O antigen of *Franconibacter pulveris* G3872 (O1) is a 4-deoxy-\(\beta\)-arabino-hexose-containing polysaccharide synthesized by the ABC-transporter-dependent pathway

Min Wang,\(^1,2\) Nikolay P. Arbatsky,\(^3\) Lingling Xu,\(^1,2\) Alexander S. Shashkov,\(^3\) Lei Wang\(^1,2\) and Yuriy A. Knirel\(^3\)

\(^1\)TEDA School of Biological Sciences and Biotechnology, Nankai University, TEDA, 300457 Tianjin, China
\(^2\)Key Laboratory of Molecular Microbiology and Technology, Ministry of Education, College of Life Sciences, Nankai University, 300071 Tianjin, China
\(^3\)N. D. Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, 119991 Moscow, Russia

*Franconibacter (Enterobacter, Cronobacter) pulveris* bacteria share several typical characteristics with, and hence pose a challenge for the detection of, *Cronobacter sakazakii*, an emerging opportunistic pathogen, which can cause severe infections in neonates. A structurally variable O-specific polysaccharide (OPS) called O antigen provides the major basis for the typing of Gram-negative bacteria. We investigated the structure and genetics of the O antigen of *F. pulveris* G3872 (designated O1). An OPS was isolated by mild alkaline degradation of the LPS, whereas the same polysaccharide and its oligosaccharide fragments were obtained by mild acid degradation. Studies by sugar analysis and NMR spectroscopy showed that the OPS contained \(\beta\)-ribose, \(L\)-rhamnose (\(L\)-Rha) and a rarely occurring monosaccharide 4-deoxy-\(\beta\)-arabino-hexose, and the OPS structure was established. The O-antigen gene cluster of *F. pulveris* G3872 between JUMPStart and *gnd* genes includes putative genes for \(L\)-Rha, but no genes for the synthesis of 4-deoxy-\(\beta\)-arabino-hexose. A mutation test with the *wzm* gene confirmed that the OPS is synthesized and exported by the ABC-transporter-dependent pathway. A trifunctional transferase was suggested to catalyse formation of two glycosidic linkages and add a methyl group to the non-reducing end of the OPS to terminate the chain elongation. A carbohydrate-binding module that presumably recognizes the terminal methyl-modified monosaccharide was found at the C-terminus of Wzt. Primers specific for *F. pulveris* G3872 were designed based on the *wzm* gene, which has potential to be used for identification and detection of the O1 serogroup.

INTRODUCTION

*Enterobacter pulveris* sp. nov. was originally recognized based on phylogenetic analyses, DNA–DNA hybridization data, and unique physiological and biochemical characteristics (Stephan *et al*., 2008). In 2013, the taxonomy of the genus *Enterobacter* was re-evaluated based on MLSA, and it was proposed to reclassify *Ent. pulveris* into the genus *Cronobacter* as *Cronobacter pulveris* (Brady *et al*., 2013). Recently, new genome-scale analyses led to the proposal that this species should be reclassified as *Franconibacter pulveris* (Stephan *et al*., 2014). In the present work, the last name is used following the current recommendation of the
In this work, the structure of the OPS of the LPS pathway. Specific primers useful for the detection of O antigen is synthesized by the ABC-transporter-dependent pathway. It was found that the O antigen sequences and analysed. It was found that the O antigen is one of the most variable cell constituents due to variations in the types of sugars present, their arrangement and linkages, and is subject to intense selection by the host immune system and bacteriophages. The variability of the O antigen provides the major basis for the serotyping of Gram-negative bacteria. For example, at least 180 and 46 O serogroups have been recognized in Salmonella enterica and related genera. They fall into three main classes: (i) genes for synthesis of nucleotide precursors of sugars that are specific to the O antigen, (ii) glycosyltransferase genes for sugar transfer, and (iii) genes for O-antigen processing. Three O-antigen biosynthesis pathways, the flipase Wzx/polymerase Wzy-dependent pathway, the ATP-binding cassette (ABC)-transporter-dependent and the synthase-dependent pathway have been distinguished, the first two being widespread among Gram-negative bacteria (Raez & Whitfield, 2002; Greenfield & Whitfield, 2012). In the Wzx/Wzy pathway, the O unit is synthesized in the cytoplasm by sequential transfer of a sugar phosphate and one or more sugars from the nucleotide sugar precursors to the undecaprenyl phosphate (UndP) carrier. Then, the O unit is translocated across the membrane and polymerized in the periplasm to form a polysaccharide chain. In the ABC-transporter-dependent pathway, glycosyltransferases mediate the sequential addition of sugar residues to the non-reducing end of the growing polymer attached to the UndP carrier to form the complete O antigen. The polysaccharide is then translocated across the membrane by an ABC transporter. In both pathways, the mature OPS is ligated to the independently synthesized core-lipid A component to afford the complete LPS.

