JOURNAL OF BACTERIOLOGY, July 1994, p. 4144–4156
0021-9193/94/$04.00+0
Copyright © 1994, American Society for Microbiology
Vol. 176, No. 13

Structure of the O Antigen of *Escherichia coli* K-12 and the Sequence of Its *rfb* Gene Cluster

GORDON STEVENSON,[1] BRIAN NEAL,[1] DAN LIU,[1] MATTHEW HOBBS,[1] NICOLLE H. PACKER,[2] MICHAEL BATLEY,[2] JOHN W. REDMOND,[2] LENNART LINDQUIST,[3] AND PETER REEVES[1]*

Department of Microbiology, University of Sydney, Sydney, New South Wales 2006,[1] and School of Chemistry, Macquarie University, New South Wales 2109,[2] Australia, and Department of Clinical Bacteriology, Karolinska Institute, Huddinge Hospital, S-141 86 Huddinge, Sweden[3]

Received 8 February 1994/Accepted 30 April 1994

Escherichia coli* K-12 has long been known not to produce an O antigen. We recently identified two independent mutations in different lineages of K-12 which had led to loss of O antigen synthesis (D. Liu and P. R. Reeves, Microbiology 140:49–57, 1994) and constructed a strain with all *rfb* (O antigen) genes intact which synthesized a variant of O antigen O16, giving cross-reaction with anti-O17 antibody. We determined the structure of this O antigen to be →2)-β-D-Gal*f*-(1→6)-α-D-Glc*p*-(1→3)-α-L-Rha*p*-(1→3)-α-D-Glc*p*NAc-(1→, with an *O*-acetyl group on C-2 of the rhamnose and a side chain α-D-Glc*p* on C-6 of GlcNAc. O antigen synthesis is *rfe* dependent, and D-Glc*p*NAc is the first sugar of the biological repeat unit. We sequenced the *rfb* (O antigen) gene cluster and found 11 open reading frames. Four rhamnose pathway genes are identified by similarity to those of other strains, the rhamnose transferase gene is identified by assay of its product, and the identities of other genes are predicted with various degrees of confidence. We interpret earlier observations on interaction between the *rfb* region of *Escherichia coli* K-12 and those of *E. coli* O4 and *E. coli* Flexneri. All K-12 *rfb* genes were of low G+C content for *E. coli*. The rhamnose pathway genes were similar in sequence to those of (*Shigella*) Dysenteriae 1 and Flexneri, but the other genes showed distant or no similarity. We suggest that the K-12 gene cluster is a member of a family of *rfb* gene clusters, including those of Dysenteriae 1 and Flexneri, which evolved outside *E. coli* and was acquired by lateral gene transfer.

Escherichia coli K-12 was isolated in 1922 and used as a standard *E. coli* strain at Stanford University for many years; the strains which survive all derive from cultures given to E. Tatum and others in the 1940s and early 1950s, when *E. coli* K-12 was first used for the genetic studies which led to its adoption as the major strain for laboratory study. After 50 years of intensive study, *E. coli* K-12 is arguably the best understood of all organisms, having been used for studies of many facets of living organisms, outlined in a two-volume book on *E. coli* and *Salmonella enterica* (45). Currently, 50% of its genome has been sequenced (56), and completion of this task will increase the focus of attention on *E. coli* K-12 for studies which integrate genomic information and biochemical processes.

There are, however, significant gaps in our knowledge of *E. coli* K-12. During its first 25 to 30 years in the laboratory, it probably accumulated a range of mutations which improved adaptation to a laboratory environment but destroyed its ability to survive in its natural environment. Among them were mutations in the *rfb* gene cluster which led to loss of O antigen synthesis. The O antigen, a repeat unit polysaccharide which is a component of lipopolysaccharide (LPS), is the major surface antigen of many gram-negative bacteria (see reference 53 for a review) and, for *E. coli*, was present in by far the majority of strains when first isolated. However, it is often lost during culture, presumably because it offers no advantage under such conditions but has substantial requirements for synthesis. In *E. coli* K-12, *rfb* mutations appear to have occurred before any of the current stocks were established, and hence it was not known which of the more than 160 O antigen forms was present in the strain when first isolated.

We recently identified two independent *rfb* mutations in different lineages of *E. coli* K-12: most strains carry the *rfb-50* mutation, an IS*5* insertion in the last gene of the cluster; but strain WG1 is unique in carrying the *rfb-51* mutation, a deletion of TDP-rhamnose synthesis genes. We were then able to show that a strain carrying the *rfb-50* mutation, when complemented by a plasmid carrying the last gene derived from strain WG1 in its complete form, had its O antigen synthesis restored. In this paper, we report the structure of the O antigen of *E. coli* K-12 and the sequence of its *rfb* gene cluster.

MATERIALS AND METHODS

Bacterial strains and plasmids, media, and growth conditions. Bacterial strains and plasmids used in this study are shown in Table 1. Cultures for general use were grown at 37°C in L broth (constituents from Difco) or on nutrient agar plates with antibiotics added as appropriate. The culture of P4863 for LPS extraction was grown in a 50-liter batch in a fermentor with aeration. The medium contained, per liter, 10 g of bacterial peptone (Oxoid), 5 g of yeast extract (Oxoid), 5 g of glucose, and 5 g of NaCl. It was buffered by addition of 17 mM KH_2PO_4 and 72 mM K_2HPO_4; ampicillin to 25 μl ml^{-1} and 15 ml of silicone antifoaming reagent (catalog no. 33151; Sigma) were added before inoculation with 500 ml of an overnight culture and growth at 37°C for 17 h with aeration. The culture volume was reduced to 8 liters by using an ultrafiltration system (model DC10LA utilizing hollow-fiber membrane; Amicon). The cells were then harvested by centrifugation (2,600 × *g* for 15 min), washed twice with 4.5 liters of sterile deionized water, resuspended in 150 ml of 1% aqueous phenol

* Corresponding author. Mailing address: Department of Microbiology, Bldg. G08, University of Sydney, Sydney, NSW 2006, Australia. Phone: (61-2) 6922536. Fax: (61-2) 6924571. Electronic mail address: reeves@angis.su.au.oz.

TABLE 1. Strains and plasmids used in this study

Strain or plasmid	Laboratory stock no.	Characteristics	Reference or source
Strains			
C600	P334	F⁻, *thr-1 leuB6 tonA1 lacY1 tonA21 supE44 rfbC1* λ⁻	3
SØ874	P4052	*lacZ4503 trp-355 upp-12 relA rpsL150* Δ(*sbcB-rfb*)	46
WG1	P4810	Wild-type strain (CGSC5073)	4
	P4863	W1485 carrying pPR1474	37
AN21548	P4991	*thr-1 leuB6* Δ(*gpt-proA*)*66 hisG4 argE3 ara-14 lacY1 galK2 xyl-5 mtl-1 mgl-51 rpsL31 kdgK51 supE44 rfbC1 rfe*::Tn*10-48*	42
P 4	M492	*E. coli* wild-type strain O16	50
K 12a	M493	*E. coli* wild-type strain O17	50
Plasmids			
pRL100		*rfe* gene of *E. coli* K-12 in pBR322	42
pPR637		Low-copy-number cloning vector derived from pGB2 (17) containing *lacZ* and polylinker region of pUC19	This laboratory
pPR1010		*rfbB* gene of *S. enterica* LT2 in pUC9	29
pPR1167		3.1-kb *Hin*dIII fragment from K352[a] (34) in pUC18	37
pPR1168		4.1-kb *Eco*RI-*Hin*dIII fragment from K352[a] (34) in pUC18	37
pPR1461		8-kb *Eco*RI fragment from *rfb* region of WG1 in pUC18 in orientation for expression	37
pPR1472		2-kb *Hin*dIII fragment from 1G1[a] (34) in pUC18	This study
pPR1473		1.6-kb *Hin*dIII fragment from 1G1[a] (34) in pUC18	This study
pPR1474		5-kb *Eco*RI-*Sph*I fragment from WG1 in pUC18	37
pPR1538		8-kb *Eco*RI fragment from pPR1461 in pPR1555	This study
pPR1539		770-bp *Eco*RI-*Hin*dIII fragment from K351[a] in pGEM7zf(+)	This study
pPR1555		4.1-kb *Hin*dIII-*Eco*RI fragment from pPR1168 in pPR637 (29)	This study
pPR1556		2.2-kb *Hin*dIII-*Pst*I fragment from pPR1491 (37) in pPR637 (29)	This study

[a] K351, K352, and 1G1 are phage lambda clones of K-12 strain W3110 (34).

solution, stored at 4°C, and extracted within 24 h. The yield was 230 g (wet weight) of packed cells.

Terminology for *Shigella* strains. The four species of *Shigella* and *E. coli* are a single species on the basis of DNA relatedness as determined by hybridization of chromosomal DNA (12). Further, for genes which have been sequenced in *E. coli* and one of the *Shigella* species, the difference is within the range expected for variation within *E. coli*. For example, the *crp* genes of *E. coli* K-12 and *Shigella flexneri* differ at four bases, well within the range expected for variation within *E. coli*, whereas those of *E. coli* and the closely related species *S. enterica* differ at 77 bases (12.3%) (21). It is extremely difficult to justify separate species or genera for these organisms (12), and while this anomalous terminology may have advantages in some areas, it is very misleading when discussing evolutionary relationships and confusing to those outside the immediate field. We have therefore treated *Shigella* strains as strains of *E. coli* and to indicate this have not italicized the names and have capitalized the first letter of the old species names in accordance with the tradition adopted for old species names of *S. enterica* (35).

Chemicals and enzymes. Uridine diphosphate and *N*-acetyl-D-[^{14}C]glucosamine (UDP-[^{14}C]GlcNAc) were obtained from Amersham, and [^{32}P]dCTP was purchased from Bresatec. Restriction enzymes and ligase were from Boehringer Mannheim or Promega. Synthesis of dTDP-[U-^{14}C]rhamnose was as described previously (37, 41).

Preparation of polysaccharide. LPS was isolated from packed cells by phenol-water extraction (71) followed by nuclease treatment and ultracentrifugation three times (85,000 × *g* for 8 h). The polysaccharide was prepared by hydrolysis of LPS (100 mg) in 1% acetic acid (10 ml) at 100°C for 90 min. Lipid A was removed by filtration through acid-washed Celite. The filtrate was freeze-dried, redissolved in water (0.25 ml), clarified by centrifugation, and fractionated on a Biogel P6 column (30 by 2.5 cm) eluted with water (0.5 ml/min). The eluate was monitored by refractive index, and the high-molecular-weight fractions were pooled to give the polysaccharide.

Compositional analysis. Polysaccharide (1 mg) was hydrolyzed in 2 M trifluoroacetic acid (TFA) (0.2 ml) for 4 h at 100°C and evaporated to dryness with a Speed Vac concentrator. The composition was analyzed by two methods. The monosaccharides were analyzed directly by high-pH chromatography on a Dionex CarboPac PA1 column (4 by 250 mm) with an isocratic gradient of 15 mM sodium hydroxide solution with pulsed amperometric detection (Waters 464). Alditol acetate derivatives (28) of the monosaccharides were analyzed by gas chromatography with a BP-X5 column (25 m by 0.25 mm; film thickness, 0.25 μm). Identification was by comparison with authentic standards.

