

Structural and genetic characterization of the closely related O-antigens of *Escherichia coli* O85 and *Salmonella enterica* O17

17(2) (2011) 164–173
© SAGE Publications 2010
ISSN 1753-4259 (print)
10.1177/1753425910369270

Andrei V. Perepelov¹, Dan Li^{2,3}, Bin Liu^{2,3}, Sof'ya N. Senchenkova¹, Dan Guo^{2,3}, Alexander S. Shashkov¹, Lu Feng^{2,3,4}, Yuriy A. Knirel¹, Lei Wang^{2,3,4,5}

¹N.D. Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, Moscow, Russian Federation

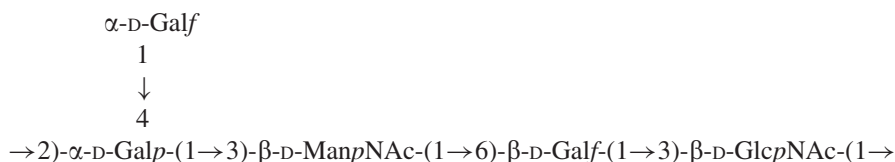
²TEDA School of Biological Sciences and Biotechnology, Nankai University, Tianjin, P. R. China

³Tianjin Key Laboratory of Microbial Functional Genomics, Tianjin, P. R. China

⁴Tianjin Research Center for Functional Genomics and Biochip, Tianjin, P. R. China

⁵The Key Laboratory of Molecular Microbiology and Technology, Ministry of Education, P. R. China

O-Antigen is a part of the lipopolysaccharide present in the outer membrane of Gram-negative bacteria, which confers major antigenic variability to the cell surface. In this study, we report on a previously undefined pair of *Escherichia coli* and *Salmonella enterica* with closely related O-antigens. The O-polysaccharides were isolated from the lipopolysaccharides of *E. coli* O85 and *S. enterica* O17 by mild acid degradation and studied by sugar analysis and NMR spectroscopy. The following structure was established for the O-unit of the *E. coli* O85-polysaccharide:



The *S. enterica* O17-polysaccharide has the same carbohydrate backbone and, in addition, contains an *O*-acetyl group at position 2 of ~80% β -Galf residues. The O-antigen gene cluster of *E. coli* O85 was found to be closely related to that of *S. enterica* O17. Screening of type strains of all *E. coli* and *S. enterica* O-serogroups revealed two genes specific to the *E. coli* O85 O-antigen gene cluster, which can be used for development of PCR-based assays for identification and detection of *E. coli* O85 strains.

Keywords: *Escherichia coli*, *Salmonella enterica*, O-antigen structure, O-antigen gene cluster

INTRODUCTION

Lipopolysaccharide (LPS) is an important component of the outer membrane of Gram-negative bacteria. It usually consists of three distinct regions – lipid A, the core oligosaccharide, and the O-specific polysaccharide (O-antigen). The O-antigen is composed of repeating O-units, typically containing 2–6 sugar residues each, and

is one of the most variable antigenic cell wall constituents. Until now, 174 and 46 O-serogroups, respectively, have been documented in the *Escherichia coli* and *Salmonella enterica* typing schemes.^{1–3} The O-antigen is subject to intense selection by the host immune system and also serves as a receptor by some bacteriophages, which may both contribute to the maintenance of diversity within species, such as *E. coli*, by intermittent

Received 18 December 2009; Revised 4 March 2010; Accepted 15 March 2010

Correspondence to: Andrei V. Perepelov, N.D. Zelinsky Institute of Organic Chemistry, Leninsky Prospekt 47, Moscow V-334, GSP-1, 119991 Russia. Tel: +7 499 1376148; Fax: +7 499 1355328; E-mail: perepel@ioc.ac.ru

selection against specific O-antigen forms.^{4,5} O-Antigen heterogeneity among respective clones is also thought to be an important advantage for bacteria colonizing specific micro-environments.⁶ Furthermore, there is evidence that the O-antigen is an important virulence factor; its loss makes many pathogens serum-sensitive or otherwise seriously impaired in virulence.⁶⁻⁹

Escherichia coli is a clonal species which includes both commensal and pathogenic strains. *Escherichia coli* serotyping is based on antigenic differences between their somatic (O) and flagellar (H) (sometimes also capsular, K) antigens. The *E. coli* O85 serogroup strains are associated with the entero-invasive *E. coli* (ETEC), which are a common cause of acute diarrhea in adults and children.¹⁰ Some serotypes (*E. coli* O85:H32 and *E. coli* O85:H49) were found to be Shiga toxin producers that can cause hemorrhagic colitis and hemolytic uremic syndrome in humans.^{11,12} *Escherichia coli* O85 strains can also be responsible for cellulitis in broilers.¹³ *Salmonella enterica* is recognized as a major pathogen of animals and humans, and strains of this species are serotyped in a similar fashion as *E. coli*.

Genes for O-antigen biosynthesis in *E. coli* and *S. enterica* are normally clustered between *galF* and *gnd* and are classified into three different groups: (i) nucleotide sugar synthesis genes used to generate O-antigen components; (ii) sugar transferase encoding genes; and (iii) O-unit processing genes, normally including those for O-unit flippase (*wzx*) and O-antigen polymerase (*wzy*). Genes of the last two groups are usually specific to particular O-antigens.

Only three common O-antigens have been identified in *E. coli* and *S. enterica*, namely in *E. coli* O111/*S. enterica* O35, *E. coli* O55/*S. enterica* O50, and *E. coli* O157/*S. enterica* O30.¹⁴ Recently, the O-antigen of *S. enterica* O:6,14 (H) has been shown to be closely related to those of the *E. coli* O77 group.¹⁵ In this study, we report on the closely related O-antigen structures of a new pair of *E. coli* O85/*S. enterica* O17. In addition, the O-antigen gene cluster of *E. coli* O85 was characterized and compared to that previously reported for *S. enterica* O17.¹⁶ Finally, two genes specific to *E. coli* O85 useful for identification of this serotype by PCR were revealed by screening representative strains of all *E. coli* and *S. enterica* O-serogroups (see Table 3).

