conferred a positive *gnd* phenotype (see Materials and Methods), suggesting that it carries the *gnd* gene. A 0.5-kb *Kpn*I fragment in pMAV57 and pJHCV32 (Fig. 1) is similar to a *Kpn*I fragment of *gnd* in *E. coli* K-12 W3110 (3). Hybridization with the K-12 *gnd* probe showed homology with a 5.5-kb *Hin*dIII fragment (fragment B in Fig. 1a), a 5-kb *Bgl*II fragment, and the 0.5-kb *Kpn*I fragment of pMAV57 (Fig. 1a, and data not shown), placing *gnd*$_{EcO7}$ adjacent to the right end of the O7-LPS genes as shown in Fig. 1a. A copy of the IS5 insertion element is located upstream of *gnd* in *E. coli* K-12 W3110 (3). A hybridization experiment with the IS5 probe did not reveal any detectable homology with pMAV57, indicating that IS5 is not present in the corresponding region of the physical map of the VW187 chromosome.

A 0.5-kb *Eco*RV-*Kpn*I fragment was cloned into pGEM3, giving rise to pCM70 (Fig. 1b), and it was sequenced. The DNA sequence revealed the existence of a truncated open reading frame (ORF) of 125 amino acids with 95% identity to the corresponding region of the *gnd*$_{EcO7}$, identifying it as part of the *gnd* gene of strain VW187.

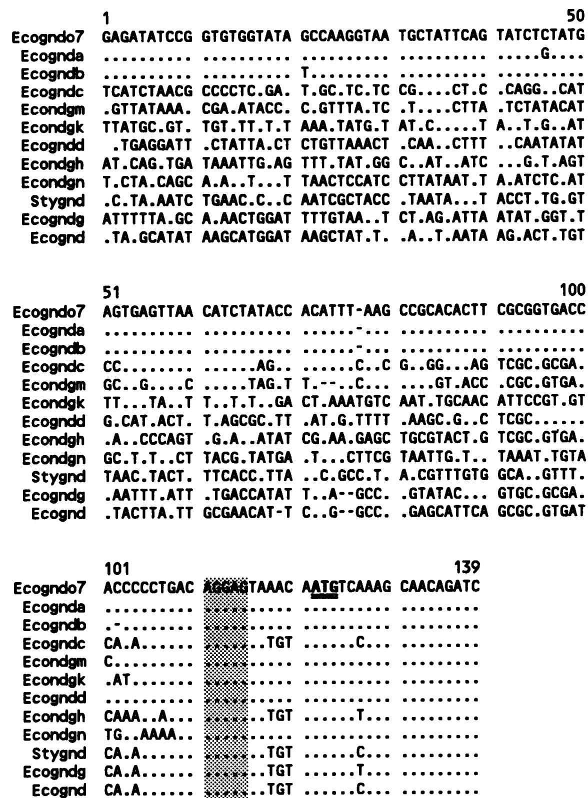

FIG. 2. Alignment of sequences upstream of various *gnd* alleles (4, 12). The alignment was done with the GCG program PILEUP. Dots indicate bases similar to those of the *gnd*$_{EcO7}$. Shaded areas denote the ribosome binding sites. Double underlining indicates the initiation codon of *gnd*$_{EcO7}$. The following *gnd* alleles were examined (GenBank entries are given in parentheses): Ecognda from *E. coli* B/r (M18956); Ecogndb from *E. coli* 740 (M18957); Ecogndc from *E. coli* 567 (M18958); Econdgm from *E. coli* RM224H (M64330); Econdgk from *E. coli* RM45E (M64328); Ecogndd from *E. coli* 558 (M18960); Econdgh from *E. coli* RM191F (M64325); Econdgn from *E. coli* RM2021 (M64331); Stygnd from *S. enterica* LT2 (M18959); Ecogndg from *E. coli* RM39A (M64324); Ecognd from *E. coli* K-12 (K02072).

Despite the high conservation found in the coding regions of *gnd*$_{EcO7}$ and *gnd*$_{EcK12}$, the sequences located upstream of the initiation codon, and including the ribosome binding site, the promoter region, and other regulatory elements involved in the control of the expression of *gnd* (4), displayed only 62% identity. A search of other *gnd* alleles (4, 12) showed that the sequence of *gnd*$_{EcO7}$ upstream of the initiation codon was 99 and 98% identical to the corresponding sequences of the *gnd* alleles found in *E. coli* B/r and *E. coli* F740, respectively (Fig. 2). In contrast, other *E. coli gnd* alleles examined showed lower identity with *gnd*$_{EcO7}$ in this region, ranging from 59 to 87% (Fig. 2). In these cases, the best alignments occurred in the region encompassing the *gnd* ribosome binding site and the sequences downstream (Fig. 2). These results suggest that *gnd*$_{EcO7}$ is closely related to the *gnd* genes in *E. coli* B/r and *E. coli* F740. In addition, the physical map of a 4-kb region upstream of the *gnd* promoter in pGB310B/r (Table 1) which carries the *gnd* allele derived