Methylation analysis. The polysaccharide was methylated as described by Ciucanu and Kerek (18). Before extraction with dichloromethane, the pH of the aqueous solution was adjusted to below 3 in order to depress the ionization of carboxyl groups and improve the solubility of the methylated polysaccharide in the organic solvent. The methylated polysaccharide was hydrolyzed in TFA (0.5 ml) at 100°C for 4 h. Reduction with NaB_2H_4 and acetylation were done conventionally (16). The partially methylated alditol acetates were analyzed by gas chromatography-mass spectrometry (Fisons MD800) using a BP-X5 column (25 m by 0.25 mm; film thickness, 0.25 μm). The oven temperature was held at 45°C for 1 min, increased at 15°C/min to 140°C, increased at 4°C/min to 300°C, and held for 10 min.

Partial acid hydrolysis. A solution of polysaccharide (20 mg) in D_2O (0.6 ml) was hydrolyzed by the addition of 20 μl of 8 M TFA, followed by heating at 80°C in a nuclear magnetic resonance (NMR) tube. The hydrolysis was monitored by observation of the anomeric proton NMR peaks. After 2 h, the pH was adjusted to 7.0 with sodium hydroxide. Oligosaccha-

ride fractions were eluted from a Biogel P2 column (50 by 1 cm) with water (0.15 ml/min) by refractive index detection.

NMR spectroscopy. Spectra were obtained at 30°C in D_2O solution by using a Varian XL-400 spectrometer. Proton chemical shifts are expressed relative to internal acetone at 2.230 ppm, and ^{13}C chemical shifts are expressed relative to internal acetone at 31.45 ppm. One-dimensional (1D) proton spectra were recorded by using volume-selective 90°C pulses (8) to minimize the water hump. Proton-proton correlation spectra (COSY) and proton-carbon correlation spectra (HETCOR) were obtained by using standard Varian software, which permits both removal and retention of proton-proton splitting in the heteronuclear correlation spectrum (7, 9, 57). Modification of the pulse amplifier power supply to reduce the transmitter power enabled the observation of total correlation spectra (TOCSY) by using MLEV-17 spin-locking (6.3-kHz field strength) and 2.5-ms trim pulses (8). One-dimensional TOCSY spectra were observed by replacing the initial pulse with a DANTE sequence.

Assay for transferases. Transferases were assayed essentially as described by Liu and Reeves (37), with 30 μl of membrane fraction in a total reaction mixture of 100 μl. The reactions involved transfer of radioactive sugar from UDP-[U-^{14}C]GlcNAc to the added decaprenylphosphate (DP) or endogenous undecaprenol or from dTDP-[U-^{14}C]rhamnose to the product of that reaction using nonradioactive UDP-GlcNAc, and transfer was assayed by extraction of radioactivity into chloroform-methanol (2:1).

SDS-PAGE of LPS. LPS preparations from whole cell membranes, sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), and silver staining were carried out as previously described by Brown et al. (13).

Construction of plasmid pPR1538. The 4.1-kb *Eco*RI-*Hin*dIII insert from pPR1168 (originally from Kohara phage 352 [34]) was cloned into plasmid pPR637, which provided a different replicon and antibiotic selection, and the resulting plasmid was named pPR1555. This 4.1-kb *Eco*RI-*Hin*dIII fragment carries the *rfbA*, *rfbC*, and *rfbD* genes and the 3′ end of the *rfbB* gene (see Fig. 4). The 8-kb *Eco*RI insert from pPR1461 (originally from K-12 strain WG1) was then cloned into the *Eco*RI site of pPR1555, and the resulting plasmid was named pPR1538. The *rfb* gene clusters in strain W3110 (Kohara phage 352) and strain WG1 carried the *rfb-50* and *rfb-51* mutations, respectively (37). The 12.1-kb insert in pPR1538 combines wild-type DNA from the two sources to reconstruct the presumptive original K-12 sequence for the *rfb* region from the 3′ end of *rfbB* (the first *rfb* gene) through the end of the *rfb* gene cluster to the *gnd* gene and thus has an *rfb* cluster with only *rfbB* not present.

Serology. The O and H antigens of the isolates were determined by standard serological methods (49). Briefly, O antigens were prepared by heating overnight broth cultures for 1 h on a boiling water bath, and they were then identified by agglutination in 171 standard O antisera. The H antigens were prepared by adding formalin (0.3%, vol/vol) to motile broth cultures, and then they were identified by agglutination in 56 standard H antisera. Tests for the presence of polysaccharide capsular K antigen were performed by the cetavlon method

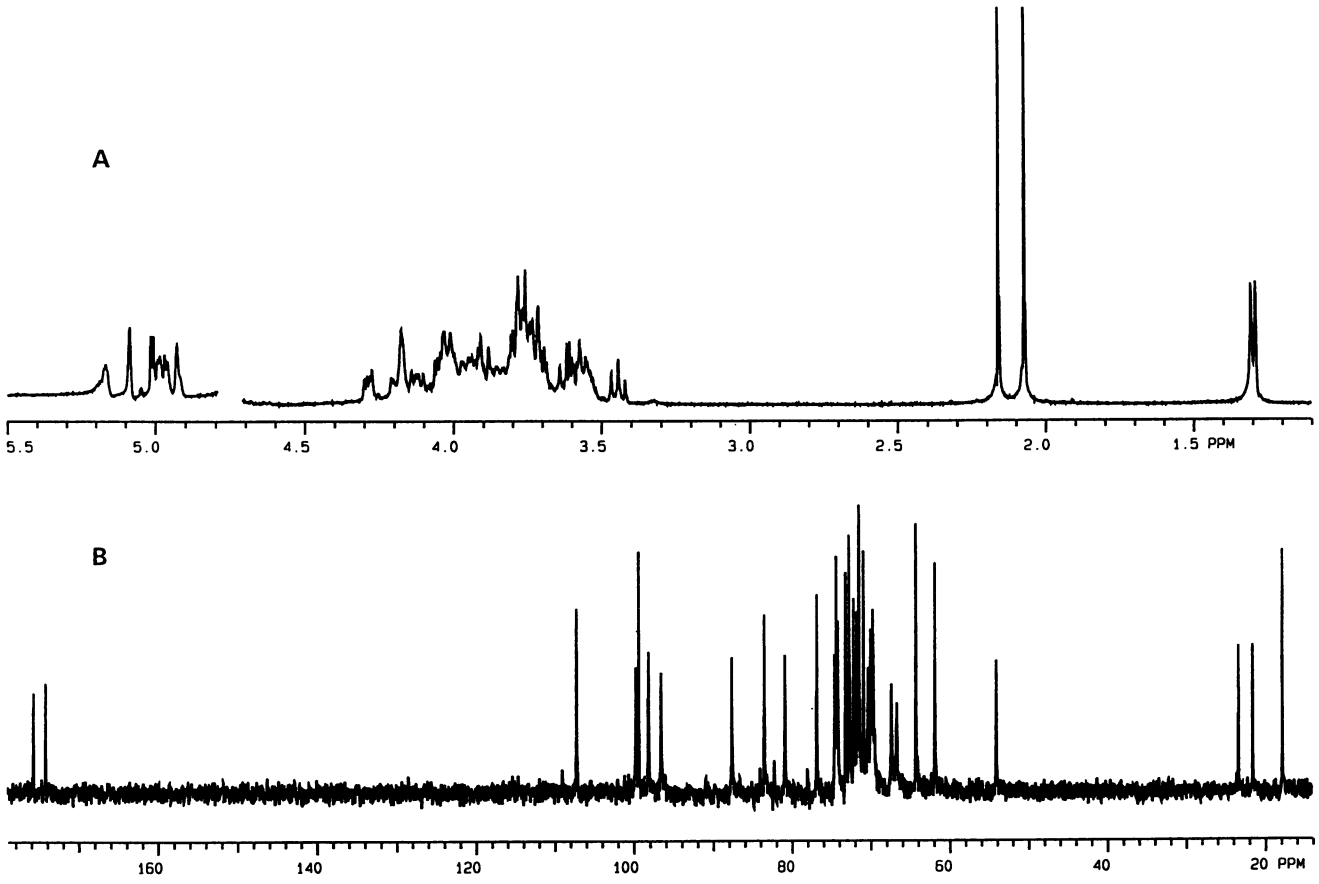

FIG. 1. ^1H NMR spectrum (A) and ^{13}C spectrum (B) of the O antigen from *E. coli* K-12 LPS.

(48) and by countercurrent immunoelectrophoresis against 74 standard K antisera (49).

DNA methods. The preparations of chromosomal and plasmid DNA were as described by Bastin et al. (6), and Southern hybridization was as described by Sambrook et al. (59). Sequencing was performed by the chain termination method (60). Plasmids were sequenced by using the universal priming site (or the reverse universal priming site). Exonuclease III deletion families of plasmids were constructed by using a Pharmacia Erase-a-base kit. Double-stranded cycle sequence reactions were performed with a Perkin-Elmer Cetus DNA Thermal Cycler according to an Applied Biosystems (AB) protocol; for automated sequencing, an AB protocol using dye-labelled primer was used, and the sequence was determined with an AB 370A Sequencer. PCR (58) was employed to amplify DNA fragments.

Computer analysis. Computer programs were accessed through the Australian Genomic Information Service (54). Sequence data were assembled into contiguous sequences with the XDAP program (25). Sequence data bases were searched by using the NCBI BLAST network server (1). Pairwise sequence comparisons were performed with the COMPARE option of the Genetics Computer Group suite of programs (23).

Nucleotide sequence accession number. The sequence of the *rfb* region of *E. coli* K-12 has been submitted to GenBank, together with details of open reading frames. The accession number is U09876.

RESULTS AND DISCUSSION

The structure of the O antigen of *E. coli* K-12. O antigen was extracted from cells of strain P4863 as described in Materials and Methods. The polysaccharide consisted of glucose, *N*-acetylglucosamine, galactose, and rhamnose in the ratio 1.8:1.0:0.7:0.6 as determined by hydrolysis followed by Dionex chromatography and 2.2:1.0:0.5:0.8 as determined by gas chromatographic-mass spectrometric analysis of the alditol acetates. The proton NMR spectrum (Fig. 1) contained five distinct anomeric resonances with integrated areas of 1.00, 1.07, 0.93, 0.90, and 1.00 (relative to the area of the methyl resonance of rhamnose taken as 3.00). The composition of the O antigen repeat unit (O-unit) is, therefore, glucose, *N*-acetylglucosamine, galactose, and rhamnose in the ratio 2:1:1:1. While some destruction of the rhamnose during hydrolysis was expected, the low figure for galactose content was unexpected. No destruction of galactose was observed when monosaccharide mixtures were subjected to the same treatment and analysis, and we conclude that the loss of galactose is a consequence of its substitution at the 2 position. The NMR spectrum eliminated the possibility that the low value for galactose content was due to heterogeneity of the polysaccharide.