MATERIALS AND METHODS

Cultivation of bacteria and isolation of LPSs

Bacterial cells of *E. coli* O85 and *S. enterica* O17, strains G1189 and G1460 (both of which were obtained from the Institute of Medical and Veterinary Science, Adelaide, Australia), respectively, were grown to late log phase in 8 l of LB using a 10-l fermentor (Biostat

C-10; B. Braun Biotech International, Germany) under constant aeration at 37°C and pH 7.0. Bacterial cells were washed and dried as described.¹⁷ Dried cells (5 g and 4.7 g, respectively) were extracted with a phenol–water mixture as described.¹⁸ After dialysis of combined phenol and water layers, contaminants were precipitated by adding 50% aqueous trichloroacetic acid at 4°C to pH 2, the supernatant was dialyzed and lyophilized to give the corresponding LPSs (500 mg and 430 mg, respectively).

Preparation of O-polysaccharides

Delipidation of the LPSs of *E. coli* O85 and *S. enterica* O17 (~100 mg and 120 mg, respectively) was performed with aqueous 2% acetic acid (6 ml) at 100°C until lipid precipitation (~2 h). 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 Superfine in 0.05 M pyridinium acetate buffer, pH 4.5, using a differential refractometer as detector (Knauer, Germany). The yield of both O-polysaccharides was ~30% of the LPS mass.

Chemical analyses

The O-polysaccharides were hydrolyzed with 2 M trifluoroacetic acid (120°C, 2 h); monosaccharides were reduced with 0.25 M NaBH₄ in aqueous 1 M ammonia (20°C, 1 h), acetylated with a 1:1 (v/v) mixture of pyridine and acetic anhydride (120°C, 30 min) and analyzed by GLC of the alditol acetates using a Hewlett-Packard 5890 chromatograph (USA) equipped with an Ultra-2 column (Agilent) using a temperature gradient from 160–290°C at 3°C/min. The absolute configuration of the monosaccharides was determined by GLC of the acetylated (+)-2-octyl glycosides as described.^{19,20}

Nuclear magnetic resonance (NMR) spectroscopy

Samples were deuterium-exchanged by freeze-drying twice from D₂O and then dissolved in 0.5 ml of 99.95% D₂O and examined at 30°C. All NMR spectra were recorded on a Bruker DRX-500 spectrometer (Germany) using internal sodium 3-(trimethylsilyl)propionate-2,2,3,3-d₄ (δ_H 0.00) and acetone (δ_C 31.45) as references. Bruker XWINNMR 2.6 program was used to acquire and process NMR data. Mixing times of 200 ms and 100 ms were used in TOCSY and ROESY experiments, respectively. The HMBC experiment was optimized for the coupling constant 8 Hz.

Construction of the DNase I shotgun bank

Chromosomal DNA was prepared as described previously.²¹ Long-range PCR was performed with the Expand Long Template PCR system (Roche Applied Science) with primers wl-1098 (5'-ATT GGT AGC TGT AAG CCA AGG GCG GTA GCG T-3') and wl-2211 (5'-CAC TGC CAT ACC GAC GAC GCC GAT CTG TTG CTT GG-3'), based on sequences of the JUMPstart site and the *gnd* gene sequence, respectively.⁴ The PCR was performed as follows: denaturation at 95°C for 30 s, annealing at 60°C for 45 s, and extension at 68°C for 15 min for 30 cycles. The PCR products were digested with DNase I, and the resulting DNA fragments were cloned into the pGEM-T Easy (Promega, Madison, WI, USA) vector to produce a shotgun bank as described previously.⁴

Sequencing and analysis

Sequencing was carried out by the Tianjin Biochip Corporation using an ABI 3730 (Applied Biosystems Inc., Foster City, CA, USA) automated DNA sequencer. Sequence data were assembled using the Staden Package.²² The program Artemis²³ was used for annotation. Blast and PSI-BLAST²⁴ were used for searching databases including GenBank, COG and Pfam protein motif databases.²⁵ The program TMHMM 2.0 (<<http://www.cbs.dtu.dk/services/TMHMM-2.0/>>) was used to identify potential transmembrane segments.

PCR specificity testing of E. coli O85

Chromosomal DNA was prepared from each of the 186 representative *E. coli* strains (including *Shigella*) corresponding to all O-serogroups in addition to 46 representative *S. enterica* strains of all O-serogroups (Table 3). Thirty-four pools of DNA (each containing DNA from 6–10 strains) were used for specificity testing. Pools were screened using primers designed based on the specific genes of *E. coli* O85 (Table 4). The PCR cycles used were as follows: denaturation at 95°C for 30 s, annealing for 30 s and extension at 72°C for 1 min. The PCR was carried out in a total volume of 25 µl and 10 µl were subjected to agarose gel electrophoresis to assess product size and amplification specificity.

Nucleotide sequence accession number

The DNA sequence of the *E. coli* O85 O-antigen gene cluster has been deposited in GenBank under the accession number GU299798.

RESULTS AND DISCUSSION*Isolation and sugar composition of the O-polysaccharides*

The LPSs were obtained from dried bacterial cells of *E. coli* O85 and *S. enterica* O17 by the Westphal procedure¹⁸ and degraded with mild acid. High molecular mass O-polysaccharides were isolated by gel-permeation chromatography on Sephadex G-50. Sugar analysis by GLC after full acid hydrolysis of both O-polysaccharides revealed galactose, *N*-acetylglucosamine and *N*-acetylmannosamine (ManNAc). Gas-liquid chromatography of the acetylated glycosides with a chiral alcohol showed that all monosaccharides have the D-configuration.

Structure elucidation of the E. coli O85-polysaccharide

The ¹³C-NMR spectrum of the *E. coli* O85-polysaccharide (Fig. 1A) showed signals for five monosaccharide residues, including those for five anomeric carbons at δ 101.3–109.7, two nitrogen-bearing carbons (C-2 of ManNAc and GlcNAc) at δ 54.0 and 56.5, five CH₂O groups (C-6 of monosaccharides) at δ 61.5, 61.8, 62.1, 64.4 and 72.4 (data of attached-proton test), other sugar carbons in the region δ 68.3–84.2 and two *N*-acetyl groups at δ 23.5, 23.7 (both methyl), 176.2 and 176.5 (both carbonyl). The ¹H-NMR spectrum of the O-polysaccharide contained, *inter alia*, signals for five anomeric protons at δ 4.62–5.43 and two *N*-acetyl groups at δ 2.05 and 2.07.

