A group of *Escherichia coli* and *Salmonella enterica* O antigens sharing a common backbone structure

Wei Wang,[1,2]† Andrei V. Perepelov,[3]† Lu Feng,[1,2]† Sergei D. Shevelev,[3] Quan Wang,[1,2] Sof'ya N. Senchenkova,[3] Weiqing Han,[1,2] Yayue Li,[1,2] Alexander S. Shashkov,[3] Yuriy A. Knirel,[3] Peter R. Reeves[4] and Lei Wang[1,2]

Correspondence
Lei Wang
wanglei@nankai.edu.cn

[1]TEDA School of Biological Sciences and Biotechnology, Nankai University, 23 Hongda Street, TEDA, Tianjin 300457, P. R. China

[2]Tianjin Key Laboratory of Microbial Functional Genomics, Nankai University, 23 Hongda Street, TEDA, Tianjin 300457, P. R. China

[3]N. D. Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, 119991 Moscow, Russian Federation

[4]School of Molecular and Microbial Biosciences (G08), University of Sydney, Sydney, NSW 2006, Australia

The O-antigen moiety of the LPS is one of the most variable cell surface components of the Gram-negative bacterial outer membrane. Variation is due to the presence of different sugars and sugar linkages. Here, it is reported that a group of *Escherichia coli* O serogroups (O17, O44, O73, O77 and O106), and the *Salmonella enterica* serogroup O : 6,14 (H), share a common four-sugar backbone O-subunit structure, and possess almost identical O-antigen gene clusters. Whereas the *E. coli* O77 antigen does not have any substitutions, the other O antigens in this group differ by the addition of one or two glucose side branches at various positions of the backbone. The O-antigen gene clusters for all members of the group encode only the proteins required for biosynthesis of the common four-sugar backbone. The identification of three genes within a putative prophage in the *E. coli* O44 genome is also reported; these genes are presumably involved in the glucosylation of the basic tetrasaccharide unit. This was confirmed by deletion of one of the genes, which encodes a putative glucosyltransferase. Structural analysis of the O antigen produced by the mutant strain demonstrated the absence of glucosylation. An O-antigen structure shared by five *E. coli* and one *S. enterica* serogroups, all of which have a long evolutionary history, suggests that the common backbone may be important for the survival of *E. coli* strains in the environment, or for their pathogenicity.

INTRODUCTION

The O antigen, which consists of many repeats of an oligo-saccharide unit (O unit), is part of the LPS in the outer membrane of Gram-negative bacteria, and it contributes to major antigenic variability to the cell surface (Reeves & Wang, 2002). In *Escherichia coli* and *Salmonella*, genes involved in O-antigen synthesis are usually clustered

†These authors contributed equally to this paper.

Abbreviations: GlcNAc, *N*-acetyl-D-glucosamine; ROESY, rotating Overhauser effect spectroscopy; TOCSY, total correlation spectroscopy.

The GenBank/EMBL/DDBJ accession numbers for the sequences of *E. coli* O73, O77 and O106 O-antigen gene clusters are DQ000313, DQ000314 and DQ000315, respectively.

between *galF* and *gnd* in the chromosome. These genes encode proteins for the biosynthesis of nucleotide sugars, sugar transferases and proteins for the processing of O antigen, and they include genes encoding flippase (Wzx) and polymerase (Wzy) (Reeves & Wang, 2002). The different O-antigen forms are almost entirely due to genetic variations in their respective O-antigen gene clusters. Sequencing and comparison of various O-antigen gene clusters support the notion that interspecies and intraspecies gene transfers play an important role in the formation of new O-antigen forms (Shepherd *et al.*, 2000; Tarr *et al.*, 2000).

Shigella flexneri, a major cause of bacillary dysentery worldwide, is atypical, as 12 of its 13 serotypes share a
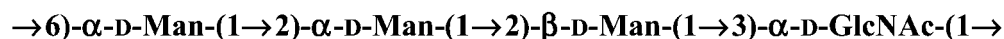
common O-antigen backbone. The differences between serotypes are due to glucosyl and/or O-acetyl groups added to sugars in the backbone of the O unit by transferases encoded by genes mapping outside of the common O-antigen gene cluster (Allison & Verma, 2000).

In a study comparing restriction patterns of *E. coli* O-antigen gene clusters, it has been found that serogroups O17, O44, O73, O77 and O106 have almost identical patterns after *Mbo*II digestion of amplified PCR products of O-antigen gene clusters (Coimbra *et al.*, 2000), suggesting that they may also share a common gene cluster. Also, structures of *E. coli* O17, O44 and O77 (Masoud & Perry, 1996; Staaf *et al.*, 1995; Yildirim *et al.*, 2001) share the same O-repeat backbone (Fig. 1), which is a tetrasaccharide composed of one *N*-acetyl-D-glucosamine (GlcNAc) and three D-mannose residues. The *E. coli* O17
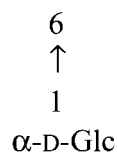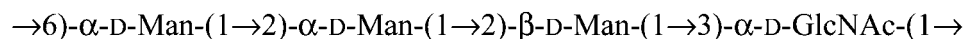
and O44 O antigens are modified by a single side-branch glucose residue attached at different positions of the backbone. The *Salmonella enterica* O:6,14 (H) O antigen also has the same backbone, with a glucose side chain at another position (di Fabio *et al.*, 1988). The O-antigen gene cluster of *S. enterica* has been sequenced (Fitzgerald *et al.*, 2003), showing that it encodes genes for the biosynthesis of tetrasaccharide backbone only, and not those for the glucose side-branch.

