The lipopolysaccharide of the mastitis isolate *Escherichia coli* strain 1303 comprises a novel O-antigen and the rare K-12 core type

Katarzyna A. Duda,¹ Buko Lindner,² Helmut Brade,³ Andreas Leimbach,^{4,5} Elżbieta Brzuszkiewicz,⁵ Ulrich Dobrindt^{4,6} and Otto Holst¹

Correspondence

Katarzyna Duda
kduda@fz-borstel.de

¹Division of Structural Biochemistry, Research Center Borstel, Leibniz-Center for Medicine and Biosciences, D-23845 Borstel, Germany

²Division of Immunochemistry, Research Center Borstel, Leibniz-Center for Medicine and Biosciences, D-23845 Borstel, Germany

³Division of Medical and Biochemical Microbiology, Research Center Borstel, Leibniz-Center for Medicine and Biosciences, D-23845 Borstel, Germany

⁴Institute for Molecular Infection Biology, Julius-Maximilians-University of Würzburg, D-97070 Würzburg, Germany

⁵Göttingen Genomics Laboratory, Institute of Microbiology and Genetics, Georg-August-University of Göttingen, D-37077 Göttingen, Germany

⁶Institute for Hygiene, University of Münster, D-48149 Münster, Germany

Mastitis represents one of the most significant health problems of dairy herds. The two major causative agents of this disease are *Escherichia coli* and *Staphylococcus aureus*. Of the first, its lipopolysaccharide (LPS) is thought to play a prominent role during infection. Here, we report the O-antigen (OPS, O-specific polysaccharide) structure of the LPS from bovine mastitis isolate *E. coli* 1303. The structure was determined utilizing chemical analyses, mass spectrometry, and 1D and 2D NMR spectroscopy methods. The O-repeating unit was characterized as $[-\rightarrow 4)-\beta\text{-D-Quip3NAc-(1}\rightarrow 3)-\alpha\text{-L-Fucp2OAc-(1}\rightarrow 4)-\beta\text{-D-Galp-(1}\rightarrow 3)-\alpha\text{-D-GalpNAc-(1}\rightarrow -]$ in which the O-acetyl substitution was non-stoichiometric. The nucleotide sequence of the O-antigen gene cluster of *E. coli* 1303 was also determined. This cluster, located between the *gnd* and *galF* genes, contains 13 putative open reading frames, most of which represent unknown nucleotide sequences that have not been described before. The O-antigen of *E. coli* 1303 was shown to substitute O-7 of the terminal LD-heptose of the K-12 core oligosaccharide. Interestingly, the non-OPS-substituted core oligosaccharide represented a truncated version of the K-12 outer core – namely terminal LD-heptose and glucose were missing; however, it possessed a third Kdo residue in the inner core. On the basis of structural and genetic data we show that the mastitis isolate *E. coli* 1303 represents a new serotype and possesses the K-12 core type, which is rather uncommon among human and bovine isolates.

Received 17 November 2010

Revised 23 February 2011

Accepted 1 March 2011

Abbreviations: 14 : 0(3-OH), 3-hydroxymyristic acid; COSY: correlation spectroscopy; ESI FT-ICR MS: electrospray ionization Fourier-transformed ion cyclotron resonance mass spectrometry; LD-Hep: L-glycero-D-manno-heptose; Hex: hexose; HexN: hexosamine; HSQC-DEPT: heteronuclear single-quantum correlation-distortionless enhancement by polarization transfer; Kdo: 3-deoxy-D-manno-oct-2-ulosonic acid; NOE: nuclear Overhauser enhancement; OPS: O-specific polysaccharide; P: phosphate; PEtN: 2-aminoethanol phosphate; ROESY: rotating-frame Overhauser-effect spectroscopy; SEC: size-exclusion chromatography; TOCSY: total correlation spectroscopy.

The GenBank accession number for the DNA sequence of the O-antigen gene cluster of *E. coli* strain 1303 described in this paper is FN995094.

A supplementary table is available with the online version of this paper.

INTRODUCTION

Mastitis is one of the major diseases of cattle, causing high economic losses; bacteria are the main aetiological agents. The outcome of the disease depends on the type of pathogen. Infections induced by *Escherichia coli* often result in an acute mastitis with severe clinical consequences (Petzl *et al.*, 2008). A specific set of virulence-associated genes has not yet been identified for *E. coli* mastitis isolates. Consequently, it has been hypothesized that the cow's genetic predisposition, immune status and lactation stage as well as environmental factors determine the severity of *E. coli* mastitis. Thus, pathogen-associated molecular

patterns (PAMPs), e.g. LPS, could be sufficient to elicit mastitis by *E. coli* (Burvenich *et al.*, 2003). LPS expression contributes to serum resistance and virulence (Raetz & Whitfield, 2002) and may thus offer a selective advantage for *E. coli* during infection of the bovine mammary gland. In many wild-type bacteria LPS consists of the highly antigenic O-specific polysaccharide and the more conserved core oligosaccharide, further divided into outer and inner part and lipid A, the latter of which represents the toxic moiety in toxic LPS (Holst *et al.*, 2009). More than 180 different O-antigens, defining different serogroups, have been described for *E. coli* (Stenutz *et al.*, 2006); however, none has so far been characterized of LPS from a strain causing mastitis. The core region of *E. coli* LPS is represented by five types, R1, R2, R3, R4 and K-12, the chemical structures of which have been published (Jansson *et al.*, 1981; Holst *et al.*, 1991; Haishima *et al.*, 1992; Vinogradov *et al.*, 1999; Müller-Loennies *et al.*, 2002, 2003). The structure of the inner core in case of all these types is very similar and contains the common sequence LD-Hep-(1→7)-[Glc-(1→3)]-LD-Hep-(1→3)-LD-Hep-(1→5)-[Kdo-(2→4)]-Kdo. It is mainly the structure of the outer core that differentiates the mentioned core types. The unique feature of the K-12 core type is the presence of a fourth Hep residue in the outer core (Holst, 1991). The major core glycoform isolated after complete deacylation of the K-12 LPS possessed the structure shown in Fig. 1 (Müller-Loennies *et al.*, 2003).

The distribution of different *E. coli* core types in the environment is very heterogeneous. Predominantly, the R1 core type is detected in human and cattle populations whereas the K-12 core type is only rarely identified (Heinrichs *et al.*, 1998; Amor *et al.*, 2000; Gibbs *et al.*, 2004). *E. coli* K-12 is commonly used in the laboratory and has ever since its first description in 1944 (Gray & Tatum, 1944) expressed an R-form LPS, which unlike an S-form LPS lacks the O-antigen. Two independent mutations in the *wbbL* O-specific polysaccharide gene cluster were identified in different lineages of *E. coli* K-12. Most strains carry the IS5 insertion in the last gene of the biosynthetic cluster. It was shown that complementation of the IS5 mutation leads to the production of an O-antigen in LPS of *E. coli* K-12 which was typed as O16 (Liu & Reeves, 1994). The structure of the O16 antigen was determined (Jann *et al.*, 1994; Stevenson *et al.*, 1994) and it was shown to be

attached to O-7 of LD-Hep of the outer core (Feldman *et al.*, 1999).

In this work, the structure of the O-specific polysaccharide (OPS) of the bovine mastitis isolate *E. coli* strain 1303 was elucidated and was shown to be interestingly linked to the K-12 core type. Also, it was proven that the K-12 core type was substituted by the OPS at O-7 of the terminal LD-heptose.