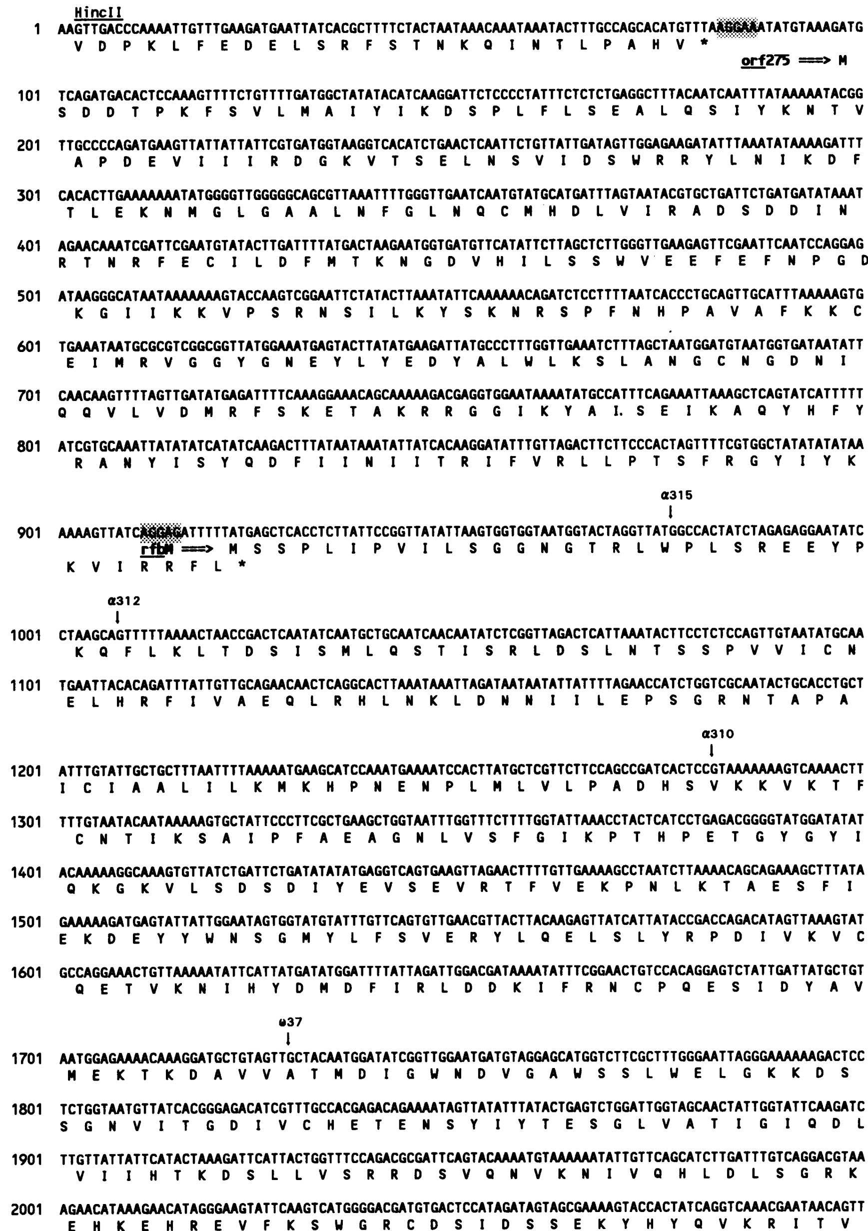

FIG. 3. Nucleotide sequence of the 4,274-bp *Hin*cII-*Kpn*I fragment (Fig. 1) and its deduced amino acid sequence. The DNA sequence is numbered in the 5'-to-3' direction. The bases corresponding to the recognition sites for *Hin*cII and *Kpn*I are indicated. Shaded boxes denote the location of putative ribosome binding sites. ↓ indicates the location of Tn*3*HoHo1 transposon insertions (for the nomenclature of the insertions, refer to Table 1).

from *E. coli* B/r (3) was found to be very similar to the map of the corresponding region in rfb_{EcO7} containing the GDP-mannose biosynthesis genes (see below). Since the O-type of *E. coli* B is not known because of the inability of this strain to express O-specific side-chain LPS, it is tempting to speculate that *E. coli* B could be an O7 *E. coli* carrying one or several mutations affecting the expression of O-specific LPS.

Sequence of the rfb_{EcO7} gene cluster containing the GDP-mannose biosynthesis genes. DNA sequencing was continued upstream of gnd_{EcO7} into the rfb_{EcO7} region, revealing the presence of three ORFs, all transcribed toward *gnd*, coding for 275, 464, and 453 amino acids, respectively, and designated ORF275, ORF464, and ORF453 (Fig. 1b and Fig. 3). No promoter regions were identified upstream of ORF275, where the DNA sequence encodes the carboxy terminus of another ORF (37). Also, no promoter appeared to be present in the 63 bp separating the end of ORF464 from the beginning of ORF453 (Fig. 3). Therefore, it is possible that the three ORFs are the last genes of a larger transcriptional unit.