The O antigens of F. pulvères have not been studied so far. In this work, the structure of the OPS of F. pulvères G3872 (designated O1) was established, and its O-antigen gene cluster was sequenced and analysed. It was found that the O antigen is synthesized by the ABC-transporter-dependent pathway. Specific primers useful for the identification and detection of the O1 serogroup of F. pulvères were designed.

**METHODS**

**Bacterial strain, cultivation and isolation of the LPS.** F. pulvères strain G3872 was from tea from the Xinjiang Entry–Exit Inspection and Quarantine Bureaus of China. It was identified through 16S rRNA gene sequencing. Bacteria were grown in 81 Luria–Bertani medium using a 10 l fermenter (BIOSTAT C10; B. Braun Biotech International) under constant aeration at 37 °C and pH 7.0. Bacterial cells were washed and dried as described by Robbins & Uchida (1962). LPS was isolated with a yield of 6-5 % by the phenol/water procedure (Westphal & Jann, 1965), followed by dialysis of the extract without layer separation, and purification as described by Sun et al. (2014).

**Degradation of the LPS.** An LPS sample (95 mg) in aqueous 12 % ammonia (3 ml) was heated for 17 h at 37 °C; after evaporation, the residue was dissolved in 2 ml water, and the lipid precipitate was removed by centrifugation. Gel-permeation chromatography of the supernatant on a column (683.4 cm) of Sephadex G-50 Superfine (Amersham Biosciences) in 0.1 % HOAc monitored using a Uvicord detector (LKB, Sweden) at 206 nm afforded a polysaccharide (49 mg).

An LPS sample (83 mg) was treated with 0.1 M NaOAc pH 4.3 (3 ml) for 4 h at 100 °C, the precipitate was removed by centrifugation, and the supernatant was fractionated on Sephadex G-50 as described above to give polysaccharide (23 mg) and oligosaccharide (16 mg) fractions. A part of the latter (8 mg) was fractionated by HPLC on a Zorbax C18 column (25×1 cm) in water (1 ml min⁻¹) to yield disaccharide 1. The other part (8 mg) was reduced with 1 M NaBH₄ in water (16 h, 20 °C) and desalted with a Dowex 50×4 (H⁺-form) cation-exchange resin. After evaporation three times with 10 % HOAc/MEOH, the residue was fractionated by HPLC as above to give trisaccharide 2.

**Sugar analyses.** A polysaccharide sample (2 mg) was hydrolysed with 3 M CF₂CO₂H (120 °C, 3 h), and monosaccharides were analysed by using a Biotronik LC 2000 sugar analyser on a column (10×0.4 cm) of Dionex AX8 anion-exchange resin in 0.2 M sodium borate buffer pH 8.0 at 70 °C. 4-Deoxy-β-arabino-hexose from the OPS of Citrobacter braakii PCM 1531 (Katzenellenbogen et al., 2003) was used as the authentic sample. The absolute configurations of the monosaccharides were determined by GLC of the acetylated (S)-2-ctyl glycosides (Leontin & Lönnberg, 1993) on a Hewlett-Packard 5890 GLC instrument equipped with a HP-5ms column (25 m×0.25 mm) using a temperature gradient of 160 °C (3 min) to 290 °C at 7 °C min⁻¹.