In this study, we determined the structure of *E. coli* O73 O-repeat unit, which has the same backbone as the O17, O44 and O77 O units, and two D-glucose side-branch residues. We also sequenced the O-antigen gene clusters of *E. coli* O73, O77 and O106. A comparison of these, and those of *E. coli* O17 and O44, and *S. enterica* O:6,14 (H), showed that the six gene clusters are almost identical. In addition,
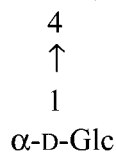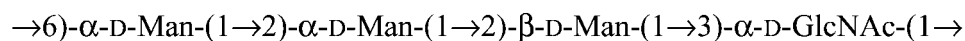

E. coli O77

→6)-α-D-Man-(1→2)-α-D-Man-(1→2)-β-D-Man-(1→3)-α-D-GlcNAc-(1→

E. coli O17

→6)-α-D-Man-(1→2)-α-D-Man-(1→2)-β-D-Man-(1→3)-α-D-GlcNAc-(1→

$$6$$
$$\uparrow$$
$$1$$
α-D-Glc

E. coli O44

→6)-α-D-Man-(1→2)-α-D-Man-(1→2)-β-D-Man-(1→3)-α-D-GlcNAc-(1→

$$4$$
$$\uparrow$$
$$1$$
α-D-Glc

E. coli O73

α-D-Glcp^{I} α-D-Glcp^{II}

$$1 \qquad\qquad 1$$
$$\downarrow \qquad\qquad \downarrow$$
$$4 \qquad\qquad 3$$

→6)-α-D-Manp^{I}-(1→2)-α-D-Manp^{II}-(1→2)-β-D-Manp^{III}-(1→3)-α-D-Glc$p$NAc-(1→
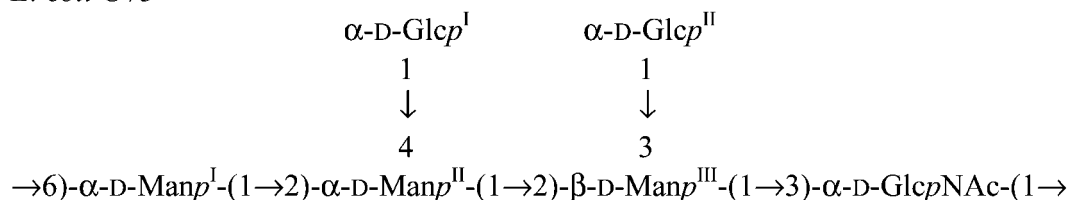
Fig. 1. Structures of the O polysaccharides of *E. coli* O77, O17, O44 and O73. Substitution with the side-branch glucose residues in *E. coli* O73 is non-stoichiometric. The structure of *E. coli* O77 in bold is the core structure of the O77 group.

we identified the glucosyltransferase gene for transfer of the side-branch glucose in *E. coli* O44. Our results support the proposal that these serotypes have a common sugar backbone encoded by a conserved O-antigen gene cluster, with variation due to side branches encoded by genes located elsewhere in the genome, as for *Shi. flexneri*. We named the serotypes the 'O77 group'.

METHODS

Bacteria and plasmids. Plasmids used in this study were maintained in *E. coli* K-12 strain DH5α, which was purchased from the Beijing DingGuo Biotechnology Development Center. The *E. coli* O17, O44, O73, O77 and O106 type strains K12a, H702c, P12a, E10, H521a (laboratory stock numbers G1298, G1291, G1057, G1075 and G1255, respectively) were obtained from the Institute of Medical and Veterinary Science, Adelaide, Australia. Plasmid pKD20 was kindly provided by Patrick Higgins, Department of Biochemistry and Molecular Genetics, University of Alabama at Birmingham, Birmingham, AL, USA.

Bacterial cultivation and isolation of LPS. Bacterial cells were grown to late exponential phase in 8 l Luria–Bertani medium, using a 10 l fermenter (Biostat C-10; B. Braun Biotech International) under constant aeration at 37 °C and pH 7.0. Bacterial cells were washed and dried as described by Robbins & Uchida (1962). LPS of *E. coli* O73 and the *E. coli* O44-derived mutant strain H1408 (920 and 1000 mg, respectively) was isolated from dried cells (16.3 and 10.9 g, respectively) by the phenol/water method (Wang *et al.*, 2001a), and purified by precipitation of nucleic acids and proteins with aqueous 50 % TCA, as described previously (Wang & Reeves, 2000).

Isolation of O polysaccharide. Delipidation of the LPS of *E. coli* O73 and *E. coli* O44 mutant strain H1408 (124 and 130 mg, respectively) was performed with aqueous 2 % acetic acid (6 ml) at 100 °C, until lipid A precipitation. The precipitate was removed by centrifugation (13 000 *g*, 20 min), and the supernatant was fractionated on a column (56 × 2.6 cm) of Sephadex G-50 (S) (Amersham Biosciences) in 0.05 M pyridinium acetate buffer, pH 4.5, and monitored using a differential refractometer (Knauer). High-molecular-mass polysaccharides were obtained in yields of 48 and 27 % of the LPS weight of *E. coli* O73 and *E. coli* O44 mutant strain H1408, respectively.

Chemical analyses. The polysaccharide of *E. coli* O73 was hydrolysed with 2 M CF_3CO_2H (120 °C, 2 h), and sugars were identified by GLC of the alditol acetates, using a Hewlett Packard 5890 chromatograph equipped with an Ultra-2 column (Supelco), and a temperature gradient of 160–290 °C at 3 °C min^{-1}. The absolute configurations of the monosaccharides were determined by GLC of the acetylated (−)-2-octyl glycosides, according to methods described by Gerwig *et al.* (1979) and Leontein *et al.* (1978). Methylation of the polysaccharide was performed with CH_3I in DMSO, in the presence of sodium methylsulfinylmethanide (Conrad, 1972). Partially methylated monosaccharides were derived by hydrolysis under the same conditions as those in the sugar analysis, and then reduced with $NaBH_4$, acetylated, and analysed by GLC-MS on a Hewlett Packard HP 5989A instrument equipped with a 30 m HP-5ms column (Hewlett Packard), under the same chromatographic conditions as described for GLC.

NMR spectroscopy. Samples were deuterium-exchanged by freeze-drying twice from D_2O, and then examined as solutions in 99.96 % D_2O at 27 °C. NMR spectra were recorded on a Bruker DRX-500 spectrometer (Germany) using internal acetone (δ_H 2.225, δ_C 31.45)

as a reference. Two-dimensional NMR spectra were obtained using standard Bruker software, and the Bruker XWINNMR 2.6 program was used to acquire and process the NMR data. Mixing times of 200 and 100 ms were used in total correlation spectroscopy (TOCSY) and rotating Overhauser effect spectroscopy (ROESY) experiments, respectively.