METHODS

Bacterial strain, isolation and degradation of the LPS.

Escherichia coli 1303, a well-characterized mastitis model strain, was isolated from udder secretions of a cow with clinical mastitis (Petzl *et al.*, 2008). Bacteria were grown in a 10 l fermenter (BIOFLO 110, New Brunswick Scientific) in Luria-Bertani medium, at pH 7.2, 40% dissolved oxygen and agitation between 300 and 900 r.p.m. The LPS was isolated utilizing the hot phenol/water procedure (Westphal & Jann, 1965), and purified by incubation with DNase and RNase (37 °C, 16 h, with gentle mixing) and proteinase K (56 °C, 6 h with gentle mixing) followed by ultracentrifugation (three times at 105 000 g, 4 °C, 4 h). Subsequently, the LPS (72 mg) was treated with 0.1 M sodium acetate buffer, pH 4.4, for 5 h at 100 °C, and the polysaccharide fraction was separated by size-exclusion chromatography (SEC) on a column of Toyo Pearl HW-40 in 0.05 M pyridinium acetate buffer, pH 4.5 (9 mg). The OPS fraction was further O-deacylated utilizing abs. hydrazine (37 °C, 30 min, 6.1 mg; Haishima *et al.*, 1992). Two other fractions were isolated, namely core substituted by an O-antigen (5.6 mg) and the core (4.5 mg). Additionally, another portion of LPS (100 mg) was directly O-deacylated and fractionated on Sephacryl 200 eluted with a buffer containing 0.25% sodium deoxycholate, 0.2 M NaCl, 1 mM EDTA and 10 mM Tris/HCl (pH 9.2). The fraction containing the O-deacylated LPS with a short O-antigen (32.6 mg) was used for mass spectrometry analyses.

General and analytical methods. The composition of the isolated fractions was determined by methanolysis (2 M HCl/MeOH, 85 °C, 2 h), followed by acetylation (85 °C, 10 min) and detection by GLC-MS [Hewlett Packard HP 5890 (series II) gas chromatograph equipped with a fused-silica SPB-5 column (Supelco, 30 m × 0.25 mm × 0.25 µm film thickness), FID and MS 5989A mass spectrometer with vacuum gauge controller 59827A]. The temperature programme was 150 °C for 3 min, then 5 °C min⁻¹ to 330 °C. Sugars were identified as their alditol acetates after hydrolysis (2 M trifluoroacetic acid, 120 °C, 2 h), reduction (NaBH₄, 16 h in the dark) and acetylation (85 °C for 10 min) (Sawardeker *et al.*, 1965) by GLC [HP 5890 (series II) gas chromatograph with FID and a column (30 m × 2.5 mm × 0.25 µm, Agilent Technologies) of polysilican SPD-5]. Helium was used as carrier gas (70 kPa). The temperature programme was 150 °C for 3 min, then

Fig. 1. Structure of the major core glycoform isolated after complete deacylation of the K-12 LPS (Müller-Loennies *et al.*, 2003).

3 °C min⁻¹ to 320 °C. The absolute configurations of the sugars were determined as described by Gerwig *et al.* (1979). Methylation was carried out according to Ciucanu & Kerek (1984).

SDS-PAGE and Western blotting. LPS (200 µg) was applied to one large slot (12.5 cm) and separated by SDS-PAGE on a 5% stacking and 15% separating gel at a constant voltage of 150 V. The gels were transferred overnight onto PVDF membranes (pore size 0.45 µm, Millipore) by tank blotting (Bio-Rad). Prior to use, the membranes were wetted in methanol for 10 s, after which they were washed in distilled water for at least 5 min. Following transfer, the blots were cut into strips (0.5 cm width) and placed in Mini-incubation trays (Bio-Rad). The following steps were performed at room temperature. After blocking in blotting buffer (50 mM Tris/HCl, 0.2 M NaCl, pH 7.4) supplemented with 10% non-fat dry milk for 1 h, the antibodies diluted in blocking buffer were added, incubated for 16 h and washed six times (5 min each) in blotting buffer. Alkaline-phosphatase-conjugated goat anti-mouse IgG (heavy and light chain specific, Dianova) was added (diluted 1:1000 in blotting buffer) and incubation was continued for another 2 h. After washing as before, 5-bromo-4-chloro-3-indoyl phosphate and *p*-toluidine *p*-nitro blue tetrazolium chloride (Bio-Rad) were added as substrates according to the supplier's instruction. After 15 min the reaction was stopped by the addition of distilled water. The monoclonal antibodies used were FDP-11 (specific for the R1 core type), FDP-3 (reacting with the R2 and K-12 core type) (both antibodies were obtained from F. Di Padova, Novartis Basel), S37-20 (specific for the R3 core type; unpublished data of H. Brade and others), S31-14 (specific for the K-12 core type; Brade *et al.*, 1996) and WN1 222-5 (reacting with all five *E. coli* core types; Di Padova *et al.*, 1993).

Mass spectrometry. Electrospray ionization Fourier-transformed ion cyclotron resonance (ESI FT-ICR) MS was performed in the negative-ion mode using an APEX Qe instrument (Bruker Daltonics) equipped with a 7 T magnet and a dual Apollo ion source. Mass spectra were acquired in broad band modes. The samples (~10 ng µl⁻¹) were dissolved in a 50:50:0.001 (by vol.) mixture of 2-propanol, water and triethylamine, and were sprayed at a flow rate of 2 µl min⁻¹. Capillary entrance voltage was set to 3.8 kV, and drying gas temperature to 150 °C. Mass spectra were calibrated externally by lipids of known structure, charge deconvoluted, and the given mass numbers were referred to the monoisotopic masses of neutral molecules.

NMR spectroscopy. NMR spectroscopy experiments were carried out after H-²H exchange of the samples utilizing 99.9% ²H₂O. All 1D (¹H, ¹³C), and 2D homonuclear (¹H, ¹H) correlation spectroscopy (COSY), total correlation spectroscopy (TOCSY), and rotating-frame Overhauser-effect spectroscopy (ROESY), as well as heteronuclear (¹H, ¹³C) single-quantum correlation-distortionless enhancement by polarization transfer (HSQC-DEPT) experiments of native OPS, core fraction with short O-antigen (core-short OPS) and non-substituted core fraction were recorded at 300 K with a Bruker DRX Avance 600 MHz spectrometer (operating frequencies 600.31 MHz for ¹H NMR, 150.96 MHz for ¹³C NMR), equipped with a 5 mm QXI multinuclear-inverse probehead with a *z* gradient, and applying standard Bruker software. Spectra of O-deacylated OPS were recorded at 300 K with a Bruker DRX Avance 700 MHz spectrometer (operating frequencies 700.75 MHz for ¹H NMR, 176.2 MHz for ¹³C NMR), equipped with a 5 mm CPQCI multinuclear-inverse cryo-probehead with a *z* gradient, and applying standard Bruker software. Chemical shifts were reported relative to an internal standard of acetone (δ_H 2.225, δ_C 31.45). Mixing times of 100 and 250 ms were used in TOCSY and ROESY experiments, respectively.

