JOURNAL OF BACTERIOLOGY, Aug. 2005, p. 5259–5266 Vol. 187, No. 15
0021-9193/05/$08.00+0 doi:10.1128/JB.187.15.5259–5266.2005

Identification of an *Escherichia coli* Operon Required for Formation of the O-Antigen Capsule

Adi Peleg,[1]† Yulia Shifrin,[1]† Ophir Ilan,[1] Chen Nadler-Yona,[1] Shani Nov,[1] Simi Koby,[1] Kobi Baruch,[1] Shoshy Altuvia,[1] Maya Elgrably-Weiss,[1] Cecilia M. Abe,[2] Stuart Knutton,[2] Mark A. Saper,[3] and Ilan Rosenshine[1]*

Department of Molecular Genetics and Biotechnology, The Hebrew University, Faculty of Medicine, POB 12272, Jerusalem 91120, Israel[1]; Institute of Child Health, University of Birmingham, Whittall Street, Birmingham B4 6NH, United Kingdom[2]; and Biophysics Research Division and Department of Biological Chemistry, University of Michigan, 930 North University Ave., Ann Arbor, Michigan 48109-1055[3]

Received 3 December 2004/Accepted 9 May 2005

Escherichia coli **produces polysaccharide capsules that, based on their mechanisms of synthesis and assembly, have been classified into four groups. The group 4 capsule (G4C) polysaccharide is frequently identical to that of the cognate lipopolysaccharide O side chain and has, therefore, also been termed the O-antigen capsule. The genes involved in the assembly of the group 1, 2, and 3 capsules have been described, but those required for G4C assembly remained obscure. We found that enteropathogenic *E. coli* (EPEC) produces G4C, and we identified an operon containing seven genes, *ymcD*, *ymcC*, *ymcB*, *ymcA*, *yccZ*, *etp*, and *etk*, which are required for formation of the capsule. The encoded proteins appear to constitute a polysaccharide secretion system. The G4C operon is absent from the genomes of enteroaggregative *E. coli* and uropathogenic *E. coli*. *E. coli* K-12 contains the G4C operon but does not express it, because of the presence of IS*1* at its promoter region. In contrast, EPEC, enterohemorrhagic *E. coli*, and *Shigella* species possess an intact G4C operon.**

Escherichia coli strains exhibit variations in their cell surface polysaccharides. The primary surface polysaccharides, lipopolysaccharide (LPS) and capsule, are serotype specific (28, 33). Based on their mechanisms of synthesis and assembly, the *E. coli* capsules have been classified into four groups (33). Since in some cases the polysaccharide of the group 4 capsule (G4C) is identical to that of the cognate LPS O side chain, G4C was also termed O-antigen capsule (13). *E. coli* O111 and O127, serotypes which are typical also of enteropathogenic *E. coli* (EPEC) and enterohemorrhagic *E. coli*, were reported to form G4C (13, 27). The specific genes required for the assembly of group 1, 2, and 3 capsules have been identified and characterized (33). However, until now, the genes necessary for the assembly of the G4C were not known.

LPS is an amphipathic glycoconjugate whose hydrophobic domain, lipid A, forms the outer leaflet of the outer membrane. The core oligosaccharide links lipid A to the oligosaccharide side chain, known as the O antigen. The structure of lipid A and of the core oligosaccharide is conserved, but the structure of the O antigen is variable (28). The *E. coli* O antigen may be synthesized by several alternative mechanisms, including the Wzy-dependent pathway. The O antigen, which is synthesized via the Wzy-dependent pathway, is made up of repeat units typically consisting of three to five sugar residues (28). The structural diversity of the O antigens stems mainly from variations in sugar composition, sequence, and linkage. Synthesis and assembly of LPS require numerous proteins, which are encoded by genes clustered in three loci (28): the lipid A biosynthesis region, the core oligosaccharide synthesis locus, and the O-antigen synthesis locus. Lipid A linked to the core oligosaccharide is assembled on the inner leaflet of the plasma membrane and then translocated to the outer leaflet. The O-antigen repeat unit is also synthesized on the inner leaflet of the plasma membrane, on a lipid carrier (undecaprenol phosphate, und-PP). Und-PP, linked to repeat units, is then translocated to the outer leaflet of the plasma membrane (28). Translocation of the individual und-PP-linked O units to the periplasm requires Wzx, the O-unit transporter ("flippase") (23). The und-PP-linked O units are polymerized on the periplasmic face of the plasma membrane. The reaction involves transfer of the nascent polymer from its und-PP carrier to the nonreducing end of the new und-PP-linked O repeat. The net effect is a chain-length increase by 1 repeat unit. Polymerization is catalyzed by the O-polysaccharide polymerase Wzy (2), an integral membrane protein exhibiting specificity for the cognate O unit. The polymerized O antigen is then ligated to the core oligosaccharide. WaaL, an integral membrane protein, is required for this process (28). Upon ligation of the O side chain, the complete LPS is translocated to the external leaflet of the outer membrane.

Polymerization of G4C polysaccharide is also dependent on the O-antigen polymerase Wzy (1, 32). However, in contrast to LPS, G4Cs are composed of a high-molecular-weight polysaccharide attached to the bacterial surface not via lipid but probably by interaction with a specific surface protein(s) (29). We previously showed that EPEC O127:H6 expresses the *etk* gene, which encodes a protein tyrosine kinase (17). We now report that *etk* is encoded within an operon containing six additional genes and that the products of each of these genes are required for the assembly of G4C. The G4C operon appears to encode

* Corresponding author. Mailing address: Department of Molecular Genetics and Biotechnology, Faculty of Medicine, The Hebrew University, POB 12272, Jerusalem 91120, Israel. Phone: 972 2 6758754. Fax: 972 2 6757308. E-mail: ilanro@cc.huji.ac.il.

† Y. Shifrin and A. Peleg contributed equally to this study.