Putative initiation codons and ribosome binding sites are also indicated in Fig. 3. The sequence AGGA was localized 11 bp upstream of the proposed start of ORF275. ORF464 was preceded by the sequence AGGAG localized 6 bp upstream of its start codon. The sequence AGAGA was localized 9 bp upstream of the start codon of ORF453. All

```
2101  AATCCAAGTGAAAATTATCGTTGCAATTACATCATCACCGTGCGGAACATTGGGGTTGTTGTAATGGGGATTGCTAAACTTACAGTTGCAGAAGAAATAA
      N  P  S  E  N  Y  R  C  N  Y  I  I  T  V  R  N  I  G  V  V  V  M  G  I  A  K  L  T  V  A  E  E  I  K

                                                                                          ω31
                                                                                           |
2201  AAATTTTAAAAGAGAATGAGTCTGTATATATTCCTGCAGGTATTAAGCATAGTTTGAAAATATTGGACAATACCACTTGTGTTAATAGAAGTTTGGACCG
       I  L  K  E  N  E  S  V  Y  I  P  A  G  I  K  H  S  L  K  I  L  D  N  T  T  C  V  N  R  S  L  D  R

      α406
       |
2301  GTTCTTATCTTGCTGATGATGATATCCTTCGATTTGAAGATTGATATGGTAGAGCTTAGAAATAGAGAGTCTTAATAATGCTAACTTGCTTTAAAGCTTA
      F  L  S  C  *  *  *
                                                                          rfbK ==> M  L  T  C  F  K  A  Y

2401  TGATATTCGCGGGAAACTAGGCGAAGAACTGAATGAAGATATCGCCTGGCGCATTGGGCGTGCCTATGGCGAATTTCTCAAACCGAAAACCATTGTGTTA
       D  I  R  G  K  L  G  E  E  L  N  E  D  I  A  W  R  I  G  R  A  Y  G  E  F  L  K  P  K  T  I  V  L

2501  GGCGGTGATGTCCGCCTCACCAGCGAAACCTTAAAACTGGCGCTGGCGAAAGGTTTACAGGATGCGGGCGTCGATGTGCTGGATATCGGTATGTCCGGCA
      G  G  D  V  R  L  T  S  E  T  L  K  L  A  L  A  K  G  L  Q  D  A  G  V  D  V  L  D  I  G  M  S  G  T

2601  CCGAAGAGATCTATTTCGCCACGTTCCATCTCGGCGTGGATGGTGGCATTGAAGTTACCGCCAGCCATAATCCGATGGATTATAACGGCATGAAACTCGT
        E  E  I  Y  F  A  T  F  H  L  G  V  D  G  G  I  E  V  T  A  S  H  N  P  M  D  Y  N  G  M  K  L  V

2701  GCGCGAGGGGGCTCGCCCGATCAGCGGGGATACCGGACTGCGCGATGTCCAGCGTCTGGCAGAAGCCAACGACTTCCCTCCCGTTGATGAAACCAAACGC
       R  E  G  A  R  P  I  S  G  D  T  G  L  R  D  V  Q  R  L  A  E  A  N  D  F  P  P  V  D  E  T  K  R

2801  GGTCGTTATCAGCAAATCAATCTGCGTGACGCTTACGTTGATCACCTGTTCGGTTATATCAACGTCAAAAACCTCACGCCGCTCAAGCTGGTGATTAACT
      G  R  Y  Q  Q  I  N  L  R  D  A  Y  V  D  H  L  F  G  Y  I  N  V  K  N  L  T  P  L  K  L  V  I  N  S

2901  CCGGGAATGGCGCGGCGGGGCCGGTGGTGGACGCCATTGAAGCCCGCTTTAAAGCCCTCGGCGCACCGGTGGAATTAATCAAAGTACACAACACGCCGGA
        G  N  G  A  A  G  P  V  V  D  A  I  E  A  R  F  K  A  L  G  A  P  V  E  L  I  K  V  H  N  T  P  D

3001  CGGCAATTTCCCCAACGGTATTCCTAACCCGTTGCTGCCGGAATGCCGCGACGACACCCGTAATGCGGTCATCAAACACGGCGCGGATATGGGCATTGCC
       G  N  F  P  N  G  I  P  N  P  L  L  P  E  C  R  D  D  T  R  N  A  V  I  K  H  G  A  D  M  G  I  A

3101  TTTGATGGCGATTTTGACCGCTGTTTCCTGTTTGACGAAAAAGGGCAGTTTATCGAGGGCTACTACATTGTCGGCCTGCTGGCAGAAGCATTCCTCGAAA
      F  D  G  D  F  D  R  C  F  L  F  D  E  K  G  Q  F  I  E  G  Y  Y  I  V  G  L  L  A  E  A  F  L  E  K

3201  AAAATCCCGGCGCGAAGATCATCCACGATCCGCGTCTCTCCTGGAACACCGTTGATGTGGTGACCGCCGCGGGCACCCCGGTAATGTCGAAAACCGGACA
        N  P  G  A  K  I  I  H  D  P  R  L  S  W  N  T  V  D  V  V  T  A  A  G  T  P  V  M  S  K  T  G  H

3301  CGCCTTTATTAAAGAACGTATGCGCAAGGAAGACGCCATCTACGGTGGCGAAATGAGCGCCCACCACTATTTCCGTGATTTCGCTTACTGCGACACGGGC
       A  F  I  K  E  R  M  R  K  E  D  A  I  Y  G  G  E  M  S  A  H  H  Y  F  R  D  F  A  Y  C  D  T  G

3401  ATGATCCCGTGGCTGCTGGTGGCCGAACTGGTGTGTCTGAAAGGAAAAACGCTGGGCGAACTGGTGCGCGACCGGATGGCGGCGTTTCCGGCAAGCGGTG
      M  I  P  W  L  L  V  A  E  L  V  C  L  K  G  K  T  L  G  E  L  V  R  D  R  M  A  A  F  P  A  S  G  E

3501  AGATCAACAGCAAACTGGCGCACCCCGTTGAGGCGATTAATCGCGTCGAACAGCATTTTAGCCGCGACGCTGGCGGTGGATCGCACCGATGGCATCAGCA
        I  N  S  K  L  A  H  P  V  E  A  I  N  R  V  E  Q  H  F  S  R  D  A  G  G  G  S  H  R  W  H  Q  H

3601  TGACCTTTGCCGACTGGCGGCTTTAACGTGCGCCTCCTCCAACACCGAACCGGTGGTGCGGTTGAATGTGGAATCGCGCGGTGATGTACCGCTGATGGAA
       D  L  C  R  L  A  A  L  T  C  A  S  S  N  T  E  P  V  V  R  L  N  V  E  S  R  G  D  V  P  L  M  E

3701  GAAAAGACAAAACTTATCCTTGAGTTACTGAACAAGTAATTCAGTAATTTCATATAAATGGTTTTTAAAAAACGGAAAAGATGAGATATCCGGTGTGGTA
      E  K  T  K  L  I  L  E  L  L  N  K  *

3801  TAGCCAAGGTAATGCTATTCAGTATCTCTATGAGTGAGTTAACATCTATACCACATTTAAGCCGCACACTTCGCGGTGACCACCCCCTGACAGGAGTAAA

3901  CAATGTCAAAGCAACAGATCGGCGTCGTCGGTATGGCAGTGATGGGGCGCAACCTCGCGCTCAACATCGAAAGTCGTGGTTATACCGTCTCTATTTTCAA
gnd' ==> M  S  K  Q  Q  I  G  V  V  G  M  A  V  M  G  R  N  L  A  L  N  I  E  S  R  G  Y  T  V  S  I  F  N

4001  CCGTTCCCGTGAAAAGCAGGAAGAAGTGATTGCCGAAAATCCAGGCAAGAAACTGGTTCCTTACTATACGGTGAAAGAGTTTGTTGAATCTCTGGAAACG
       R  S  R  E  K  Q  E  E  V  I  A  E  N  P  G  K  K  L  V  P  Y  Y  T  V  K  E  F  V  E  S  L  E  T

4101  CCTCGTCGCATCCTGTTAATGGTGAAAGCAGGTGCAGGCACGGATGCTGCTATTGATTCCCTCAAGCCATACCTCGATAAAGGTGACATCATCATTGATG
      P  R  R  I  L  L  M  V  K  A  G  A  G  T  D  A  A  I  D  S  L  K  P  Y  L  D  K  G  D  I  I  I  D  G

                                                                          KpnI
4201  GTGGTAATACCTTCTTCCAGGACACCATTCGTCGTAACCGTGAGCTTTCTGCCGAAGGCTTTAATTTTTATCGGTACC
       G  N  T  F  F  Q  D  T  I  R  R  N  R  E  L  S  A  E  G  F  N  F  Y  R  Y
```

FIG. 3—*Continued.*

three of these sequences resemble ribosome binding sites with respect to nucleotide composition and distance upstream of the start codons (15). The start codon of ORF464 overlaps with the stop codon of ORF275, whereas three stop codons in tandem were found at the end of ORF453. No ORF of significant length was identified in the opposite strand of the 4,271-bp *Hin*cII-*Kpn*I fragment.

The coding sequences of ORF275 and ORF464 have G+C contents (31.8 and 34.8%, respectively) lower than that typical of *E. coli* (55%), whereas the G+C content of ORF453 is 55.4%. Similar observations were reported for the *rfb* genes of *S. enterica* (6, 19, 58) and the *rfa* genes of both *E. coli* K-12 and *S. enterica* serovar *typhimurium* (49). There is a high preference for unusual codons in ORF275 and ORF464. The low G+C content and the atypical codon usage argue that at least part of the rfb_{EcO7} region is the result of a relatively recent acquisition of genes from other microorganisms of low G+C content, as has been suggested for *S. enterica* serogroups A, B, C2, and E1 (6, 19, 58).