Therefore, the polysaccharide has a pentasaccharide O-unit containing three D-Gal residues and one residue each of D-GlcNAc and D-ManNAc.

The ¹H- and ¹³C-NMR spectra of the polysaccharide were assigned (Table 1) using two-dimensional correlation spectroscopy, including ¹H/¹H COSY, TOCSY, H-detected ¹H, ¹³C-HSQC and HMQC-TOCSY experiments, and five sugar spin systems were revealed. Based on coupling constant values estimated from the two-dimensional spectra, the spin systems were assigned to two Galf residues (**A** and **D**) and one residue each of Galp (**B**), ManpNAc (**C**) and GlcpNAc (**E**). The α-configuration of units **A** and **B** and β-configuration of units **C**, **D**, and **E** were inferred by ¹³C-NMR chemical shifts compared with published data.^{26,27} The *J*_{1,2} coupling constant values of ~3 Hz and ~7 Hz, respectively, confirmed that unit **B** is α-linked and unit **E** is β-linked.

Significant down-field displacements, due to α-glycosylation effects,²⁷ of the signals for C-2 and C-4 of unit **B**, C-3 of unit **C**, C-6 of unit **D** and C-3 of unit **E** to δ 79.6, 81.4, 79.7, 72.4 and 81.9, respectively, as

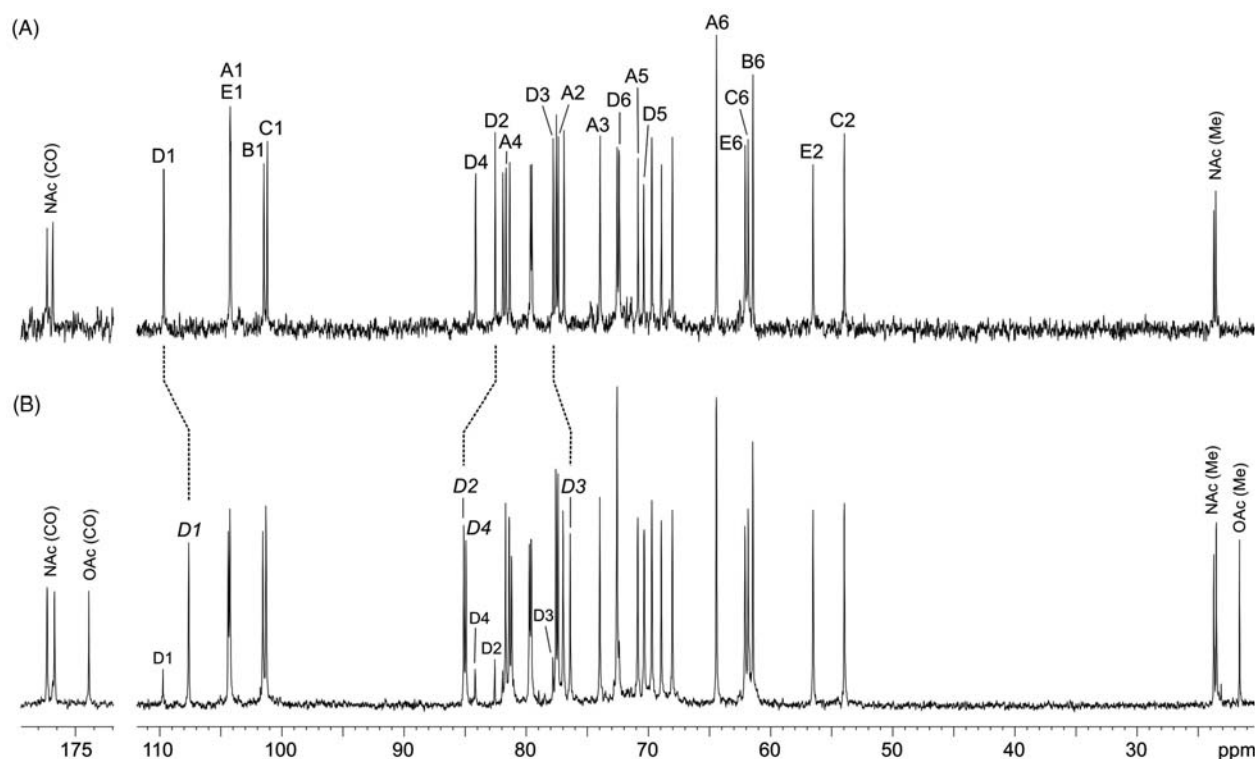


Fig. 1. 125-MHz ^{13}C -NMR spectra of the O-polysaccharides of *E. coli* O85 (A) and *S. enterica* O17 (B). Numbers refer to carbons in sugar residues denoted as shown in Table 1. Peak annotations for 2-O-acetylated residue **D** in *S. enterica* O17 are shown in italics.