DNA sequence analysis. Total DNA of *E. coli* 1303 was prepared with the MasterPure DNA Purification kit (Epicentre) according to

the manufacturer's instructions. Sequencing was performed with a Genome Sequencer FLX system (Roche Applied Science). The resulting whole-genome shotgun reads were *de novo* assembled with the Roche Newbler assembly software (Margulies *et al.*, 2005).

The program ARTEMIS, version 11 (Rutherford *et al.*, 2000), was used for annotation. Homopolymer stretches leading to frameshifts in the O-antigen gene cluster were resolved by amplifying the region utilizing PCR and subsequent Sanger sequencing. BLAST and PSI-BLAST (Altschul *et al.*, 1997) were used for searching databases including GenBank, COG and the Pfam protein motif database (Bateman *et al.*, 2002). The program TMHMM 2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>) was used to identify potential transmembrane segments.

The DNA sequence of the O-antigen gene cluster of *E. coli* strain 1303 has been deposited in GenBank under the accession number FN995094.

RESULTS AND DISCUSSION

The structure of the OPS

The LPS was isolated from bacterial cells utilizing hot phenol/water extraction and subsequently purified by enzymic treatment and ultracentrifugation (yield: 1.23% of bacterial dry weight). Part of it was hydrolysed under mild acidic conditions to give the native OPS, which was purified by SEC (yield: 12.5% of the LPS) and further O-deacylated (O-deacylated OPS) by mild hydrazine treatment (37 °C, 30 min, yield: 8.5% of the LPS).

Compositional analyses of the native OPS fraction revealed the presence of fucose (Fuc), 3-amino-6-deoxyhexose (Qui3N), Gal and GalN. The absolute configurations of Gal and GalN were identified as D and that of Fuc as L. Methylation analysis of the O-deacylated OPS identified 1,3,5-tri-O-acetyl-6-deoxy-2,4-di-O-methylgalactose, 1,4,5-tri-O-acetyl-2,3,6-tri-O-methylgalactose, 1,4,5-tri-O-acetyl-3,6-dideoxy-2-O-methyl-3-methylamidoglucose and 1,3,5-tri-O-acetyl-2-deoxy-4,6-di-O-methyl-2-methylamidogalactose, indicating the OPS to be composed of the four units of 3-substituted L-Fuc, 4-substituted D-Gal, 4-substituted Qui3N and 3-substituted D-GalN. All residues were pyranoses.

These results were further confirmed by high-resolution ESI FT-ICR MS of the native OPS, which possessed groups of molecules representing the core oligosaccharides differing by one O-chain repeating unit. The measured mass difference of 740.285 Da was in excellent agreement with the calculated mass of 740.286 Da of a repeating unit consisting of 1 HexN, 1 deoxy-Hex, 1 Hex, 1 deoxy-HexN and 3 acetyl groups (C₃₀H₄₈O₁₉N₂) (mass spectrum not shown). Similar data with one acetyl group less (C₂₈H₄₆O₁₈N₂, 698.274 Da) were obtained for the O-deacylated OPS.

The structure of the O-repeating unit was established by NMR spectroscopy. The complete assignment of the OPS ¹H and ¹³C resonances (Table 1) was achieved by combining the information obtained from COSY, TOCSY and ROESY, as well as HSQC-DEPT experiments. The ¹H NMR spectrum of the native OPS (Fig. 2a) identified five signals in the anomeric region at δ_H 5.24,

Table 1. ^1H and ^{13}C NMR chemical shifts (δ , p.p.m.) of the native and O-deacylated OPS of *E. coli* 1303

Spectra were recorded at 27 °C in $^2\text{H}_2\text{O}$ relative to internal acetone (δ_{H} 2.225; δ_{C} 31.45). Underlined chemical shifts indicate substituted positions. Values of $^3J_{(1,2)}$ are in Hz; ND, not determined.

			1	$^3J_{(1,2)}$	2	3	4	5	6a	6b
Native OPS										
α -D-GalpNAc	A	H	5.24	(2.5)	4.36	3.91	4.29	4.11	3.75	
		C	98.08	ND	48.77	<u>78.50</u>	69.63	72.03	62.05	
α -L-FucpOAc	B	H	5.20	(2.2)	5.10	4.33	4.05	4.36	1.19	–
		C	94.07	ND	70.38	<u>77.40</u>	72.87	67.62	16.17	–
β -D-Quip3NAc	C	H	4.69	(6.0)	3.23	4.05	3.48	3.66	1.36	–
		C	105.12	ND	73.13	57.66	<u>77.11</u>	73.08	19.28	–
β -D-Galp	D	H	4.45	(6.9)	3.64	4.00	3.65	3.60	3.75	
		C	105.81	ND	69.97	66.02	<u>78.58</u>	75.91	62.05	
O-Deacylated OPS										
α -D-GalpNAc	A	H	5.25	ND	4.36	3.91	4.29	4.11	3.74	
		C	98.42	(178)	49.07	<u>78.88</u>	69.91	72.41	62.36	
α -L-Fucp	B'	H	5.06	ND	3.97	4.11	4.02	4.33	1.18	–
		C	96.89	(173)	68.38	<u>80.02</u>	72.96	67.94	16.36	–
β -D-Quip3NAc	C	H	4.74	(6.5)	3.31	4.07	3.51	3.66	1.36	–
		C	105.06	(165)	73.86	57.89	<u>77.47</u>	73.38	19.25	–
β -D-Galp	D	H	4.47	(6.6)	3.62	4.11	3.65	3.63	3.74	
		C	106.13	(160)	70.31	66.33	<u>78.93</u>	75.91	62.36	

5.20, 5.10, 4.69, 4.45. Based on the ^1H NMR spectrum of the O-deacylated OPS (Fig. 2b), the signal at δ_{H} 5.10 originated from H-2 of 2-O-Ac-Fuc. The rest of the anomeric protons were sequentially labelled from **A** to **D** in order of decreasing chemical shifts. The high-field region contained one O-acetyl signal at δ_{H} 2.12, two N-acetyl signals at δ_{H} 2.02 and 1.95, revealing that the amino sugars were N-acetylated, as well as two upfield signals characteristic of the 6-deoxy functions of Fuc (at δ_{H} 1.19) and of Qui3N (at δ_{H} 1.36).

By an HSQC experiment (Fig. 3) the direct correlation of all assigned ^1H signals with ^{13}C signals could be achieved.

The first spin system with the anomeric signal **A** (δ_{H} 5.24) originated from an α -GalpNAc residue. Strong intra-residual NOEs **A** H-3/H-4 and H-4/H-5 (data not shown) identified its *galacto*-configuration, whereas its α -configuration was established on the basis of a small $^3J_{1,2}$ coupling constant (2.5 Hz). Nitrogen substitution at C-2 was proven by a cross-peak H-2/C-2 on HSQC, where C-2 had a chemical shift characteristic of a carbon atom bearing an acetamido function (δ_{C} 48.77 for native OPS).

Residue **B** was identified as α -FucpOAc. The high-field signals of H-6 and C-6 of the methyl function proved the 6-deoxy group (δ_{H} 1.19, δ_{C} 16.17, native OPS). The *galacto*-configuration was established on the basis of the intra-residual NOE connectivities **B** H-3/H-4 and **B'** (O-deacylated sample) H-4/H-5. The upfield shifts of H-1 and H-2 of residue **B** from δ_{H} 5.20 and δ_{H} 5.10, respectively, to δ_{H} 5.04 and δ_{H} 3.97, respectively, in **B'** of the O-deacylated sample indicated that residue **B** carried an O-acetyl group at position 2. Additionally, the α -linkage was corroborated

by the intra-residual NOE signal **B'** H-1/H-2 and the small value of $^3J_{1,2}$ (2.2 Hz, native OPS). The O-acetyl substitution was non-stoichiometric.