TABLE 1. Bacterial strains and plasmids used in this study

Strain or plasmid	Description	Reference or source
Non-EPEC strains		
MC4100	F$^-$ *araD139* Δ(*argF-lac*)*U169 rpsL150 relA1 flbB5301 deoC1 pstF25 rbsR*]	S. Altuvia
XL1-Blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F′ *priAB lacI*^{q}ZΔM15 Tn*10*(Tet^r)]	Stratagene
XL2-Blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F′ *priAB lacI*^{q}ZΔM15 Tn*10*(Tet^r) Amy Cam^r]	Stratagene
SY327 λ*pir*	Permissive host for transformation	M. Donnenberg
SM10 λ*pir*	Permissive host for conjugation	M. Donnenberg
EPEC strains		
E2348/69	Wild-type EPEC O127:H6	22
OI899	E2348/69 *etk*::*kan*	This study
AD1104	E2348/69 Δ*etp*::*cm*	This study
AD2417	E2348/69 Δ*etp*::*kan*	This study
AD1661	E2348/69 Δ*yccZ*::*kan*	This study
AD1596	E2348/69 Δ*ymcA*::*kan*	This study
AD1663	E2348/69 Δ*ymcB*::*kan*	This study
AD1963	E2348/69 Δ*ymcC*::*kan*	This study
AD1691	E2348/69 Δ*ymcD*::*kan*	This study
AD2119	E2348/69 Δ*ymcD*::*pAP2125*	This study
MS1651	E2348/69 Δ*Pcps*::*kan*	This study
SN40	E2348/69 *ymcA*::mini-Tn*10kan*	This study
Plasmids		
pQE31	Expression vector	Qiagen
pGEM-T	Cloning vector	Promega
pACYC184	Cloning vector	New England Biolabs
pSA10	Expression vector containing *lacI*^{q}	30
pKD3	Template for amplification of Cm cassette	5
pKD4	Template for amplification of Kan cassette	5
pKD13	Template for amplification of Kan cassette	5
pKD46	Helper plasmid expressing λ Red genes	5
pNDMI1	Containing *lacI* on SalI fragment	S. Altuvia
pOI194	*etk* cloned in pQE31	17
pUC4Kan	Containing *kan* cassette	Pharmacia
pOI277	pACYC expressing *etk*	17
pAP406	*etp* cloned in pQE31	This study
pAP2064	*etp* cloned in pSA10	This study
pAP1710	*ymcA* cloned in pSA10	This study
pAP1720	*ymcB* cloned in pSA10	This study
pAP1711	*ymcC* cloned in pSA10	This study
pAP1712	*ymcD* cloned in pSA10	This study
pAP1713	*yccZ* cloned in pSA10	This study
pOI231	pOI194 containing *kan* cassette in the BamHI site of *etk*	This study
pGP704	Suicide vector	25
pAP2125	*ymcD* cloned in pGP704	This study
pAP2133	*ymcD-cspH* intergenic region cloned in pGEM-T	This study

components of a polysaccharide secretion and capsule assembly apparatus. This operon was found in *E. coli* K-12, but it is not expressed. It is also present in a specific subset of *E. coli* and *Shigella* sp. strains.

MATERIALS AND METHODS

Bacterial strains and plasmids. The bacterial strains and plasmids used in this study are listed in Table 1. The primers are listed in Table 2.

Construction of mutants. The Lambda Red system was used to replace the wild-type alleles with inactivated genes. To inactivate *etk*, the *kan* cassette of pUC4Kan (Pharmacia) was inserted into the BglII site of *etk* cloned in pQE31 (pOI194), resulting in the plasmid pOIK. The *etk*::*kan* gene was then amplified and used to replace the wild-type allele. Most of the coding regions of *etp*, *ymcA*, *ymcB*, *ymcC*, *ymcD*, and *yccZ* were replaced with a *kan* or chloramphenicol (Cm) cassette, amplified from pKD3 or pKD4, respectively. Similarly, the promoter region of the colanic acid (CA) operon and *wza*, the first gene of the operon, were deleted and replaced with a *kan* cassette. To construct a mutant containing a polar cassette downstream of the G4C promoter, the *ymcD* coding region was cloned into the *pir*-dependent suicide vector pGP704 (25), to give pAP2125. The plasmid was then mobilized into EPEC (Str^r) by conjugation from *E. coli* SM10 λ*pir*. Transconjugants containing chromosomally integrated plasmid were selected on plates containing ampicillin (50 μg/ml) and streptomycin (100 μg/ml). Plasmid integration via homologous recombination between the *ymcD* alleles was confirmed by PCR.

Plasmid construction. A SalI fragment from pNDMI1 carrying *lacI*^{q} was cloned into the SalI site of pACYC184, generating pACYC234. The SalI site distal to the HindIII site was

TABLE 2. Oligonucleotide primers (in 5′-3′ orientation) used for PCR

Primer and use	Sequence[a]
Gene inactivation	
ymcA KO F	ATCGCACCGGAAGGTGAATTCAGCGTCAACTACCGGGATAACGATCAATACgtgtaggctggagctgcttc
ymcA KO R	ACCTTTACTAAAGCCGCCTTCGCCGTAGTCATCTTTAGATACGTTACTcatatgaatactcctccttag
ymcB KO F	GCGCAAGGAACGGTGACTATTTATCTGCCTGGCGAACAACAGACACTTgtgtaggctggagctgcttc
ymcB KO R	TACATGCGCAGAAAAGCCCAACCAAAGCTGGCTGCCAGGCGGTGGTTCcatatgaatatcctccttag
ymcC KO F	GTGCGCCCTCTTATTTTATCGATTTTCGCACTATTTCTTGCGGGATGTACGTGAgtgtaggctggagctgcttcg
ymcC KO R	TCATGGTGCGGGTTTAAGAAACGTCATTTCTACGGGAATAACCCCTGCGCCcatatgaatatcctccttag
ymcD KO F	GCTGCGTTTGCTGCCCCTGAAGCAACAAACGGGACCGAGGCAACGACGGGTgtgtaggctggagctgcttc
ymcD KO R	AACAACTGCCAACGTCGCCACACCAACAGCCGTTGCCACACCGAAACcatatgaatatcctccttag
yccZ KO F	ACCCTTGTTCCTGGTCAGAATTTATCGACCAGTAACAAGGATGTCATTgtgtaggctggagctgcttc
yccZ KO R	CGGTACCAGCAGAGAAATCACACGGTTCCAACGAGCCAGCGGCGCGGTcatatgaatatcctccttag
Etk KO F	ACCGTCAATGGTATGGTC
Etk KO R	CGTTTTACCACTGTCTGG
Etp KO F	GTTTGTACCGGCAACATTTGCCGTTCGCCCATTGGCGAGCGCTTGCTGCGTTgtgtaggctggagctgcttc
Etp KO R	GACTTTTACGATAGGGATCCGGGATCTCTTTCTGTTCCAGCCATTGCCCcatatgaatatcctcctta
Caps$^-$ F	CCTAAAGAAACTCCTAAAAACCATATTGAATGACACTTAATATAATTCTTgtgtaggctggagctgcttc
Caps$^-$ R	GCTTATCAAGGTTACTGACACCAATAATGGCATCAATTTCATTTTGGATTTATTCCGGGGATCCGTCGACC
Cloning	
ymcAcomp F	GCCAATTGATGAAGAAGAATTCTTATCT
ymcAcomp R	GCCTGCAG**TTA**GTGATGGTGATGGTGATGTTGTCCAACCGGTACTTC
ymcBcomp F	GCGAATTCATGAATAAATTACAGTCGTATTTC
ymcBcomp R	GCCTGCAG**TTA**GTGATGGTGATGGTGATGGTCAGGAACACGTTGCGT
ymcCcomp F	GCGAATTCATGGTCGATACATTTCGCG
ymcCcomp R	GCCTGCAG**TTA**GTGATGGTGATGGTGATGTGGTGCGGGTTTAAGAAAC
ymcDcomp F	GCGAATTCATGAATAAAGGAAAAGTTATGAAAC
ymcDcomp R	GCCTGCAG**TTA**GTGATGGTGATGGTGATGACTTTGCTGGCTGTGCGT
yccZcomp F	GCCAATTGATGAAAAAGAACATTTTTAAGTTTTC
yccZcomp R	GCCTGCAG**TTA**GTGATGGTGATGGTGATGGTTGGGCCATGTTTGAATC
Etk F	ACCGTCAATGGTATGGTC
Etk R	CGTT**TTA**CCACTGTCTGG
Etp F	AGAAGATCTGGTGGATTCAAACATGGC
Etp R	GTAAGC**TTA**CCGGCTGAGGCGCT
pSAEtp F	GCGAATTCGTGGATTCAAACATGGCCCAAC
pSAEtp R	GCGTCGAC**TTA**GTGATGGTGATGGTGATGCCG
2133F	GCGGATCCTAAATGGCCTGCGAGGAC
2133R	GCTCTAGATTAAGTGGTCGTTGTGGTGCC
Sequencing and analysis	
pSA10 F	GCTCGTATAATGTGTGGAATTG
pSA10 R	CGATGGTGTCAACGTAAATGC
M13 F	CGCCAGGGTTTTCCCAGTCACGAC
M13 R	AGCGGATAACAATTTCACACAGGA
k1	CAGTCATAGCCGGAATAGCCT
k2	CGGTGCCCTGAATGAACTGC
828	GATTGACGAAAGCTTGTG