The putative polypeptide encoded by ORF275 has a predicted molecular mass of 31.8 kDa and a theoretical pI of 9.8. No striking homology with any other gene was found by searching the GenBank, EMBL, and SwissProt data bases. The hydropathicity profile shows no significant hydrophobic domains indicative of transmembrane peptide structures. However, two hydrophobic regions are evident in both the amino and carboxy termini which could be involved in membrane association (data not shown). The high pI of this protein suggests that it could correspond to a peripheral membrane protein that can be associated with the membrane by virtue of its positive charge. A gene encoding a predicted 35.5-kDa polypeptide with a pI of 9.42 has been identified in a similar position in *S. enterica* LT2 and was thought to correspond to a nucleotide sugar transferase (ORF16.5 [19]).

TABLE 2. Comparisons of $RfbM_{EcO7}$ with other bacterial GDP-mannose pyrophosphorylases[a]

Protein	Size (kDa)	No. of amino acids	Species	Involvement	% Identity (no. of amino acids)[b]
$RfbM_{EcO7}$	52.5	464	*E. coli*	O unit	100 (464)
CpsB	53.3	480	*S. enterica* LT2	Colanic acid	53 (452)
$RfbM_{STY}$	52.9	473	*S. enterica* M67	O unit	50 (459)
$RfbM_{STY}$	54	479	*S. enterica* LT2	O unit	48 (457)
XanB[c]	50	466	*X. campestris*	Xanthan	47 (458)
AlgA[c]	52.9	480	*P. aeruginosa*	Alginate	42 (456)

[a] Comparisons were done with BESTFIT (gap weight of 5.00 and length weight of 0.5), using the entire $RfbM_{EcO7}$ amino acid sequence as a probe.
[b] Indicates the number of amino acids of the regions overlapping with $RfbM_{EcO7}$ in each protein.
[c] Bifunctional enzyme with both PMI and GDP-mannose pyrophosphorylase activities (21, 50).

Alignment of the predicted amino acid sequence of this gene with that of ORF275 by using the program BESTFIT revealed 45% similarity and only 17.5% identity, with five gaps. ORF275 may encode the GDP-mannose transferase or another nucleotide sugar transferase needed for the formation of the O7 repeating unit.

ORF464 encodes a polypeptide with predicted molecular mass of 52.5 kDa and pI of 6.9. The hydropath profile suggested a hydrophilic protein (data not shown). The predicted polypeptide displayed 49 and 54% amino acid identity with RfbM and CpsB, respectively, of *S. enterica* group B (19, 52), 42.6% identity with AlgA from *Pseudomonas aeruginosa* (50), and 46.6% identity with XanB from *Xanthomonas campestris* (21), all GDP-mannose pyrophosphorylases (Table 2). ORF464 was designated *rfbM*$_{EcO7}$ and presumably encodes the GDP-mannose pyrophosphorylase activity required for the biosynthesis of the GDP-mannose precursor of the O7-specific side chain.

ORF453 encodes a polypeptide with predicted molecular mass of 49.7 kDa and pI of 6.1. It exhibits 83.5% amino acid identity with CpsG of *S. enterica* (52) and 55% identity with XanA of *X. campestris* (21), both having mannose-6-phosphate:phosphomannomutase activities involved in the biosynthesis of colanic acid and xanthan, respectively (Table 3). In contrast, ORF453 lacks any significant identity with RfbK of *S. enterica* LT2 (19), the phosphomannomutase encoded by the *rfb* region in this strain (Table 3). ORF453 was designated *rfbK*$_{EcO7}$ and presumably encodes a phosphomannomutase, the other component of the GDP-mannose biosynthesis pathway. $RfbK_{EcO7}$, CpsG, and XanA proteins contain a conserved motif for catalase-1 (2) and also display high homology with 12 amino acids located surrounding the active center of the rabbit muscle phosphoglucomutase (Fig. 4) (21, 45). These sequences also have the motif Gly-x-Gly-x-x-Gly, found in protein kinases (16, 21) (Fig. 4). None of these features are evident in the *Salmonella* RfbK proteins. Thus, these proteins seem to belong to a family of enzymes different from the classical phosphomannomutase identified in *S. enterica* groups A, B, and D (6, 19, 58).

Functional studies. Experiments were done to confirm that *rfbM*$_{EcO7}$ and *rfbK*$_{EcO7}$ expressed their predicted enzymatic activities, using cell extracts prepared from *E. coli* CLM4 carrying appropriate plasmids. This strain is a *recA* derivative of strain SØ874 which contains a deletion eliminating the chromosomal *rfb* region and therefore lacks GDP-mannose pyrophosphorylase and phosphomannomutase activities. Plasmids pCM104 and pCM108 (Table 1 and Fig. 1) contain *rfbM*$_{EcO7}$ and *rfbK*$_{EcO7}$, respectively, subcloned into the expression vector pEX1 (Table 1), where transcription of these genes is under the control of the p*tac* promoter. Cell extracts prepared from CLM4(pCM104) and CLM4 (pCM108) exhibited 3.77×10^4 mU of GDP-mannose pyrophosphorylase and 4.6×10^2 mU of phosphomannomutase activities, respectively, confirming that these plasmids carry the corresponding *rfbM*$_{EcO7}$ and *rfbK*$_{EcO7}$ genes. The enzymatic activities were detected in the supernatant fraction, confirming that $RfbM_{EcO7}$ and $RfbK_{EcO7}$ are soluble proteins as suggested by the calculated hydropath profiles.

SDS-polyacrylamide gel electrophoresis (PAGE) of cell lysates prepared from IPTG-induced cultures of CLM4 containing pCM104 and then either Coomassie blue or silver staining revealed no specific polypeptides (Fig. 5a, lanes C and D). In contrast, lysates of CLM4 carrying pCM108 had two unique polypeptides of approximately 52 and 42.5 kDa (Fig. 5a, lane F), presumably corresponding to $RfbK_{EcO7}$ and the truncated Gnd_{EcO7}, respectively. The apparent molecular mass of $RfbK_{EcO7}$ is in good agreement with the molecular mass of the polypeptide predicted from its deduced amino acid sequence. The 116-kDa LacZ expressed by the control plasmid pEXZ was also detected (Fig. 5a, lane B) and served as a control for the efficiency of IPTG induction. These results demonstrate that the expression of the *rfbM*$_{EcO7}$ gene in pCM104 is less efficient relative to that of the *rfbK*$_{EcO7}$ and *lacZ* genes. This could be due to the high number of rare codons in *rfbM*$_{EcO7}$, which may result in

TABLE 3. Comparisons with bacterial phosphomannomutases[a]

Protein	Size (kDa)	No. of amino acids	Species	Involvement	% Identity (no. of amino acids)[b]
$RfbK_{EcO7}$	49.7	453	*E. coli*	O unit	100 (453)
CpsG	50	456	*S. enterica* LT2	Colanic acid	84 (453)
XanA[c]	48.9	448	*X. campestris*	Xanthan	55 (451)
$RfbK_{STY}$	52	477	*S. enterica* LT2	O unit	19 (449)