Construction of DNase I shotgun bank. Chromosomal DNA was prepared as described by Bastin & Reeves (1995). Primers no. 1523 and no. 1524 (Wang *et al.*, 2001b), based on *galF* and *gnd* genes, respectively, were used to amplify the DNA of *E. coli* O73, O77 and O106 O-antigen gene clusters, using the Expand Long Template PCR system (Roche). The PCR cycles used were as follows: denaturation at 94 °C for 10 s, annealing at 60 °C for 30 s, and extension at 68 °C for 15 min. The PCR products were digested with DNase I, and the resulting DNA fragments were cloned into pGEM-T Easy to produce a bank, using the method described previously (Wang & Reeves, 1998).

Sequencing and analysis. Sequencing was carried out by the Tianjin Biochip Corporation, using an ABI 3730 automated DNA sequencer (Applied Biosystems). Sequence data were assembled using the Staden Package (Staden, 1996). Artemis (Rutherford *et al.*, 2000) was used to identify ORFs, and for annotation. BlockMaker was used to search for conserved motifs. BLAST and PSI-BLAST (Altschul *et al.*, 1997) were used for searching databases, including GenBank, COG and Pfam protein motif databases (Bateman *et al.*, 2002; Tatusov *et al.*, 2001). The SOSUI system (Hirokawa *et al.*, 1998) was used to identify potential transmembrane segments. Sequence alignment and comparisons were performed using the program CLUSTALW (www.ebi.ac.uk/clustalw). The synonymous substitution rate (K_s) and non-synonymous rate (K_a) were calculated using K-Estimator 6.0 (Comeron, 1999).

Deletion and complementation of the *E. coli* O77 *wzy* gene and the *E. coli* O44 *wbbG* gene. The Red recombination system of λ phage was used to replace the *wzy* gene in *E. coli* O77 and the *wbbG* gene in *E. coli* O44; in both cases the gene was replaced by the *cat* gene, using the methods described by Datsenko & Wanner (2000). The *cat* gene was PCR amplified from plasmid pKK232-8 using primers binding to the 5′ and 3′ ends of the *cat* gene, with each primer carrying 40 bp based on the *E. coli* O77 DNA that flanks *wzy* (wl_4696/wl_4697, 5′-ATGCATAAGCAATATTATTCATTATTGT-GCAATAAAATGGAGAAAAAAATCACTGG-3′/5′-CTTACTGACC-ATGATAAATTTTATTTTTAACAAGTAATGCCTTAAAAAAATTA-CGC-3′), or the *E. coli* O44 DNA that flanks *wbbG* (wl_5183/wl_5184, 5′-TAAGCAGCGTCCAAAGTATGTACTTAAGGGTAAAAAATGA-TTCAGGAGCTAAGGAAGCTAAA-3′/5′-AAAGTGAATCTATTCC-ATTTTTCCAAATAAGAACTTTTTTTTTACGCCCCGCCCTGCCAC-TCA-3′). The PCR products were transformed, respectively, into *E. coli* O77 and *E. coli* O44 type strains carrying pKD20, and chloramphenicol-resistant transformants were selected after induction of the Red genes. The replacements were confirmed by PCR using one primer specific to the *cat* gene, and the other specific to the DNA flanking the *E. coli* O77 *wzy* gene, or the DNA flanking the *E. coli* O44 *wbbG* gene. Cloning of the *wzy* and *wbbG* genes was performed by PCR amplification from the *E. coli* O77 and *E. coli* O44 type strains, respectively, using primer pairs wl_4698/wl_4699 (5′-ATTACCATGGATAAGCAATATTATTCATTATTG-3′/5′-CTGGGA-TCCAAATAAACCACTTACTGACC-3′) binding to the ends of *wzy*, and wl_5185/wl_5186 (5′-ACGCCATGGATGAGCAAGAAGCTTG-AGCTTT-3′/5′-CTGGGATCCCTAAAGTGAATCTATTCCATTT-3′) binding to the ends of *wbbG*. The PCR products were cloned into pTRC99A to make plasmid pLW1062 containing the *wzy* gene of *E. coli* O77, and plasmid pLW1180 containing the *wbbG* gene of *E. coli* O44. Membrane preparation, SDS-PAGE, and silver staining for

visualization of the LPS of the various constructs, were carried out as described by Wang & Reeves (1994).

RESULTS AND DISCUSSION

Elucidation of the structure of the *E. coli* O73 O antigen

Structures for *E. coli* O17, O44 and O77 antigens, and *S. enterica* O:6,14 (H) antigen, are known (Masoud & Perry, 1996; Staaf *et al.*, 1995; Yildirim *et al.*, 2001). The structure of the *E. coli* O73 O antigen was determined using O73 polysaccharide obtained by mild acid degradation of LPS isolated from dried cells by the phenol/water procedure (Wang *et al.*, 2001a). Sugar analysis after full acid hydrolysis of the polysaccharide revealed glucose, mannose and glucosamine in the ratio 1.2:1:0.1. GLC of the acetylated glycosides with (−)-2-octanol showed that all sugars had the D configuration.

The ^{13}C NMR spectrum of the polysaccharide (Fig. 2a) contained major signals for six monosaccharide residues, including those for six anomeric carbons at δ 98.3–102.9, one nitrogen-bearing carbon (C-2 of GlcN) at δ 54.4, and one *N*-acetyl group at δ 23.3 (CH$_3$) and 174.8 (CO). Accordingly, the ^1H NMR spectrum of the polysaccharide

contained, *inter alia*, signals for six anomeric protons at δ 4.81–5.35, and one *N*-acetyl group at δ 2.03. The spectrum also contained minor signals, indicating one or more monosaccharides in non-stoichiometric amounts.