Residue **C** was identified as 3,6-dideoxy-3-acetamidoglucose (Qui3NAc), having characteristic 6-deoxy sugar high-field signals of H-6/C-6 (δ_{H} 1.36, δ_{C} 19.28, native OPS). Additionally, C-3 possessed a chemical shift characteristic of a carbon atom bearing an acetamido function (δ_{C} 57.66, native OPS). The *gluco*-configuration was proven based on the intra-residual NOE signal **C** H-2/H-4 and the β -configuration on the NOE connectivities **C** H-1/H-3 and **C** H-1/H-5 as well as on the large $^3J_{1,2}$ value (6.0 Hz, native OPS). Its D-configuration was deduced after comparison of the obtained chemical shift values with those published (MacLean & Perry, 1997).

Residue **D** was assigned as β -Galp. Its β -configuration was deduced from the large $^3J_{1,2}$ coupling constant (6.9 Hz, native OPS) and the *galacto*-configuration from the intra-residual NOE contact **D** H-3/H-4.

The monosaccharide sequence of the sugars in the OPS was established from the observed inter-residual NOE cross-peaks in the ROESY spectra, i.e. **A** H-1/C H-4, **B'** H-1/D H-4, **C** H-1/B' H-3, **D** H-1/A H-3.

On the basis of the above data the OPS of *E. coli* 1303 had the structure

Fig. 2. ^1H NMR spectrum of the native OPS (600 MHz, 300K, $^2\text{H}_2\text{O}$) (a) and O-deacylated OPS (700 MHz, 300 K, $^2\text{H}_2\text{O}$) (b) of *E. coli* 1303.

The structure of the OPS of the mastitis isolate *E. coli* 1303 showed great similarities with, but was not identical to, those present in the LPS of two subtypes of the *E. coli* O5 serotype known to date, namely O5ab (MacLean & Perry, 1997) and O5ac (strain 180/C3) (Urbina *et al.*, 2005). In OPS of *E. coli* 1303 α -L-FucpOAc was present instead of β -D-Ribf as in OPS of O5ab and O5ac, and additionally β -D-Quip3NAc was substituted at position 4 and not at position 2 as in the O5ac strain.

E. coli O5ab:

$[-\rightarrow 4)-\beta\text{-D-Quip3NAc}-(1\rightarrow 3)-\beta\text{-D-Ribf}-(1\rightarrow 4)-\beta\text{-D-Galp}-(1\rightarrow 3)-\alpha\text{-D-GalpNAc}-(1\rightarrow)]-$

E. coli O5ac:

$[-\rightarrow 2)-\beta\text{-D-Quip3NAc}-(1\rightarrow 3)-\beta\text{-D-Ribf}-(1\rightarrow 4)-\beta\text{-D-Galp}-(1\rightarrow 3)-\alpha\text{-D-GalpNAc}-(1\rightarrow)]-$

Sequence analysis of the O-antigen gene cluster

The O-antigen determinant located between *galF* and *gnd* on the *E. coli* strain 1303 chromosome, as also reported for

Fig. 3. Heteronuclear 2D ^{13}C - ^1H chemical shift correlation of the anomeric and ring region resonances of the O-deacylated OPS of *E. coli* 1303. The spectrum was recorded at 700 MHz and 300 K.

other O-antigen clusters (Reeves & Wang, 2002), was sequenced and the genetic structure of this 13 095 bp DNA region was analysed in detail. The O-antigen gene cluster has an overall G+C content of 36.9 mol% (Fig. 4a). As also described for other *E. coli* O antigen gene clusters, all predicted ORFs, with the exception of one transposase-encoding ORF, have a lower G+C content than the average *E. coli* genome, suggesting that they may have been acquired by horizontal transfer from other species (Reeves & Wang, 2002).

Thirteen putative open reading frames (ORFs) with the same transcriptional direction were identified as shown in Fig. 4(b). The majority of the putative ORFs represented unknown nucleotide sequences that have not been described before. Whereas the *rmlB* gene (ORF 1) and an IS4-family transposase-encoding gene (ORF 12) have been described in other *E. coli* isolates as well, only *rmlA* (ORF 2) and ORF 3 exhibited similarity (72 % and 74 % identity) to fragments of the glucose-1-phosphate thymidyltransferase-encoding gene Abu_1817 of *Arcobacter butzleri* RM4018 (accession no. CP000361) and *fdtA* of *Salmonella enterica* subsp. *enterica* serovar Pomona strain NML 07-0213 (accession no. EU805803), respectively.

The gene products of ORF 1 (*rmlB*) and ORF 2 (*rmlA*) showed 92 % and 82 % identity to other known RmlB and RmlA homologues, respectively (Table 2; see also Supplementary Table S1, available with the online version of this paper). The *rmlA* and *rmlB* genes have been well characterized in *E. coli*. RmlA converts D-glucose 1-phosphate to dTDP-D-glucose, which is then converted by RmlB to dTDP-4-dehydro-6-deoxy-D-glucose. The latter

Fig. 4. Genetic structure of the antigen-encoding determinant of the mastitis isolate *E. coli* 1303. (a) G+C content of the O-antigen gene cluster relative to the average G+C content of the *E. coli* core chromosome (50.8 %, horizontal line). (b) The positions and transcriptional directions of identified putative ORFs are indicated by arrows. The *galF* and *gnd* genes flanking the O-antigen gene cluster are indicated in black. ORFs with high similarity to known bacterial DNA sequences are indicated in grey.

compound is a common intermediate of many different sugars (Graninger *et al.*, 2002; Samuel & Reeves, 2003). Similar to this OPS and the O5 oligosaccharide variants, the *E. coli* O91 oligosaccharide contains, among other sugar residues, a modified D-Quip3N. For the corresponding *wb*_{O91}* gene cluster the *wbsB* gene has been predicted to code for an isomerase catalysing the conversion of dTDP-4-dehydro-6-deoxy-D-glucose into dTDP-3-dehydro-6-deoxy-D-glucose (Perelle *et al.*, 2002). Also in *Thermoanaerobacterium thermosaccharolyticum* dTDP- α -D-Quip3NAc biosynthesis has been shown to include an isomerase which catalyses the formation of dTDP-3-

dehydro-6-deoxy-D-glucose from the RmlB product dTDP-4-dehydro-6-deoxy-D-glucose. In the case of the O-antigen determinant of *E. coli* 1303, the ORF 3-encoded protein contained a C-terminal isomerase domain of WxcM-like proteins (PF05523) and was 69 % and 46 % identical to the dTDP-4-oxo-6-deoxy-D-glucose-3,4-oxoisomerase WbsB (accession no. AAK60451) of *E. coli* O91 and the QdtA isomerase (accession no. AAR85518) of *T. thermosaccharolyticum* E207-71, respectively. Therefore, ORF 3 (*qdtA*) may encode an isomerase responsible for the formation of dTDP-3-dehydro-6-deoxy-D-glucose. In *T. thermosaccharolyticum*, 3-acetamido-3,

Table 2. Characteristics of the ORFs located in the O-antigen gene cluster of *E. coli* 1303

An extended version of this table is available as Supplementary Table S1 with the online version of this paper.