[a] Comparisons were done with BESTFIT (gap weight of 5.00 and length weight of 0.5), using the entire $RfbK_{EcO7}$ amino acid sequence as a probe.
[b] Indicates the number of amino acids of the regions overlapping with $RfbK_{EcO7}$ in each protein.
[c] Bifunctional enzyme with both phosphoglucose mutase and phosphomannomutase activities (21).

```
         1                                                       50
O7rfbk   --MLTCFKAY DIRGKLGEEL NEDIAWRIGR AYGEFLKPKT IVLGGDVRLT
Stycpsg  MTK....... ....R..... .......... ....Y..... V.........
   Xana  .MT.PA.... ....RVPD.. ...L.R...V .LAAQ.DQGP V...H....A

         51                                                     100
O7rfbk   SETLKLALAK GLQDAGVDVL DIGMSGTEEI YFATFHLGVD GGIEVTASHN
Stycpsg  ..A.NV.... .......... .......... .......... ..........
   Xana  .PA.QE..SA ..RAS.RE.I ...LC....V ..Q.D..KAA ..VM......
                                                 AI GGIILTASHN

         101                                                    150
O7rfbk   PMDYNGMKLV REGARPISGD TGLRDVQRLA EANDFPPVDE TKRGRYQQIN
Stycpsg  .......... .......... .......... ..G.....N. AA..S.R..S
   Xana  .......... ..Q.....S. ...FAI-.DT V.A.TAAAG. PTAAEHSRTD
         PGGPNG

         151                                                    200
O7rfbk   LRDAYVDHLF GYINVKNLTP LKLVINSGNG AAGPVVDAIE ARFKALGAPV
Stycpsg  .....I...L ...S.N.... ....F.A... ...I.... ..L.......
   Xana  -KT..LE..L S.VDRST.K. ....V.A... G..LIV.LLA PHL-----PF
                                     GxG xxG

         201                                                    250
O7rfbk   ELIKVHNTPD GNFPNGIPNP LLPECRDDTR NAVIKHGADM GIAFDGDFDR
Stycpsg  .F..I..... .T........ .......... K...E..... ..........
   Xana  .FVRVFHE.. .......... ..Q.N..ATA K..KE....F ...W......

         251                                                    300
O7rfbk   CFLFDEKGQF IEGYYIVGLL AEAFLEKNPG AKIIHDPRLS WNTVDVVTAA
Stycpsg  .......... .......... .......H.. .........T ...EA.....
   Xana  ..F..HT.R. .....L.... .Q.I.A.Q.. G.VV.....T ...VEM.ED.

         301                                                    350
O7rfbk   -GTPVMSKTG HAFIKERMRK EDAIYGGEMS AHHYFRDFAY CDTGMIPWLL
Stycpsg  G......... .........T .......... .......... ..S.......
   Xana  G.IP.LC.S. ......K..S .N.V...... .....E.... A.S.......
                                              RxFAY xDA
                                                S     S
                                                T     T

         351                                                    400
O7rfbk   VAELVCLKGK TLGELVRDRM AAFPASGEIN SKLAHPVEAI NRVEQHFSRD
Stycpsg  ........RQ S......... .......... .R..E.AA.. A...A..AEE
   Xana  I....SQSGR S.AD..EA.. QK..C..... F.VDDAKA.V A..MA.YGDQ

         401                                                    450
O7rfbk   AGGGSHRWHQ HDLCRLAA-- -LTCASSNTE PVVRLNVESR GDVPLMEEKT
Stycpsg  .QAVD---RT DG.SMSF.DW RFNLR..... .......... ..I....AR.
   Xana  SPELDY---T DGISADFGQW RFNLR..... .LL.....T. ..AA.L.TR.

         451    460
O7rfbk   KLILELLNK*
Stycpsg  RTL.A...Q*
   Xana  QE.SN..RG*
```

FIG. 4. Alignment of $RfbK_{EcO7}$ (O7rfbk), CpsG (Stycpsg), and XanA (Xana) amino acid sequences. Alignment was done with the GCG program PILEUP. Dots indicate amino acids identical to those of rfb_{EcO7}. Shaded areas in positions 88 to 106, 178 to 183, and 336 to 343 indicate the homologies with the region containing the active center of rabbit muscle phosphoglucomutase (45), a motif found in protein kinases (16), and the catalase-1 motif, respectively. The amino acid sequences of the various motifs are shown beneath the aligned sequences.

poor translation of its mRNA. The $RfbM_{EcO7}$ polypeptide encoded by pCM104 was seen after in vivo labelling of maxicells. Figure 5b, lane C, shows a polypeptide of about 52 kDa as predicted for $RfbM_{EcO7}$ that is not present in control lanes containing extracts from cells with no plasmid (lane A) and cells with pEXZ (lane B). To see the $RfbM_{EcO7}$ polypeptide, we exposed the autoradiograph for a long time, resulting in an overexposure of other polypeptides. Protein extracts from cells containing pEXZ revealed a polypeptide with a mass of approximately 116 kDa that corresponds to LacZ, and bands of lower molecular weight which presumably are due to partially degraded LacZ protein or partially translated products (Fig. 6b, lane B). None of these bands

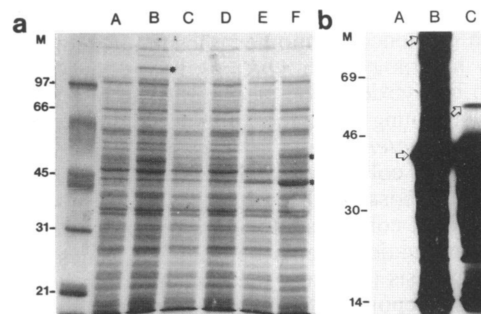

FIG. 5. Expression of $rfbM_{EcO7}$ and $rfbK_{EcO7}$ genes. (a) Cell lysates were prepared as described in the text from cells grown in the presence (lanes B, D, and F) or in the absence (lanes A, C, and E) of IPTG, and samples were examined by SDS-PAGE and then staining with Coomassie blue. Cells contained pEXZ (lanes A and B), pCM104 (lanes C and D), and pCM108 (lanes E and F). Asterisks indicate β-galactosidase subunit (lane B), $RfbK_{EcO7}$ (lane F, top), and Gnd (lane F, bottom), which have apparent molecular masses of 116, 52, and 42.5 kDa, respectively. M, molecular size standards expressed in kilodaltons: phosphorylase *b*, 97 kDa; albumin, 66 kDa; ovalbumin, 45 kDa; carbonic anydrase, 31 kDa; trypsin inhibitor, 21 kDa. (b) Autoradiograph of polypeptides expressed in a maxicell system. Cells were labelled in vivo with $[^{35}S]$methionine and processed as described previously (5). Lane A, strain HB101; lane B, strain HB101(pEXZ); lane C, HB101(pCM104). Arrows indicate β-galactosidase subunit (lane B, top), RfbM polypeptide (lane C), and $LacI^q$ (present in lanes B and C), which have apparent molecular masses of 116, 52, and 40 kDa, respectively. M, ^{14}C-labeled molecular mass markers: bovine serum albumin, 69 kDa; ovalbumin, 46 kDa; carbonic anhydrase, 30 kDa; lysozyme, 14 kDa.

were present in cell extracts carrying pCM104. The 40-kDa polypeptide present in lanes B and C is probably $LacI^q$. The maxicell experiment showed that it is possible to visualize the product of the $rfbM_{EcO7}$ gene and confirmed that its expression is rather poor relative to that of *lacZ*.