^1H and ^{13}C NMR spectra were analysed by 2D correlation spectroscopy, total correlation spectroscopy (TOCSY), H-detected ^1H, ^{13}C heteronuclear single quantum correlation, and heteronuclear multiple quantum correlation-TOCSY (Table 1). Based on the coupling constant values estimated from the 2D spectra, the spin systems were assigned to three mannose residues (Man^I–Man^{III}), two glucose residues (Glc^I and Glc^{II}), and one GlcNAc residue, all of which were in the pyranose form. The integral intensities of the signals for Glc^I and Glc^{II} were lower than those of the other sugar residues, and it was concluded that they were present in non-stoichiometric quantities (~80% of each residue; also, see methylation analysis data below).

The $J_{1,2}$ coupling constant values of ~3 Hz indicated that GlcNAc and both Glc residues were α linked. The NMR chemical shifts for H-5 and C-5, compared with published data (Jansson et al., 1989), showed that Man^I and Man^{II} were also α linked, whereas Man^{III} was β linked. Significant down-field displacements, due to α-glycosylation effects (Lipkind et al., 1988), of the signals for C-6 of Man^I, C-2 and C-4 of Man^{II}, C-2 and C-4 of Man^{III}, and C-3 of
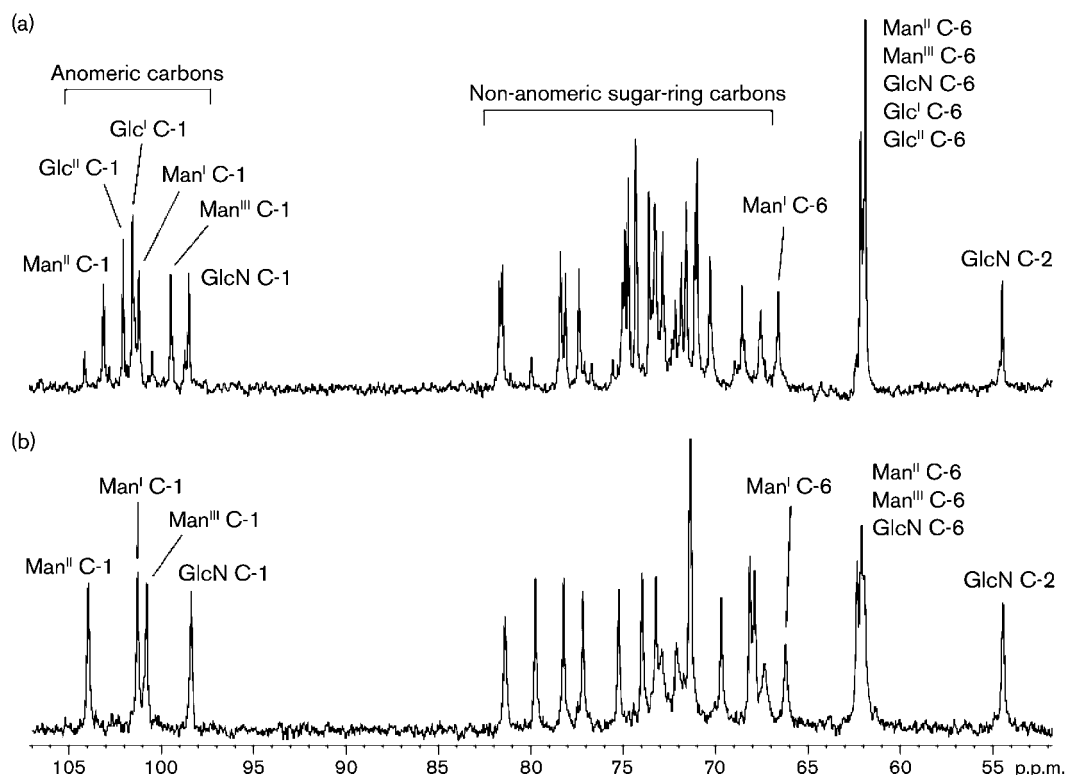

Fig. 2. ^{13}C NMR spectra (125 MHz) of the O polysaccharides of *E. coli* O73 (a) and *E. coli* O44 mutant strain H1408 (b). The region of CO resonances is not shown. The spectra differ in the presence (*E. coli* O73) or absence (*E. coli* strain H1408) of signals for two side-branch non-stoichiometric glucose residues.

Table 1. ^{1}H and ^{13}C NMR data of the *E. coli* O73 and *E. coli* O44 mutant strain H1408 O polysaccharides (δ, p.p.m.)

The chemical shifts for NAc are δ_H 2.03 (Me), δ_C 23.3 (Me) and 174.8 (CO) in *E. coli* O73; δ_H 2.04 (Me), δ_C 23.1 (Me) and 175.2 (CO) in H1408.

Sugar residue	H-1	H-2	H-3	H-4	H-5	H-6a	H-6b
***E. coli* O73**							
→6)-α-D-Man*p*^{I}-(1→	5.15	4.12	3.89	3.99	3.98	3.60	4.12
→2,4)-α-Man*p*^{II}-(1→	5.35	4.09	4.30	3.94	3.85	3.78	3.87
→2,3)-β-Man*p*^{III}-(1→	4.81	4.07	3.85	3.87	3.44	3.75	3.94
→3)-α-Glc*p*NAc-(1→	4.89	4.03	3.95	3.53	3.76	3.78	3.87
α-Glc*p*^{I}-(1→	5.24	3.56	3.65	3.43	3.76	3.78	3.87
α-Glc*p*^{II}-(1→	5.23	3.58	3.54	3.43	3.76	3.78	3.87
***E. coli* O44 mutant strain H1408**							
→6)-α-D-Man*p*^{I}-(1→	5.03	4.10	3.84	3.95	3.80	3.53	4.09
→2)-α-Man*p*^{II}-(1→	5.30	4.06	4.01	3.75	3.97	3.82	3.82
→2)-β-Man*p*^{III}-(1→	4.75	3.95	3.70	3.64	3.40	3.74	3.93
→3)-α-Glc*p*NAc-(1→	4.86	4.07	3.93	3.54	3.74	3.78	3.86

