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 Dobrindt4,6 and Otto Holst1

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 \( \beta-D-Quio3NAC-(1\rightarrow3)-\alpha-L-Fucp2OAc-(1\rightarrow4)\beta-D-Galp-(1\rightarrow3)-\alpha-D-GalpNAC-(1\rightarrow) \) 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.

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 \[\text{LD}-\text{Hep}\rightarrow (1\rightarrow 7)-[\text{Glc}\rightarrow (1\rightarrow 3)]-\text{LD}-\text{Hep}\rightarrow (1\rightarrow 3)-\text{LD}-\text{Hep}\rightarrow (1\rightarrow 5)-[\text{Kdo}\rightarrow (2\rightarrow 4)]-\text{Kdo}\]. 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, 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) (Savardeker et al., 1965) by GLC [HP 5890 (series II) gas chromatograph with FID and a column (30 m × 0.25 mm × 0.25 μm, Agilent Technologies) of polysilicon SP-5]. Helium was used as carrier gas (70 kPa). The temperature programme was 150 °C for 3 min, then

<table>
<thead>
<tr>
<th>1-α-D-Hep</th>
<th>β-D-Gal</th>
<th>1-α-D-Hep</th>
<th>α-Kdo</th>
</tr>
</thead>
<tbody>
<tr>
<td>↓</td>
<td>↓</td>
<td>↓</td>
<td>↓</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>4</td>
<td>P</td>
</tr>
<tr>
<td>α-D-GlcP-(1→2)α-D-GlcP-(1→3)α-D-GlcP-(1→3)1-α-D-Hep-(1→3)1-α-D-Hep-(1→5)-α-Kdo-(2→6)-lipid A</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**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 core type; Di Padova, Novartis Basel), S31-20 (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 WNI 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⁻¹⁻H exchange of the samples utilizing 99.9% H₂O. 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 XQ1 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 (δH 2.225, δC 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-methylgalactolactose, 1,4,5-tri-O-acetyl-2,3,6-tri-O-methylgalactolactose, 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 (C₂₅H₄₆O₁₈N₂, 698.274 Da) was in excellent agreement with the calculated mass of 740.285 Da was in excellent agreement with the calculated mass of 740.285 Da was in excellent agreement with the calculated mass of 740.286 Da. The program TMHMM 2.0 (http://www.cbs.dtu.dk/services/TMHMM/) was used to identify potential transmembrane segments.

**RESULTS AND DISCUSSION**

The structure of the OPS was determined using 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. 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.
Table 1. $^1$H and $^{13}$C NMR chemical shifts ($\delta$, p.p.m.) of the native and O-deacylated OPS of E. coli 1303