Putative ORF no.	Designation	Length (bp)	G+C content (mol%)	No. of aa	Putative function of protein
1	<i>rmlB</i> (GI:315461754)	1086	43.4	361	dTDP-glucose 4,6-dehydratase
2	<i>rmlA</i> (GI:315461755)	870	37.8	289	Glucose-1-phosphate thymidyltransferase
3	<i>qdtA</i> (GI:315461756)	408	36.3	135	dTDP-4-oxo-6-deoxy-D-glucose-3,4-oxoisomerase
4	<i>qdtB</i> (GI:315461757)	1104	36.8	367	Transaminase
5	<i>wzx</i> (GI:315461758)	1251	34.9	416	O-antigen flippase
6	<i>wbnC</i> (GI:315461759)	513	30.01	170	Acetyltransferase
7	ORF 7 (GI:315461760)	921	30.8	306	Glycosyltransferase
8	<i>wzy</i> (GI:315461761)	1281	30.1	426	O-antigen polymerase
9	ORF 9 (GI:315461762)	1083	30.8	360	Glycosyltransferase
10	<i>wcaG</i> (GI:315461763)	924	30.8	307	Nucleoside-diphosphate-sugar epimerase
11	ORF 11 (GI:315461764)	810	34.2	269	Glycosyltransferase
12	<i>tnp</i> (GI:315461765)	1119	41.1	372	Transposase
13	<i>qdtC</i> (GI:315461766)	483	32.7	160	Transacetylase

6-dideoxy- α -D-glucose formation involves subsequent reaction steps catalysed by a transaminase and a transacetylase (Pfösl *et al.*, 2008). The corresponding enzymes of the tested O-antigen gene cluster may be encoded by ORF 4 and ORF 13. The ORF 4 (*qdtB*)-encoded gene product belonged to the DegT/DnrJ/EryC1/StrS aminotransferase family and showed 49 % identity to the transaminase QdtB (accession no. AAR85519) of *T. thermosaccharolyticum* E207-71. ORF 13 (*qdtC*) coded for a putative transacetylase with a WcxM-like, left-handed parallel β -helix (LbH) N-terminal domain responsible for the transacetylation function. This protein exhibited 76 % identity to the transacetylase QdtC of *Salmonella choleraesuis* (accession no. D7PFB5) and 44 % identity to the QdtC transacetylase of *T. thermosaccharolyticum* E207-71 (accession no. AAR85517).

The OPS of *E. coli* 1303 also contained α -L-FucpOAc. Fucose biosynthesis involves a three-step pathway for converting GDP-D-mannose to GDP-L-fucose (Ginsburg, 1961). First, the GDP-mannose dehydratase Gmd converts GDP-mannose into GDP-4-keto-6-deoxymannose. This is followed by epimerase and reductase reaction steps to give GDP-L-fucose. In *E. coli*, the latter two steps are catalysed by a single bifunctional enzyme, the GDP-fucose synthase Fcl (Samuel & Reeves, 2003). The *gmd* and *fcl* genes are usually adjacently located within the colanic acid gene cluster of *E. coli*, as has been shown for *E. coli* MG1655 (Andrianopoulos *et al.*, 1998). In the *wb*₁₃₀₃* gene cluster, the protein encoded by ORF 10 exhibited 42 % and 22 % identity to the putative UDP-glucose-4-epimerase EDWATA_01329 (accession no. D4F3L9) of *Edwardsiella tarda* ATCC 23685 and the GDP-fucose synthase Fcl (accession no. NP_416556) of *Escherichia coli* MG1655. The presence of a conserved domain of the NAD-dependent epimerase/dehydratase family (PF01370), which also included the conserved WcaG nucleoside-diphosphate-sugar epimerase domain (COG0451), further supports the idea that this protein may be involved in fucose biosynthesis. The protein encoded by ORF 6 (*wbnC*) belonged to the trimeric LpxA-like enzymes superfamily of acetyltransferases from a wide range of bacteria and is 39 % identical to the O-acetyltransferase WbnC (Q9RP59) of *E. coli* O113. Thus, WbnC is likely to be responsible for O-acetylation of Fuc in the tested O-polysaccharide, although further experimental determination is needed. The O-antigen flippase (Wzx) and O-antigen polymerase (Wzy) are hydrophobic membrane proteins involved in O-antigen processing. In *E. coli* strain 1303, 12 transmembrane helices were predicted for the deduced amino acid sequence of ORF 5 (*wzx*), which shared 42 % identity with the *E. coli* O114 Wzx protein (Q697E1). ORF 8 (*wzy*) showed no marked similarity to other *wzy* genes, but as the deduced amino acid sequence exhibited some homology to the putative O-antigen polymerases it was considered as the putative *wzy* gene. Furthermore, 11 transmembrane helices have been predicted from the deduced amino acid sequence. This is also the case for the putative Wzy proteins of an *E. coli* O4 strain (accession no. AAC43898) and of *E. coli* K-12

(accession no. AAB88404). Generally, the number of transmembrane helices of Wzy proteins of other *E. coli* serogroups is variable, ranging from 8 to 12. A large number of transmembrane segments and a large periplasmic loop are typical topological characteristics of Wzy proteins (Daniels *et al.*, 1998).

Four additional ORFs involved in O-antigen biosynthesis have been predicted (ORFs 7, 9, 11 and 12). Their precise function will have to be studied experimentally. ORF 7 coded for a member of the glycosyltransferase family 2 which is 41 % and 34 % identical to the glycosyltransferase ESA_01184 (accession no. A9Y3E9) of *Enterobacter sakazakii* (Mullane *et al.*, 2008) or putative glycosyltransferase WbtE (accession no. Q6QNC3) of *Escherichia coli* O103. The ORF 9-encoded protein belonged to the group 1 glycosyltransferases and showed 30 % identity to the glycosyltransferase WclF of *E. coli* O155 (accession no. AAV74551) and 27 % identity to WbuB (accession no. AAT28929), a protein encoded in the *E. coli* O26 O-antigen gene cluster (D'Souza *et al.*, 2002). The fucosyltransferase WbuB has been proposed to be a transferase for the linkage α -L-FucpNAc-(1 $\rightarrow$ 3)- α -D-GlcpNAc (D'Souza *et al.*, 2002). This linkage is absent in the *E. coli* 1303 O-antigen, but as an α -L-Fucp2OAc-(1 $\rightarrow$ 4)- β -D-Gal linkage is present, ORF 9 may encode a FucpOAc transferase. The ORF 11 gene product belonged to the glycosyltransferase family 2 and shares 60 % or 48 % amino acid identity with glucosyltransferase WbeD (accession no. B8R1W7) of *Salmonella enterica* serovar Pomona and *E. coli* O107 (accession no. B8QSK1), respectively. In addition, the ORF 11-encoded protein showed 49 % identity to WbwC of *E. coli* O104, which is a galactosyltransferase responsible for the β -D-Gal-(1 $\rightarrow$ 3)- β -D-GalNAc linkage (Wang *et al.*, 2009). Accordingly, ORF 11 encoded a galactosyltransferase which may link the Galp and GalpNAc moieties present in the 1303 OPS.