Table 1. ^1H - and ^{13}C -NMR chemical shifts (δ , ppm) of the signals from O-antigens of *E. coli* O85 and *S. enterica* O17

Sugar residue	Nucleus	1	2	3	4	5	6
<i>E. coli</i> O85 ^a							
α -D-Galf-(1→	^1H	4.96	4.13	4.29	3.86	3.79	3.65; 3.65
A	^{13}C	104.3	77.4	74.0	81.7	70.9	64.4
2,4)- α -D-Galp-(1→	^1H	5.43	3.82	3.82	3.99	4.14	3.76; 3.76
B	^{13}C	101.5	79.6	69.0	81.4	72.6	61.5
3)- β -D-ManpNAc-(1→	^1H	4.82	4.58	3.96	3.76	3.44	3.83; 3.93
C	^{13}C	101.3	54.0	79.7	68.0	77.6	61.8
6)- β -D-Galf-(1→	^1H	5.05	4.02	4.04	4.04	3.93	3.70; 3.93
D	^{13}C	109.7	82.6	77.8	84.2	70.4	72.4
3)- β -D-GlcpNAc-(1→	^1H	4.62	3.86	3.68	3.51	3.51	3.76; 3.91
E	^{13}C	104.3	56.5	81.9	69.7	76.9	62.1
<i>S. enterica</i> O17 ^b (data given for O-acetylated O-units)							
α -D-Galf-(1→	^1H	4.95	4.13	4.28	3.85	3.77	3.64; 3.64
A	^{13}C	104.3	77.3	74.1	81.7	70.9	64.5
2,4)- α -D-Galp-(1→	^1H	5.42	3.81	3.80	3.98	4.14	3.75; 3.75
B	^{13}C	101.7	79.6	68.9	81.4	72.6	61.5
3)- β -D-ManpNAc-(1→	^1H	4.82	4.58	3.95	3.76	3.44	3.83; 3.93
C	^{13}C	101.3	53.9	79.8	68.1	77.6	61.9
6)- β -D-Galf2Ac-(1→	^1H	5.18	4.89	4.24	4.12	3.95	3.70; 3.94
D	^{13}C	107.6	85.2	76.5	85.0	70.4	72.5
3)- β -D-GlcpNAc-(1→	^1H	4.60	3.85	3.68	3.50	3.50	3.76; 3.92
E	^{13}C	104.3	56.6	81.2	69.8	77.0	62.2

^aThe chemical shifts for the *N*-acetyl groups are δ_{H} 2.05 and 2.07; δ_{C} 23.5, 23.7 (both Me), 176.2 and 176.5 (both CO).

^bThe chemical shifts for the *N*-acetyl groups are δ_{H} 2.05 and 2.08; δ_{C} 23.5, 23.7 (both Me), 176.1 and 176.5 (both CO) and for the *O*-acetyl group are δ_{H} 2.12; δ_{C} 21.6 (Me), 174.3 (CO).

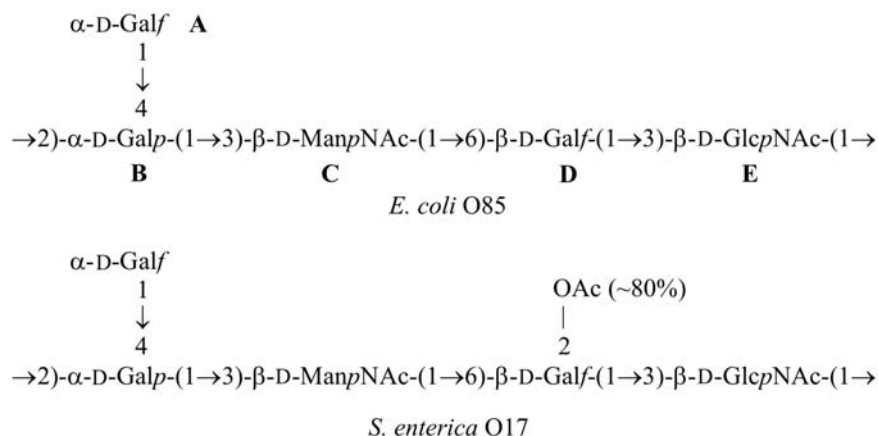


Fig. 2. Structures of the O-polysaccharides of *E. coli* O85 and *S. enterica* O17.

compared with their positions in the spectra of the corresponding non-substituted monosaccharides,^{26,27} demonstrated the glycosylation pattern in the O-unit. The sequence of the sugar residues was determined by a two-dimensional ROESY experiment, which showed inter-residue cross-peaks between the following anomeric protons and protons at the linkage carbons: **A** H-1/**B** H-4; **B** H-1/**C** H-3; **C** H-1/**D** H-6a,6b; **D** H-1/**E** H-3 and **E** H-1/**B** H-2. The monosaccharide sequence was confirmed by a ¹H,¹³C-HMBC experiment, which revealed correlations between the anomeric protons and linkage carbons and *vice versa* (data not shown).

Therefore, the O-polysaccharide of *E. coli* O85 has the structure shown in Figure 2.

Structure elucidation of the *S. enterica* O17-polysaccharide

The ¹H- and ¹³C-NMR (Fig. 1B) spectra of the O-polysaccharide of *S. enterica* O17 showed two series of signals of different intensities, evidently owing to non-stoichiometric *O*-acetylation as there were signals for methyl of an *O*-acetyl group at δ_{H} 2.12 and δ_{C} 21.6. The minor series of signals was fully assigned and the same carbohydrate backbone structure as described above for the O-polysaccharide of *E. coli* O85 was established.

Position of the *O*-acetyl group was determined by a ¹H,¹³C-HSQC experiment, which showed a displacement of ~80% of the **D** H-2/C-2 cross-peak from δ 4.02/82.6 in the O-polysaccharide of *E. coli* O85 to 4.89/85.2 in that of *S. enterica* O17. This was evidently caused by a deshielding effect of the *O*-acetyl group (α -effect of *O*-acetylation) and indicated a partial *O*-acetylation of unit **D** at position 2. This conclusion was confirmed by up-field shifts by 2.1 ppm and 1.3 ppm of the signals for C-1 and C-3 of unit **D** (Fig. 1) caused by β -effects of 2-*O*-acetylation.²⁸ Therefore, the only difference

between the two O-polysaccharides studied is *O*-acetylation in *S. enterica* O17.

Thus, the O-polysaccharide of *S. enterica* O17 has the structure shown in Figure 2.

Sequencing and characterization of *E. coli* O85 *O*-antigen gene cluster

A sequence of 11,203 base pairs (bp) between JUMPStart and *gnd* was examined and 8 open reading frames (ORFs) were identified and found to possess the same transcriptional direction (Fig. 3). All ORFs functions were assigned on the basis of sequence similarity to previously described ORF sequences available from databases (Table 2).