A polysaccharide sample (20 mg) was heated with 0.5 M CF₂CO₂H (2 ml) for 1 h at 100 °C, and the products were fractionated by gel-permeation chromatography on a column (120×1.2 cm) of Sephadex G-25 in 0.1 % HOAc. Monosaccharides were fractionated by HPLC as described above to give 1,6-anhydro-4-deoxy-β-β-arabinohexose (0.7 mg). Specific optical rotation of this compound, [α]D = 156° (about 0.07, water), was measured with a Jasco DIP-360 polarimeter (Japan) [compare with published data in Horton & Wander (1970) [α]D = 168° (water)].

**NMR spectroscopy.** Samples were freeze-dried twice from 99.9 % D₂O and dissolved in 99.95 % D₂O. 1H- and 13C-NMR spectra were recorded on a Bruker Avance II 600 MHz spectrometer at 30 °C using internal sodium 3-(trimethylsilyl)propanoate-2,2,3,3-D₄ (δH 0, δC –1.6) as a reference for calibration. 2D NMR spectra were obtained using standard pulse sequences, and the Bruker TopSpin 2.1 program was employed to acquire and process the NMR data. A mixing time of 150 and 200 ms was used in 1H,1H-TOCSY (total correlation...
agarose gel electrophoresis to allow examination of the amplified products.

Construction of a random DNase I shotgun bank. Chromosomal DNA was prepared as described by Bastin & Reeves (1995). Long-range PCR was performed with the Expand Long Template PCR system (Roche Applied Science) using the primers wl-10324 (5’-GCACCTGGTAGCTATGGAAGGGGCGATCATC-3’) and wl-2211 (5’-ACTGCGCATACCGAGGCGCCATCTGTGTGTCTGG-3’), based on the sequences of the JUMPStart site and gnd gene, respectively, which flank the O-antigen gene cluster in Cronobacter turicensis. PCR was performed over 32 cycles of denaturation at 94°C for 30 s, annealing at 61°C for 45 s and extension at 68°C for 15 min. The PCR products were digested with DNase I, and the resulting DNA fragments were cloned into pGEM-T Easy (Promega) to produce a shotgun bank as described by Wang & Reeves (1998).

Sequencing and analysis. Sequencing was carried out using an ABI 3730 automated DNA sequencer (Applied Biosystems). Sequencing data were assembled using the Staden Package software (Staden, 1996) and annotated with the Artemis program (Rutherford et al., 2000). BLAST software (http://blast.ncbi.nlm.nih.gov/Blast.cgi) was used to search the available databases including GenBank (http://www.ncbi.nlm.nih.gov/Genbank), the Clusters of Orthologous Groups (COG; http://www.ncbi.nlm.nih.gov/COG/) and Pfam (http://pfam.sanger.ac.uk) protein motif databases. TMHMM 2.0 (http://www.cbs.dtu.dk/services/TMHMM) was used to identify potential transmembrane segments.

Deletion of the wzm gene from F. pulveris G3872. The wzm gene was replaced with a chloramphenicol acetyltransferase (CAT) gene using the RED recombination system of phage lambda. The CAT gene was PCR amplified from plasmid pK232-8 (Pharmacia) by using primers wl-12364 (5’-TGACAGCTTTGAGGCTCTA TTGAGAAAAAGGTACATCGGAGGGAAAAACTCGTG-3’) and wl-12365 (5’-AATTTCAATCTTATATCAGTCAACGGGGATATGAAAA TCAAAAAATGCCC-CG-3’) bound to the 5’ and 3’ ends of the wzm gene, respectively. Each primer carried 39 bp based on the F. pulveris G3872 DNA that flanked the corresponding gene. The PCR product was used to transform the F. pulveris G3872 strain carrying pBluescribe-MCS-α2, and chloramphenicol-resistant transformants were selected after induction of the RED genes. PCR performed with primers specific for the CAT gene and F. pulveris G3872 DNA flanking the wzm gene was used to confirm the replacement.