Complementation of transposon mutants. We reported Tn*3*HoHo1 insertions in the rfb_{EcO7} region of cosmids pJHCV31 and pJHCV32 (36). Insertions 315, 312, 310, 37, 31, and 406 mapped within the sequenced *Hin*cII-*Kpn*I fragment now identified as containing the GDP-mannose biosynthesis genes and part of gnd_{EcO7} (35), and all of them, except for 406, gave an O7-negative phenotype. The location of these insertions was determined by sequencing (Fig. 3), and revised names are given in Table 1.

rfbM::Tn*3*HoHo1-α406 causes the deletion of the three terminal carboxy amino acids of RfbM (Fig. 3). Since this insertion gives an O7-positive phenotype, we conclude that they are not essential for enzymatic activity. The O7-deficient phenotype of the remaining insertion mutants was restored by complementation in *trans* with the rfb_{EcO7} gene on pCM104, to give a full-length O7-specific LPS as detected by immunoblotting (data not shown). Since $rfbK_{EcO7}$ is downstream of $rfbM_{EcO7}$ and complementation was observed with just pCM104, which does not carry a complete copy of $rfbK_{EcO7}$, it is possible that either the Tn*3*HoHo1 insertions are not creating a polar effect on the transcription of $rfbK_{EcO7}$ or there is another gene encoding a phosphomannomutase activity elsewhere in the $rfbK_{EcO7}$ region. Unfortunately, we cannot completely rule out this latter possibility because Tn*3*HoHo1 insertions into $rfbK_{EcO7}$ were not isolated, but a hybridization with a fragment internal to the $rfbK_{EcO7}$ gene did not detect any homology with either pJHCV32 or pJHCV31 other than in the region correspond-

ing *rfbK* (data not shown). This experiment does not preclude the existence of another phosphomannomutase gene with low homology with $rfbK_{EcO7}$ below the sensitivity of detection by hybridization. The similarity of $RfbK_{EcO7}$ and CpsG is intriguing since the latter is found outside the *rfb* region of *S. enterica* and is presumably involved in the biosynthesis of colanic acid. Since VW187 can form colanic acid (20), it is possible that $rfbK_{EcO7}$ is also involved in the biosynthesis of colanic acid and perhaps arose from one or more recombination events between the genes for colanic acid biosynthesis (*cps*) and the *rfb* genes. Recently, Lee et al. (24) have reported that in *S. enterica* group C1 a gene similar to *cpsG* appears to be located near *gnd* in a position similar to that of $rfbK_{EcO7}$. This finding supports the idea that $rfbK_{EcO7}$ may have indeed resulted from a recombination event between *cps* and *rfb* genes.

Expression of O7-LPS requires the product of the *manA* gene. The biosynthesis pathway for GDP-mannose in *Salmonella* species and *E. coli* requires the isomerization of fructose-6-phosphate to mannose-6-phosphate mediated by the PMI encoded by the *manA* gene which maps outside the *rfb* region (1, 32). Mannose-6-phosphate is converted into mannose-1-phosphate by RfbK and subsequently into GDP-mannose by RfbM (41). Mutations in *manA* cause not only the inability to form O side chains that have mannose in the O repeating unit but also the inability of cells to grow with mannose as the only carbon source (34). In contrast, in *P. aeruginosa* and *X. campestris*, the utilization of mannose does not seem to require a PMI activity, but this activity is necessary for the synthesis of capsular polysaccharides (21). In these strains, a bifunctional enzyme encoded by *algA* and *xanB* displays both PMI and GDP-mannose pyrophosphorylase activities (21, 50). The *manA* strain GMS343 carrying pJHCV32 did not form O7-specific LPS in the absence of mannose (data not shown), indicating that, like RfbM of *S. enterica* but unlike AlgA and XanB, $RfbM_{EcO7}$ encoded by pJHCV32 lacks PMI activity.

Comparisons with GDP-mannose biosynthesis genes of *Salmonella* species. In the past few years, *rfb* genes of several serovars of *S. enterica* have been characterized by DNA sequencing (6, 19, 58). In all these cases, *rfbM* is located immediately upstream from *rfbK* and in the proximity of the *gnd* gene. The finding that $rfbM_{EcO7}$ and $rfbK_{EcO7}$ are present in the same relative order and also in the vicinity of the *gnd* gene is intriguing, especially considering the relatively low level of homology between the O7 and *Salmonella rfb* gene clusters (36). The only gene of the rfb_{EcO7} cluster with strong homology to the *rfb* genes of *S. enterica* is $rfbM_{EcO7}$, and no significant similarities with other *Salmonella rfb* genes have been detected in the VW187 DNA sequence spanning about 9 kb of the rfb_{EcO7} region (35). Furthermore, a gene, *galP* (UDP-galactosyl transferase), located downstream from *rfbK* in the *rfb* clusters of *S. enterica* groups A, B, C2, and D is not found in the same location in VW187, although the O7 repeating unit contains galactose, necessitating a UDP-galactosyl transferase function for its biosynthesis. The *rfbM* and *rfbK* coding regions overlap in *S. enterica* groups A, B, C2, D, and E1, whereas the $rfbM_{EcO7}$ and $rfbK_{EcO7}$ coding regions are separated by 63 bases (Fig. 3). This would be expected if $rfbK_{EcO7}$ resulted from a recombination event involving a gene(s) from the *cps* cluster, as discussed above.

The similar general organization of the *rfb* regions in the proximity to the *gnd* gene suggests a common origin, perhaps from an ancestral *rfb* cluster. A succession of further recombination events in the presence of different types of selective pressures presumably have determined a separate evolution of the *rfb* gene clusters in different bacterial species and even in different bacterial clones within members of the same species, increasing the variation of gene structure but still maintaining a relatively similar organization in terms of functional conservation. The analysis of more *rfb* gene clusters from *E. coli* and other enteric bacteria will permit a better understanding of the molecular nature of the genetic variations observed in these clusters.