Sugar biosynthetic pathway genes

Orf1 belongs to the UDP-*N*-acetylglucosamine 2-epimerase family (PF02350, *E* value = 1.9×10^{-166}) and shares 64% identity and 79% similarity with UDP-*N*-acetyl glucosamine-2-epimerase (MnaA) of *Citrobacter* species. MnaA catalyzes the C2-epimerization of UDP-GlcNAc to UDP-ManNAc, suggesting that *orf1* is responsible for the synthesis of UDP-ManNAc and was, therefore, named *mnaA*.

Orf7 belongs to the UDP-galactopyranose mutase (UGM) family (PF03275, *E* value = 1.7×10^{-125}) and shares 69% identity and 82% similarity with UGM of *E. coli* 53638. The flavoenzyme UGM that is encoded by *glf* is a cell-wall biosynthesis mediator in many pathogenic micro-organisms and catalyzes a unique ring contraction reaction that results in conversion of UDP-Galp to UDP-Galf.²⁹ The latter is an essential precursor of Galf found in many different cell-wall glycoconjugates of pathogenic bacteria, such as *E. coli*, *Klebsiella pneumoniae* and *Mycobacterium*

<i>S. enterica</i> O17		<i>mnaA</i>	<i>wzy</i>	<i>wfbQ</i>	<i>wzx</i>	<i>wfbR</i>	<i>wfbS</i>	<i>wfbT</i>	<i>wfbU</i>
DNA identity%		71	68	66	66	65	67	76	68
Protein identity%		70	59	57	54	57	60	80	62
<i>E. coli</i> O85		<i>mnaA</i>	<i>wzy</i>	<i>wfbQ</i>	<i>wzx</i>	<i>wfbR</i>	<i>wfbS</i>	<i>glf</i>	<i>wfbU</i>

Fig. 3. Comparison of the O-antigen gene clusters of *E. coli* O85 and *S. enterica* O17.

tuberculosis.^{30–32} Therefore, we proposed that *orf7* is involved in the synthesis of UDP-Galf and named it *glf* accordingly.

Other monosaccharides present in the *E. coli* O85 O-unit are common sugars and synthesized by house-keeping genes located outside the O-antigen cluster.³³

Genes for O-unit processing

For most *E. coli* and *S. enterica* O-antigens, the O-unit is translocated across the membrane by Wzx protein and then expressed on the periplasmic surface followed by Wzy-mediated polymerization into a long-chain O-antigen.⁵ Both Wzx and Wzy are highly hydrophobic membrane proteins sharing little sequence identity with their respective homologues. *orf2* and *orf4* were the only two genes predicted to encode proteins with several transmembrane regions. Orf4 has 12 predicted transmembrane segments, which is a feature of Wzx proteins.³⁴ This ORF shows 23% identity and 41% similarity to the LPS biosynthesis related flippase of *Bacteroides fragilis* NCTC 9343. Orf2 has 8 predicted transmembrane segments with a large periplasmic loop of 60 amino-acid residues, which is a typical topology of Wzy proteins.³⁵ It shows 26% identity and 50% similarity to the Wzy protein of *E. coli* S88. Therefore, *orf2* and *orf4* were identified as the O-antigen polymerase gene (*wzy*) and O-unit flippase gene (*wzx*), respectively, and named accordingly.

Genes encoding sugar transferases

Glycosyltransferases are specific to respective sugar donors and acceptors and also in creating linkages between sugars. The *wecA* gene encodes the glycosyl-phosphate transferase responsible for transfer of GlcNAc1P or GalNAc1P from UDP-GlcNAc or UDP-GalNAc to an undecaprenolphosphate carrier as the first

step in the O-unit synthesis. *wecA* is located outside the O-antigen gene cluster.⁵ For the remaining four monosaccharides in the *E. coli* O85 O-unit, four glycosyl-transferase genes were expected to be present in the *E. coli* O85 O-antigen cluster.

Orf3, Orf5, Orf6 belong to the glycosyl transferases group 1 family (PF00534, *E* value = $3 \times e^{-37}$, $1.1 \times e^{-28}$, and 0.073, respectively), members of which transfer UDP-, ADP-, GDP- or CMP-linked sugars to a variety of substrates, including glycogen, fructose-6-phosphate and LPSs. Orf8 belongs to the glycosyl transferases group 2 family (PF00535, *E* value = $5.1 \times e^{-29}$). The enzymes belonging to this family catalyze the transfer of sugar moieties from activated donor molecules to specific acceptor molecules, forming glycosidic bonds. Orf3, Orf5, Orf6 and Orf8 share various degrees of identity and similarity with glycosyltransferases of *Serratia proteamaculans* 568, *Caulobacter crescentus* NA1000, *Vibrio cholerae* MZO-2 and *Escherichia albertii* TW07627, respectively (Table 2). Therefore, *orf3*, *orf5*, *orf6* and *orf8* were proposed to be glycosyl-transferase genes and named *wfbQ*, *wfbR*, *wfbS* and *wfbU*, respectively.

O-antigen gene clusters of *E. coli* O85 and *S. enterica* O17 are evolutionarily related

Escherichia coli O85 and *S. enterica* O17 have similar O-antigen structures, the only difference being the presence or absence of an O-acetyl group on one of the Galf residues. A comparison of the two gene clusters showed that both possess eight ORFs with the same order (Fig. 3). The level of amino-acid identity was between 54–62% for glycosyltransferase and O-antigen processing genes, and between 70–80% for nucleotide sugar pathway genes. It is entirely possible that the O-antigen gene clusters of *E. coli* O85 and *S. enterica* O17 diverged from a common ancestor. A slightly