ORF 12 encoded a protein unrelated to O-antigen biosynthesis. Together with its sequence context, ORF 12 was identified as a remnant of an H-repeat-type transposable element (positions 1863–3744). H-repeats may be involved in horizontal gene transfer and in the generation of polymorphisms in O-antigen gene clusters (Xiang *et al.*, 1994).

In *E. coli*, the genes for the synthesis of nucleotide precursors of common sugars, e.g. GlcpNAc, Glcp and Galp, are usually located outside the O-antigen gene cluster (Samuel & Reeves, 2003). UDP-GalNAc is synthesized from UDP-GlcNAc by the UDP-GlcNAc-4-epimerase encoded by the *gne* gene, which can be part of the O-antigen determinant located between *galF* and *gnd*. Alternatively, *gne* can be immediately upstream of *galF* (Wang *et al.*, 2005). In *E. coli* strain 1303, the *gne* gene is absent from the O-antigen gene cluster and is likely to be upstream of *galF*. O-repeating unit synthesis in enterobacteria is often initiated by transferring GlcNAc 1-phosphate or GalNAc 1-phosphate to an undecaprenol

phosphate carrier. The corresponding enzyme WecA is encoded outside the O-antigen gene cluster as well as other O-antigen processing proteins, such as the chain-length determinant Wzz (Samuel & Reeves, 2003).

Identification of the core oligosaccharide type and the linkage site of the OPS

Western blot analysis of LPS from *E. coli* 1303 was performed with monoclonal antibodies specific for the different *E. coli* core types; this revealed that strain 1303 carried the K-12 core type in its LPS (Fig. 5). This result was rather unexpected, since (i) this *E. coli* core type had been detected earlier in only 4% of faecal human and bovine isolates (Gibbs *et al.*, 2004), (ii) *E. coli* K-12 strains that are widely used in laboratories produce an R-form LPS lacking OPS repeating units (Feldman *et al.*, 1999), and (iii) an *E. coli* K-12 strain in which O-antigen

assembly was restored exhibited serotype O16 (Liu & Reeves, 1994).

The ESI MS spectrum of O-deacylated LPS (Fig. 6) comprised four groups of molecules showing heterogeneity originating from non-stoichiometric substitutions with PEtN, sodium and potassium adducts (not labelled). The first complex group of molecular ions around 2957.93 u referred to the O-deacylated lipid A+core whereas the three other groups around 3790.26, 4488.55 and 5186.81 u represented the O-deacylated lipid A, the core+one, two and three repeating units, respectively, in which each repeating unit consisted of HexN, deoxy-Hex, Hex, deoxy-HexN and two *N*-acetyl groups, with a total mass of 698.27 u. The most prominent molecular peak at 2957.93 u represented a molecule composed of 4 P, 1 PEtN, 2 HexN, 2 14:0(3-OH), 3 Kdo, 3 Hep and 3 Hex corresponding to O-deacylated lipid A plus a truncated K-12 core region with an additional third Kdo residue, as compared to the published data (Holst *et al.*, 1991; Müller-Loennies *et al.*, 2003).

The component belonging to the molecular peak at 2737.86 u was composed of 4 P, 1 PEtN, 2 HexN, 2 14:0(3-OH), 3 Hep, 3 Hex and 2 Kdo, and was the core species to which OPS was attached after previous addition of 1 Hex moiety and 1 Hep moiety. The molecular ion corresponding to the complete core (Müller-Loennies *et al.*, 2003) was not observed; however, the molecule consisting of the complete core and one O-deacylated O-repeating unit (HexN, deoxy-Hex, Hex, deoxy-HexN, two *N*-acetyl groups, calculated mass 698.27 u) was found at 3790.26 u. Thus, the mass at 2957.93 u originated from the core oligosaccharide that was not substituted by OPS, and that at 3790.26 u from a core oligosaccharide substituted by one O-antigen repeating unit. Interestingly, the non-substituted core oligosaccharide differed in structure from those substituted with the OPS, i.e. it represented a truncated version with an additional, third Kdo residue. Moreover, the substituted core corresponded to the most prominent glycoform, namely glycoform 1 of the K-12 core type (Müller-Loennies *et al.*, 2003), and the non-substituted oligosaccharide represented a novel glycoform of the K-12 structure, i.e. a truncated core lacking any rhamnose residues. Previously the occurrence of three Kdo moieties in the truncated K-12 core was associated with simultaneous presence of one rhamnose (Rha) residue (Müller-Loennies *et al.*, 2003).

The NMR study of the fraction core-short OPS isolated from the LPS after acetate buffer hydrolysis and separation on Toyo Pearl HW-40 identified the biological O-repeating unit of the OPS from *E. coli* 1303 and the linkage site to the core region. The spin systems of two OPS repeating units, one terminal (with a terminal β -D-Quip3NAc residue) and one connected to the core (where the 3-substituted D-GalpNAc was shown to be β -configured, and not α -configured as in the other O-repeating units) were identified. Due to the high heterogeneity of the sample, the complete spin systems of only the three first sugars of the

Fig. 5. Western blot of LPS from *E. coli* 1303 strain (amounts given on the figure) with monoclonal antibodies specific for different *E. coli* core types: WN1 222-5, binds to all five *E. coli* core types (Di Padova *et al.*, 1993); FDP-11, specific for R1 core type; FDP-3, specific for R2 and K-12 core types; S37-20, specific for R3 core type; and S31-14, specific for K-12 core type (Brade *et al.*, 1996).

Fig. 6. Charge-deconvoluted ESI FT-ICR MS spectrum of O-deacylated LPS from *E. coli* 1303 recorded in the negative-ion mode. The molecular peak at 2957.93 u corresponds to a molecule composed of 2 P, 2 HexN, 2 14:0(3-OH) (O-deacylated lipid A), 2 P, 1 PEtN, 3 Kdo, 3 Hep, 3 Hex (core region). The component with molecular mass 3790.26 u lacked 1 Kdo residue and possessed in addition 1 Hep, 1 Hex and the first O-repeating unit, as compared to the non-substituted core. P, phosphate; HexN, hexosamine; 14:0(3-OH), 3-hydroxymyristic acid; PEtN, 2-aminoethanol phosphate; Kdo, 3-deoxy-D-manno-oct-2-ulosonic acid; Hep, L-glycero-D-manno-heptose; Hex, hexose. Δm 698.27 u corresponded to the mass of one O-deacylated O-repeating unit.

outer core region (7-substituted L- α -D-Hep and 6- and 2-substituted α -D-Glc) could be resolved (data not shown), as compared with published data (Müller-Loennies *et al.*, 2003). Thus, the disaccharide β -D-GalpNAc-(1 $\rightarrow$ 7)-L- α -D-Hep was proven, which identified the site of attachment of the OPS at core oligosaccharide:

This is consistent with the data obtained for *E. coli* K-12/O16 MFF1, a mutant strain expressing LPS consisting

Fig. 7. Overlay of the anomeric regions of HSQC-DEPT spectra of the non-substituted core (red) and core-short OPS fraction (green) of *E. coli* 1303. The spectra were recorded at 700 MHz and 300 K. The signals differing between the two fractions are marked in grey. The structures of the non-substituted, truncated core (red) and substituted (green) are given below, drawn according to Müller-Loennies *et al.* (2003).

of lipid A-core region plus the first sugar of the O-repeating unit (serotype O16), namely β -D-GlcpNAc. The latter was shown to be linked to the position O-7 of the terminal L- α -D-Hepp moiety of the outer core (Feldman *et al.*, 1999).