ACKNOWLEDGMENTS

This work was supported by grant MA10206 from the Medical Research Council.

We thank colleagues mentioned or referenced in Table 1 for providing strains and plasmids. We also extend our gratitude to Lydia Dafoe and Diane Moyles for technical assistance and to Joanna S. Brooke for performing the experiment with the *manA* mutant and critically revising the manuscript. We also thank the reviewers for helpful editorial comments.

REFERENCES

1. **Bachmann, B. J.** 1990. Linkage map of *Escherichia coli* K-12, edition 8. Microbiol. Rev. **54:**130–197.
2. **Bairoch, A.** 1991. Prosite: a dictionary of sites and patterns in proteins. Nucleic Acids Res. Suppl. **19:**2241–2245.
3. **Barcak, G. J., and R. E. Wolf, Jr.** 1988. Growth-rate-dependent expression and cloning of *gnd* alleles from natural isolates of *Escherichia coli*. J. Bacteriol. **170:**365–371.
4. **Barcak, G. J., and R. E. Wolf, Jr.** 1988. Comparative nucleotide sequence analysis of growth-rate-regulated *gnd* alleles from natural isolates of *Escherichia coli* and from *Salmonella typhimurium* LT-2. J. Bacteriol. **170:**372–379.
5. **Boulnois, G. J., and K. N. Timmis.** 1984. Synthesis of plasmid-encoded polypeptides in maxicells, p. 204–211. *In* A. Puhler and K. N. Timmis (ed.), Advanced molecular genetics. Springer-Verlag KG, Berlin.
6. **Brown, P. K., L. K. Romana, and P. R. Reeves.** 1992. Molecular analysis of the *rfb* cluster of *Salmonella* serovar *muenchen* (strain M67): the genetic basis of the polymorphism between groups C2 and B. Mol. Microbiol. **6:**1385–1394.
7. **Chang, A. C. Y., and S. N. Cohen.** 1978. Construction and characterization of amplifiable DNA cloning vehicles derived from P15A cryptic plasmid. J. Bacteriol. **134:**1141–1156.
8. **Cohen, S. N., A. C. Y. Chang, and L. Hsu.** 1972. Nonchromosomal antibiotic resistance in bacteria: genetic transformation of *Escherichia coli* by R-factor DNA. Proc. Natl. Acad. Sci. USA **69:**2110–2114.
9. **Csonka, L. A., and A. J. Clarke.** 1979. Deletions generated by the transposon Tn*10* in the *srl rec*A region of the *Escherichia coli* K-12 chromosome. Genetics **93:**321–343.
10. **Curtiss, R.** 1981. Gene transfer, p. 243–265. *In* P. Gerhardt, R. G. E. Murray, R. N. Costilow, E. W. Nester, W. A. Wood, N. R. Krieg, and G. B. Phillips (ed.), Manual of methods for general bacteriology. American Society for Microbiology, Washington, D.C.
11. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:**387–395.
12. **Dykhuizen, D. E., and L. Green.** 1991. Recombination in *Escherichia coli* and the definition of biological species. J. Bacteriol. **173:**7257–7268.
13. **Feinberg, A. P., and B. Vogelstein.** 1983. A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132:**6–13.
14. **Fraenkel, D. G.** 1987. Glycolysis, pentosephosphate pathway, and Entner-Doudoroff pathway, p. 142–150. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
15. **Gold, L., and G. Stormo.** 1987. Translational initiation, p.