The L configuration was assigned to rhamnose on the basis of the ^{13}C NMR chemical shifts for rhamnose and glucose, following the method of Shashkov et al. (62) and assuming the D configuration for glucose. While the α effects were not within the suggested limits for either configuration, they were closer to those for the L configuration. The β effects were in excellent agreement with those expected for the L configuration and in poor agreement with those expected for the D configuration.

The chemical structure of the polysaccharide (Fig. 2) was determined by a combination of methylation analysis, NMR spectroscopy, and partial acid hydrolysis. The ^1H and ^{13}C NMR chemical shifts (Tables 2 and 3) were assigned by means of proton-proton correlation spectroscopy (COSY), relayed proton-proton correlation spectroscopy (TOCSY), and proton-carbon correlation spectroscopy. Proton-proton coupling constants were obtained from the 1D spectrum, when possible, or from a 2D J-spectrum.

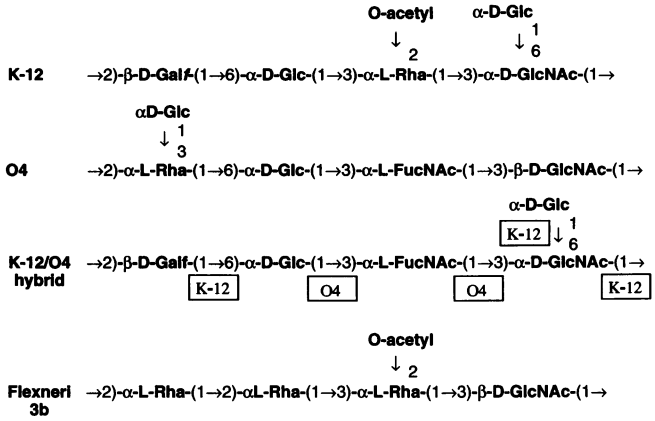

FIG. 2. Structure of the O-units of selected *E. coli* strains. The O-units of O antigens of *E. coli* K-12 (this paper), *E. coli* O4 (33), a hybrid (GH58) between *E. coli* K-12 and O4 (33), and *E. coli* Flexneri (31), as determined by NMR, are shown. The linkages in the hybrid molecule correspond to those found in strains K-12 and O4 as indicated. The Flexneri structure shown includes the basic RhaI-RhaII-RhaIII-GlcNAc main chain and the *O*-acetyl substitution on RhaIII encoded by the *oac* gene of phage Sf6, which is in effect the Flexneri 3b structure.

The sets of coupled proton resonances for each residue were completely and unambiguously identified from the COSY and TOCSY spectra, except for the H-6 protons of the glucose in the main chain. No peak that could be assigned to these protons was observed in either the COSY or TOCSY spectrum, but their proton chemical shift was eventually determined from the heteronuclear correlation spectrum (see be-

TABLE 2. ^1H NMR chemical shifts of O antigen of *E. coli* K-12

Sugar residue	Expt[a]	Chemical shift for:					
		H-1	H-2	H-3	H-4	H-5	H-6
→2-β-Gal*f*	A	5.09	4.18	4.29	4.05	3.91	3.71
	B	5.08	4.17	4.29	4.05	3.91	3.70
	C	+0.01	+0.01	0.00	0.00	0.00	+0.01
→6-α-Glc*p*	A	4.97	3.56	3.72	3.56	4.02	3.95[b]
	B	5.02	3.59	3.77	3.56	4.10	
	C	−0.05	−0.03	−0.05	0.00	−0.09	
→3-α-Rha*p*	A	4.93	5.17	3.96	3.62	4.13	1.30
	B	4.92	3.94	3.77	3.55	4.04	1.28
	C	+0.01	+1.23	+0.19	+0.07	+0.09	+0.02
→3-α-GlcNAc*p*	A	4.99	4.20	3.78	3.72	4.02	3.84
	B	4.99	4.16	3.80	3.70	4.03	3.83
	C	0.00	+0.04	−0.02	+0.02	−0.01	+0.01
α-Glc*p*	A	5.02	3.59	3.78	3.45	3.78	3.90
	B	5.02	3.59	3.78	3.45	3.78	3.90
	C	0.00	0.00	0.00	0.00	0.00	0.00

[a] A, O antigen in D_2O; B, de-O-acetylated O antigen in D_2O; C, differences between the chemical shifts for A and B.

[b] The chemical shift for H-6 of this residue was determined by using the ^1H-^{13}C heteronuclear correlation spectrum.

TABLE 3. ^{13}C NMR chemical shifts of O antigen of *E. coli* K-12

Sugar residue	Expt[a]	Chemical shift for:					
		C-1	C-2	C-3	C-4	C-5	C-6
→2-β-Gal*f*	A	107.2	87.5	76.8	83.4	71.4	64.2
	B	107.2	87.6	76.7	83.4	71.5	64.2
	C	0.0	−0.1	+0.1	0.0	−0.1	0.0
→6-α-Glc*p*	A	98.1	72.1	74.1	70.2	71.4	66.7
	B	98.1	72.5	74.1	70.4	71.5	66.8
	C	0.0	−0.4	0.0	−0.2	−0.1	−0.1
→3-α-Rha*p*	A	99.7	69.7	74.5	71.7	69.9	17.7
	B	102.0	69.2	78.2	71.5	70.1	17.8
	C	−2.3	+0.5	−3.7	+0.2	−0.2	−0.1
→3-α-GlcNAc*p*	A	96.5	54.0	80.8	69.6	72.6	67.4
	B	98.0	54.1	80.5	69.6	72.6	67.2
	C	−1.5	−0.1	+0.3	0.0	0.0	−0.2
α-Glc*p*	A	99.3	72.7	74.3	70.8	73.1	61.8
	B	99.3	72.7	74.3	70.8	73.1	61.8
	C	0.0	0.0	0.0	0.0	0.0	0.0

[a] A, O antigen; B, de-O-acetylated O antigen; C, differences between the chemical shifts for A and B.

low). The H-3 and H-5 signals of the branching glucose residue have the same chemical shift. This was confirmed by decoupling H-3/H-5 and observing the H-4 signal, which collapsed to a singlet. The chemical shifts of H-2 and H-4 of the branching glucose residue, which are identical at 30°C, differ by 0.02 ppm at 80°C.

Most of the ^{13}C resonances were assigned unambiguously from the heteronuclear correlation spectrum. The only significant exceptions were the glucose and glucosamine C-6 signals. The ^{13}C spectrum in the region between 60 and 70 ppm showed clearly that there were two 6-substituted sugars (rather broad peaks at 67.4 and 66.7 ppm) and two sugars not substituted at the 6 position (64.2 and 61.8 ppm), but determining which residues were 6 substituted was more difficult. The heteronuclear correlation spectrum proved that the peak at 64.2 ppm belonged to the galactose residue, but the chemical shifts of the remaining H-6 protons were rather similar. The difficulty was caused, in part, by heteronuclear virtual coupling (43, 44). When the heteronuclear correlation spectrum was repeated by using the older procedure in which proton-proton couplings in the f_1 domain are retained (9), it became apparent that virtual coupling between C-6 and H-5 of the branching glucose resulted in cross peaks between the C-6 resonance and both the H-5 and H-6 resonances. In the original spectrum, the BIRD sequence used to remove proton-proton coupling resulted in a diffuse cross peak covering the whole region between H-5 and H-6. This interpretation of the spectra was verified by simulation of the virtual coupling effects by using the ANTIOPE program (66) generously supplied by J. S. Waugh.

Three other ^{13}C assignments that might have been ambiguous because of overlapping proton resonances (C-2 and C-4 of the main-chain glucose, C-3 and C-5 of the branching glucose, and the C-5 signals of glucosamine and the branching glucose) were easily resolved by reference to the spectra of standard compounds (10). The proton and carbon chemical shifts obtained from these correlation spectra were also in very good agreement with predictions based on known compounds (10) and with the results reported for a very similar O antigen (33).

The anomeric configuration for each residue was deduced from the chemical shifts of the C-1 and H-1 signals (10). Proton-proton coupling constants were also consistent with the assigned anomeric configurations of all the pyranose rings. Furthermore, the coupling constants indicated the axial or equatorial orientations of the other ring protons, thus confirming that rhamnose, glucose, and glucosamine were present.

The linkage position on each sugar was deduced from the results of methylation analysis and, independently, from the ^{13}C chemical shift information. The methylation product contained a 1,3,5-tri-*O*-acetyl-(1-deuterio)-2,4-di-*O*-methyl-6-deoxyhexitol; a 1,3,5-tri-*O*-acetyl-(1-deuterio)-(2-*N*-methylacetamido)-4,6-di-*O*-methylhexitol; a 1,5,6-tri-*O*-acetyl-(1-deuterio)-2,3,4-tri-*O*-methylhexitol; and a 1,5-di-*O*-acetyl-(1-deuterio)-2,3,4,6-tetra-*O*-methylhexitol, identified by their mass spectrometric fragmentation pattern and chromatographic behavior (28). There were no standards available for the galactofuranoside, but a fifth chromatographic peak was observed, for which the fragmentation pattern (*m/e* = 306, 205, 190, 130, 129, 101, 89, and 88) was as expected for 1,2,4-tri-*O*-acetyl-(1-deuterio)-3,5,6-tri-*O*-methylhexitol.

The location of the *O*-acetyl substituent, whose presence was indicated by a proton singlet at 2.14 ppm and ^{13}C resonances at 22.7 and 173.5 ppm, was deduced from the chemical shift for H-2 of rhamnose and confirmed by de-O acetylation with dilute sodium hydroxide. A 20-mg sample of polysaccharide in 600 μl of D_2O was hydrolyzed by the addition of 20 μl of 1 M sodium hydroxide solution and heating at 100°C for 10 min followed by adjustment of the pH to 6.2. The resulting changes in ^1H and ^{13}C chemical shifts (Tables 2 and 3) confirmed that the acetylation was at O-2 of rhamnose, and smaller changes for carbons in the main-chain glucose residue and the glucosamine strongly suggested that these residues were adjacent to rhamnose.

The sequence of the sugar residues in the chain was determined by NMR spectroscopy and partial acid hydrolysis. Two-dimensional NOESY spectra with mixing times of 300 and 500 ms were used to locate close approaches between protons. The region of the NOESY spectrum containing interactions involving anomeric protons is shown in Fig. 3 and summarized in Table 4. The observed close approaches are consistent with the proposed structure, as were the fragments obtained by partial hydrolysis with dilute TFA. On gel chromatography, six oligosaccharide fractions were obtained, as well as a fraction containing the monosaccharides rhamnose and glucose. Two important fractions were examined by NMR spectroscopy. The proton NMR spectrum of the first selected fraction showed that it consisted almost entirely of α-Glc*p*-Rha. Anomeric proton signals of both α- and β-rhamnose were observed (5.16 ppm, J = 2.1 Hz; 4.87 ppm, J = 1.0 Hz), but only α-glucose H-1 peaks were present (5.11 ppm, J = 3.9 Hz; 5.09 ppm, J = 3.9 Hz). The constitutional analysis showed glucose and rhamnose in the ratio 1.0:0.7. This confirmed the relative positions of glucose and rhamnose in the O antigen.