The comparison of the anomeric regions of the core and core-short OPS fractions in the HSQC spectrum (Fig. 7) revealed that in the first fraction signals originating from the O-antigen, from 7-substituted L- α -D-Hepp and from 6-substituted and 2-substituted α -D-Glcp were missing; however, an additional signal at δ_H 5.38 p.p.m. was seen, which was identified as terminal α -D-Glcp. This confirmed the data obtained from ESI MS analysis, i.e. that the non-substituted core represented a shorter K-12 core oligosaccharide which lacked the terminal L- α -D-Hepp and 6-substituted α -Glcp. Since the samples were obtained after acetate buffer hydrolysis that cleaved any branching Kdo residue(s), no conclusions concerning the number of Kdo residues could be drawn from these data.

The presence of different glycoforms of *E. coli* K-12 core was already well established with the identification of four core structures differing in length, and amount of Kdo, P and PEtN residues (Holst *et al.*, 1991; Müller-Loennies *et al.*, 2003). However, the previous studies were performed on R-form LPS and nothing could be concluded about the core structure(s) in the S-form. Here, we have shown that the core substituted with OPS from the LPS of a mastitis *E. coli* was longer than the non-substituted, having additional Hep and Hex residues. Such core heterogeneity based on the presence/absence of an O-chain was also observed in LPS from *Bordetella parapertussis*, in which the core oligosaccharide from the R-form LPS possessed three sugar residues more than the core OS from the R-form LPS (Zarrouk *et al.*, 1997). In LPS from *Pseudomonas aeruginosa* serotype O5 the addition of the OPS to the core followed the translocation of a Rha residue and loss of a Glc residue, as compared to the non-substituted core (Sadovskaya *et al.*, 2000). Interestingly, the same alteration in the core structure as found in *E. coli* 1303, namely the addition of the third Kdo moiety in the inner core and the truncation of the outer core, was also observed in an *E. coli* strain which possessed the *waaZ* gene encoding the transferase of the third Kdo. WaaZ is present only in *E. coli* having the K-12 core type; however, if a copy of *waaZ* was added by a plasmid to *E. coli* with the R1 core type, novel truncated LPS with the additional third Kdo residue was expressed, which is otherwise not found in this chemotype (Firdich *et al.*, 2003).

ACKNOWLEDGEMENTS

We thank Herman Moll for help with GC/MS analysis, Heiko Käbner for NMR recording, Dr Patricia Sánchez Carballo for valuable discussions and Dr F. Di Padova for providing monoclonal antibodies WN1 222-5, FDP-3 and FDP-11. This work was supported by the Deutsche Forschungsgemeinschaft (FOR 585, project 8, and DO789/3-1 and DO789/4-1).

REFERENCES

- 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.
- Amor, K., Heinrichs, D. E., Firdich, E., Ziebell, K., Johnson, R. P. & Whitfield, C. (2000). Distribution of core oligosaccharide types in lipopolysaccharides from *Escherichia coli*. *Infect Immun* **68**, 1116–1124.
- Andrianopoulos, K., Wang, L. & Reeves, P. R. J. (1998). Identification of the fucose synthetase gene in the colanic acid gene cluster of *Escherichia coli* K-12. *J Bacteriol* **180**, 998–1001.
- Bateman, A., Birney, E., Cerruti, L., Durbin, R., 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.
- Brade, L., Grimmecke, H. D., Holst, O., Brabetz, W., Zamojski, A. & Brade, H. (1996). Specificity of monoclonal antibodies against *Escherichia coli* K-12 lipopolysaccharide. *J Endotoxin Res* **3**, 39–47.
- Burvenich, C., Van Merris, V., Mehrzad, J., Diez-Fraile, A. & Duchateau, L. (2003). Severity of *E. coli* mastitis is mainly determined by cow factors. *Vet Res* **34**, 521–564.
- Ciucanu, I. & Kerek, F. (1984). A simple and rapid method for the permethylation of carbohydrates. *Carbohydr Res* **131**, 209–217.
- D'Souza, J. M., Wang, L. & Reeves, P. (2002). Sequence of the *Escherichia coli* O26 O antigen gene cluster and identification of O26 specific genes. *Gene* **297**, 123–127.
- Daniels, C., Vindurampulle, C. & Morona, R. (1998). Overexpression and topology of the *Shigella flexneri* O-antigen polymerase (Rfc/Wzy). *Mol Microbiol* **28**, 1211–1222.
- Di Padova, F. E., Brade, H., Barclay, G. R., Poxton, I. R., Liehl, E., Schuetze, E., Kocher, H. P., Ramsay, G., Schreier, M. H. & other authors (1993). A broadly cross-protective monoclonal antibody binding to *Escherichia coli* and *Salmonella* lipopolysaccharides. *Infect Immun* **61**, 3863–3872.
- Feldman, M. F., Marolda, C. L., Monteiro, M. A., Perry, M. B., Parodi, A. J. & Valvano, M. A. (1999). The activity of a putative polyisoprenol-linked sugar translocase (Wzx) involved in *Escherichia coli* O antigen assembly is independent of the chemical structure of the O repeat. *J Biol Chem* **274**, 35129–35138.
- Firdich, E., Lindner, B., Holst, O. & Whitfield, C. (2003). Overexpression of the *waaZ* gene leads to modification of the structure of the inner core region of *Escherichia coli* lipopolysaccharide, truncation of the outer core, and reduction of the amount of O polysaccharide on the cell surface. *J Bacteriol* **185**, 1659–1671.
- Gerwig, G. J., Kamerling, J. P. & Vliegenthart, J. F. (1979). Determination of the absolute configuration of mono-saccharides in complex carbohydrates by capillary G.L.C. *Carbohydr Res* **77**, 10–17.
- Gibbs, R. J., Stewart, J. & Poxton, I. R. (2004). The distribution of, and antibody response to, the core lipopolysaccharide region of *Escherichia coli* isolated from the faeces of healthy humans and cattle. *J Med Microbiol* **53**, 959–964.
- Ginsburg, V. (1961). Studies on the biosynthesis of guanosine diphosphate L-fucose. *J Biol Chem* **236**, 2389–2393.
- Graninger, M., Kneidinger, B., Bruno, K., Scheberl, A. & Messner, P. (2002). Homologs of the Rml enzymes from *Salmonella enterica* are responsible for dTDP- β -L-rhamnose biosynthesis in the gram-positive thermophile *Aneurinibacillus thermoaerophilus* DSM 10155. *Appl Environ Microbiol* **68**, 3708–3715.