Specificity assay with PCR. The primers wl-44241 (5’-GCGAAA TCTTATCTTGACGACATC-3’) and wl-44242 (5’-CAGAAATAG- CAACGGGTAGC-3’) were designed based on the sequence of the wzy gene using the Primer Premier 5.0 software program (Premier Biosoft International), and used to amplify the DNA templates. PCR reactions were performed in a total volume of 25 µl containing 2.5 µl 10× PCR buffer (50 mM KCl, 10 mM Tris-HCl, pH 8.3), 0.5 µl 10 mM NTPs, 0.5 µl 10 µM each primer, 1.5 U Taq DNA polymerase (Takara Biotechnology) and 50–100 ng DNA template. Thermal PCR conditions were as follows: an initial denaturation step at 95°C for 5 min; 30 cycles of 94°C for 30 s, 53°C for 30 s and 72°C for 1 min; and a final extension at 72°C for 5 min. The reaction mixture (5 µl) was subjected to agarose gel electrophoresis to allow examination of the amplified products.

RESULTS

Isolation and composition of the O polysaccharide

LPS was obtained from cells of F. pulveris G3872 by extraction with hot aqueous phenol. To enable analysis of the O-antigenic structure, the LPS was O-deacylated by treatment with 12% aqueous ammonia. Monosaccharide analysis after full acid hydrolysis of the resultant polysaccharide revealed rhamnose (Rha), ribose (Rib) and 4-deoxy-arabino-hexose (ara4Hex), which were identified by comparison with authentic standards. GLC analysis of the acetylated (5)-octyl glycosides showed that Rha has the L-configuration and Rib has the d-configuration. As reported elsewhere (Romanowska et al., 1981), during hydrolysis ara4Hex was converted to a large extent into a 1,6-anhydro derivative, which was isolated by reversed-phase HPLC and identified by NMR spectroscopy (for assigned 1H- and 13C-NMR chemical shifts see Table 1). The d-configuration of ara4Hex was established by the specific optical rotation of the 1,6-anhydro derivative.

Structure elucidation of the O polysaccharide by NMR spectroscopy

The 1H-NMR spectrum of the OPS showed signals for five anomic protons at δ 4.98–5.30 (all broadened singlets), two C-CH2–C groups (H-4 of ara4dHex) at δ 1.54 (2H), 1.74 and 1.76, and two CH2–C groups (H-6 of Rha) at δ 1.32 and 1.35. The 13C-NMR spectrum of the OPS contained signals for five anomic carbons at δ 100.7–107.8, two C-CH2–C groups (C-4 of ara4dHex) at δ 29.7 and 30.3, two CH2–C groups (C-6 of Rha) at δ 18.4 and 18.5, and three HOCH2–C groups (C-5 of Rib and C-6 of ara4dHex) at δ 62.5, 65.4 and 65.5. Together with sugar analysis data, these data demonstrated a pentasaccharide repeating unit containing one residue of d-Rib and two residues each of l-Rha and d-ara4dHex.

The NMR spectra of the OPS were assigned using 2D 1H,1H-COSY (correlation spectroscopy), 1H,1H-H-TOCSY, 1H,1H-ROESY (Fig. 1), 1H,13C-HSQC (heteronuclear single-quantum coherence) (Fig. 2), and 1H,13C-HSQC-NOESY (nuclear Overhauser enhancement spectroscopy) experiments, and spin-systems for five sugar residues were identified (Table 1). A comparison of 13C-NMR chemical shifts of the monosaccharide residues (Table 1) with published data (Bock & Pedersen, 1983; Gamian et al., 1985) indicated that Rib is in the β-furanose form, both Rha residues are in the α-pyranose form and both ara4dHex residues are in the β-pyranose form. The β-configuration of the ara4dHex residues (units d and E) was confirmed by intraroside H-1/H-2 and H-1/H-5 correlations in the 1H,1H-ROESY spectrum (Fig. 1).

Low-field positions at δ 78.0–83.7 of the signals for C-2 of Rib (unit C) and one of the ara4dHex residues (unit D), C-3 of one of the Rha residues (unit A), C-3 and C-4 of the
Table 1. $^1$H- and $^{13}$C-NMR chemical shifts (δ, p.p.m.)