[a] For gene inactivation primers, sequences complementing the resistance cassette are in lowercase; for cloning primers, sequences encoding 6His are underlined and stop codons are in boldface.

disrupted. The pOI194 plasmid was digested with XhoI and HindIII. The fragment containing *6his-etk* under the *tac* promoter was cloned into the SalI and HindIII sites of pACYC234, giving rise to pOI277, which expresses 6His-Etk. The *ymcA*, *ymcB*, *ymcC*, *ymcD*, and *yccZ* genes were amplified with primers fitted with suitable restriction sites and encoding six additional His residues at the 3′ end of the genes (Table 1) and cloned into the pSA10 expression vector digested with EcoRI and PstI or SalI. The *etp* gene was amplified and cloned in pSA10 as described above (to generate pAP2064) or in pQE31 to generate pAP406, which encodes *6his-etp*. The primers which were used are listed in Table 2.

Agglutination assay. Bacteria were grown overnight in LB without NaCl, and when needed 0.1 mM IPTG (isopropyl-β-D-thiogalactopyranoside) was added. The bacteria then remained untreated or were heated to 100°C for 2 min and cooled for 1 h at room temperature. Twofold dilutions of anti-O antiserum in phosphate-buffered saline were prepared, and 50 μl of each dilution was added to 50 μl of heated and unheated bacteria in 96-well plates. The plates were then incubated overnight at 50°C and examined for bacterial agglutination.

Electron microscopy. Bacteria grown overnight in LB at 37°C were washed with 0.1 M cacodylate buffer, pH 7.0; labeled with cationized ferritin (2 mg/ml; Sigma) for 30 min; and fixed with 5% glutaraldehyde for 1 h at room temperature (21). Fixed samples were washed, postfixed with 1% osmium tetroxide for 2 h at room temperature, dehydrated in a graded series of ethanol solutions, and embedded in Epon resin. Thin sections were stained with 2% uranyl acetate and lead citrate and examined

under a JEOL 1200EX transmission electron microscope operating at 80 kV. To quantify capsule expression, low-power micrographs were recorded and the percentage of bacteria expressing cationized ferritin staining material was determined.

Buoyancy assay. Bacterial strains were grown without shaking in 5 ml LB to an optical density at 600 nm (OD_{600}) of 0.3 at 37°C in transparent tubes. The medium was underlaid with 2 ml Percoll (55% Percoll [Amersham], 25 mM phosphate buffer, pH 6.5) and centrifuged in an Eppendorf 5403 centrifuge swing bucket rotor at 1,000 × *g* at room temperature for 20 min. The wild-type encapsulated EPEC cells formed a band at the Percoll-medium interface, and the noncapsulated strains formed pellets.

Preparation of total cellular polysaccharide. Bacteria were grown overnight in 50 ml LB without NaCl. The OD of the cultures was adjusted to 1.0, and the cultures were centrifuged and resuspended in 500 μl of phosphate-buffered saline. An equal volume of saturated phenol (pH 8.0) was added, and the mixture was incubated for 30 min at 70°C with occasional mixing followed by centrifugation (1 h, 10,000 × *g*). The top, aqueous phase was recovered, 2 volumes of 100% ethanol was added to each sample, and the polysaccharides were allowed to precipitate for 1 h at −70°C. The samples were centrifuged at 12,000 × *g* for 30 min, and the pellets were washed with 500 μl of 70% ethanol, recentrifuged, and lyophilized.

Separation of capsule polysaccharide from LPS. The lyophilized total polysaccharide preparations were resuspended in 500 μl of water. The resuspended LPS, which formed micelles, was precipitated by ultracentrifugation (1 h, 86,000 × *g*) (27). The supernatant containing the capsule polysaccharide was recovered, and residual LPS contamination was removed by phase partition with Triton X-114 as described previously (24). Briefly, Triton X-114 was added to the supernatant to a final concentration of 1%. The mixture was incubated at 4°C for 1 h with constant stirring to ensure a homogenous solution. The mixture was then transferred to a 37°C water bath, incubated for 10 min, and centrifuged (1 h, 1,000 × *g*) at 25°C. The resulting aqueous phase containing purified capsule polysaccharide was carefully aspirated. We confirmed that the preparation was LPS free by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and Western analysis using anti-O127 (data not shown).

Analysis of LPS and capsule polysaccharide. Total cell polysaccharide was separated on 10% polyacrylamide gels containing 0.5% SDS (SDS-PAGE), and the gels were used for immunoblot analysis. The SDS-PAGE gels allowed visualization of the LPS without the interference of capsular polysaccharide, which cannot enter the gel because of its low net charge and high molecular weight (13, 27). Samples containing purified capsule polysaccharide were applied directly to nitrocellulose membranes for dot blot analysis. Blots were developed with anti-O127 monospecific rabbit antiserum (Statens Serum Institut, Denmark) and secondary anti-rabbit immunoglobulin G conjugated to alkaline phosphatase (Sigma).