The other selected fraction contained glucose, *N*-acetylglucosamine, and galactose in the ratio 1.0:0.9:0.4, but the proton NMR spectrum showed that the material was α-Glc*p*-α-GlcNAc*p*-Gal and that galactose was the only reducing terminal residue. (Again, the low value for galactose in the constitutional analysis is due to preferential destruction.) The ^{13}C spectrum confirmed that the trisaccharide contained a 1→6 glycosidic linkage.

Defining the end of the *rfb* gene cluster. In our earlier study (37), we had located the *rfbB* gene and mutations *rfb-50* and *rfb-51*. One end of the *rfb* cluster is defined by the *gnd* gene immediately downstream of the gene with the *rfb-50* mutation,

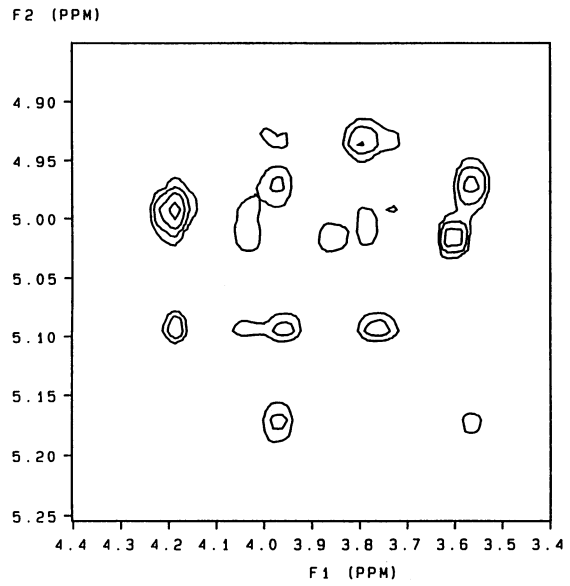

FIG. 3. Partial ^{1}H NOESY spectrum of O antigen from *E. coli* K-12 LPS showing interactions between anomeric and nonanomeric protons.

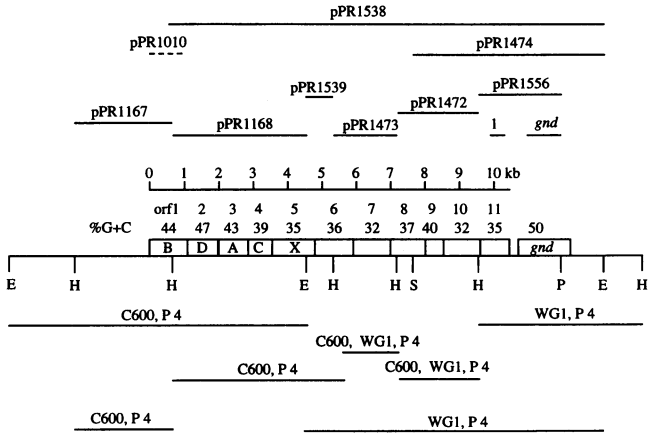

FIG. 4. Map of the original K-12 *rfb* region. Boxes indicate ORFs or genes as shown by sequencing. The 11 *rfb* ORFs are numbered above the boxes, and for those which have been identified, the letter in the box indicates the *rfb* gene involved. Restriction sites used in cloning are shown below the boxes. A scale is shown above the gene symbols. Plasmids and DNA fragments used in this study are indicated at the top. Fragments 1 and *gnd* (37) were made by PCR. Below the restriction sites are shown the fragments detected by Southern hybridization in DNA from strains C600 and WG1 and P 4 as indicated (see text). E, *Eco*RI; H, *Hin*dIII; P, *Pst*I; S, *Sph*I.

but the other end was not defined in that study. *rfbB* is the first gene in the *E. coli* Dysenteriae *rfb* cluster (32), and we looked first to see whether that was also the case for K-12. Plasmid pPR1538 (Fig. 4) was constructed by using fragments from strain WG1, which carries the *rfb-51* mutation, and strain W3110, which carries the *rfb-50* mutation, in such a way as to reconstruct the original operon from *rfbD* to its downstream end. This plasmid was transformed into the *rfb* deletion strain SØ874 together with pPR1010, which carries the *rfbB* gene of *S. enterica* LT2. This combination of plasmids was sufficient to confer polymerized O antigen synthesis, as shown by SDS-PAGE (data not shown). We conclude that *rfbB* is the first gene of the cluster, as a substantial region upstream of *rfbB* is missing in this strain.

***rfe* dependence and GlcNAc transferase.** The enterobacterial common antigen (ECA), thought to be present in all members of the family *Enterobacteriaceae*, has *N*-acetyl-D-glucosamine (GlcNAc) as a constituent sugar (38). The *rfe* gene in the ECA gene cluster is thought to encode the UDP-GlcNAc:undecaprenylphosphate GlcNAc-1-phosphate (GlcNAc-P) transferase which transfers GlcNAc-P from UDP-GlcNAc to undecaprenylphosphate (UndP) to give GlcNAc-pyrophosphorylundecaprenol (GlcNAc-PP-Und) and thereby initiate synthesis of the ECA repeat unit (42). The synthesis of several O antigens has been shown to require the presence of a functional *rfe* gene, and many of these now appear to fall into two groups. The first group comprises the O antigens of *E. coli* O8 and O9, which contain only mannose (40), and that of *Klebsiella pneumoniae* O1 (20), which contains only galactose, and the atypical T1 antigen, produced as an alternative O antigen by some mutants of *S. enterica*, which contains polygalactose and polyribose components (40). The second group contains O antigens, such as those of *S. enterica* group C1 and *E. coli* Dysenteriae 1, with a typical multisugar O-unit which includes GlcNAc. The role of Rfe in the first group is not understood, as the repeat units do not contain GlcNAc (see reference 53 for a review). With regard to the second group, it has been shown indirectly for *E. coli* Dysenteriae 1, by SDS-PAGE analysis of the LPS produced by mutants, that GlcNAc is the first sugar in the biological repeat unit of the O antigen and is added to UndP by Rfe to initiate O-unit synthesis (32). As the O antigen of *E. coli* K-12 contains GlcNAc, we investigated the possibility that its synthesis is *rfe* dependent. To strain AN21548, which is an *rfe* mutant, was added one or both of plasmids pRL100, which carries the *rfe* gene, and pPR1538, which carries a reconstruction of the original *rfb* gene cluster apart from *rfbB*. The host strain AN21548 is assumed to carry the *rfbC1* and *rfb-50*::IS*5* mutations, as it is from the same lineage as C600. Both mutations should be complemented by pPR1538 to give *rfb* function. We found that AN21548 carrying both plasmids made LPS with O antigen (Fig. 5, lane c) but that plasmid pPR1538 alone (lane b) did not confer O antigen synthesis on strain AN21548. Lanes d and e of Fig. 5 are positive controls, containing strains in which the *rfb-50* mutation is complemented by pPR1474. We conclude that synthesis of the K-12 O antigen is *rfe* dependent, with the implication that O-unit synthesis starts with addition of GlcNAc-P to UndP by Rfe. We also showed by use of [^{14}C]UDP-GlcNAc that, as expected for a strain synthesizing ECA (55), GlcNAc was transferred to lipid and this transfer required the presence of pRL100 (Table 5). We confirmed that UDP-GlcNAc was the first sugar added by showing that it was the only nucleotide

TABLE 4. Close approaches between anomeric and other protons as determined by NOESY spectroscopy

Anomeric proton	Adjacent proton(s)
Galactose	Glc I: H-5, H-6; GlcNAc: H-3[a]; Gal: H-2, H-3
Glucose I[b]	Rha: H-3; Glc I: H-2, H-3[c]
Rhamnose	GlcNAc: H-3; Rha: H-3
N-Acetylglucosamine	Gal: H-2; GlcNAc:H-2, H-3, H-5
Glucose II[d]	GlcNAc: H-5, H-6; Glc II: H-2

[a] Other possible, but less likely, assignments are Glc II: H-3 or H-5.
[b] Main-chain glucose residue.
[c] Other possible, but less likely, assignments are Gal: H-6 or GlcNAc: H-4.
[d] Branching glucose residue.

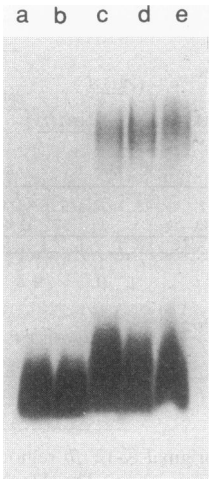

FIG. 5. Requirement of the *rfe* gene for expression of the *E. coli* K-12 O antigen. Membrane extracts were run on SDS-PAGE gels and stained by silver staining. Lanes: a, AN21548 (Δ*rfb rfe*); b, AN21548 carrying pPR1538 (*rfb* genes); c, AN21548 carrying pPR1538 (*rfb* genes) and pRL100 (*rfe* gene); d, EMG2 (*rfb-50*) carrying pPR1474 (*orf9 orf10 orf11*); e, W1485 (*rfb-50*) carrying pPR1474 (*orf9 orf10 orf11*).

sugar required to allow rhamnose to be added to the O-unit (see below).

Sequence of the *rfb* region. We sequenced the *rfb* region over an area other than the short segment previously sequenced (37). The ends of plasmids pPR1472, pPR1473, pPR1167, pPR1168, and pPR1539 were sequenced with the universal primer and reverse primer. Exonuclease III deletion families of these plasmids and also of derivatives with the inserts inverted were constructed. The universal priming site was used for sequencing selected strains from these families. Synthetic oligomers were made to sequence the region 300 bp upstream from *rfbB* to a position 400 bp downstream of the start codon of this gene. As *rfbB* appears to be the first gene of the cluster, we did not sequence upstream of this segment.

As we were nearing completion of the sequencing, we learned that M. Valvano and colleagues had sequenced much of the region in their determination of the *E. coli* K-12 genes required for complementation of the *E. coli* Flexneri (note explanation of terminology in Materials and Methods) *rfb* region to give group 6 specificity (see below). We contacted them and exchanged information. They had complete sequence from position 2780 in *rfbA* (see reference 73), and we used their sequence to fill in gaps in that part of our sequence. Analysis of the sequence revealed 11 open reading frames (ORFs) (Fig. 4) in the *rfb* gene cluster, all transcribed towards the *gnd* gene; the *rfb* gene cluster covered 10,462 bp from the start of *orf1* to the end of *orf11*. Only one discrepancy with the sequence obtained by Yao and Valvano (73) remains after discussion of all discrepancies and is given in the GenBank entry; it has no effect on gene length, and it does not affect the conclusions.

TDP-Rha biosynthetic pathway genes. The four genes required for TDP-Rha synthesis, *rfbB*, *-D*, *-A*, and *-C*, have been shown previously to be present in *E. coli* K-12 and to map in the *rfb* region (5, 67). They are easily identified as *orf1* to *orf4* by their similarity to the homologous genes of *S. enterica* (29), *E. coli* Dysenteriae (32), and *E. coli* Flexneri (50a) (GenBank entry SHFRFBAB).