Spectra were recorded at 27 °C in $^2$H$_2$O relative to internal acetone ($\delta_{H}$ 2.225; $\delta_{C}$ 3.145). Underlined chemical shifts indicate substituted positions. Values of $^3$J$_{(1,2)}$ are in Hz; ND, not determined.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>3J$_{(1,2)}$</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6a</th>
<th>6b</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Native OPS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha$-GalpNAc</td>
<td>A</td>
<td>H</td>
<td>5.24</td>
<td>(2.5)</td>
<td>4.36</td>
<td>3.91</td>
<td>4.29</td>
<td>4.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C</td>
<td>98.08 ND</td>
<td>48.77</td>
<td>78.50</td>
<td>69.63</td>
<td>72.03</td>
<td>62.05</td>
</tr>
<tr>
<td>$\alpha$-FucpOAc</td>
<td>B</td>
<td>H</td>
<td>5.20</td>
<td>(2.2)</td>
<td>5.10</td>
<td>4.33</td>
<td>4.05</td>
<td>4.36</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C</td>
<td>94.07 ND</td>
<td>70.38</td>
<td>77.40</td>
<td>72.87</td>
<td>67.62</td>
<td>16.17</td>
</tr>
<tr>
<td>$\beta$-Qui3NAc</td>
<td>C</td>
<td>H</td>
<td>4.69</td>
<td>(6.0)</td>
<td>3.23</td>
<td>4.05</td>
<td>3.48</td>
<td>3.66</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C</td>
<td>105.12 ND</td>
<td>73.13</td>
<td>57.66</td>
<td>77.11</td>
<td>73.08</td>
<td>19.28</td>
</tr>
<tr>
<td>$\beta$-Galp</td>
<td>D</td>
<td>H</td>
<td>4.45</td>
<td>(6.9)</td>
<td>3.64</td>
<td>4.00</td>
<td>3.65</td>
<td>3.60</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C</td>
<td>105.81 ND</td>
<td>69.97</td>
<td>66.02</td>
<td>78.58</td>
<td>75.91</td>
<td>62.05</td>
</tr>
<tr>
<td><strong>O-Decylated OPS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha$-GalpNAc</td>
<td>A</td>
<td>H</td>
<td>5.25</td>
<td>ND</td>
<td>4.36</td>
<td>3.91</td>
<td>4.29</td>
<td>4.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C</td>
<td>98.42 (178)</td>
<td>49.07</td>
<td>78.88</td>
<td>69.91</td>
<td>72.41</td>
<td>62.36</td>
</tr>
<tr>
<td>$\alpha$-Fucp</td>
<td>B'</td>
<td>H</td>
<td>5.06</td>
<td>ND</td>
<td>3.97</td>
<td>4.11</td>
<td>4.02</td>
<td>4.33</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C</td>
<td>96.89 (173)</td>
<td>68.38</td>
<td>80.02</td>
<td>72.96</td>
<td>67.94</td>
<td>16.36</td>
</tr>
<tr>
<td>$\beta$-Qui3NAc</td>
<td>C</td>
<td>H</td>
<td>4.74</td>
<td>(6.5)</td>
<td>3.31</td>
<td>4.07</td>
<td>3.51</td>
<td>3.66</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C</td>
<td>105.06 (165)</td>
<td>73.86</td>
<td>57.89</td>
<td>77.47</td>
<td>73.38</td>
<td>19.25</td>
</tr>
<tr>
<td>$\beta$-Galp</td>
<td>D</td>
<td>H</td>
<td>4.47</td>
<td>(6.6)</td>
<td>3.62</td>
<td>4.11</td>
<td>3.65</td>
<td>3.63</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C</td>
<td>106.13 (160)</td>
<td>70.31</td>
<td>66.33</td>
<td>78.93</td>
<td>75.91</td>
<td>62.36</td>
</tr>
</tbody>
</table>

5.20, 5.10, 4.69, 4.45. Based on the $^1$H NMR spectrum of the O-deacylated OPS (Fig. 2b), the signal at $\delta_{H}$ 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 $\delta_{H}$ 2.12, two N-acetyl signals at $\delta_{H}$ 2.02 and 1.95, revealing that the amino sugars were N-acetylated, as well as two upfield characteristic of the 6-deoxy functions of Fuc (at $\delta_{H}$ 1.19) and of Qui3N (at $\delta_{H}$ 1.36).