Complement sensitivity assay. Blood was taken from healthy subjects after receiving their informed consent. The blood was left to clot at 37°C for 45 min, and the serum was separated from the clot by centrifugation (10 min, 700 × *g*). Aliquots containing 0.2 ml of bacterial culture in early log phase (OD_{600} of 0.2) were mixed with 0.6 ml serum and incubated at 37°C. Samples of 0.1 ml were diluted and plated for viable count at different time points. Complement was inactivated by incubating the serum for 20 min at 56°C before mixing it with the bacterial culture.

TABLE 3. Agglutination with anti-O127 antibody of wild-type EPEC O127, *etk*::*kan* and Δ*etp*::*kan* mutants, and mutants complemented with plasmids carrying wild-type *etk* or *etp* alleles

Inactivated/complementing gene	Anti-O127 titer[a]
Wild type	4 (64)
etk::*kan*	64 (128)
etk::*kan*/*etk*	8 (32)
Δ*etp*::*kan*	64 (64)
Δ*etp*::*kan*/*etp*	8 (64)

[a] The titer is the reciprocal value of the highest antiserum dilution causing agglutination. The values obtained after removing the capsule by heating for 2 min at 100°C are in parentheses.

Transposon mutagenesis and isolation of noncapsulated mutant. Transposon mutagenesis with mini-Tn*10kan* was carried out as described previously (6). Mutants deficient in capsule formation were isolated as pellet-forming bacteria in the Percoll buoyancy assay.

RNA analysis. Primer extension assays were carried out as described previously (10). Briefly, bacterial cultures grown to an OD_{600} of 0.3 or 1.0 were pelleted and resuspended in 10 mM Tris, pH 7.5, and 1 mM EDTA. Lysozyme was added to 0.9 mg/ml, and the samples were subjected to three freeze-thaw cycles. Total RNA was isolated using an Ultraspec RNA kit according to the manufacturer's (BIOTECX Laboratories) instructions, except that 1 ml of reagent was used for 12 to 16 OD_{600} units of cells. The RNA samples (3 μg) were subjected to primer extension at 42°C for 45 min, using avian myeloblastosis virus reverse transcriptase (CHIMWEx) and end-labeled primer 828. The extension products and the sequencing reaction mixtures primed with end-labeled primer 828 were separated on a 6% sequencing gel.

Genome comparison. We used the coliBASE server (http://colibase.bham.ac.uk/) to access the different databases and for comparison of the G4C operon region in different genomes.

RESULTS

Etk and Etp are required for capsule formation by EPEC. Homologs of Etk (YccC) and Etp (YccY) are frequently involved in capsule formation (17, 31, 33). We, therefore, tested the role of Etk and Etp in capsule formation by EPEC O127:H6. A common definition of the *E. coli* capsule is "a heat-sensitive structure that interferes with bacterial agglutination upon incubation with anti-O antibody" (12, 13). We, therefore, used the agglutination test to compare wild-type EPEC with the Δ*etp*::*kan* and *etk*::*kan* mutants. The titer of the anti-O127 antiserum required to agglutinate the *etp* and *etk* mutants was 16-fold lower than that required to agglutinate the wild-type EPEC (Table 3). This indicates that LPS is masked by a capsule in the wild type but not in the mutants. LPS masking was removed upon heat destruction of the capsule (Table 3). Complementation with plasmids expressing *etk* or *etp* (pOI277 and pAP2064, respectively) restored the LPS-masking effect.

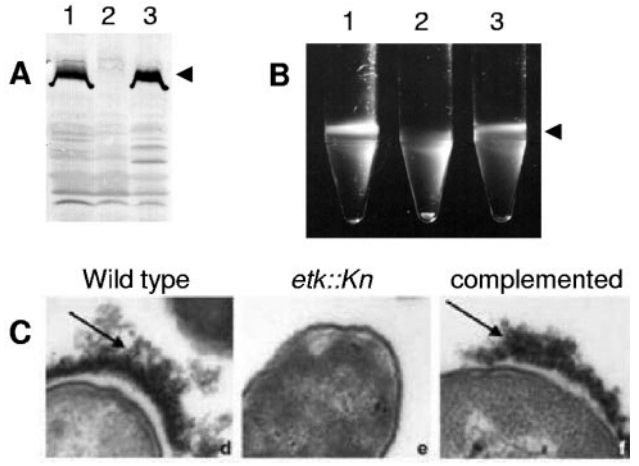

FIG. 1. Expression of Etk and its relation to buoyant density and capsule formation. Bacterial proteins were extracted and used for immunoblotting to which an anti-Etk antibody was applied (A). Etk is expressed by wild-type EPEC (lane 1) and by the EPEC *etk*::*kan* mutant complemented with a plasmid expressing *etk* (lane 3), but not by the EPEC *etk*::*kan* mutant (lane 2). The bacterial buoyancy assay is shown in panel B. Bacteria were grown to an OD_{600} of 0.3, and 5 ml of the cultures was underlaid with 2 ml Percoll in a transparent tube and centrifuged. Wild-type EPEC (tube 1) and an *etk*::*kan* mutant containing a plasmid expressing *etk* (tube 3) appear at the Percoll-medium interface (arrow), whereas the *etk*::*kan* cells (tube 2) form a pellet. Panel C shows electron microscopy images of the Etk-dependent capsule formed by EPEC O127. The *etk*::*kan* mutant appears noncapsulated, whereas the wild-type strain and the *etk*::*kan* mutant complemented with an *etk*-expressing plasmid form capsules (indicated by arrows).

Taken together, these results indicate that Etk and Etp are required for capsule formation by EPEC.

To further support this notion, we examined the bacteria by transmission electron microscopy, following staining with cationized ferritin. A thick layer of ferritin-stained material was seen at the surface of most (80.5%) of the wild-type bacteria, although this layer was not always uniformly dense over the bacterial surface (Fig. 1C). The appearance of the *etk* mutant complemented with plasmid expressing *etk* was similar to that of the wild type (77.5% encapsulated bacteria), whereas very few *etk*::*kan* mutants were stained with ferritin (8%). In the latter case, the staining was sparse and generally appeared filamentous, suggesting that the ferritin might be staining a glycosylated surface pilus structure rather than capsular material. Colonies of heavily encapsulated bacteria are usually mucoid. In contrast, the wild-type EPEC, as well as the *etp* and *etk* mutants, exhibited a nonmucoid colony morphology. Differences in bacterial buoyancy on a Percoll step density gradient were exploited to detect capsules that do not produce mucoid colonies (16). Using this assay, we found that the EPEC *etk*::*kan* mutant exhibited a lower buoyancy than did the wild-type strain (Fig. 1B). The buoyancy of the *etk* mutant was restored upon complementation with a plasmid expressing *etk*. Similar results were obtained using the EPEC *etp* mutant (data not shown). These findings support the notion that *etk* and *etp* are required for capsule formation.