The rhamnose transferase gene. If GlcNAc is the first sugar added to UndP, it becomes clear from the structure that rhamnose is the second. We suspected that *orf1* could be the rhamnose transferase, as the *rfb-50* mutation in it is very stable. It is known for *S. enterica* LT2 that mutations which can initiate but not complete O-unit synthesis are unstable, presumably because they lead to accumulation of intermediates which reduce the pool of UndP needed for other syntheses, including that of peptidoglycan (40). Mutations which block GlcNAc or Rha addition are likely to be stable, as addition of the first sugar to UndP is reversible, and it is only if two or more sugars are added that O-unit intermediates on UndP accumulate, making it unavailable for other syntheses (40). If the same applies to K-12, the only *rfb* mutations likely to be stable are those which block rhamnose addition, and since *orf11* is clearly not one of the TDP-Rha biosynthetic pathway genes, it may well be the gene for rhamnose transferase. We tested the strains used in the assay of GlcNAc transferase for their ability to add rhamnose to UndP in the presence or absence of plasmids pRL100 and pPR1556, which carry the *rfe* gene and *orf11*, respectively. We found (Table 5) that both plasmids

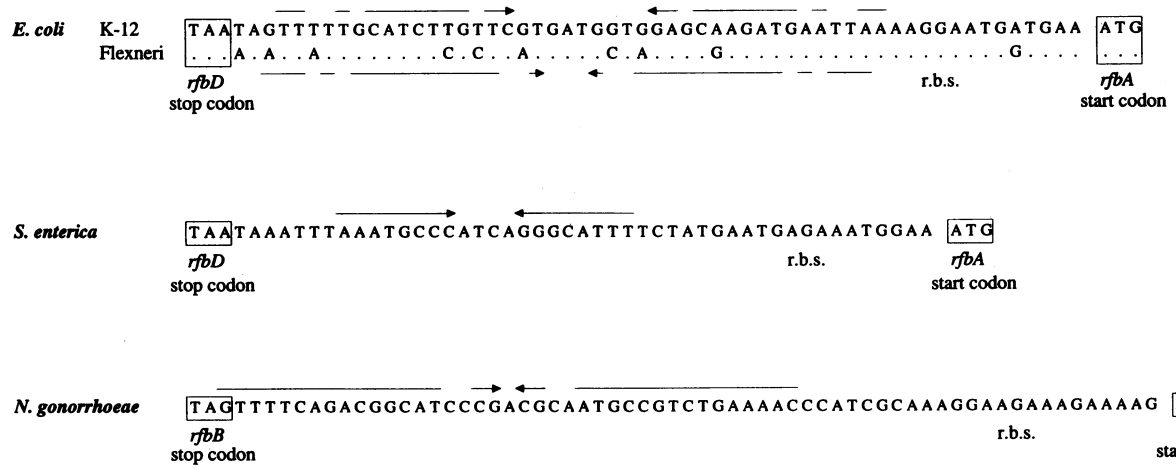

FIG. 6. Stem-loop structures found in the intergenic sequence between *rfbD* and *rfbA* in *E. coli* K-12, *E. coli* Flexneri (50a) (GenBank SHFRFBAB), and *S. enterica* LT2 (29) and between homologs of *rfbB* and *rfbA* of *N. gonorrhoeae* (55a) (GenBank NGOCPSA). Positions where the Flexneri sequence is the same as in K-12 are indicated by dots. Arrows show the regions of dyad symmetry. r.b.s., ribosomal binding site.

were required for transfer of radioactive rhamnose from TDP-L-[U-^{14}C]rhamnose. We obtained the same results with SØ874 as the host strain, showing not only that *orf11* was required but also that it is the only gene required in the *rfb* cluster. Transfer in the absence of cold UDP-GlcNAc was reduced by about 65% (data not shown), with residual transfer presumably due to the presence of GlcNAc-PP-Und acting as a substrate.

These experiments show that *orf11* is the gene for the rhamnose transferase and confirm that GlcNAc is the first sugar added, with Rha being the second. In both the GlcNAc and Rha transferase assays, the level of transfer was low compared with that observed for other transferases and there was no difference in those assays in which DP was omitted, suggesting that the GlcNAc phosphate transferase, unlike the galactose phosphate transferase, cannot utilize DP as an acceptor.

***rfbX* and *rfc*.** All *rfb* gene clusters sequenced thus far have included a gene, named *rfbX*, for which one can predict a protein product with about 12 transmembrane segments. *orf5* is the only gene which fits this pattern and is 32% identical to the *rfbX* gene of *E. coli* Dysenteriae I, and it was named accordingly. Orf7 is the only other predicted gene product with multiple potential transmembrane segments and fits the pattern expected for Rfc, and the gene is named accordingly, although we have not attempted to show directly that Orf7 is the O-antigen polymerase.

***O*-acetyl transferase.** The O antigens of most Flexneri strains of *E. coli* have the same basic RhaI-RhaII-RhaIII-GlcNAc structure but with substitutions which can include an *O*-acetyl group on the C-2 position of RhaIII (64). This *O*-acetyl group in Flexneri strains is in a similar chemical environment to that of strain K-12 (Fig. 2). The *O*-acetyl transferase in Flexneri is known to be encoded by the *oac* gene, which has been cloned and sequenced (19, 69).

Yao et al. (72) cloned the Flexneri 2a and 3a *rfb* regions into *E. coli* K-12 and found that the O antigens synthesized reacted with Flexneri group 6 antiserum, which is specific for the *O*-acetyl group. The Flexneri *O*-acetyl group is determined by the *oac* gene present on bacteriophage Sf6. This phage is not present in Flexneri 2a, which lacks the *O*-acetyl group, so the presence of group 6 activity in the Flexneri 2a clone was unexpected. Furthermore, the integration site for phage Sf6 is far from the *rfb* region, so the clones of Flexneri 3a *rfb* genes would not carry the *oac* gene. It was concluded that K-12 had an *O*-acetyl transferase of appropriate specificity which conferred group 6 specificity on Flexneri O antigen (72) and, as Flexneri clones in a K-12 strain carrying a deletion of the *his rfb* region did not confer group 6 specificity, that the gene was in or near the K-12 *rfb* gene cluster. Furthermore, a clone of the K-12 *rfb* region conferred group 6 specificity on Flexneri strains (72) which did not carry phage Sf6, supporting this suggestion. More recent work has complicated this theory (see below), but it remains true that a single *rfb* gene, which is presumably the gene responsible for the O acetylation of the K-12 O antigen, is expected for this *O*-acetyl transfer. Orf9 is a good candidate for the transferase as it is very similar to choramphenicol acetyltransferase and lactose acetyltransferase, with the segment of greatest similarity being in a recognized acetyltransferase motif.

The glucose transferase gene. Of proteins in the data base, Orf10 has the greatest similarity to RfbU of *S. enterica* group B, which encodes a mannose α(1-4) transferase, and we suggest that the Glc α(1-3) transferase gene is *orf10*.

Side chain glucose transfer. The gene for the transferase for the side chain glucose may not be in the *rfb* cluster, as there are several cases known of side chain glucose being added from Und-P-Glc after synthesis and translocation of the O-unit and, in those cases for which the genetics are known, the genes for synthesis and translocation of Und-P-Glc and for the Glc transferase map away from the *rfb* cluster (40). The conservative expectation, then, is for no Glc transferase gene in the *rfb* cluster for the side chain Glc. This is supported by the observation that recombinant *E. coli* K-12 strains with the *his rfb* region replaced by that from an O4 *E. coli* strain make O antigen with an O4 main chain but with glucose substituted 1→6 onto GlcNAc (33), as in strain K-12 (see below). This side branch glucose is not present in the O4 structure, and the simplest explanation is that the K-12 glucose transferase gene maps away from the *rfb* region and remains when the *rfb* region is replaced (see below).

TABLE 5. Assay of GlcNAc transferase and rhamnose transferase by incorporation of radioactive sugar from UDP-[U-^{14}C]GlcNAc or dTDP-[U-^{14}C]rhamnose into chloroform-methanol-extractable material

Strain	Amt of radioactively labelled sugar incorporated (cpm)[a]	
	[^{14}C]UDP-GlcNAc	[^{14}C]dTDP-Rha[b]
AN21548	86	171
AN21548/pRL100	4,258	225
AN21548/pPR1556	76	390
AN21548/pPR1556/pRL100	3,442	3,744
SØ874/pRL100	ND	73
SØ874/pRL100/pPR1556	ND	2,143

[a] ND indicates that the test was not done.
[b] UDP-GlcNAc was added for all reactions with [^{14}C]dTDP-Rha.

Synthesis and transfer of galactofuranose. We now have to identify the Gal*f* transferase and Gal*f* pathway genes. There are only two genes, *orf6* and *orf8*, available, and one of these must be the gene for the Gal*f* transferase, leaving one gene for the pathway. Gal*f* synthesis and transfer to bacterial polysaccharides have been studied with *S. enterica* and *Penicillium charlesii*. These processes are generally quite well conserved and, for speculative purposes, it is reasonable to combine the two sets of data. It has been shown convincingly that in *S. enterica*, UDP-Gal*p* is a precursor of the Gal*f* in T1 polysaccharide (61), but these in vivo experiments could not indicate how many steps were involved. It has also been shown that UDP-Gal*f* can be made from glucose in vitro with a cell extract of *P. charlesii* and that the Gal*f* can be transferred to the polysaccharide galactocarylose (68). These results are consistent with the simple hypothesis that UDP-Gal*p* is converted to UDP-Gal*f*, which acts as precursor for transfer of Gal*f* to polysaccharides. In further support of this hypothesis, an oxidoreductase which can form a 2-keto group on both UDP-Gal*p* and ADP-ribose has been found and isolated from *P. charlesii* (24, 30). The authors of those reports point out the similarity in configuration of ADP-ribose and UDP-Gal*f* and suggest that there may be a 2-keto intermediate in conversion of UDP-Gal*p* to UDP-Gal*f*. As there are only sufficient genes in the *E. coli rfb* region for one gene to be involved in synthesis of a Gal*f* precursor, we suggest that a single enzyme can convert UDP-Gal*p* to UDP-Gal*f* through a 2-keto intermediate in a manner parallel to the interconversion of UDP-Gal*p* and UDP-Glc*p* by galactose-4-epimerase. We propose that *orf6* is the gene involved, as the protein product has the Rossman fold found in many sugar pathway enzymes, including Gal epimerase, but not reported to be present in sugar transferases. The remaining gene, *orf8*, is likely to encode the

Gal*f* transferase. Although the identification of these two genes is tentative, the hypothesis can be tested by studying the biosynthesis of UDP-Gal*f* or other intermediates and transferring Gal*f* by using cloned gene products and the genetic tools available in *E. coli* K-12.