Table 2. Characteristics of the ORFs in the *E. coli* O85 O-antigen gene cluster

Gene name	Position of gene	G + C content (%)	Conserved domain(s)	Similar protein(s), strain(s) (Genbank accession no.)	%Identical aa/%similar aa (no. of aa overlap)	Putative function of protein
<i>mmaA</i>	1103.2239	35.7	UDP- <i>N</i> -acetylglucosamine 2-epimerase (PF02350) E value = $1.9 \times e^{-166}$	UDP- <i>N</i> -acetylglucosamine-2-epimerase, <i>Citrobacter</i> sp. 30_2 (EEH96276)	64/79(368)	UDP- <i>N</i> -acetyl glucosamine-2-epimerase
<i>wzy</i>	2266.3384	29.3		O-Antigen polymerase, <i>E. coli</i> S88 (CAR03419)	26/50(230)	O-Antigen polymerase
<i>wfbQ</i>	3359.4411	30.8	Glycosyl transferases group 1 (PF00534) E value = $3 \times e^{-37}$	Glycosyl transferase, <i>Serratia proteamaculans</i> 568 (ABV41990)	27/50(352)	Glycosyl transferase
<i>wzx</i>	4420.5640	28.6	Polysaccharide biosynthesis protein (PF01943) E value = 0.00043	Wzx, <i>Bacteroides fragilis</i> NCTC 9343 (AAK68914)	23/41(424)	O-Antigen flippase
<i>wfbR</i>	5630.6796	29.1	Glycosyl transferases group 1 (PF00534) E value = $1.1 \times e^{-28}$	Glycosyltransferase, <i>Caulobacter crescentus</i> NA1000 (ACL93931)	26/48(401)	Glycosyl transferase
<i>wfbS</i>	6793.7908	31.4	Glycosyl transferases group 1 (PF00534) E value = 0.073	Glycosyl transferase, <i>Vibrio cholerae</i> MZO-2 (EDM53706)	27/47(329)	Glycosyl transferase
<i>glf</i>	7908.9008	31.3	UDP-galactopyranose mutase (PF03275) E value = $1.7 \times e^{-125}$	UDP-galactopyranose mutase, <i>E. coli</i> 53638 (EDU65802)	69/82(365)	UDP-galactopyranose mutase
<i>wfbU</i>	9005.9919	32.6	Glycosyl transferase group 2 (PF00535) E value = $5.1 \times e^{-29}$	WfbU, <i>E. albertii</i> TW07627 (EDS90282)	54/70(301)	Glycosyl transferase

Table 3. Complete 186 *E. coli* (including *Shigella*) and 46 *S. enterica* type strains and PCR pools used for testing of *E. coli* O85-specific primers

Pools No.	Chromosomal DNAs included in the pool	Source ^a
1	<i>E. coli</i> type strains for O-serogroups 1,2,3,4,10,16,18,39	IMVS
2	<i>E. coli</i> type strains for O-serogroups 40,41,48,49,71,73,88,100	IMVS
3	<i>E. coli</i> type strains for O-serogroups 102,109,119,120,121,125,126,137	IMVS
4	<i>E. coli</i> type strains for O-serogroups 138,139,149,7,5,6,11,12	IMVS
5	<i>E. coli</i> type strains for O-serogroups 13,14,15,17,19ab,20,21,22	IMVS
6	<i>E. coli</i> type strains for O-serogroups 23,24,25,26,27,28,29,30	IMVS
7	<i>E. coli</i> type strains for O-serogroups 32,33,34,35,36,37,38,42	IMVS
8	<i>E. coli</i> type strains for O-serogroups 43,44,45,46,50,51,52,53	IMVS
9	<i>E. coli</i> type strains for O-serogroups 54,55,56,57,58,59,60,61	IMVS
10	<i>E. coli</i> type strains for O-serogroups 62,63,64,65,66,68,69,70	IMVS
11	<i>E. coli</i> type strains for O-serogroups 74,75,76,77,78,79,80,81	IMVS
12	<i>E. coli</i> type strains for O-serogroups 82,83,84,96,86,87,89,90	IMVS
13	<i>E. coli</i> type strains for O-serogroups 91,92,95,97,98,99,101	IMVS
14	<i>E. coli</i> type strains for O-serogroups 112,162,113,114,115,116,117,118	IMVS
15	<i>E. coli</i> type strains for O-serogroups 123,165,166,167,168,169,170,171	IMVS ^b
16	<i>E. coli</i> type strains for O-serogroups 146,172,127,128,129,130,131,132	IMVS ^c
17	<i>E. coli</i> type strains for O-serogroups 133,134,135,136,140,141,142,143	IMVS
18	<i>E. coli</i> type strains for O-serogroups 144,145, 147,148,150,151,152,173	IMVS
19	<i>E. coli</i> type strains for O-serogroups 153,154,155,156,157,158,159,164	IMVS
20	<i>E. coli</i> type strains for O-serogroups 160,161,163,8,9,124,111,31	IMVS
21	<i>E. coli</i> type strains for O-serogroups 103,104,105,106,107,108,110	IMVS
22	<i>S. boydii</i> type strains for O-serotypes B2,B4,B5,B6,B8,B9,B11,B12,B14	ICDC
23	<i>S. boydii</i> type strains for O-serotypes B1,B3,B7,B10,B13,B15,B16,B17,B18	ICDC
24	<i>S. dysenteriae</i> type strains for O-serotypes D1,D2,D3,D4,D5,D6,D7,D8	ICDC
25	<i>S. dysenteriae</i> type strains for O-serotypes D9,D10,D11,D12,D13	ICDC
26	<i>S. flexneri</i> type strains for O-serotypes F6a,F1a,F1b,F2a,F2b,F3,F4a,F4b,F5(v:7),F5(v:4)	ICDC
27	<i>S. sonnei</i> type strains for O-serotypes DS, DR	ICDC
28	<i>S. enterica</i> type strains for O-serogroups A,B,C1,C2,D1,D2,D3,E1	IMVS
29	<i>S. enterica</i> type strains for O-serogroups E4,11,13,H,16,17,18,21	IMVS
30	<i>S. enterica</i> type strains for O-serogroups 28,30,35,38,39,40,41,42	IMVS
31	<i>S. enterica</i> type strains for O-serogroups 43,44,45,47,48,50	IMVS
32	<i>S. enterica</i> type strains for O-serogroups 51,52,53,54,55,56,57,58	IMVS
33	<i>S. enterica</i> type strains for O-serogroups 59,60,61,62,63,65,66,67	IMVS
34	Strains in No. 12 pool with <i>E. coli</i> O85 type strain	IMVS

^aIMVS, Institute of Medical and Veterinary Science, Adelaide, Australia; ICDC, Institute for Infectious Disease Control and Prevention, Chinese Center for Disease Control and Prevention.