The CA operon is not required for production of the Etk-dependent capsule. We identified in the EPEC genome database two gene clusters encoding enzymes required for polysaccharide biosynthesis. One is the CA biosynthesis operon, and downstream of it resides the LPS O side chain biosynthesis region. To test the possible involvement of the CA operon in the formation of the *etk*/*etp*-dependent capsule, we deleted the CA operon promoter region and part of the operon's first gene (*wza*). To confirm inactivation of the CA operon, we introduced into the mutant (MS1651) and into wild-type EPEC a plasmid overexpressing RcsA, a regulator that activates expression of the CA operon (14). As expected, RcsA-expressing EPEC, grown in LB at 37°C, produced mucoid colonies, but strain MS1651 lost this ability upon RcsA overexpression (data not shown). In contrast, the buoyancy of this mutant was similar to that of wild-type EPEC when grown in LB at 37°C. These results indicate that the CA operon is not required for the production of the *etp*/*etk*-dependent capsule. It also suggests that *etp* and *etk* may be involved in the formation of G4C, composed of polymers similar to those of the O side chain polysaccharide.

Etp and Etk are required for the formation of G4C. Some *E. coli* strains of serogroups O127 and O111 form G4C, which is composed of a polysaccharide similar to that of the LPS O side chain (13, 27). G4C polymers extracted from these strains reacted with anti-O antibody, but the encapsulated O111 bacteria do not undergo agglutination by anti-O111 antibodies (13). To investigate whether Etp and Etk are involved in the production of G4C, the capsule and LPS of EPEC strain E2348/69, serogroup O127, were extracted and separated by differential centrifugation and phase partition (24, 27). SDS-PAGE and Western blotting with anti-O127 were used to assess the amount of LPS side chain without the possible interference of the capsular polysaccharide, which cannot enter the gel because of its low net charge and high molecular weight (13, 27, 34) (Fig. 2A). The purified capsular polysaccharide was analyzed by immuno-dot blot assay with anti-O127 (Fig. 2B). We found that anti-O127 antibody reacted with the purified capsular polysaccharide, indicating that this is an O-antigen capsule. We also found that synthesis of the capsular polysaccharide was abolished upon inactivation of *etk* or *etp* and was associated with increased synthesis of the LPS O side chain (Fig. 2A and B). Introduction of complementing plasmids partially restored capsule production, which was associated with a reduction in LPS O side chain synthesis (Fig. 2A and B). These results indicate that *etp* and *etk* are required for the assembly of G4C and are consistent with the hypothesis that the capsule synthesis system and the LPS O side chain synthesis system are competing for the same repeat-unit precursor (27). Some strains producing G4C are sensitive to complement; spontaneous variants, deficient in capsule formation, exhibit increased O-antigen synthesis, which may lead to changes in O-antigen conformation (12, 18, 19). These changes in O-antigen synthesis correlated with increased bacterial complement resistance (12, 18, 19). Given our previous results, we predicted that the *etk* mutant would be complement resistant. To test this hypothesis, cultures in early logarithmic growth phase were mixed with fresh human serum and survival was monitored. The human serum rapidly killed wild-type EPEC (Fig. 2C), but the EPEC *etk*::*kan* mutant was barely affected. Complementation with plasmid expressing *etk* partially restored complement sensitivity. Inactivation of the complement by heat treatment abol-

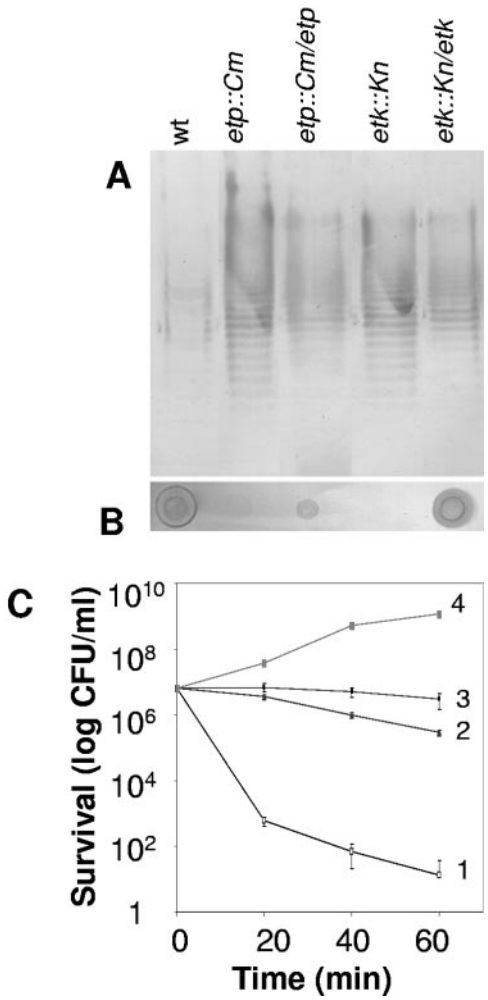

FIG. 2. Etk and Etp are required for synthesis of G4C. Total polysaccharides (LPS and capsular polysaccharide) were extracted, and the capsular polysaccharide was separated from the LPS. Production of LPS O antigen was analyzed by Western blotting (A), and the amount of purified capsular polysaccharide was estimated by dot blot assay (B). In both cases, care was taken to apply amounts of material equivalent to the same amount of bacteria in all the lanes. The Western blot and the dot blots were developed with anti-O127 antiserum. The genotypes of the analyzed strains are indicated above the lanes. The *etk*::*kan* and the Δ*etp*::Cm mutants did not produce capsules (B) but synthesized greater amounts of the LPS side chain (A). Complementation of the mutants with plasmids expressing *etk* or *etp* partially restored capsule production (B) and reduced synthesis of the LPS side chain (A). Panel C shows the resistance of the *etk*::*kan* mutant to complement-mediated killing. Early-log cultures were incubated with fresh or inactivated human serum. Wild-type EPEC cells were killed rapidly by the fresh serum (line 1) but not by the heat-inactivated serum (line 4). The EPEC *etk*::*kan* mutant (line 3) was highly resistant to fresh serum; complementation of the mutant with plasmid expressing *etk* (line 2) resulted in a partial increase in serum sensitivity. All experiments were carried out in triplicate, and the standard error is indicated.

ished the serum's ability to kill the bacteria (Fig. 2C). The tested serum did not contain anticapsule antibody (data not shown), excluding the possibility that the complement was specifically recruited to the encapsulated bacteria by anticapsule antibodies. These results indicate that in EPEC O127 the interplay between G4C and LPS synthesis is similar to that described for *E. coli* O111 (12).