<table>
<thead>
<tr>
<th>Sugar residue</th>
<th>H-1 (1a, 1b)</th>
<th>H-2</th>
<th>H-3</th>
<th>H-4 (4eq, 4ax)</th>
<th>H-5 (5a, 5b)</th>
<th>H-6 (6a, 6b)</th>
<th>C-1</th>
<th>C-2</th>
<th>C-3</th>
<th>C-4</th>
<th>C-5</th>
<th>C-6</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>OPS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>→3)-α-1-Rhap-1→ A</td>
<td>5.02</td>
<td>4.13</td>
<td>3.84</td>
<td>3.58</td>
<td>3.76</td>
<td>1.35</td>
<td>103.0</td>
<td>71.3</td>
<td>79.9</td>
<td>72.5</td>
<td>70.7</td>
<td>18.5</td>
</tr>
<tr>
<td>→3,4)-α-1-Rhap-1→ B</td>
<td>5.04</td>
<td>4.26</td>
<td>4.04</td>
<td>3.76</td>
<td>3.94</td>
<td>1.32</td>
<td>103.2</td>
<td>71.0</td>
<td>82.1</td>
<td>78.0</td>
<td>69.1</td>
<td>18.4</td>
</tr>
<tr>
<td>→2)-β-1-Ribf-1→ C</td>
<td>5.30</td>
<td>4.13</td>
<td>4.46</td>
<td>4.06</td>
<td>3.72, 3.88</td>
<td>–</td>
<td>107.8</td>
<td>83.7</td>
<td>70.8</td>
<td>83.6</td>
<td>62.5</td>
<td>–</td>
</tr>
<tr>
<td>→2)-β-1-ara4dHexp-1→ D</td>
<td>5.07</td>
<td>3.75</td>
<td>4.31</td>
<td>1.54, 1.76</td>
<td>3.92</td>
<td>*</td>
<td>100.7</td>
<td>78.3</td>
<td>68.7</td>
<td>30.3</td>
<td>73.6</td>
<td>65.5</td>
</tr>
<tr>
<td>β-1-ara4dHexp-1→ E</td>
<td>4.98</td>
<td>3.70</td>
<td>4.09</td>
<td>1.54, 1.74</td>
<td>3.95</td>
<td>*</td>
<td>101.8</td>
<td>69.8</td>
<td>68.8</td>
<td>29.7</td>
<td>73.0</td>
<td>65.4</td>
</tr>
<tr>
<td><strong>Disaccharide 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>→2)-α-1-ara4dHexp D</td>
<td>5.11</td>
<td>3.68</td>
<td>4.13</td>
<td>1.70, 2.00</td>
<td>4.25</td>
<td>3.66, 3.68</td>
<td>94.1</td>
<td>78.7</td>
<td>67.4</td>
<td>30.6</td>
<td>68.1</td>
<td>64.4</td>
</tr>
<tr>
<td>β-1-ara4dHexp-1→ Eu</td>
<td>5.00</td>
<td>3.70</td>
<td>4.11</td>
<td>1.56, 1.78</td>
<td>3.98</td>
<td>*</td>
<td>99.8</td>
<td>70.0</td>
<td>68.7</td>
<td>29.7</td>
<td>72.9</td>
<td>65.3</td>
</tr>
<tr>
<td>→2)-β-1-ara4dHexp D</td>
<td>5.13</td>
<td>3.72</td>
<td>4.36</td>
<td>1.57, 1.83</td>
<td>3.99</td>
<td>*</td>
<td>92.9</td>
<td>78.3</td>
<td>68.2</td>
<td>29.7</td>
<td>72.9</td>
<td>65.3</td>
</tr>
<tr>
<td>β-1-ara4dHexp-1→ E</td>
<td>4.97</td>
<td>3.78</td>
<td>4.12</td>
<td>1.56, 1.78</td>
<td>3.98</td>
<td>*</td>
<td>101.6</td>
<td>69.7</td>
<td>68.6</td>
<td>29.7</td>
<td>72.9</td>
<td>65.3</td>
</tr>
<tr>
<td><strong>Trisaccharide 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>→2)-β-Ribitol C</td>
<td>3.80, 3.86</td>
<td>3.93</td>
<td>3.89</td>
<td>3.77</td>
<td>3.65, 3.81</td>
<td>–</td>
<td>61.8</td>
<td>80.2</td>
<td>71.2</td>
<td>72.8</td>
<td>64.0</td>
<td>–</td>
</tr>
<tr>
<td>→3)-α-1-Rhap-1→ A</td>
<td>4.92</td>
<td>4.07</td>
<td>3.90</td>
<td>3.57</td>
<td>3.97</td>
<td>1.29</td>
<td>100.5</td>
<td>71.4</td>
<td>79.3</td>
<td>72.5</td>
<td>70.2</td>
<td>17.8</td>
</tr>
<tr>
<td>α-1-Rhap-1→ B</td>
<td>5.06</td>
<td>4.08</td>
<td>3.86</td>
<td>3.48</td>
<td>3.85</td>
<td>1.3</td>
<td>103.6</td>
<td>71.3</td>
<td>71.2</td>
<td>73.2</td>
<td>70.3</td>
<td>17.8</td>
</tr>
<tr>
<td>1,6-Anhydro-4-deoxy-β-1-arabino-hexose</td>
<td>5.39</td>
<td>3.46</td>
<td>3.86</td>
<td>1.77, 2.15</td>
<td>4.73</td>
<td>3.76, 3.90</td>
<td>102.8</td>
<td>76.2</td>
<td>69.3</td>
<td>37.2</td>
<td>74.8</td>
<td>69.3</td>
</tr>
</tbody>
</table>