	C-1	C-2	C-3	C-4	C-5	C-6
***E. coli* O73**						
→6)-α-D-Man*p*^{I}-(1→	102.9	71.5	77.0	67.3	72.9	66.3
→2,4)-α-Man*p*^{II}-(1→	99.3	78.2	71.7	77.2	73.2	62.0
→2,3)-β-Man*p*^{III}-(1→	101.1	74.8	81.6	68.3	78.0	62.3
→3)-α-Glc*p*NAc-(1→	98.3	54.4	81.4	70.1	73.2	62.0
α-Glc*p*^{I}-(1→	101.4	73.4	74.5	70.8	74.2	62.0
α-Glc*p*^{II}-(1→	101.9	72.7	74.7	70.9	74.2	62.0
***E. coli* O44 mutant strain H1408**						
→6)-α-D-Man*p*^{I}-(1→	103.8	71.1	71.9	67.1	72.7	66.0
→2)-α-Man*p*^{II}-(1→	100.6	79.6	71.1	67.7	73.8	61.9
→2)-β-Man*p*^{III}-(1→	101.1	77.0	75.1	67.9	78.0	62.1
→3)-α-Glc*p*NAc-(1→	98.2	54.2	81.2	69.5	73.0	61.9

α-GlcNAc, demonstrated the glycosylation pattern in the O unit. The sequence of the sugar residues in the O unit was determined by a 2D ROESY experiment, which revealed strong inter-residue cross-peaks between the following anomeric protons and protons at the linkage carbons: Man^I H-1/Man^{II} H-2, Man^{II} H-1/Man^{III} H-2, Man^{III} H-1/GlcNAc H-3, GlcNAc H-1/Man^I H-6a, Glc^I H-1/Man^{II} H-4, and Glc^{II} H-1/Man^{III} H-3. This pattern was independently confirmed by a ^{1}H,^{13}C heteronuclear multiple band correlation experiment.

Methylation analysis identified derivatives from terminal glucose, and 2,3-disubstituted, 2,4-disubstituted, 2-substituted and 6-substituted mannose, in the ratio 1.5:1:1:0.3:0.5, and a derivative from 3-substituted GlcNAc. A lower than stoichiometric content of the derivative from terminal glucose, and the occurrence of the derivative from 2-substituted mannose, confirmed incomplete substitution of the main chain with the side-branch glucose residues. A lower than expected content of a derivative from 6-substituted mannose could be accounted for by incomplete cleavage of the glycosidic linkage in the GlcNAc→Man disaccharide fragment.

Therefore, the O polysaccharide of *E. coli* O73 has a branched hexasaccharide repeating unit, with the structure shown in Fig. 1. The structure includes the O77 tetrasaccharide backbone and two side-branch glucose residues.

Based on the structural data, we conclude that *E. coli* O17, O44, O73 and O77, and *S. enterica* O:6,14 (H), share the O77 O-antigen backbone structure. Therefore, we refer to this group as the O77 group, which also includes *E. coli* O106 (see below).

The O-antigen gene clusters of the O77 group

The DNA sequences of the O-antigen gene clusters from *E. coli* O73, O77 and O106 type strains are almost identical (>99% DNA identity), and they encode six genes transcribed from *galF* to *gnd* (Fig. 3). The sequence of the *E. coli* O44 O-antigen gene cluster, obtained from the genome sequence of *E. coli* strain 042 (www.sanger.ac.uk/Projects/Escherichia_Shigella/), also shares >99% DNA identity with each of the O73, O77 and O106 sequences. The *E. coli* O17 gene cluster was not sequenced, but

Fig. 3. Comparison of O-antigen gene clusters of *S. enterica* O:6,14 (H) (AY334017) and *E. coli* O17/44/73/77/106 (this study). All genes are transcribed in a direction from *galF* to *gnd*. Dashed lines align genes or DNA segments that are homologous in these O-antigen gene clusters. Genes were assigned functions based on homology: *wbaC*, mannosyltransferase; *wbaD*, mannosyltransferase; *manC*, mannose-1-phosphate guanylyltransferase; *manB*, phosphomannomutase; *wzy*, O-antigen polymerase; *wzx*, O-antigen flippase.

adjacent-gene PCR was carried out on the type strain using primers based on the genes (including *galF* and *gnd*) of *E. coli* O77. The PCR products produced by each primer pair had the same size in both *E. coli* O17 and O77 strains (data not shown), indicating that *E. coli* O17 has the same O-antigen gene cluster as the others. We also found that the O-antigen gene cluster of *S. enterica* O:6,14 (H) (GenBank accession no. AY334017) has the same genes in the same order as that of *E. coli* O77 (Fitzgerald *et al.*, 2003) (Fig. 3). Details of each gene in *E. coli* O77 O-antigen gene cluster are given below.

Orf1 and Orf2 belong to glycosyltransferase family 1 (PF00534). Orf1 shares 49 % identity with the mannosyltransferase WbaC of *S. enterica* C1 (Lee *et al.*, 1992). *S. enterica* C1 O unit has two adjacent D-Man-$\alpha(1{\rightarrow}2)$-D-Man linkages, as does the *E. coli* O77 O unit, and both have one fewer transferase gene than linkages. Therefore, *orf1* is proposed to encode a mannosyltransferase for both of the D-Man-$\alpha(1{\rightarrow}2)$-D-Man linkages, and is named *wbaC*. Orf2 shares 52 % identity with a putative glycosyltransferase involved in *E. coli* O6 O-antigen assembly. Since the D-Man-$\beta(1{\rightarrow}3)$-D-GlcNAc linkage is the only linkage present in both *E. coli* O77 and O6 O antigens, *orf2* is proposed to encode the transferase for the D-Man-$\beta(1{\rightarrow}3)$-D-GlcNAc linkage, and given the name *wbaD*.

Orf3 and Orf4 were identified as ManC (mannose-1-phosphate guanylyltransferase) and ManB (phosphomannomutase), respectively, which are involved in the synthesis of GDP-mannose, based on their high level identity to many known ManC and ManB proteins (identity between 60 and 83 %). Therefore, *orf3* and *orf4* were named *manC* and *manB*, respectively.