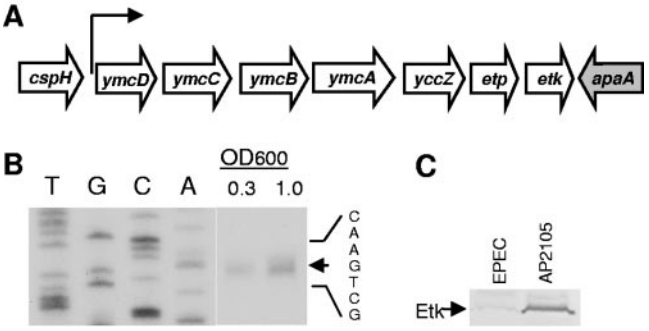

FIG. 3. The G4C operon. Schematic of the G4C operon and flanking genes (*cspH* and *appA*) (A). The location of the G4C operon promoter is indicated by an arrow. Mapping of the G4C promoter (B). Total RNA was extracted from exponential- and stationary-phase EPEC cultures (OD_{600} of 0.3 and 1, respectively). Primer extension analysis (3 μg, total RNA) of the *ymcD* promoter region and the sequencing reaction were carried out using the same primer. Plasmid pAP2133 served as a template for the sequencing reaction. The transcription start site was mapped to a G residue located 166 bases upstream of the *ymcD* translation start site. In panel C, Western analysis with anti-Etk antibody was used to compare wild-type EPEC with strain AD2119, in which a polar cassette between *ymcD* and the promoter was inserted. The polar insertion abolished Etk expression, confirming that *ymcD* through *etk* are expressed by the same transcriptional unit.

The G4C operon. To identify additional genes involved in capsule formation, we screened a mini-Tn*10* mutant library and isolated an EPEC mutant, SN40, which exhibits decreased buoyancy in the Percoll buoyancy assay. The mini-Tn*10* insertion site in SN40 was mapped to *ymcA*, a gene located three open reading frames upstream of *etk* (Fig. 3A). Western analysis showed that SN40 does not produce Etk (Fig. 4A), nor is Etk production restored upon complementation with *ymcA*-expressing plasmid (data not shown). These results indicate that the mini-Tn*10* insertion in SN40 introduced a polar effect and suggest that *ymcA* and *etk* are encoded within an operon

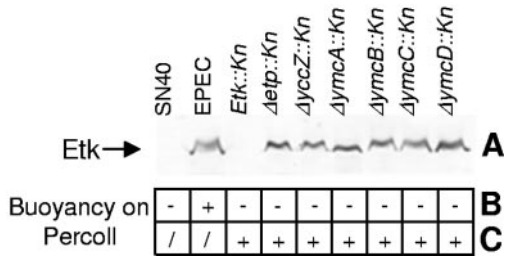

FIG. 4. Inactivation of each of the G4C operon genes results in lack of buoyancy on a Percoll step gradient. Each of the G4C operon genes was inactivated. To confirm nonpolarity, expression of Etk by the different mutants was determined using Western blot analysis with anti-Etk antibody (A). The genotype of the tested EPEC is indicated above each lane. SN40 is an EPEC strain containing a mini-Tn*10kan* polar insertion in *ymcA*. The wild-type strain is indicated as EPEC. Below the blot, the results of the buoyancy assays using the corresponding mutants (B) and the mutants complemented with plasmid expressing the corresponding gene (C) are shown. Buoyancy on Percoll is indicated as a plus sign, lack of buoyancy is indicated as a minus sign, and a slash indicates that buoyancy was not tested.

containing seven genes: *ymcD*, *ymcC*, *ymcB*, *ymcA*, *yccZ*, *etp*, and *etk* (Fig. 3A). To test this prediction, we performed primer-extension analysis of RNA extracted from EPEC, using a primer complementing the *ymcD* mRNA (Fig. 3). A transcription start site was localized to a G residue located 166 bp upstream of the *ymcD* translation start site (Fig. 3B). To test whether this is the only promoter of the G4C operon, we inserted a suicide plasmid, which served as a polar cassette, between the G4C promoter and *ymcD*. The insertion abolished the production of Etk (Fig. 3C), confirming that *ymcD* and *etk* are the first and last genes, respectively, of the G4C operon.

All the genes in the G4C operon are required for capsule formation. We next introduced a nonpolar mutation into each of the G4C operon genes and confirmed the lack of polarity by Western analysis, using anti-Etk antibody (Fig. 4A). All the mutants proved to be deficient in capsule formation, as determined by the buoyancy assay (Fig. 4B). We then cloned each of the G4C genes in the pSA10 expression vector (Table 1) and used these plasmids in complementation assays. Capsule production was restored by complementation with plasmids expressing the corresponding genes (Fig. 4C). These results indicate that each of the seven proteins encoded by the operon is required for G4C formation.

The G4C operon in different *E. coli* and *Shigella* sp. strains. Using genomic databases, we compared the chromosomal region containing the G4C operon with the corresponding region in several *E. coli* and *Shigella* sp. strains. We found an intact G4C operon in seven genomes. These include EPEC O127 strain E2348/69; two EHEC strains, EDL933 and Sakai; and three *Shigella* strains, *S. sonnei* 53G, *S. flexneri* strains 2a 301 and 2a 2457T, and *S. dysenteriae* M131649. These *E. coli* and *Shigella* strains are expected to form G4C. Intact G4C operons were found also in two *E. coli* K-12 strains, MG1655 and W3110. But in both cases the promoter was inactivated by an IS*1* element inserted 15 bp upstream of the *ymcD* translation start site, a finding that explains the lack of *etk* expression in these *E. coli* K-12 strains (17, 31). The entire G4C operon is absent from uropathogenic *E. coli* strain CFT073 and from enteroaggregative *E. coli* strain 042.