*Signals for H-6a, 6b of β-1-ara4dHexp residues are at δ 3.60–3.66.*
other Rha residue (unit B) (Table 1), as compared with published data of non-substituted Rib and Rha (Bock & Pedersen, 1983) and data of the non-substituted ara4dHex residue (unit E) (Table 1), defined the glycosylation pattern in the branched OPS. The positions of substitution of the monosaccharides were confirmed and their sequence was determined by 2D $^1$H,$^1$H-ROESY (Fig. 1) and $^1$H,$^{13}$C-HMBC experiments, which showed interresidue correlations between the anomeric protons and protons at the linkage carbons (ROESY) or linkage carbons (HMBC).

**Fig. 1.** Part of a 2D ROESY spectrum of the OPS of *F. pulveris* G3872. The corresponding parts of the $^1$H-NMR spectrum are shown along the axes. Numbers refer to protons in sugar residues denoted by letters as shown in Table 1 and Fig. 3.

**Fig. 2.** Parts of a 2D $^1$H,$^{13}$C-HSQC spectrum of the OPS of *F. pulveris* G3872. The corresponding parts of the $^1$H- and $^{13}$C-NMR spectra are shown along the horizontal and vertical axes, respectively. Numbers refer to H/C pairs in sugar residues denoted by letters as shown in Table 1 and Fig. 3.
In addition to the major signals for the repeating pentasaccharide shown in Fig. 3, the revised structure of the OPS of *F. pulveris* G3872 has the same repeating unit as that obtained by mild alkaline degradation of the polysaccharide, which may belong to an H-2/C-2 correlation of a hexosamine and an H/C correlation of a methyl group, respectively, in non-repetitive regions of the OPS. Owing to the low intensity of these signals, no further information could be inferred from the NMR spectra.

### Analysis of the products of mild acid degradation of the LPS

To cleave the lipid moiety, the LPS was hydrolysed at pH 4.3 to give polysaccharide and oligosaccharide fractions, which were separated by gel-permeation chromatography on Sephadex G-50. As judged by the 1H- and 13C-NMR spectra, the polysaccharide had the same repeating unit (Table 2). Therefore, the OPS of *F. pulveris* G3872 has the structure shown in Fig. 3.

In addition to the major signals for the repeating pentasaccharide unit, the 1H,13C-HSQC spectrum of the polysaccharide showed minor signals at δH/δC 3.64/54.2 and 3.63/58.6, which may belong to an H-2/C-2 correlation of a hexosamine and an H/C correlation of a methyl group, respectively, in non-repetitive regions of the OPS. Owing to the low intensity of these signals, no further information could be inferred from the NMR spectra.