By an HSQC experiment (Fig. 3) the direct correlation of all assigned $^1$H signals with $^{13}$C signals could be achieved. The first spin system with the anemonic signal A ($\delta_{H}$ 5.24) originated from a $\alpha$-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 $\alpha$-configuration was established on the basis of a small $^3$J$_{H-1,H-2}$ coupling constant (2.5 Hz). Nitrogen substitution at C-3 possessed a chemical shift characteristic of a carbon atom bearing an acetamido function (CH$_3$CO2 at C-3) was identified as 3,6-dideoxy-3-acetamidoglucose (Qui3NAc), having characteristic 6-deoxy sugar high-field signals of H-6/C-6 ($\delta_{H}$ 1.36, $\delta_{C}$ 19.28, native OPS). Additionally, C-3 possessed a chemical shift characteristic of a carbon atom bearing an acetamido function (CH$_3$CO2 at C-3) was identified as 3,6-dideoxy-3-acetamidoglucose (Qui3NAc), having characteristic 6-deoxy sugar high-field signals of H-6/C-6 ($\delta_{H}$ 1.36, $\delta_{C}$ 19.28, native OPS). Its $\beta$-configuration was proven based on the intra-residual NOE signal C H-2/H-4 and the $\beta$-configuration on the NOE connectivities C H-1/H-3 and C H-1/H-5 as well as on the large $^3$J$_{H-1,H-2}$ value (6.0 Hz, native OPS). Its $\beta$-configuration was deduced after comparison of the obtained chemical shift values with those published (MacLean & Perry, 1997).

Residue C was assigned as $\beta$-Galp. Its $\beta$-configuration was deduced from the large $^3$J$_{H-1,H-2}$ 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-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

\[
\text{C} \rightarrow \text{B} \rightarrow \text{A}
\]
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.

**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 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 thymidylyltransferase-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. 2.](image)

**Fig. 2.** ^1^H NMR spectrum of the native OPS (600 MHz, 300 K, ^2^H_2_0) (a) and O-deacylated OPS (700 MHz, 300 K, ^2^H_2_0) (b) of *E. coli* 1303.

![Fig. 3.](image)

**Fig. 3.** Heteronuclear 2D ^13^C—^1^H 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.
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-Qui\(^3\)N. For the corresponding \(wbsB\) 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-\(\alpha\)-D-Qui\(^3\)NAc 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-

![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.](http://mic.sgmjournals.org)

### 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.

<table>
<thead>
<tr>
<th>Putative ORF no.</th>
<th>Designation (GI:code)</th>
<th>Length (bp)</th>
<th>G+C content (mol%)</th>
<th>No. of aa</th>
<th>Putative function of protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>rmlB</td>
<td>1086</td>
<td>43.4</td>
<td>361</td>
<td>dTDP-glucose 4,6-dehydratase</td>
</tr>
<tr>
<td>2</td>
<td>rmlA</td>
<td>870</td>
<td>37.8</td>
<td>289</td>
<td>Glucose-1-phosphate thymidylyltransferase</td>
</tr>
<tr>
<td>3</td>
<td>qdtA</td>
<td>408</td>
<td>36.3</td>
<td>135</td>
<td>dTDP-4-oxo-6-deoxy-D-glucose-3,4-oxoisomerase</td>
</tr>
<tr>
<td>4</td>
<td>qdtB</td>
<td>1104</td>
<td>36.8</td>
<td>367</td>
<td>Transaminase</td>
</tr>
<tr>
<td>5</td>
<td>wzx</td>
<td>1251</td>
<td>34.9</td>
<td>416</td>
<td>O-antigen flippase</td>
</tr>
<tr>
<td>6</td>
<td>wbnC</td>
<td>513</td>
<td>30.01</td>
<td>170</td>
<td>Acetyltransferase</td>
</tr>
<tr>
<td>7</td>
<td>ORF 7</td>
<td>921</td>
<td>30.8</td>
<td>306</td>
<td>Glycosyltransferase</td>
</tr>
<tr>
<td>8</td>
<td>wzy</td>
<td>1281</td>
<td>30.1</td>
<td>426</td>
<td>O-antigen polymerase</td>
</tr>
<tr>
<td>9</td>
<td>ORF 9</td>
<td>1083</td>
<td>30.8</td>
<td>360</td>
<td>Glycosyltransferase</td>
</tr>
<tr>
<td>10</td>
<td>wcaG</td>
<td>924</td>
<td>30.8</td>
<td>307</td>
<td>Nucleoside-diphosphate-sugar epimerase</td>
</tr>
<tr>
<td>11</td>
<td>ORF 11</td>
<td>810</td>
<td>34.2</td>
<td>269</td>
<td>Glycosyltransferase</td>
</tr>
<tr>
<td>12</td>
<td>tnp</td>
<td>1119</td>
<td>41.1</td>
<td>372</td>
<td>Transposase</td>
</tr>
<tr>
<td>13</td>
<td>qdtC</td>
<td>483</td>
<td>32.7</td>
<td>160</td>
<td>Transacetylase</td>
</tr>
</tbody>
</table>
6-dideoxy-α-D-glucose formation involves subsequent reaction steps catalysed by a transaminase and a transacetylase (Pföstl 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. thermosacharolyticum 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. D7PF5) and 44% identity to the QdtC transacetylase of T. thermosacharolyticum 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 reactions to give GDP-D-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 gene 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 UDP-fucose synthase Fcl (accession no. NP_416566) 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. AA74551) 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→3)-α-L-DglcpNAc (D’Souza et al., 2002). This linkage is absent in the E. coli 1303 O-antigen, but as an α-L-Fucp2OAc(1→4)-β-D-Gal linkage is present, ORF 11 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 WclG of E. coli O104, which is a galactosyltransferase responsible for the β-D-Gal-(1→3)-β-D-GalNAc linkage (Wang et al., 2009). Accordingly, ORF 11 encoded a galactosyltransferase which may link the Gal 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 gneF. O-repeating unit synthesis in enterobacteria is often initiated by transferring GlcNAc 1-phosphate or GalNAc 1-phosphate to an undecaprenol.
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).
outer core region (7-substituted L-α-D-Hepp and 6- and 2-
substituted α-D-Glcp) could be resolved (data not shown),
as compared with published data (Müller-Loennies et al.,
2003). Thus, the disaccharide β-D-GalpNac-(1→7)-L-α-D-
Hepp was proven, which identified the site of attachment
of the OPS at core oligosaccharide:

\[
\beta-D-\text{QuisNac}-(1\rightarrow3)-\alpha-L-\text{FucOAc}-(1\rightarrow4)-\beta-D-\text{Gal}-(1\rightarrow3)-
\alpha-D-\text{GalNac}-(1\rightarrow4)-\beta-D-\text{QuisNac}-(1\rightarrow3)-\alpha-L-\text{FucOAc}-(1\rightarrow4)-
\beta-D-\text{Gal}-(1\rightarrow3)-\beta-D-\text{GalNac}-(1\rightarrow7)-\alpha-L-\text{D-Hep}
\]

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

---

**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-ulono-
sic acid; Hep, L-glycero-D-manno-heptose;
Hex, hexose. Δ m 698.27 u corresponded to
the mass of one O-deacylated O-repeating
unit.

**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-GlcNAc. The latter was shown to be linked to the position O-7 of the terminal 1→2-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 1→2-D-Hepp and from 6-substituted and 2-substituted x-D-GlcP were missing; however, an additional signal at δ H 5.38 p.p.m. was seen, which was identified as terminal x-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 1→2-D-Hepp and 6-substituted x-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 on R-form LPS and nothing could be concluded about the al.

A Glc residue, as compared to the non-substituted core followed the translocation of a Rha residue and loss of serotype O5 the addition of the OPS to the aeruginosa was already well established with the identification of four Kdo residues could be drawn from these data.

Kdo residue(s), no conclusions concerning the number of a Glc residue, as compared to the non-substituted core.

ACKNOWLEDGEMENTS

We thank Herman Moll for help with GC/MS analysis, Heiko Käffner for NMR recording, Dr Patricia Sanchez Carballo for valuable discussions and Dr F. Di Padova for providing monoclonal antibodies WNI 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


Vinogradov, E. V., Van Der Drift, K., Thomas-Oates, J. E., Meshkov, S., Brade, H. & Holst, Q. (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.


Edited by: G. H. Thomas