DISCUSSION

E. coli employs several alternative mechanisms to synthesize the O side chain, of which the Wzy-dependent pathway is used also for synthesis of the G4C polymers (1, 32). In this report we describe the identification and characterization of an operon containing seven genes that are required for G4C formation. Thus, the genes required for G4C formation are organized in two separate operons. One operon contains genes required for the Wzy-dependent synthesis of the O side chain and the G4C polymers. The other, the G4C operon, is required for secretion of the G4C polymers and for capsule assembly. The G4C operon is conserved between different *E. coli* serogroups, indicating that the encoded secretion apparatus is not specific to the cognate O polymers. Wzy-dependent O-antigen synthesis requires Wzy, the O side chain polymerase (2), and WaaL, which ligates the polymerized O antigen to the core oligosaccharide (28). The Wzz protein is required for determination of O side chain length, which is typically 15 to 25 repeat units (7, 11). It has been speculated that Wzy, WaaL, and Wzz are components in an O side chain synthesis complex (28). We are currently testing the hypothesis that the products of the G4C operon form a complex that competes with the O side chain synthesis complex for the repeat unit precursor. The G4C operon encodes seven proteins: Etk, Etp, YzzZ, YmcA, YmcB, YmcC, and YmcD. Etk, Etp, and YccZ are similar to Wzc, Wzb, and Wza, respectively, which are required for secretion of polysaccharide including colanic acid and of K30 polymers (9, 31, 35). Members of the Etk protein family are inner membrane proteins possessing two transmembrane helices and tend to oligomerize (8, 26). They are composed of three distinct domains: a short cytoplasmic N-terminal domain, a large periplasmic loop, and a large C-terminal cytoplasmic domain (8). Since the periplasmic domains of Etk and Wzz show some similarity, we are currently testing whether Etk uses its Wzz-like domain to recruit the Wzy to the G4C complex and to modulate it to polymerize very long capsular O polymers. The C-terminal domain of Etk and related proteins contains a protein tyrosine kinase catalytic site. These proteins undergo autophosphorylation at several conserved tyrosines clustered at the C terminus. The phosphorylation is essential for secretion of the cognate polysaccharides (26, 31, 35). An additional, minor substrate of Wzc is Ugd (UDP-glucose dehydrogenase) (15). Etk was reported also to phosphorylate RpoH and RseA (anti-RpoE) (20), and therefore, it might also be involved in genetic regulation. Etp is a low-molecular-weight protein tyrosine phosphatase (31). It is predicted to reside in the cytoplasm and mediate Etk dephosphorylation both in vitro and in vivo (reference 31 and our unpublished results). We speculate that Etp and Etk might energize the G4C assembly system by repeated phosphorylation and dephosphorylation of the Etk C-terminal cytoplasmic domain. YccZ is an outer membrane lipoprotein. Octamers of Wza-K30, and probably also of other Wza-like proteins, including YccZ, form pore-like structures in the outer membrane (3, 9). It is believed that the polysaccharides are delivered to the surface via these pores (3, 9). We predict, therefore, that the G4C polymer is secreted via the YccZ channel. Based on their deduced sequence, YmcA, YmcB, YmcC, and YmcD contain an N-terminal signal sequence and should be exported proteins. Two of them, YmcA and YmcC, are putative lipoproteins. YmcB appears to reside in the periplasm in a soluble form or is associated with the inner or outer membrane. YmcD is a small protein containing two putative transmembrane domains. It is unusually rich in threonine: 28 of its 107 residues are threonines. *E. coli* contains paralogs of the *ymcABCD* genes, the *yjbHGFE* genes, which form a putative operon localized in the *xylE malG* region. The function of the *yjbHGFE* genes is unknown. Interestingly, *Vibrio cholerae* O139 contains homologs of the *ymcABC* genes, the *otnGFE* genes (4). The function of these genes has not been determined, but they are located in the vicinity of *otnAB*, which are transcribed in the opposite direction and required for the formation of the O139 capsule (4).

In conclusion, we here identify seven genes required for G4C formation by EPEC O127. We recently found that all the tested EPEC serotypes, including O111, O119, O142, and O55, and the closely related enterohemorrhagic *E. coli* serotype O157, form G4C and/or express Etk (reference 17 and unpublished data). This correlation suggests that G4C may play a role in the virulence of these closely related pathogens. We are now

testing this hypothesis. G4C might also be involved in the virulence of *Shigella* strains containing intact G4C operons.

ACKNOWLEDGMENTS

We thank S. Gottesman for the *rcsA*-expressing plasmid.

This work was supported by grants from the United States-Israel Binational Science Foundation, The Israel Science Foundation founded by The Israel Academy of Science and Humanities (grant 223/99-2), the European Union Fifth Framework Quality of Life Program (QLK2-2000-00600), and the Center of Study of Emerging Disease. I.R. is the Etta Rosensohn Professor in Bacteriology, and M.A.S. was a Lady Davis Visiting Professor, at The Hebrew University.