### Characterization of the O-antigen gene cluster

DNA sequences of 13 428 bp between JUMPStart and *gnd* were obtained from *F. pulveris* G3872. The cluster contained nine ORFs with the same transcriptional direction from JUMPStart to *gnd* (Fig. 4). The putative functions of all ORFs summarized in Table 3 were assigned based on the similarities of the predicted proteins to those from available databases. From them, Orf1–Orf6 share high level identity to many known RmlB, A, D and C proteins (67–87%) responsible for the synthesis of dTDP-L-Rha by a well-known pathway (Hao & Lam, 2011). Orf6 and Orf7 were proposed to be responsible for the O-antigen processing by the ABC-transporter-dependent pathway (next section), and three remaining genes, Orf5, Orf8 and Orf9, to encode glycosyltransferases.

Predicted Orf5 belongs to the glycosyl transferase family 2 (PF00535, e-value=8.7×10^-23) and is a homologue of many *Pseudomonas* rhamnosyltransferases. Predicted Orf8 is a...
Table 3. Characteristics of the ORFs in the O-antigen gene cluster of *F. pulveris* G3872.

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Position of gene (length of predicted product, aa)</th>
<th>G+C content (mol%)</th>
<th>Conserved domain(s) (Pfam no./e value)</th>
<th>Similar protein(s), strain(s) (GenBank accession no.)</th>
<th>% identity/% similarity (range of aa sequence overlap)</th>
<th>Putative function of protein</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>rmlB</em></td>
<td>118 : 1182 (354)</td>
<td>42.44</td>
<td>NAD-dependent epimerase/dehydratase family (PF01370/1.2e^{-25})</td>
<td>dTDP-glucose 4,6-dehydratase, <em>Escherichia coli</em> HS (YP_001458854.1)</td>
<td>80/89 (285)</td>
<td>dTDP-glucose 4,6-dehydratase</td>
</tr>
<tr>
<td><em>rmlA</em></td>
<td>1193 : 2065 (290)</td>
<td>44.32</td>
<td>Nucleotidyl transferase (PF00483/3.9e^{-66})</td>
<td>Glucose-1-phosphate thymidylyltransferase, <em>Raoultella terrigena</em> (AAQ82933.1)</td>
<td>83/89 (200)</td>
<td>Glucose-1-phosphate thymidylyltransferase</td>
</tr>
<tr>
<td><em>rmlD</em></td>
<td>2090 : 2980 (296)</td>
<td>43.20</td>
<td>RmlD substrate binding domain (PF04321/7.4e^{-105})</td>
<td>dTDP-4-dehydrorhamnose reductase, <em>Klebsiella pneumoniae</em> subsp. <em>pneumoniae</em> MGH 78578 (YP_001336137.1)</td>
<td>77/87 (228)</td>
<td>dTDP-4-dehydrorhamnose reductase</td>
</tr>
<tr>
<td><em>rmlC</em></td>
<td>2985 : 3533 (182)</td>
<td>40.61</td>
<td>dTDP-4-dehydrorhamnose 3,5-epimerase (PF0908/8.1e^{-79})</td>
<td>dTDP-4-keto-6-deoxyglucose 3,5-epimerase, <em>Raoultella terrigena</em> (AAQ82927.1)</td>
<td>78/86 (143)</td>
<td>dTDP-4-keto-6-deoxyglucose 3,5-epimerase</td>
</tr>
<tr>
<td><em>orf5</em></td>
<td>3541 : 4461 (306)</td>
<td>38.11</td>
<td>Glycosyl transferase family 2 (PF00353/5/8.7e^{-23})</td>
<td>Glycosyl transferase family protein, <em>Pseudomonas fluorescens</em> Pf-1 (YP_349780.1)</td>
<td>46/70 (142)</td>
<td>Glycosyl transferase</td>
</tr>
<tr>
<td><em>wzm</em></td>
<td>4495 : 5325 (276)</td>
<td>38.74</td>
<td>ABC-2 type transporter (PF01061/1.1e^{-33})</td>
<td>Putative ABC-transporter system integral membrane protein, <em>Serratia marcescens</em> (AAC01811.1)</td>
<td>57/87 (152)</td>
<td>ABC transporter system</td>
</tr>
<tr>
<td><em>wzt</em></td>
<td>5315 : 6670 (451)</td>
<td>39.74</td>