The *rfb* gene cluster of *E. coli* K-12 and its evolutionary relationships. The genes of the *rfb* cluster of K-12, like those of all other *rfb* clusters from *E. coli* and *S. enterica* sequenced thus far, all transcribe in the same direction, which is now known to be the direction preferred in relation to replication (15). The position of the rhamnose pathway genes at the start of the gene cluster and their order, *rfbB*, *-D*, *-A*, *-C*, is the same as in *E. coli* Dysenteriae (32) and in *S. enterica* group B (and related groups A, C2, D, and E) (see reference 52 for a review, but note that in that review, the identification of *rfbC* and *rfbD* was in error and the order was given as *rfbB*, *-C*, *-A*, *-D*; a correction will be published shortly). These genes are followed by *rfbX*, as in *E. coli* Dysenteriae and *S. enterica* group E (the dideoxyhexose pathway genes intervene in the related *S. enterica* groups A, B, D, and C2). Also as for *S. enterica* and *E. coli* Dysenteriae, the transferase genes and other group-specific genes, *rfbX* and, when present in the *rfb* locus, *rfc*, are grouped together.

The order of the glycosyl transferase genes, if our predictions are correct, is the inverse of that in which the gene products act, as observed for *S. enterica* groups B (and related groups) (29), E1 (70), and C2 (14) and *E. coli* Dysenteriae (32). In both *E. coli* Dysenteriae and *E. coli* K-12, the transferase which acts first (of those encoded in the *rfb* gene cluster) is the last gene of the whole cluster. This resembles the situation in *S. enterica* group B (and related groups), in which the galactose transferase gene is the last gene of the cluster. This pattern has now been observed sufficiently often to suggest that either it represents an early arrangement which has been conserved or it offers some selective advantage.

As seems to be general for *rfb* genes, the G+C content is atypical for the host species. In this case (Fig. 4), it varies considerably within the cluster but is lower for all genes than the 0.5% characteristic of *E. coli*, suggesting that the cluster was assembled from more than one source before lateral transfer to *E. coli*.

The rhamnose pathway genes in Flexneri (50a) (GenBank entry SHFRFBAB) and, in part, in Dysenteriae 1 have been sequenced. The first three, *rfbB*, *-D*, and *-A*, show considerable similarity to those of K-12, but there is much greater difference for *rfbC*. For Dysenteriae, the remaining four genes of the *rfb* cluster have also been sequenced and show very little similarity to those of K-12 (only 32% amino acid identity for *rfbX*). A short segment of Flexneri sequence downstream of *rfbC* shows no obvious similarity to other *rfb* genes, and most other Flexneri genes must differ substantially from those of the other two, as there are considerable differences in the transferases needed. This situation is similar to that found with *S. enterica* groups B, D, C2, and E1, in which the rhamnose and mannose pathway genes are in general extremely similar, whereas other genes may show no detectable sequence similarity at all. We have suggested that the common gene order of these *S. enterica* *rfb* clusters represents the gene order of an ancestral cluster, the extreme variation in *rfbX* and the transferases indicating the very long time since divergence. The high level of similarity in the flanking rhamnose and (in general) mannose pathway genes, with about 1% nucleotide base differences, is attributed to a form of concerted evolution whereby there is sufficient genetic exchange in these regions to allow them to undergo random genetic drift together.

For the *E. coli* groups under consideration, the first three rhamnose pathway genes of Flexneri and K-12 differ by about 15% at nucleotide level and about 5.5% at amino acid level, equivalent to the general level of divergence between genes of species such as *S. enterica* and the *E. coli*. Segments of sequence available for these Dysenteriae genes have a similar level of divergence from those of both K-12 and Flexneri. Clearly, the rhamnose pathway genes of these related *E. coli* *rfb* gene clusters have been diverging for much longer than their *S. enterica* counterparts, indicating that the level of genetic exchange between them has been much less than in the *S. enterica* strains studied, and indeed the divergence is now at a level which will dramatically reduce recombination between them (51). The much greater difference observed for *rfbC* (64% amino acid identity) for K-12 and Flexneri, when data are available, parallels the situation in *S. enterica*, where the *rfbC* genes of groups B, D, and C2 are very similar but that of group E1 has diverged substantially. We suggest that the same explanation applies and that in both species, random genetic drift in *rfbC* is reduced by the proximity of the central region, for which variation is presumably maintained as a balanced polymorphism giving different forms of O antigen. If drift takes sequence divergence of *rfbC* beyond the level at which recombination readily occurs, then *rfbC* is no longer subject to counterbalancing concerted evolution, as recombination cannot occur within the gene, and each form of *rfbC* is linked to a different (maintained) form of the central region. As for *S. enterica*, the very high level of presumably neutral difference in *rfbC* indicates a very long period of divergence since the concerted evolution ceased.

Inverted repeats upstream of the *rfbA* gene and the Y function. In *E. coli* K-12, *E. coli* Flexneri, and *S. enterica* LT2, there is a stem-loop structure in the intergenic region between *rfbD* and *rfbA* which is conserved despite considerable changes in sequence (Fig. 6). These structures may represent the regulatory site called Y which was located in *S. enterica* in this general vicinity by Nikaido et al. (47), who observed that deletions which extend into this region showed derepression of an upstream gene, then known as *rfbL* but now shown to be *cpsG*, a gene of the capsule GDP-fucose pathway (65). The ends of the deletions which defined the Y site were later approximately located by Brahmbhatt et al. (11), and knowledge of the sequence (29) now allows us to locate the site somewhere between a *Sph*I site at position 1676 and a *Kpn*I site at position 4488. The only intergenic region in this segment is that between *rfbD* and *rfbA*, and it is tempting to suggest that the inverted repeats confer the Y function in LT2 and that the same function is present in *E. coli* K-12 and Flexneri.

We thus have indirect evidence which allows us to speculate that in *S. enterica* LT2, the stem-loop structure between *rfbD* and *rfbA* functions in repression of a gene(s) in the *cps* locus. We have little knowledge of any regulatory role it may play in *rfb* expression, as the deletion mutants involved had only *rfbB* left intact of the genes being assayed. With regard to *E. coli* K-12, we know that a comparable *cps* cluster is present (2), and it is quite possible that the Y function operates in K-12, although much more work will be needed to establish the details of any interactions between the *cps* and *rfb* gene clusters of *E. coli* and *S. enterica*.

The inverted repeat is also found upstream of an *rfbA* homolog within the capsule (*cps*) gene clusters of *Neisseria meningitidis* (26) and *Neisseria gonorrhoeae* (55a), where it lies in an intergenic segment between homologs of *rfbB* and *rfbA* (Fig. 6).

Isogenes of *rfbA* and *rfbB* in the *E. coli rff* gene cluster. TDP-4-keto-6-deoxy-glucose (TDP-KDG) is an intermediate in the synthesis of both TDP-rhamnose and TDP-fucosamine, required for *S. enterica* group B O antigen synthesis and for

ECA synthesis, respectively. The enzymes for TDP-KDG synthesis are thymidylyl transferase and dTDP-glucose oxidoreductase encoded by *rfbA* and *rfbB*, respectively, in the *rfb* cluster. We searched the sequence of the K-12 *rff* (ECA) locus (22) and found that ORFs o292 and o355 have 63 and 73% amino acid identity, respectively, to RfbA and RfbB of K-12, with one or two short internal insertion or deletion events. The two genes are adjacent, and we suggest that they encode a thymidylyl transferase and dTTP-glucose oxidoreductase, respectively, although o355 was tentatively identified as *rffE*, which encodes UDP-GlcNAc epimerase (22). It appears that *E. coli* K-12 has two sets of genes for synthesis of TDP-KDG, one in the *rfb* gene cluster and the other in the *rff* gene cluster, although it remains to be confirmed that the genes in the latter cluster are functional.

Nikaido et al. (47) observed in LT2 a second dTDP-glucose oxidoreductase additional to that encoded by *rfbB*. We suggest that it is encoded by the *S. enterica* homolog of o355. It should be noted that although ECA synthesis in *S. enterica* LT2 is sometimes assumed to include a requirement for *rfbB* and *rfbA* (for example, see reference 42), the evidence is for a requirement for *rfbA* and mute on a requirement for *rfbB*. This His-399 strain with *rfbA* but not *rfbB* deleted lacks ECA (39), and a mutant with a point mutation in *rfbA* also lacked ECA (36). However, there have to our knowledge been no studies of ECA synthesis in strains carrying mutations in *rfbB* but not *rfbA*. If we are correct in our identification of the second dTDP-glucose oxidoreductase, then clearly only *rfbB* is required for synthesis of ECA in LT2.

Synthesis of this second dTDP-glucose oxidoreductase is derepressed in LT2 in strains with deletions extending beyond *cpsB*, which define a site, X (47), in a region now known to be in the *cps* gene cluster (65). There appears to be a complex set of interactions between the *rfb*, *cps*, and *rff* gene clusters.

***E. coli* K-12 has an O16 O antigen which cross-reacts with O17.** The O and H antigens of P4863 were kindly determined by Uffe Skov Sørenson as described in Materials and Methods and found to be O16 and H48. There was, as expected, no K antigen detected. P4863 also reacted with O17 antiserum. The combination O16:H48 is quite common, but the combinations O16:O17 and O17:H48 are not present in the data base at the International Escherichia and Klebsiella Centre, and in particular, the cross-reaction with O17 antiserum exhibited by *E. coli* K-12 is not exhibited by the type O16 strain, P 4 (50). Presumably, there is a group on the O antigen of *E. coli* K-12 which is not present on the O antigen of strain P 4 but gives cross-reaction with the O17 antigen of strain M493 (this is strain K 12a in the work of Ørskov et al. [50], but we use our laboratory name to avoid confusion, as the strain is not related to K-12). We compared the restriction maps of two O16 strains and the O17 strain by Southern hybridization with six probes (pPR1167, pPR1168, pPR1473, pPR1474, 1 and gnd [Fig. 4]) which covered most of the *rfb* gene cluster and some flanking DNA at each end. The probes all hybridized to P 4 and gave fragments of the size expected from the *E. coli* K-12 sequence for internal fragments. *E. coli* K-12 strains WG1 and C600 were included to represent the two mutant forms known for K-12. Each of the fragments seen in P 4, including those extending into flanking DNA, was seen in one or both of the K-12 strains, missing or replaced only when expected from our knowledge of the *rfb-50* and *rfb-51* mutations carried by C600 and WG1, respectively (Fig. 4). We conclude that the restriction maps of strain P 4 and the parental K-12 strain were indistinguishable and therefore that the O antigens of K-12 and P 4 are the same in their major features and, in particular, must have the same main chain, as any differences sufficient to change that situation would be evident on Southern hybridization. The failure of pPR1167 to probe strain WG1 indicates that the *rfb-51* deletion extends through *orf2.8* and *orf1.3*.

Only the two probes carrying flanking DNA, pPR1167 and the *gnd* gene probe, hybridized to M493, the O17 strain (data not shown). Clearly, the O16 and O17 strains do not have closely related *rfb* gene clusters and may well have very different O-antigen structures. The most probable explanation for the serological cross-reaction is that either the *O*-acetyl group or the branch glucose residue is absent in the standard O16 strain but that a similar residue is present in an otherwise rather different O17 O antigen. The branch glucose is the more likely candidate for the cross-reaction, as it is not encoded in the *rfb* region, which appears identical in P 4 and K-12. We propose to examine the O antigens of the strains concerned to determine the basis for the cross-reaction.