^bO123 from Statens Serum Institute, Copenhagen, Denmark, the rest from IMVS.

^cO146 and O172 from Statens Serum Institute, Copenhagen, Denmark, the rest from IMVS.

more significant divergence in the glycosyltransferase and O-antigen processing genes than in the nucleotide sugar pathway genes is similar to the situation found in other paired *E. coli*/*S. enterica* O-antigen gene clusters for the same basal O-antigen structure. It is assumed that the pressures on different groups of O-antigen synthesis genes exerted by natural selection or drift and resulting in this consistent pattern of divergence in the genes of O-antigen gene clusters were different.¹⁴

It seems that, after the species divergence, *S. enterica* O17 acquired a prophage gene for O-acetylation of the O-antigen, which was located outside the O-antigen gene cluster. This modification may have aided adaptation of the bacterium to its environmental niche.

Identification of *E. coli* O85 specific genes

Serotyping is the traditional typing method utilized to classify bacterial strains based on antigen agglutination mediated by specific antisera raised in rabbits against respective surface polysaccharide antigens. However, conventional serotyping is time-consuming, labor-intensive and is often non-specific since cross-reactivity between different O-antigens occurs frequently.³⁶ Based on the O-antigen structural data presented here, cross-reactivity between antiserum raised against either *E. coli* O85 or *S. enterica* O17 could be expected since they share a high degree of similarity. Moreover, this diagnostic method is often

Table 4. PCR specificity test with *E. coli* O85 genes

Gene	Base positions	Forward primer/reverse primer	Annealing temperature (°C)
wzx	4420–5640	wl-2174 (4589–4606) 5'-GTTGTAGATTTTCGGGATA-3'	50
		/wl-2175 (5332–5347) 3'-TTAGCAGCCAGACAAA-5'	
		wl-2176 (4452–4468) 5'-GGATTTTATTTTACGCT-3'	56
		/wl-2177 (5104–5120) 3'-CCGTTTACCATAGAATA-5'	
wzy	2266–3384	wl-2170 (2711–2726) 5'-AGTGGTCATCGCTCAT-3'	56
		/wl-2171 (3048–3063) 3'-CCTTACTACCTTCGTC-5'	
		wl-2172 (2749–2765) 5'-AACTGGAGTGTTTGACG-3'	60
		/wl-2173 (3346–3363) 3'-CCAGAAGTCACCATAGTA-5'	

subjective, and a significant proportion of isolates are non-typeable. Recently, PCR assays based on O-antigen-specific genes have been developed for the identification of a number of *E. coli* and *S. enterica* serogroups, including *S. enterica* O17.¹⁶

To identify the O-antigen-specific genes of *E. coli* O85, we designed PCR primers based on the O-unit processing genes (*wzx* and *wzy*) of *E. coli* O85 (Table 4). Four pairs of primers (two pairs for each gene) were used to screen DNA pools constituting 186 representative strains of *E. coli* O-serogroups (including *Shigella*) and 46 representative strains of *S. enterica* O-serogroups (including the representative strain of *S. enterica* O17). All four primer pairs produced only bands of predicted sizes when pools containing *E. coli* O85 chromosomal DNA were used as templates. No bands were detected when pools not containing *E. coli* O85 DNA were used as template. These data demonstrated that using the *E. coli* O85 *wzx* and *wzy* genes as PCR amplification targets is a viable approach in identification and detection of *E. coli* O85 strains belonging to different clonal groups, serotypes or pathotypes. Therefore, this approach provides faster and more accurate means of identifying respective *E. coli* isolates.

ACKNOWLEDGEMENTS

Andrei V. Perepelov, Dan Li and Bin Liu contributed equally to this work.

This work was supported by the Tianjin Municipal Special Fund for Science and Technology Innovation Grant 05FZZDSH00800, the National Natural Science Foundation of China (NSFC) Key Programs Grants 30530010 and 20536040, the Chinese National Science Fund for Distinguished Young Scholars (30788001), NSFC General Program Grant 30670038, 30870078, 30771175 and 30900041, the National 863 program of China grants 2006AA020703 and 2006AA06Z409, the National 973 program of China grant 2009CB522603, and National Key Programs for Infectious Diseases of

China 2008ZX10004-002, 2008ZX10004-009 and 2009ZX10004-108, and the Russian Foundation for Basic Research (Projects 08-04-01205 and 08-04-92225-NNSF).

REFERENCES

- Centers for Disease Control and Prevention. The Public Health Role of Clinical Laboratories. In: *Laboratory methods for the diagnosis of epidemic dysentery and cholera*. Atlanta: GA; CDC, 1999; 1–6.
- Lior H. Classification of *Escherichia coli*. In: Gyles CL. (ed) *Escherichia coli in domestic animals and humans*. Wallingford: CAB International, 1994; 31–72.
- Popoff MY, Minor LL. *Antigenic formulas of the Salmonella serovars, 7th revision*. WHO Collaborating Centre for Reference and Research on *Salmonella*. Paris: Pasteur Institut, 1997.
- Wang L, Reeves PR. Organization of *Escherichia coli* O157 O antigen gene cluster and identification of its specific genes. *Infect Immun* 1998; **66**: 3545–3551.
- Reeves PR, Wang L. Genomic organization of LPS-specific loci. *Curr Top Microbiol Immunol* 2002; **264**: 109–135.
- Liu B, Knirel YA, Feng L *et al.* Structure and genetics of *Shigella* O antigens. *FEMS Microbiol Rev* 2008; **32**: 627–653.
- Achtman M, Pluschke G. Clonal analysis of descent and virulence among selected *Escherichia coli*. *Annu Rev Microbiol* 1986; **40**: 185–210.
- Pluschke G, Mayden J, Achtman M *et al.* Role of the capsule and the O-antigen in resistance of O18: K1 *Escherichia coli* to complement-mediated killing. *J Bacteriol* 1983; **142**: 907–913.
- Gemski PJ, Sheahan DG, Washington O *et al.* Virulence of *Shigella flexneri* hybrids expressing *Escherichia coli* somatic antigens. *Infect Immun* 1972; **6**: 104–111.
- Penaranda ME, Evans DG, Murray BE *et al.* ST: LT: CFA/II plasmids in enterotoxigenic *Escherichia coli* belonging to serogroups O6, O8, O80, O85, and O139. *J Bacteriol* 1983; **154**: 980–983.
- Djordjevic SP, Hornitzky MA, Bailey G *et al.* Virulence properties and serotypes of Shiga toxin-producing *Escherichia coli* from healthy Australian slaughter-age sheep. *J Clin Microbiol* 2001; **39**: 2017–2021.
- Aslani MM, Bouzari S. Characterization of virulence genes of non-O157 Shiga toxin-producing *Escherichia coli* isolates from two provinces of Iran. *Jpn J Infect Dis* 2009; **62**: 16–19.