Orf5 has 12 predicted transmembrane segments, with a large periplasmic loop of 61 aa residues which is a characteristic topology for O-antigen polymerases (Daniels *et al.*, 1998). A mutant strain, H1236, in which *orf5* was replaced by a *cat* gene, produced LPS that consisted of lipid A-core, and lipid A-core with one O unit only, while the *E. coli* O77 wild-type strain, and H1236 carrying plasmid pLW1062, which contains *orf5*, both produced the normal LPS (Fig. 4a). Therefore, *orf5* was designated *wzy*, which is the O-antigen polymerase gene. Orf6 has 12 predicted transmembrane segments, with no large loops; this is a feature of Wzx proteins (Liu *et al.*, 1996). Orf6 also shares 43 % similarity to the Wzx protein of *Yersinia enterocolitica* O:8, and it was designated *wzx*, which is the O-antigen flippase gene.

Genetic analysis of the O-antigen clusters of the O77 group demonstrated that the only genes present are those involved in the synthesis of the common O unit backbone, including genes for GDP-mannose biosynthesis, sugar transfer, O-antigen flipping (Wzx) and O-antigen polymerization (Wzy). Genes for the synthesis of UDP-GlcNAc are located elsewhere in the chromosome (Reeves & Wang, 2002). GlcNAc is usually the first residue in *E. coli* O units, and is transferred by the product of the *wecA* gene located outside of the O-antigen gene cluster (Alexander & Valvano, 1994).

Location of the genes responsible for the side-branch glucosyl residues in *E. coli* O44

In *Shi. flexneri*, the addition of glucosyl residues to the O antigen is due to temperate bacteriophages that contain three genes: *gtrA*, *gtrB* and a serotype-specific glucosyltransferase gene (Allison & Verma, 2000). GtrB is thought to be a bactoprenol glucosyltransferase that catalyses the transfer of glucose from UDP-glucose to bactoprenol phosphate (UndP) to give UndP-glucose, and GtrA is
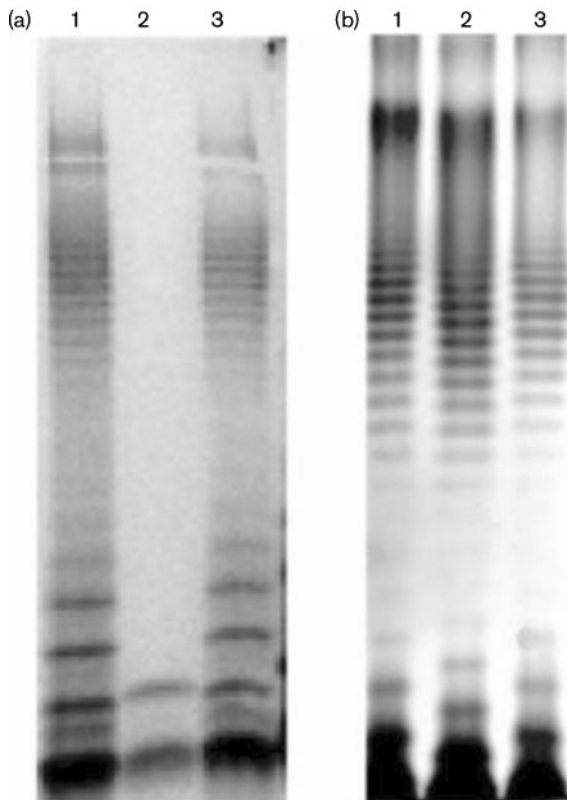
Fig. 4. Effects of deletion and complementation of the *E. coli* O77 *wzy* gene (a) and *E. coli* O44 *wbbG* gene (b). LPS was run by SDS-PAGE, and visualized by silver staining. (a) Lanes: 1, *E. coli* O77 type strain E10; 2, H1236 (E10 missing the *wzy* gene); 3, H1237 (H1236 with plasmid pLW1062 carrying the *E. coli* O77 *wzy* gene). (b) Lanes: 1, *E. coli* O44 type strain H702c; 2, H1408 (H702c missing the *wbbG* gene); 3, H1409 (H1408 with plasmid pLW1180 carrying the *E. coli* O44 *wbbG* gene).

thought to be a flippase that translocates UndP-glucose to face the periplasmic space. Both GtrA and GtrB are highly conserved. The protein encoded by the third gene, which catalyses the linkage of a glucosyl group to the O-antigen backbone, is serotype specific (Allison & Verma, 2000).

For O44, we found a set of three genes in the *E. coli* strain 042 genome that fit the pattern of the *Shi. flexneri* genes. These genes are located in a prophage adjacent to *attL*, between *ybhB* and *ybhC* (Fig. 5), the normal insertion site of λ phage (Campbell, 2003). The products of the first two genes share 96 and 86 % identity to GtrA and GtrB, respectively, of *E. coli* K-12, as well as high-level identity

(from 60 to 93 %) to GtrA and GtrB proteins of *Shi. flexneri* and *S. enterica*. The products are therefore proposed to be GtrA and GtrB.

The product of the third gene shares 22 % identity and 40 % similarity with GtrI of *Shi. flexneri* 1a; GtrI is responsible for the formation of the $\alpha(1\rightarrow4)$ linkage between the added glucosyl group and GlcNAc. The product was also predicted to have 11 transmembrane segments, with a hydrophilic carboxy-terminal region of 110 aa; this is the typical topological character of GtrI. We replaced this gene with the *cat* gene to make strain H1408. When compared with its parent on an SDS-PAGE gel, the bands of the ladder representing LPS with different numbers of O units of strain H1408 were closer together, as expected for a smaller O unit. H1408 was transcomplemented to restore the LPS phenotype of the wild-type by plasmid pLW1180 containing the third gene, and the spacing returned to that of the O44 strain (Fig. 4b). The O polysaccharide was obtained from strain H1408, and studied by 1D ^{1}H and ^{13}C spectroscopy, as described above for the *E. coli* O73 O polysaccharide (Fig. 2b). The ^{1}H and ^{13}C spectra of the mutant polysaccharide were identical to those reported earlier for the *E. coli* O77 O polysaccharide (Yildirim *et al.*, 2001), showing that the *E. coli* O44 mutant strain H1408 possesses the same O antigen as that of *E. coli* O77, with no side-branch glucose substituents.