REFERENCES

1. **Amor, P. A., and C. Whitfield.** 1997. Molecular and functional analysis of genes required for expression of group IB K antigens in *Escherichia coli*: characterization of the his-region containing gene clusters for multiple cell-surface polysaccharides. Mol. Microbiol. **26:**145–161.
2. **Bastin, D. A., G. Stevenson, P. K. Brown, A. Haase, and P. R. Reeves.** 1993. Repeat unit polysaccharides of bacteria: a model for polymerization resembling that of ribosomes and fatty acid synthetase, with a novel mechanism for determining chain length. Mol. Microbiol. **7:**725–734.
3. **Beis, K., R. F. Collins, R. C. Ford, A. B. Kamis, C. Whitfield, and J. H. Naismith.** 2004. Three-dimensional structure of Wza, the protein required for translocation of group 1 capsular polysaccharide across the outer membrane of *Escherichia coli*. J. Biol. Chem. **279:**28227–28232.
4. **Bik, E. M., A. E. Bunschoten, R. J. Willems, A. C. Chang, and F. R. Mooi.** 1996. Genetic organization and functional analysis of the otn DNA essential for cell-wall polysaccharide synthesis in *Vibrio cholerae* O139. Mol. Microbiol. **20:**799–811.
5. **Datsenko, K. A., and B. L. Wanner.** 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc. Natl. Acad. Sci. USA **97:**6640–6645.
6. **de Lorenzo, V., and K. N. Timmis.** 1994. Analysis and construction of stable phenotypes in gram-negative bacteria with Tn5- and Tn10-derived minitransposons. Methods Enzymol. **235:**386–405.
7. **Dodgson, C., P. Amor, and C. Whitfield.** 1996. Distribution of the *rol* gene encoding the regulator of lipopolysaccharide O-chain length in *Escherichia coli* and its influence on the expression of group I capsular K antigens. J. Bacteriol. **178:**1895–1902.
8. **Doublet, P., C. Grangeasse, B. Obadia, E. Vaganay, and A. J. Cozzone.** 2002. Structural organization of the protein-tyrosine autokinase Wzc within *Escherichia coli* cells. J. Biol. Chem. **277:**37339–37348.
9. **Drummelsmith, J., and C. Whitfield.** 2000. Translocation of group 1 capsular polysaccharide to the surface of *Escherichia coli* requires a multimeric complex in the outer membrane. EMBO J. **19:**57–66.
10. **Elgrably-Weiss, M., S. Park, E. Schlosser-Silverman, I. Rosenshine, J. Imlay, and S. Altuvia.** 2002. A *Salmonella enterica* serovar Typhimurium *hemA* mutant is highly susceptible to oxidative DNA damage. J. Bacteriol. **184:** 3774–3784.
11. **Franco, A. V., D. Liu, and P. R. Reeves.** 1996. A Wzz (Cld) protein determines the chain length of K lipopolysaccharide in *Escherichia coli* O8 and O9 strains. J. Bacteriol. **178:**1903–1907.
12. **Goldman, R. C., K. Joiner, and L. Leive.** 1984. Serum-resistant mutants of *Escherichia coli* O111 contain increased lipopolysaccharide, lack an O antigen-containing capsule, and cover more of their lipid A core with O antigen. J. Bacteriol. **159:**877–882.
13. **Goldman, R. C., D. White, F. Orskov, I. Orskov, P. D. Rick, M. S. Lewis, A. K. Bhattacharjee, and L. Leive.** 1982. A surface polysaccharide of *Escherichia coli* O111 contains O-antigen and inhibits agglutination of cells by O-antiserum. J. Bacteriol. **151:**1210–1221.
14. **Gottesman, S., P. Trisler, and A. Torres-Cabassa.** 1985. Regulation of capsular polysaccharide synthesis in *Escherichia coli* K-12: characterization of three regulatory genes. J. Bacteriol. **162:**1111–1119.
15. **Grangeasse, C., B. Obadia, I. Mijakovic, J. Deutscher, A. J. Cozzone, and P. Doublet.** 2003. Autophosphorylation of the *Escherichia coli* protein kinase Wzc regulates tyrosine phosphorylation of Ugd, a UDP-glucose dehydrogenase. J. Biol. Chem. **278:**39323–39329.
16. **Hakansson, S., and S. Holm.** 1986. Influence of polysaccharide capsule and ionic strength on buoyant density of group B streptococci. Acta Pathol. Microbiol. Immunol. Scand. B **94:**139–143.
17. **Ilan, O., Y. Bloch, G. Frankel, H. Ullrich, K. Geider, and I. Rosenshine.** 1999. Protein tyrosine kinases in bacterial pathogens are associated with virulence and production of exopolysaccharide. EMBO J. **18:**3241–3248.
18. **Joiner, K. A., M. A. Schmetz, R. C. Goldman, L. Leive, and M. M. Frank.** 1984. Mechanism of bacterial resistance to complement-mediated killing: inserted C5b-9 correlates with killing for *Escherichia coli* O111B4 varying in O-antigen capsule and O-polysaccharide coverage of lipid A core oligosaccharide. Infect. Immun. **45:**113–117.
19. **Kastowsky, M., T. Gutberlet, and H. Bradaczek.** 1992. Molecular modeling of the three-dimensional structure and conformational flexibility of bacterial lipopolysaccharide. J. Bacteriol. **174:**4798–4806.
20. **Klein, G., C. Dartigalongue, and S. Raina.** 2003. Phosphorylation-mediated regulation of heat shock response in *Escherichia coli*. Mol. Microbiol. **48:** 269–285.
21. **Knutton, S., P. H. Williams, D. R. Lloyd, D. C. Candy, and A. S. McNeish.** 1984. Ultrastructural study of adherence to and penetration of cultured cells by two invasive *Escherichia coli* strains isolated from infants with enteritis. Infect. Immun. **44:**599–608.
22. **Levine, M. M., E. J. Bergquist, D. R. Nalin, D. H. Waterman, R. B. Hornick, C. R. Young, and S. Sotman.** 1978. *Escherichia coli* strains that cause diarrhoea but do not produce heat-labile or heat-stable enterotoxins and are non-invasive. Lancet. **i:**1119–1122.
23. **Liu, D., R. A. Cole, and P. R. Reeves.** 1996. An O-antigen processing function for Wzx (RfbX): a promising candidate for O-unit flippase. J. Bacteriol. **178:**2102–2107.
24. **Liu, S., R. Tobias, S. McClure, G. Styba, Q. Shi, and G. Jackowski.** 1997. Removal of endotoxin from recombinant protein preparations. Clin. Biochem. **30:**455–463.
25. **Miller, V. L., and J. J. Mekalanos.** 1988. A novel suicide vector and its use in construction of insertion mutations: osmoregulation of outer membrane proteins and virulence determinants in *Vibrio cholerae* requires *toxR*. J. Bacteriol. **170:**2575–2583.
26. **Paiment, A., J. Hocking, and C. Whitfield.** 2002. Impact of phosphorylation of specific residues in the tyrosine autokinase, Wzc, on its activity in assembly of group 1 capsules in *Escherichia coli*. J. Bacteriol. **184:**6437–6447.
27. **Peterson, A. A., and E. J. McGroarty.** 1985. High-molecular-weight components in lipopolysaccharides of *Salmonella typhimurium*, *Salmonella minnesota*, and *Escherichia coli*. J. Bacteriol. **162:**738–745.
28. **Raetz, C. R., and C. Whitfield.** 2002. Lipopolysaccharide endotoxins. Annu. Rev. Biochem. **71:**635–700.
29. **Rahn, A., K. Beis, J. H. Naismith, and C. Whitfield.** 2003. A novel outer membrane protein, Wzi, is involved in surface assembly of the *Escherichia coli* K30 group 1 capsule. J. Bacteriol. **185:**5882–5890.
30. **Schlosser-Silverman, E., M. Elgrably-Weiss, I. Rosenshine, R. Kohen, and S. Altuvia.** 2000. Characterization of *Escherichia coli* DNA lesions generated within J774 macrophages. J. Bacteriol. **182:**5225–5230.
31. **Vincent, C., B. Duclos, C. Grangeasse, E. Vaganay, M. Riberty, A. J. Cozzone, and P. Doublet.** 2000. Relationship between exopolysaccharide production and protein-tyrosine phosphorylation in gram-negative bacteria. J. Mol. Biol. **304:**311–321.
32. **Wang, L., C. E. Briggs, D. Rothemund, P. Fratamico, J. B. Luchansky, and P. R. Reeves.** 2001. Sequence of the *E. coli* O104 antigen gene cluster and identification of O104 specific genes. Gene **270:**231–236.
33. **Whitfield, C., and I. S. Roberts.** 1999. Structure, assembly and regulation of expression of capsules in *Escherichia coli*. Mol. Microbiol. **31:**1307–1319.
34. **Widmalm, G., and K. Leontein.** 1993. Structural studies of the *Escherichia coli* O127 O-antigen polysaccharide. Carbohydr. Res. **247:**255–262.
35. **Wugeditsch, T., A. Paiment, J. Hocking, J. Drummelsmith, C. Forrester, and C. Whitfield.** 2001. Phosphorylation of Wzc, a tyrosine autokinase, is essential for assembly of group 1 capsular polysaccharides in *Escherichia coli*. J. Biol. Chem. **276:**2361–2371.