1302–1307. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
16. **Hanks, S. K., A. M. Quinn, and T. Hunter.** 1988. The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. Nature (London) **241**:42–52.
17. **Henikoff, S.** 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene **28**:351–359.
18. **Jann, K., and B. Jann.** 1984. Structure and biosynthesis of O-antigens, p. 138–186. *In* E. T. Rietschel (ed.), Handbook of endotoxin, vol. 1. Chemistry of endotoxin. Elsevier, Amsterdam.
19. **Jiang, X.-M., B. Neal, F. Santiago, S. J. Lee, L. K. Romana, and P. R. Reeves.** 1991. Structure and sequence of the *rfb* (O antigen) gene cluster of *Salmonella* serovar *typhimurium* (strain LT2). Mol. Microbiol. **5**:695–713.
20. **Keenleyside W. J., P. Jayaratne, P. R. MacLachlan, and C. Whitfield.** 1992. The *rscA* of *Escherichia coli* O9:K30:H12 is involved in the expression of the serotype-specific group I K (capsular) antigen. J. Bacteriol. **174**:8–16.
21. **Köplin, R., W. Arnold, B. Hötte, R. Simon, G. Wang, and A. Pühler.** 1992. Genetics of xanthan production in *Xanthomonas campestris*: the *xanA* and *xanB* genes are involved in UDP-glucose and GDP-mannose biosynthesis. J. Bacteriol. **174**:191–199.
22. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.
23. **Laemmeli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**:680–685.
24. **Lee, S. J., L. K. Romana, and P. R. Reeves.** 1992. Cloning and structure of group C1 O antigen (*rfb* gene cluster) from *Salmonella enterica* serovar *montevideo*. J. Gen. Microbiol. **138**:305–312.
25. **Linn, T., and R. St. Pierre.** 1990. Improved vector system for constructing fusions that ensures independent translation of *lacZ*. J. Bacteriol. **172**:1077–1084.
26. **Lipman, D. J., and W. R. Pearson.** 1985. Rapid and sensitive protein similarity searches. Science **227**:1435–1441.
27. **Liu, D., N. K. Verma, L. K. Romana, and P. R. Reeves.** 1991. Relationships among the *rfb* regions of *Salmonella* serovars A, B, and D. J. Bacteriol. **173**:4814–4819.
28. **Lowry, O. H., N. J. Rosebrough, A. L. Farr, and R. J. Randall.** 1951. Protein measurement with the Folin phenol reagent. J. Biol. Chem. **193**:265–275.
29. **Lüderitz, O., O. Westphal, A. M. Staub, and H. Nikaido.** 1971. Isolation and chemical and immunological characterization of bacterial lipopolysaccharides, p. 145–233. *In* G. Weinbaum, S. Kadis, and S. J. Ajl (ed.), Microbial toxins, vol. IV. Academic Press, Inc., London.
30. **Lugtenberg, B., and L. Van Alphen.** 1983. Molecular architecture and functioning of the outer membrane of *Escherichia coli* and other gram negative bacteria. Biochem. Biophys. Acta **737**:51–115.
31. **L'vov, V. L., A. S. Shashkov, B. A. Dimitriev, N. K. Kochtkov, B. Jann, and K. Jann.** 1984. Structural studies of the O-specific side chain of the lipopolysaccharide from *Escherichia coli* O:7. Carbohydr. Res. **126**:249–259.
32. **Mäkelä, P. H., and B. A. D. Stocker.** 1984. Genetics of lipopolysaccharide, p. 59–137. *In* E. T. Rietschel (ed.), The chemistry of endotoxin. Elsevier Biomedical Press, Amsterdam.
33. **Maloy, S. R., and W. D. Nunn.** 1981. Selection for loss of tetracycline resistance by *Escherichia coli*. J. Bacteriol. **145**:1110–1112.
34. **Markovits, A., R. J. Sydiskis, and M. M. Liebermann.** 1967. Genetic and biochemical studies on mannose-negative mutants that are deficient in phosphomannose isomerase in *Escherichia coli* K-12. J. Bacteriol. **94**:1492–1496.
35. **Marolda, C. L., and M. A. Valvano.** Unpublished data.
36. **Marolda, C. L., J. Welsh, L. Dafoe, and M. A. Valvano.** 1990. Genetic analysis of the O7-polysaccharide biosynthesis region from the *Escherichia coli* O7-K1 strain VW187. J. Bacteriol. **172**:3590–3599.
37. **Munch-Petersen, A.** 1962. GDPM pyrophosphorylase. Methods Enzymol. **5**:171–174.
38. **Nasoff, M. S., and R. E. Wolf, Jr.** 1980. Molecular cloning, correlation of genetic and restriction maps, and determination of the direction of transcription of *gnd* of *Escherichia coli*. J. Bacteriol. **143**:731–741.
39. **Neuhard, J., and E. Thomassen.** 1976. Altered deoxyribonucleotide pools in P2 eductants of *Escherichia coli* due to deletion of the *dcd* gene. J. Bacteriol. **126**:999–1001.
40. **Nikaido, H., K. Nikaido, and P. H. Mäkela.** 1966. Genetic determination of enzymes synthesizing O-specific sugars of *Salmonella* lipopolysaccharide. J. Bacteriol. **91**:1126–1135.
41. **Novel, G., and M. Novel.** 1973. Mutants d'*Escherichia coli* K-12 affectés pour leur croissance sur méthyl-β-D-glucuronide: localosation du gène de structure de la β-D-glucuronidase (*uid*). Mol. Gen. Genet. **120**:319–335.
42. **Ørskov, I., F. Ørskov, B. Jann, and K. Jann.** 1977. Serology, chemistry, and genetics of O and K antigens of *Escherichia coli*. Bacteriol. Rev. **41**:667–710.
43. **Passador, L., and T. Linn.** 1989. Autogenous regulation of the RNA polymerase β subunit of *Escherichia coli* occurs at the translational level in vivo. J. Bacteriol. **171**:6234–6242.
44. **Peyru, G., and D. G. Fraenkel.** 1968. Genetic mapping of loci for glucose-6-phosphate dehydrogenase, gluconate-6-phosphate dehydrogenase, and gluconate-6-phosphate dehydrase in *Escherichia coli*. J. Bacteriol. **95**:1272–1278.
45. **Ray, W. J., Jr., M. A. Hermodson, J. M. Puvathingal, and W. C. Mahoney.** 1983. The complete amino acid sequence of rabbit muscle phosphoglucomutase. J. Biol. Chem. **258**:9166–9174.
46. **Sá-Correia, I., A. Darzins, S. Wang, A. Berry, and A. M. Chakrabarty.** 1987. Alginate biosynthesis enzymes in mucoid and nonmucoid *Pseudomonas aeruginosa*: overproduction of phosphomannose isomerase, phosphomannomutase, and GDP-mannose pyrophosphorylase by overexpression of the phosphomannose isomerase (*pmi*) gene. J. Bacteriol. **169**:3224–3231.
47. **Sanderson, K. E., and J. R. Roth.** 1983. Linkage map of *Salmonella typhimurium*, edition VI. Microbiol. Rev. **47**:410–453.
48. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
49. **Schnaitman, C. A., C. T. Parker, J. D. Klena, E. L. Pradel, N. B. Pearson, K. E. Sanderson, and P. R. Maclachlan.** 1991. Physical maps of the *rfa* loci of *Escherichia coli* K-12 and *Salmonella typhimurium*. J. Bacteriol. **173**:7410–7411.
50. **Shinabarger, D., A. Berry, T. B. May, R. Rothmel, A. Fialho, and A. M. Chakrabarty.** 1991. Purification and characterization of phosphomannose isomerase-guanosine diphospho-D-mannose pyrophosphorylase. J. Biol. Chem. **266**:2080–2088.
51. **Stachel, S. E., G. An, C. Flores, and E. W. Nester.** 1985. A Tn*3* *lacZ* transposon for the random generation of β-galactosidase gene fusions: application to the analysis of gene expression in *Agrobacterium*. EMBO J. **4**:891–898.
52. **Stevenson, G., S. J. Lee, L. K. Romana, and P. R. Reeves.** 1991. The *cps* gene cluster of *Salmonella* LT2 includes a second mannose pathway: sequence of two genes and relationship to genes in the *rfb* gene cluster. Mol. Gen. Genet. **227**:173–180.
53. **Valvano, M. A.** 1992. Pathogenicity and molecular genetics of O-specific side-chain lipopolysaccharides of *Escherichia coli*. Can. J. Microbiol. **38**:711–719.
54. **Valvano, M. A., and J. H. Crosa.** 1984. Aerobactin iron transport genes commonly encoded by certain ColV plasmids occur in the chromosome of a human invasive strain of *Escherichia coli* K1. Infect. Immun. **46**:159–167.
55. **Valvano, M. A., and J. H. Crosa.** 1989. Molecular cloning and expression in *Escherichia coli* K-12 of chromosomal genes determining the O7 lipopolysaccharide antigen of a human invasive strain of *E. coli* O7:K1. Infect. Immun. **57**:937–943.
56. **Valvano, M. A., and C. L. Marolda.** 1991. Relatedness of

O-specific lipopolysaccharide side chain genes from strains of *Shigella boydii* type 12 belonging to two clonal groups and from *Escherichia coli* O7:K1. Infect. Immun. **59**:3917–3923.
57. **Valvano, M. A., R. P. Silver, and J. H. Crosa.** 1986. Occurrence of chromosome- or plasmid-mediated aerobactin iron transport systems and hemolysin production among clonal groups of human invasive strains of *Escherichia coli* K1. Infect. Immun. **52**:192–199.
58. **Wang, L., L. W. Romana, and P. R. Reeves.** 1992. Molecular analysis of a *Salmonella enterica* group E1 *rfb* gene cluster: O antigen and the genetic basis of the major polymorphism. Genetics **130**:429–443.