Both LPS size distribution and structural data indicated that the third gene is required for the attachment of the glucose residue to position 4 of the β-D-mannose residue of the O-antigen backbone. It is clear that these three genes constitute a typical set of genes for the addition of a side-branch glucose, and the third gene was named *wbbG*. Although we have data for O44 only, it is most likely that the proteins for the formation of side-branch glucose residues in O17 and O73 are encoded by similar sets of genes.

The importance of the O77 group O antigens

It has been reported that the O antigen plays an important role in the virulence of *Shi. flexneri* (Morona *et al.*, 2003; Van den Bosch & Morona, 2003), and O-antigen modification may also be an important factor in enhancing survival and pathogenicity (Allison & Verma, 2000; West *et al.*, 2005). The situation for the O antigens of the O77 group is similar to that in *Shi. flexneri*, and the same may well apply to this much less-studied group of serovars. It is noteworthy that *E. coli* O17 and O73 antigens have been found
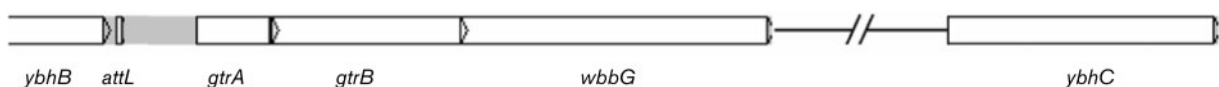

Fig. 5. Organization of the putative O-antigen modification genes in *E. coli* O44. *attL* is the phage-attachment site.

in enterohaemorrhagic *E. coli* (Bielaszewska *et al.*, 2004), and O44 and O77 antigens have been found in enteroadhesive *E. coli* (Nataro & Kaper, 1998), and in both cases adaptation to virulence may be important.

It has been proposed that *E. coli* and *S. enterica* diverged from a common ancestor about 140 million years ago (Ochman & Wilson, 1987a, b), and proteins encoded by housekeeping genes in the two species share 93 % identity on average, ranging from 76.3 to 100 % (Sharp, 1991). The sequence differences between *E. coli* O77 and *S. enterica* O:6,14 (H) are consistent with the O77-group gene cluster having been in the common ancestor, and diverged as the species diverged, rather than having been independently acquired, or acquired by one species from the other. The apparent long history of this O antigen in *E. coli* and *S. enterica* is not typical, as most O antigens in each species are not found in the other. The survival of the O antigen indicates that it is highly adaptive in *E. coli* and *S. enterica*, with their rather different niches.

ACKNOWLEDGEMENTS

REFERENCES

Alexander, D. C. & Valvano, M. A. (1994). Role of the *rfe* gene in the biosynthesis of the *Escherichia coli* O7-specific lipopolysaccharide and other O-specific polysaccharides containing *N*-acetylglucosamine. *J Bacteriol* **176**, 7079–7084.

Allison, G. E. & Verma, N. K. (2000). Serotype-converting bacteriophages and O-antigen modification in *Shigella flexneri*. *Trends Microbiol* **8**, 17–23.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402.

Bastin, D. A. & Reeves, P. R. (1995). Sequence and analysis of the O antigen gene (*rfb*) cluster of *Escherichia coli* O111. *Gene* **164**, 17–23.

Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S. R., Griffiths-Jones, S., Howe, K. L., Marshall, M. & Sonnhammer, E. L. (2002). The Pfam protein families database. *Nucleic Acids Res* **30**, 276–280.

Bielaszewska, M., Fell, M., Greune, L., Prager, R., Fruth, A., Tschape, H., Schmidt, M. A. & Karch, H. (2004). Characterization of cytolethal distending toxin genes and expression in shiga toxin-producing *Escherichia coli* strains of non-O157 serogroups. *Infect Immun* **72**, 1812–1816.

Campbell, A. (2003). Prophage insertion sites. *Res Microbiol* **154**, 277–282.

Coimbra, R. S., Grimont, F., Lenormand, P., Burguiere, P., Beutin, L. & Grimont, P. A. (2000). Identification of *Escherichia coli* O-serogroups by restriction of the amplified O-antigen gene cluster (rfb-RFLP). *Res Microbiol* **151**, 639–654.

Comeron, J. M. (1999). *K*-Estimator: calculation of the number of nucleotide substitutions per site and the confidence intervals. *Bioinformatics* **15**, 763–764.

Conrad, H. E. (1972). Methylation of carbohydrates with methyl iodide in dimethyl sulfoxide in the presence of methylsulfinylanion. *Methods Carbohydr Chem* **6**, 361–364.

Daniels, C., Vindurampulle, C. & Morona, R. (1998). Overexpression and topology of the *Shigella flexneri* O-antigen polymerase (Rfc/Wzy). *Mol Microbiol* **28**, 1211–1222.

Datsenko, K. A. & Wanner, B. L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A* **97**, 6640–6645.

di Fabio, J. L., Brisson, J. R. & Perry, M. B. (1988). Structure of the major lipopolysaccharide antigenic O-chain produced by *Salmonella carrau* (O : 6, 14, 24). *Carbohydr Res* **179**, 233–244.

Fitzgerald, C., Sherwood, R., Gheesling, L. L., Brenner, F. W. & Fields, P. I. (2003). Molecular analysis of the *rfb* O antigen gene cluster of *Salmonella enterica* serogroup O : 6,14 and development of a serogroup-specific PCR assay. *Appl Environ Microbiol* **69**, 6099–6105.

Gerwig, G. J., Kamerling, J. P. & Vliegenthart, J. F. G. (1979). Determination of the absolute configuration of monosaccharides in complex carbohydrates by capillary GLC. *Carbohydr Res* **77**, 1–7.

Hirokawa, T., Boon-Chieng, S. & Mitaku, S. (1998). SOSUI: Classification and secondary structure prediction system for membrane proteins. *Bioinformatics* **14**, 378–379.