Interaction of the *rfb* genes of K-12 and Flexneri. We discussed above the observations of Yao et al. (72), which suggested that the K-12 *rfb* region included a gene for an *O*-acetyl transferase able to acetylate the basic Flexneri O antigen to confer group 6 specificity. Yao and Valvano have now reported (73) that *orf6*, *orf8*, *orf9*, and *orf10* are all required to express group 6 antigenic specificity. Now that we know the structure of the K-12 O antigen and the nature of its *rfb* mutations, it is apparent why four genes are needed. The strain (W3110) used by Yao et al. (72) to make the clones is from a lineage which carries the *rfb-50*::IS*5* mutation. This mutation in the rhamnose transferase gene should be complemented by the Flexneri clones, as Flexneri also has Rha α(1-3) linked to GlcNAc (Fig. 2). In this regard, it should be noted that although the GlcNAc residue in K-12 has an α linkage and that in Flexneri has a β linkage in the complete O antigen, these linkages are formed during the polymerization of O-units by the O-antigen polymerases of the respective strains. Given the similarity of the O antigen of Flexneri to those of K-12 and Dysenteriae 1, we have assumed that its synthesis is *rfe* dependent and that it starts O-unit synthesis with GlcNAc. The rhamnose is added to GlcNAc while it is still on UndPP and presumably α-D-GlcNAc-PP-Und in both cases. The change to the β anomer for the GlcNAc of K-12 would be determined later by the O-antigen polymerase (63).

The Flexneri clones carrying the *rfb* region from K-12 strain W3110 would have all the genes required for synthesis of both Flexneri and K-12 O antigens, as do K-12 strains carrying the *rfb* region of Flexneri. As all genes of the K-12 *rfb* cluster other than the O-antigen polymerase gene, *rfc* or *orf7*, are required for the expression of the group 6 specificity in Flexneri strains, it is most likely that, in these strains, expression of group 6 specificity requires synthesis, but not polymerization, of the K-12 O antigen, indicating that the cross-reaction observed is due to presence of the K-12 antigen. (This has been confirmed by Yao and Valvano [73].) It appears that the K-12 *O*-acetyl transferase is specific for the K-12 O antigen but that group 6 antiserum cross-reacts with the *O*-acetyl-Rha of the K-12 antigen. The serological cross-reaction is not surprising, as the *O*-acetyl groups of both K-12 and Flexneri are on position 2 of a rhamnose α(1→3) linked to GlcNAc. The fact that the K-12 enzyme encoded by *orf9* cannot acetylate the Flexneri O-unit suggests that acetylation occurs after addition of the next sugar residue, as it is only then that the two Rha residues differ in their local environment.

Interaction of the *E. coli* K-12 and O4 *rfb* genes. Kogan et al. (33) have shown that *E. coli* K-12 strains carrying the *E. coli* O4 *rfb* gene cluster produce a modified O4 antigen. In some cases, there is presence of side chain glucose 1→6 linked to the GlcNAc of the O4 antigen in addition to or in place of glucose

1→6 linked to rhamnose, and in strain GH58, there is a major modification to give the structure shown in Fig. 2. We can now see that all linkages present in the new structure of strain GH58, but absent in the O4 structure, are found in the K-12 O antigen: all of the new O antigen structures are in effect hybrids between the O4 and K-12 structures.

In strain HB1019, in which the *rfb* region of strain K-12 is replaced by that of the O4 strain SH1 (27), half of the O-units carry a glucose residue on GlcNAc in place of that on rhamnose; this shows both that the O4 Glc transferase gene is in the *rfb* region and that the K-12 Glc transferase gene is not. Strain GH148 has the O4 *rfb* clone pGH133 (27) in strain K-12 and has 90% of the glucose substitutions on GlcNAc; there is no obvious reason why the K-12 Glc transferase is relatively more effective in this case in which the O4 gene is on a multicopy plasmid.

GH58, which produces O antigen with the hybrid main chain, also has the O4 *rfb* region cloned into K-12, but the plasmid, pGH58 (27), has much less O4 DNA than pGH133. We can explain the observations if we assume that GH58 lacks the O4 rhamnose transferase and so cannot complete the main chain of an O4 O-unit. We suggest that in GH148, the O4 transferases, being encoded on a multicopy plasmid, dominate and lead to transfer of Rha only to Glc, followed by polymerization of the O4 O-unit, whereas in GH58, in the presumed absence of the O4 rhamnose transferase, the K-12 Gal*f* transferase acts to transfer Gal*f* to the α-D-Glc*p*-(1→3)-α-L-FucNAc-(1→3)-α-D-GlcNAc-PP-Und substrate which, with α-L-FucNAc in place of α-L-Rha, offers a substrate very similar to the K-12 substrate. Once Gal*f* is added, then only the K-12 polymerase is likely to recognize the hybrid O-unit, as it is quite similar to its authentic substrate, to give the structure observed.

General conclusions. In this paper, we present the structure of the *E. coli* K-12 O16 O antigen and the sequence of the *rfb* genes which encode it. The gene cluster shows segments of considerable similarity to those of Dysenteriae and Flexneri, which we treat as clones of *E. coli* and not as separate species of the genus *Shigella*. The pattern of relationships between these *rfb* clusters in *E. coli* is similar to that previously observed for *S. enterica* groups B, D, C2, and E. Analysis of the sequences and comparison with other sequences also revealed a conserved potential regulatory region within the *E. coli* and *S. enterica* clusters.

Like Dysenteriae, and perhaps Flexneri, O16 O antigen synthesis is *rfe* dependent because synthesis of O-unit starts with GlcNAc, which is transferred to UndP by Rfe. *E. coli* K-12 has been used as a host for cloning several *rfb* regions, and at least for *E. coli* O4 and (*Shigella*) Flexneri, the O antigen produced by the clones differs from that of the source strain. Now that we know the structure of the K-12 O antigen and something of its genetics, both interactions are readily explicable.

ACKNOWLEDGMENTS

We thank Uffe B. Skov Sørensen, International *Escherichia* and *Klebsiella* Centre, Statens Seruminstitut, Copenhagen, Denmark, for carrying out the serological work; B. Bachmann, Y. Kohara, and P. Rick for sending strains; and J. S. Waugh for supplying the ANTIOPE program. We are grateful to Miguel Valvano for sharing sequence data with us prior to publication.

The work at the University of Sydney was supported by the Australian Research Council, that done at Macquarie University was supported by a Macquarie University research grant, and that done at the Karolinska Institute was supported by the Swedish Medical Research Council (grant 16X-656).

REFERENCES

1. **Altschul, A. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
2. **Aoyama, K. M., A. M. Haase, and P. R. Reeves.** Fucose pathway genes of *Escherichia coli* K-12 *cps* gene cluster: comparison with *Salmonella enterica* LT2 and evidence for lateral transfer. Mol. Biol. Evol., in press.
3. **Appleyard, R. K.** 1953. Segregation of new lysogenic types during growth of a doubly lysogenic strain derived from *Escherichia coli* K-12. Genetics **39:**440–452.
4. **Bachmann, B. J.** 1972. Pedigrees of some mutant strains of *Escherichia coli* K-12. Bacteriol. Rev. **36:**525–557.
5. **Bachmann, B. J.** 1990. Linkage map of *Escherichia coli* K-12. Microbiol. Rev. **54:**130–197.
6. **Bastin, D. A., L. K. Romana, and P. R. Reeves.** 1991. Molecular cloning and expression in *Escherichia coli* K-12 of the *rfb* gene cluster determining the O antigen of an *E. coli* O111 strain. Mol. Microbiol. **5:**2223–2231.
7. **Bax, A.** 1983. Broadband homonuclear decoupling in heteronuclear shift correlation NMR spectroscopy. J. Magn. Reson. **53:**517–520.
8. **Bax, A., and D. G. Davis.** 1985. MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy. J. Magn. Reson. **65:**355–360.
9. **Bax, A., and G. A. Morris.** 1981. An improved method for heteronuclear chemical shift correlation by two-dimensional NMR. J. Magn. Reson. **42:**501–505.
10. **Bock, K., and C. Pedersen.** 1983. Carbon-13 nuclear magnetic resonance spectroscopy of monosaccharides. Adv. Carbohydr. Chem. Biochem. **41:**27–65.
11. **Brahmbhatt, H. N., P. Wyk, N. B. Quigley, and P. R. Reeves.** 1988. Complete physical map of the *rfb* gene cluster encoding biosynthetic enzymes for the O antigen of *Salmonella typhimurium* LT2. J. Bacteriol. **170:**98–102.
12. **Brenner, D. J.** 1984. Enterobacteriaceae, p. 408–420. *In* N. R. Krieg and J. G. Holt (ed.), Bergey's manual of systematic bacteriology, vol. 1. Williams and Wilkins, Baltimore.
13. **Brown, P. K., L. K. Romana, and P. R. Reeves.** 1991. Cloning of the *rfb* gene cluster of a group C2 *Salmonella*: comparison with the *rfb* regions of groups B and D. Mol. Microbiol. **5:**1873–1881.
14. **Brown, P. K., L. K. Romana, and P. R. Reeves.** 1992. Molecular analysis of the *rfb* gene cluster of *Salmonella* serovar Muenchen (strain M67): genetic basis of the polymorphism between groups C2 and B. Mol. Microbiol. **6:**1385–1394.
15. **Burland, V., G. Plunkett III, D. Daniels, and F. Blattner.** 1993. DNA sequence and analysis of 136 kilobases of the *Escherichia coli* genome: organisational symmetry around the origin of replication. Genomics **16:**551–561.
16. **Carpita, N. C., and E. M. Shea.** 1989. Linkage structure of carbohydrates by gas chromatography-mass spectrometry (gc-ms) of partially methylated alditol acetates, p. 157–215. *In* C. J. Biermann and G. D. McGinnis (ed.), Analysis of carbohydrates by GLC and MS. CRC Press, Boca Raton, Fla.
17. **Churchward, G., D. Belin, and Y. Nagamine.** 1984. A pSC101-derived plasmid which shows no sequence homology to other commonly used cloning vectors. Gene **31:**165–171.
18. **Ciucanu, I., and F. Kerek.** 1984. Rapid and simultaneous methylation of fatty and hydroxyfatty acids for gas-liquid chromatographic analysis. J. Chromatogr. **284:**179–185.
19. **Clark, C. A., J. Beltrame, and P. A. Manning.** 1991. The *oac* gene encoding a lipopolysaccharide O-antigen acetylase maps adjacent to the integrase-encoding gene on the genome of *Shigella flexneri* bacteriophage Sf6. Gene **107:**43–52.
20. **Clarke, B. R., and C. Whitfield.** 1992. Moleuclar cloning of the *rfb* region of *Klebsiella pneumoniae* serotype O1:K20: the *rfb* gene cluster is responsible for synthesis of the D-galactan I O polysaccharide. J. Bacteriol. **174:**4614–4621.
21. **Cossart, P., E. A. Groisman, M.-C. Serre, M. C. Casadaban, and B. Gicquel-Sazey.** 1986. *crp* genes of *Shigella flexneri*, *Salmonella typhimurium*, and *Escherichia coli*. J. Bacteriol. **167:**639–646.
22. **Daniels, D. L., G. Plunkett III, V. Burland, and F. R. Blattner.** 1992. Analysis of the *Escherichia coli* genome: DNA sequence of