- Gray, C. H. & Tatum, E. L. (1944). X-ray induced growth factor requirements in bacteria. *Proc Natl Acad Sci U S A* **30**, 404–410.
- Haishima, Y., Holst, O. & Brade, H. (1992). Structural investigation on the lipopolysaccharide of *Escherichia coli* rough mutant F653 representing the R3 core type. *Eur J Biochem* **203**, 127–134.
- Heinrichs, D. E., Yethon, J. A. & Whitfield, C. (1998). Molecular basis for structural diversity in the core regions of the lipopolysaccharides of *Escherichia coli* and *Salmonella enterica*. *Mol Microbiol* **30**, 221–232.
- Holst, O. (1999). Chemical structure of the core region of lipopolysaccharides. In *Endotoxins in Health and Disease*, pp. 115–154. Edited by D. Morrison, H. Brade, S. Opal & S. Vogel. New York: Marcel Dekker.
- Holst, O., Zähringer, U., Brade, H. & Zamojski, A. (1991). Structural analysis of the heptose/hexose region of the lipopolysaccharide from *Escherichia coli* K-12 strain W3100. *Carbohydr Res* **215**, 323–335.
- Holst, O., Moran, A. P. & Brennan, P. J. (2009). *Microbial Glycobiology. Structures, Relevance and Applications*, pp. 3–13. Edited by A. P. Moran, O. Holst, P. J. Brennan & M. von Itzstein. Amsterdam: Academic Press.
- Jann, B., Shashkov, A. S., Kochanowski, H. & Jann, K. (1994). Structure of the O16 polysaccharide from *Escherichia coli* O16:K1: an NMR investigation. *Carbohydr Res* **264**, 305–311.
- Jansson, P. E., Lindberg, A. A., Lindberg, B. & Wollin, R. (1981). Structural studies on the hexose region of the core in lipopolysaccharides from *Enterobacteriaceae*. *Eur J Biochem* **115**, 571–577.
- Liu, D. & Reeves, P. R. (1994). *Escherichia coli* K12 regains its O antigen. *Microbiology* **140**, 49–57.
- MacLean, L. L. & Perry, M. B. (1997). Structural characterization of the serotype O:5 O-polysaccharide antigen of the lipopolysaccharide of *Escherichia coli* O:5. *Biochem Cell Biol* **75**, 199–205.
- Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y. J. & other authors (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376–380.
- Mullane, N., O'Gaora, P., Nally, J. E., Iversen, C., Whyte, P., Wall, P. G. & Fanning, S. (2008). Molecular analysis of the *Enterobacter sakazakii* O-antigen gene locus. *Appl Environ Microbiol* **74**, 3783–3794.
- Müller-Loennies, S., Lindner, B. & Brade, H. (2002). Structural analysis of deacylated lipopolysaccharide of *Escherichia coli* strains 2513 (R4 core-type) and F653 (R3 core-type). *Eur J Biochem* **269**, 5982–5991.
- Müller-Loennies, S., Lindner, B. & Brade, H. (2003). Structural analysis of oligosaccharides from lipopolysaccharide (LPS) of *Escherichia coli* K12 strain W3100 reveals a link between inner and outer core LPS biosynthesis. *J Biol Chem* **278**, 34090–34101.
- Perelle, S., Dilasser, F., Grout, J. & Fach, P. (2002). Identification of the O-antigen biosynthesis genes of *Escherichia coli* O91 and development of a O91 PCR serotyping test. *J Appl Microbiol* **93**, 758–764.
- Petzl, W., Zerbe, H., Günther, J., Yang, W., Seyfert, H. M., Nürnberg, G. & Schubert, H. J. (2008). *Escherichia coli*, but not *Staphylococcus aureus* triggers an early increased expression of factors contributing to the innate immune defense in the udder of the cow. *Vet Res* **39**, 18.
- Pfösti, A., Zayni, S., Hofinger, A., Kosma, P., Schäffer, C. & Messner, P. (2008). Biosynthesis of dTDP-3-acetamido-3,6-dideoxy- α -D-glucose. *Biochem J* **410**, 187–194.
- Raetz, C. R. & Whitfield, C. (2002). Lipopolysaccharide endotoxins. *Annu Rev Biochem* **71**, 635–700.
- Reeves, P. P. & Wang, L. (2002). Genomic organization of LPS-specific loci. *Curr Top Microbiol Immunol* **264**, 109–135.
- Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M. A. & Barrell, B. (2000). Artemis: sequence visualization and annotation. *Bioinformatics* **16**, 944–945.
- Sadovskaya, I., Brisson, J. R., Thibault, P., Richards, J. C., Lam, J. S. & Altman, E. (2000). Structural characterization of the outer core and the O-chain linkage region of lipopolysaccharide from *Pseudomonas aeruginosa* serotype O5. *Eur J Biochem* **267**, 1640–1650.
- Samuel, G. & Reeves, P. R. J. (2003). Biosynthesis of O-antigens: genes and pathways involved in nucleotide sugar precursor synthesis and O-antigen assembly. *Carbohydr Res* **338**, 2503–2519.
- Sawardeker, J. S., Sloneker, J. H. & Jeanes, A. (1965). Quantitative determination of monosaccharides as their alditol acetates by gas liquid chromatography. *Anal Chem* **37**, 1602–1604.
- Stenutz, R., Weintraub, A. & Widmalm, G. (2006). The structures of *Escherichia coli* O-polysaccharide antigens. *FEMS Microbiol Rev* **30**, 382–403.
- Stevenson, G., Neal, B., Liu, D., Hobbs, M., Packer, N. H., Batley, M., Redmond, J. W., Lindquist, L. & Reeves, P. (1994). Structure of the O antigen of *Escherichia coli* K-12 and the sequence of its *rfb* gene cluster. *J Bacteriol* **176**, 4144–4156.
- Urbina, F., Nordmark, E. L., Yang, Z., Weintraub, A., Scheutz, F. & Widmalm, G. (2005). Structural elucidation of the O-antigenic polysaccharide from the enteroaggregative *Escherichia coli* strain 180/C3 and its immunochemical relationship with *E. coli* O5 and O65. *Carbohydr Res* **340**, 645–650.
- Vinogradov, E. V., Van Der Drift, K., Thomas-Oates, J. E., Meshkov, S., Brade, H. & Holst, O. (1999). The structures of the carbohydrate backbones of the lipopolysaccharides from *Escherichia coli* rough mutants F470 (R1 core type) and F576 (R2 core type). *Eur J Biochem* **261**, 629–639.
- Wang, L., Liu, B., Kong, Q., Steinerück, H., Krause, G., Beutin, L. & Feng, L. (2005). Molecular markers for detection of pathogenic *Escherichia coli* strains belonging to serogroups O 138 and O 139. *Vet Microbiol* **111**, 181–190.
- Wang, Q., Perepelov, A. V., Feng, L., Knirel, Y. A., Li, Y. & Wang, L. (2009). Genetic and structural analyses of *Escherichia coli* O107 and O117 O-antigens. *FEMS Immunol Med Microbiol* **55**, 47–54.
- Westphal, O. & Jann, K. (1965). Bacterial lipopolysaccharide extraction with phenol-water and further application of procedure. *Met Carbohydr Chem* **5**, 83–87.
- Xiang, S. H., Hobbs, M. & Reeves, P. R. (1994). Molecular analysis of the *rfb* gene cluster of a group D2 *Salmonella enterica* strain: evidence for its origin from an insertion sequence-mediated recombination event between group E and D1 strains. *J Bacteriol* **176**, 4357–4365.
- Zarrouk, H., Karibian, D., Godard, I., Perry, M. B. & Caroff, M. (1997). Use of mass spectrometry to compare three O-chain linked and free lipopolysaccharide cores: differences found in *Bordetella parapertussis*. *J Endotoxin Res* **4**, 453–458.

Edited by: G. H. Thomas