Jansson, P. E., Kenne, L. & Widmalm, G. (1989). Computer-assisted structural analysis of polysaccharides with an extended version of CASPER using 1H- and 13C-n.m.r. data. *Carbohydr Res* **188**, 169–191.

Lee, S. J., Romana, L. K. & Reeves, P. R. (1992). Sequence and structural analysis of the *rfb* (O antigen) gene cluster from a group C1 *Salmonella enterica* strain. *J Gen Microbiol* **138**, 1843–1855.

Leontein, K., Lindberg, B. & Lonngren, J. (1978). Assignment of absolute configuration of sugars by GLC of their acetylated glycosides formed from chiral alcohols. *Carbohydr Res* **62**, 359–362.

Lipkind, G. M., Shashkov, A. S., Knirel, Y. A., Vinogradov, E. V. & Kochetkov, N. K. (1988). A computer-assisted structural analysis of regular polysaccharides on the basis of 13C-n.m.r. data. *Carbohydr Res* **175**, 59–75.

Liu, D., Cole, R. & Reeves, P. R. (1996). An O-antigen processing function for Wzx(RfbX): a promising candidate for O-unit flippase. *J Bacteriol* **178**, 2102–2107.

Masoud, H. & Perry, M. B. (1996). Structural characterization of the O-antigenic polysaccharide of *Escherichia coli* serotype O17 lipopolysaccharide. *Biochem Cell Biol* **74**, 241–248.

Morona, R., Daniels, C. & Van Den Bosch, L. (2003). Genetic modulation of *Shigella flexneri* 2a lipopolysaccharide O antigen modal chain length reveals that it has been optimized for virulence. *Microbiology* **149**, 925–939.

Nataro, J. P. & Kaper, J. B. (1998). Diarrheagenic *Escherichia coli*. *Clin Microbiol Rev* **11**, 142–201.

Ochman, H. & Wilson, A. C. (1987a). Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. *J Mol Evol* **26**, 74–86.

Ochman, H. & Wilson, A. C. (1987b). Evolutionary history of enteric bacteria. In *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, pp. 1649–1654. Edited by F. C. Neidhardt and others. Washington, DC: American Society for Microbiology.

Reeves, P. P. & Wang, L. (2002). Genomic organization of LPS-specific loci. *Curr Top Microbiol Immunol* **264**, 109–135.

Robbins, P. W. & Uchida, T. (1962). Studies on the chemical basis of the phage conversion of O-antigens in the E-group *Salmonella*. *Biochemistry* **1**, 323–335.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M. A. & Barrell, B. (2000). Artemis: sequence visualisation and annotation. *Bioinformatics* **16**, 944–945.

Sharp, P. M. (1991). Determinants of DNA sequence divergence between *Escherichia coli* and *Salmonella typhimurium*: codon usage, map position, and concerted evolution. *J Mol Evol* **33**, 23–33.

Shepherd, J. G., Wang, L. & Reeves, P. R. (2000). Comparison of O-antigen gene clusters of *Escherichia coli* (*Shigella*) sonnei and *Plesiomonas shigelloides* O17: sonnei gained its current plasmid-borne O-antigen genes from *P. shigelloides* in a recent event. *Infect Immun* **68**, 6056–6061.

Staaf, M., Widmalm, G., Weintraub, A. & Nataro, J. P. (1995). Structural elucidation of the O-antigenic polysaccharide from *Escherichia coli* O44 : H18. *Eur J Biochem* **233**, 473–477.

Staden, R. (1996). The Staden sequence analysis package. *Mol Biotechnol* **5**, 233–241.

Tarr, P. I., Schoening, L. M., Yea, Y. L., Ward, T. R., Jelacic, S. & Whittam, T. S. (2000). Acquisition of the *rfb-gnd* cluster in evolution of *Escherichia coli* O55 and O157. *J Bacteriol* **182**, 6183–6191.

Tatusov, R. L., Natale, D. A., Garkavtsev, I. V., Tatusova, T. A., Shankavaram, U. T., Rao, B. S., Kiryutin, B., Galperin, M. Y., Fedorova, N. D. & Koonin, E. V. (2001). The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res* **29**, 22–28.

Van den Bosch, L. & Morona, R. (2003). The actin-based motility defect of a *Shigella flexneri rmlD* rough LPS mutant is not due to loss of IcsA polarity. *Microb Pathog* **35**, 11–18.

Wang, L. & Reeves, P. R. (1994). Involvement of the galactosyl-1-phosphate transferase encoded by the *Salmonella enterica rfbP* gene in O antigen subunit processing. *J Bacteriol* **176**, 4348–4356.

Wang, L. & Reeves, P. R. (1998). Organization of *Escherichia coli* O157 O antigen gene cluster and identification of its specific genes. *Infect Immun* **66**, 3545–3551.

Wang, L. & Reeves, P. R. (2000). The *Escherichia coli* O111 and *Salmonella enterica* O35 gene clusters: gene clusters encoding the same colitose-containing O antigen are highly conserved. *J Bacteriol* **182**, 5256–5261.

Wang, L., Briggs, C. E., Rothemund, D., Fratamico, P., Luchansky, J. B. & Reeves, P. R. (2001a). Sequence of the *E. coli* O104 antigen gene cluster and identification of O104-specific genes. *Gene* **270**, 231–236.

Wang, L., Qu, W. & Reeves, P. R. (2001b). Sequence analysis of four *Shigella boydii* O-antigen loci: implication for *Escherichia coli* and *Shigella* relationships. *Infect Immun* **69**, 6923–6930.

West, N. P., Sansonetti, P., Mounier, J., Exley, R. M., Parsot, C., Guadagnini, S., Prevost, M. C., Prochnicka-Chalufour, A., Delepierre, M. & other authors (2005). Optimization of virulence functions through glucosylation of *Shigella* LPS. *Science* **307**, 1313–1317.

Yildirim, H., Weintraub, A. & Widmalm, G. (2001). Structural studies of the O-polysaccharide from the *Escherichia coli* O77 lipopolysaccharide. *Carbohydr Res* **333**, 179–183.

Edited by: D. L. Gally