the region from 84.5 to 86.5 minutes. Science **257:**771–778.
23. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:**387–395.
24. **Fobes, W. S., and J. E. Gander.** 1972. Evidence for uridine 5′-(α-D-galactopyranosyl pyrophosphate):NAD 2-hexosyl oxidoreductase in *Penicillium charlesii*. Biochem. Biophys. Res. Commun. **49:**76–83.
25. **Gleeson, T. J., and R. Staden.** 1991. An X windows and UNIX implementation of our sequence analysis package. Comput. Appl. Biosci. **7:**398.
26. **Hammerschmidt, S., C. Birkholz, U. Zähringer, B. D. Robertson, J. van Putten, O. Ebeling, and M. Frosch.** 1994. Contribution of genes from the capsule gene complex (*cps*) to lipooligosaccharide biosynthesis and serum resistance in *Neisseria meningitidis*. Mol. Microbiol. **11:**885–896.
27. **Haraguchi, G. E., U. Zahringer, B. Jann, K. Jann, R. A. Hull, and S. I. Hull.** 1991. Genetic characterisation of the O4 polysaccharide gene cluster from *Escherichia coli*. Microb. Pathog. **10:**351–361.
28. **Harris, P. J., R. J. Henry, A. B. Blakeney, and B. A. Stone.** 1984. An improved procedure for the methylation analysis of oligosaccharides and polysaccharides. Carbohydr. Res. **127:**59–73.
29. **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.
30. **Johnson, M. T., and J. E. Gander.** 1978. Purification and properties of a nucleoside diphosphosugar: NAD^+ 2-hexosyl oxidoreductase. Biochim. Biophys. Acta **523:**9–18.
31. **Kenne, L., B. Lindberg, K. Petersson, E. Katzenellenbogen, and E. Romanowska.** 1977. Basic structure of the oligosaccharide repeating-unit of the *Shigella flexneri* O-antigens. Carbohydr. Res. **56:** 363–370.
32. **Klena, J. D., and C. A. Schnaitman.** 1993. Function of the *rfb* gene cluster and the *rfe* gene in the synthesis of O antigen by *Shigella dysenteriae* 1. Mol. Microbiol. **9:**393–402.
33. **Kogan, G., G. Haraguchi, S. I. Hull, R. A. Hull, A. S. Shashkov, B. Jann, and K. Jann.** 1993. Structural analysis of O4-reactive polysaccharides from recombinant *Escherichia coli*: changes in the O-specific polysaccharide induced by cloning of the *rfb* genes. Eur. J. Biochem. **214:**259–265.
34. **Kohara, Y., K. Akiyama, and K. Isono.** 1987. The physical map of the whole *E. coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. Cell **50:**495–508.
35. **Le Minor, L., and M. Y. Popoff.** 1987. Designation of *Salmonella enterica* sp. nov., nom. rev., as the type and only species of the genus *Salmonella*. Int. J. Syst. Bacteriol. **37:**465–468.
36. **Lew, H. C., P. H. Mäkelä, H.-M. Kuhn, H. Meyer, and H. Nikaido.** 1986. Biosynthesis of enterobacterial common antigen requires dTDPglucose pyrophosphorylase determined by a *Salmonella typhimurium rfb* gene and a *Salmonella montevideo rfe* gene. J. Bacteriol. **168:**715–721.
37. **Liu, D., and P. R. Reeves.** 1994. *Escherichia coli* regains its O antigen. Microbiology **140:**49–57.
38. **Lugowski, C., E. Romanowska, L. Kenne, and B. Lindberg.** 1983. Identification of a trisaccharide repeating unit in the enterobacterial common antigen. Carbohydr. Res. **118:**173–181.
39. **Mäkelä, P. H., G. Schmidt, H. Mayer, H. Nikaido, H. Y. Whang, and E. Neter.** 1976. Enterobacterial common antigen in *rfb* deletion mutants of *Salmonella typhimurium*. J. Bacteriol. **127:** 1141–1149.
40. **Mäkelä, P. H., and B. A. D. Stocker.** 1984. Genetics of lipopolysaccharide, p. 59–137. *In* E. T. Rietschel (ed.), Handbook of endotoxin, vol. I. Elsevier Science Publishers, Amsterdam.
41. **Marumo, K., L. Lindqvist, N. Verma, A. Weintraub, P. R. Reeves, and A. A. Lindberg.** 1992. Enzymatic synthesis and isolation of thymidine diphosphate-6-deoxy-D-xylo-4-hexulose and thymidine diphosphate-L-rhamnose. Production using cloned gene products and separation by HPLC. Eur. J. Biochem. **204:**539–545.
42. **Meier-Dieter, U., K. Barr, R. Starman, L. Hatch, and P. D. Rick.** 1992. Nucleotide sequence of the *Escherichia coli rfe* gene involved in the synthesis of enterobacterial common antigen. J. Biol. Chem. **267:**746–753.
43. **Morris, G. A., and K. I. Smith.** 1985. "Virtual coupling" in heteronuclear chemical shift correlation by two-dimensional NMR. A simple test. J. Magn. Reson. **65:**506–509.
44. **Morris, G. A., and K. I. Smith.** 1987. Analysis of virtual one-bond coupling effects in heteronuclear chemical shift correlation 2D N.M.R. Mol. Physiol. **61:**467–483.
45. **Neidhardt, F. C., J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.).** 1987. *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
46. **Neuhard, J., and E. Thomassen.** 1976. Altered deoxyribonucleotide pools in P2 eductants of *Escherichia coli* K-12 due to deletion of the *dcd* gene. J. Bacteriol. **126:**999–1001.
47. **Nikaido, H., M. Levinthal, K. Nikaido, and K. Nakane.** 1967. Extended deletions in the histidine-rough-B region of the *Salmonella* chromosome. Proc. Natl. Acad. Sci. USA **57:**1825–1832.
48. **Ørskov, F., and I. Ørskov.** 1976. Special *Escherichia coli* serotypes among enterotoxigenic strains from diarrhoea in adults and children. Med. Microbiol. Immunol. **162:**73–80.
49. **Ørskov, F., and I. Ørskov.** 1984. Serotyping of *Escherichia coli*. Methods Microbiol. **14:**43–112.
50. **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.
50a. **Rajakumar, K.** Unpublished data.
51. **Rayssiguier, C., D. S. Thaler, and M. Radman.** 1989. The barrier to recombination between *Escherichia coli* and *Salmonella typhimurium* is disrupted in mismatch-repair mutants. Nature (London) **342:**396–400.
52. **Reeves, P. R.** 1993. Evolution of *Salmonella* O antigen variation by interspecific gene transfer on a large scale. Trends Genet. **9:**17–22.
53. **Reeves, P. R.** 1994. Biosynthesis and assembly of lipopolysaccharide, p. 281–314. *In* A. Neuberger and L. L. M. van Deenen (ed.), Bacterial cell wall, new comprehensive biochemistry, vol. 27. Elsevier Science Publishers, New York.
54. **Reisner, A. H., C. A. Bucholtz, J. Smelt, and S. McNeil.** 1993. Australia's national genomic information system. Proc. Twenty-Sixth Annu. Hawaii Int. Conf. Systems Sci. **1:**595–602.
55. **Rick, P. D., H. Mayer, B. A. Neumeyer, S. Wolski, and D. Bitter-Suermann.** 1985. Biosynthesis of enterobacterial common antigen. J. Bacteriol. **162:**494–503.
55a. **Robertson, B. D., and J. P. M. Van Putten.** Personal communication.
56. **Rudd, K. E.** 1993. Maps, genes, sequences, and computers: an *Escherichia coli* case study. ASM News **59:**335–341.
57. **Rutar, V.** 1984. Suppression of homonuclear J couplings in 1H-13C chemical shift correlation maps. J. Magn. Reson. **58:**306–310.
58. **Saiki, R. K., D. H. Gelfand, S. Stofell, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Erlich.** 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science **239:**487–491.
59. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
60. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
61. **Sarvas, M., and H. Nikaido.** 1971. Biosynthesis of T1 antigen in *Salmonella*: origin of D-galactofuranose and D-ribofuranose residues. J. Bacteriol. **105:**1063–1072.
62. **Shashkov, A. S., G. M. Lipkind, N. K. Knirel, and Y. A. Kochetkov.** 1988. Stereochemical factors determining the effects of glycosylation on the 13C chemical shifts in carbohydrates. Magn. Reson. Chem. **26:**735–747.
63. **Shibaev, V. N.** 1986. Biosynthesis of bacterial polysaccharide chains composed of repeating units. Adv. Carbohydr. Chem. Biochem. **44:**277–339.
64. **Simmons, D. A. R., and E. Romanowska.** 1987. Structure and biology of *Shigella flexneri* O antigens. J. Med. Microbiol. **23:**289–302.
65. **Stevenson, G., S. J. Lee, L. K. Romana, and P. R. Reeves.** 1991. The *cps* gene cluster of *Salmonella* strain 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.

66. **Stickney de Bouregas, F., and J. S. Waugh.** 1992. ANTIOPE, a program for computer experiments on spin dynamics. J. Magn. Reson. **96:**280–289.
67. **Sunshine, M. G., and B. Kelly.** 1971. Extent of host deletions associated with bacteriophage P2-mediated eduction. J. Bacteriol. **108:**695–704.
68. **Trejo, A., G. J. F. Chittendon, J. G. Buchanon, and J. Baddily.** 1970. Uridine diphosphate α-D-galactofuranose, an intermediate in the biosynthesis of galactofuranosyl residues. Biochem. J. **117:**637–639.
69. **Verma, N. K., J. B. Brandt, K. J. Verma, and A. A. Lindberg.** 1991. Molecular characterisation of the O-acetyl transferase gene of converting phage Sf6 that adds group antigen 6 to *Shigella flexneri*. Mol. Microbiol. **5:**71–75.
70. **Wang, L., L. K. 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.
71. **Westphal, O., and K. Jann.** 1965. Bacterial lipopolysaccharides. Extraction with phenol-water and further applications of the procedure. Methods' Carbohydr. Chem. **5:**83–91.
72. **Yao, Z., H. Liu, and M. A. Valvano.** 1992. Acetylation of O-specific lipopolysaccharides from *Shigella flexneri* 3a and 2a occurs in *Escherichia coli* K-12 carrying cloned *S. flexneri* 3a and 2a *rfb* genes. J. Bacteriol. **174:**7500–7508.
73. **Yao, Z., and M. A. Valvano.** 1994. Genetic analysis of the O-specific lipopolysaccharide biosysnthesis region (*rfb*) of *Escherichia coli* K-12 W3110: identification of genes that confer group 6 specificity to *Shigella flexneri* serotypes Y and 4a. J. Bacteriol. **176:**4133–4143.