13. Gomis SM, Gomis AI, Horadagoda NU *et al.* Studies on cellulitis and other disease syndromes caused by *Escherichia coli* in broilers in Sri Lanka. *Trop Anim Health Prod* 2000; **32**: 341–351.
14. Samuel G, Hogbin JP, Wang L *et al.* Relationships of the *Escherichia coli* O157, O111, and O55 O-antigen gene clusters with those of *Salmonella enterica* and *Citrobacter freundii*, which express identical O antigens. *J Bacteriol* 2004; **186**: 6536–6543.
15. Wang W, Perepelov AV, Feng L *et al.* A group of *Escherichia coli* and *Salmonella enterica* O antigens sharing a common backbone structure. *Microbiology* 2007; **153**: 2159–2167.
16. Fitzgerald C, Gheesling L, Collins M *et al.* Sequence analysis of the *rfb* loci, encoding proteins involved in the biosynthesis of the *Salmonella enterica* O17 and O18 antigens: serogroup-specific identification by PCR. *Appl Environ Microbiol* 2006; **72**: 7949–7953.
17. Robbins PW, Uchida T. Studies on the chemical basis of the phage conversion of O-antigens in the E-group *Salmonella*. *Biochemistry* 1962; **1**: 323–335.
18. Westphal O, Jann K. Bacterial lipopolysaccharides. Extraction with phenol–water and further applications of the procedure. *Methods Carbohydr Chem* 1965; **5**: 83–91.
19. Leontein K, Lönngren J. Determination of the absolute configuration of sugars by gas-liquid chromatography of their acetylated 2-octyl glycosides. *Methods Carbohydr Chem* 1993; **9**: 87–89.
20. Leontein K, Lindberg B, Lönngren J. Assignment of absolute configuration of sugars by g.l.c. of their acetylated glycosides formed from chiral alcohols. *Carbohydr Res* 1978; **62**: 359–362.
21. Bastin DA, Reeves PR. Sequence and analysis of the O antigen gene (*rfb*) cluster of *Escherichia coli* O111. *Gene* 1995; **164**: 17–23.
22. Staden R. The Staden Sequence Analysis Package. *Mol Biotechnol* 1996; **5**: 233–241.
23. Rutherford K, Parkhill J, Crook J *et al.* Artemis: sequence visualisation and annotation. *Bioinformatics* 2000; **16**: 944–945.
24. Altschul SF, Madden TL, Schaffer AA *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997; **25**: 3398–3402.
25. Bateman A, Birney E, Cerruti L *et al.* The Pfam protein families database. *Nucleic Acids Res* 2002; **30**: 276–280.
26. Bock K, Pedersen C. Carbon-13 nuclear magnetic resonance spectroscopy of monosaccharides. *Adv Carbohydr Chem Biochem* 1983; **41**: 27–66.
27. Lipkind GM, Shashkov AS, Knirel YA *et al.* A computer-assisted structural analysis of regular polysaccharides on the basis of ¹³C-n.m.r. data. *Carbohydr Res* 1988; **175**: 59–75.
28. Jansson PE, Kenne L, Schweda E. Nuclear magnetic resonance and conformational studies on monoacetylated methyl D-glucoside and D-galactopyranosides. *J Chem Soc, Perkin Trans 1* 1987; **377**: 383.
29. Chad JM, Sarathy KP, Gruber TD *et al.* Site-directed mutagenesis of UDP-galactopyranose mutase reveals a critical role for the active-site, conserved arginine residues. *Biochemistry* 2007; **46**: 6723–6732.
30. Koplin R, Brisson JR, Whitfield C. UDP-galactofuranose precursor required for formation of the lipopolysaccharide O antigen of *Klebsiella pneumoniae* serotype O1 is synthesized by the product of the *rfbDKPO1* gene. *J Biol Chem* 1997; **272**: 4121–4128.
31. Brennan PJ, Nikaido H. The envelope of *Mycobacteria*. *Annu Rev Biochem* 1995; **64**: 29–63.
32. Nassau PM, Martin SL, Brown RE *et al.* Galactofuranose biosynthesis in *Escherichia coli* K-12: identification and cloning of UDP-galactopyranose mutase. *J Bacteriol* 1996; **178**: 1047–1052.
33. Samuel G, Reeves P. Biosynthesis of O-antigens: genes and pathways involved in nucleotide sugar precursor synthesis and O-antigen assembly. *Carbohydr Res* 2003; **338**: 2503–2519.
34. Liu D, Cole R, Reeves PR. An O-antigen processing function for Wzx(RfBX): a promising candidate for O-unit flippase. *J Bacteriol* 1996; **178**: 2102–2107.
35. Daniels C, Vindurampulle C, Morona R. Overexpression and topology of the *Shigella flexneri* O-antigen polymerase (Rfc/Wzy). *Mol Microbiol* 1998; **28**: 1211–1222.
36. Fratamico PM, DebRoy C, Strobaugh TP *et al.* DNA sequence of the *Escherichia coli* O103 O antigen gene cluster and detection of enterohemorrhagic *E. coli* O103 by PCR amplification of the *wzx* and *wzy* genes. *Can J Microbiol* 2005; **51**: 515–522.