<td>ABC transporter (PF00033/5/3.7e^{-10})</td>
<td>O-antigen ABC transporter, ATP-binding protein, putative, <em>Pseudomonas syringae</em> pv. <em>tomato</em> strain DC3000 (NP_790809.1)</td>
<td>52/68 (236)</td>
<td>ABC transporter system</td>
</tr>
<tr>
<td><em>orf8</em></td>
<td>6702 : 11105 (1467)</td>
<td>40.37</td>
<td>Glycosyl transferase family 2 (PF00353/4.4e^{-28})</td>
<td>Glycosyl transferase family protein, <em>Pseudomonas putida</em> F1 (YP_001269234.1)</td>
<td>37/53 (569)</td>
<td>Bifunctional glycosyltransferase, methyltransferase</td>
</tr>
<tr>
<td><em>orf9</em></td>
<td>11111 : 12928 (605)</td>
<td>43.45</td>
<td>Haloacid dehalogenase-like hydrolase (PF06702/1.8e^{-65})</td>
<td>Hydrolase (HAD superfamily) protein, <em>Azotobacter vinelandii</em> DJ (YP_002798789.1)</td>
<td>48/66 (288)</td>
<td>Hydrolase</td>
</tr>
</tbody>
</table>
large protein of 1467 aa. The C-terminus of Orf8 contains two glycosyltransferase motifs (regions from 943 to 1095 aa and from 1197 to 1376 aa), both sharing homology with glycosyl transferase family 2 (PF00535, e value=9.4×10^-27 and 9.2×10^-28, respectively). The N-terminus contains a methyltransferase domain (a region from 96 to 194 aa) presumably responsible for adding a methyl group to the polysaccharide. Predicted Orf9 shares 60% identity with a hypothetical protein of *Salmonella enterica* strain and belongs to the ABC-2 type transporter family (PF01061, e value=1.1×10^-35), which comprises perimase components of the ABC-type polysaccharide–polyl phosphate export system. Orf6 also shares similarity with a number of polysaccharide transporters. Orf7 belongs to the O-antigen ABC-transporter family (PF00005, e value=3.7×10^-10), which comprises ATPase components of the export system. Orf7 also showed high-level identity (100%) with the sugar ABC-transporter ATP-binding protein of some other strains. Therefore, orf6 and orf7 were proposed to be wzm and wzt and named accordingly. Wzt was predicted to possess a carbohydrate-binding module appended to the C-terminus, which presumably recognizes a modification at the non-reducing end of the polysaccharide that terminates chain elongation (Greenfield & Whitfield et al., 2012; Mann et al., 2016).

To confirm the involvement of *wzm* in biosynthesis of the O antigen, a deletion experiment with replacement of *wzm* with a CAT gene was performed. As opposed to the wild-type strain, which produced the complete LPS, the *wzm*-deficient mutant produced no O antigen on the LPS (Fig. 5). The mutation was complemented with pTrc99A plasmid containing the functional *wzm* gene from *F. pulveris* G3872.

**Identification of *F. pulveris* G3872 O-antigen-specific gene**

A specific primer pair based on *wzm* of *F. pulveris* G3872 was designed and used to screen the DNA of two *F. pulveris* strains and strains of *Cronobacter* spp. and others. The occurrence in the cluster of the functional *wzm* gene from *F. pulveris* G3872 is specific for the *G.3872* O antigen designated O1.

**DISCUSSION**

A peculiar chemical feature of the OPS of *F. pulveris* G3872 is the presence of 4-deoxy-d-arabino-hexose, which is the only representative of this class of deoxy sugars found so far in nature. It occurs rarely in natural carbohydrates, and, to our knowledge, has been found only in the OPSs of several strains of *Citrobacter* spp. (Gamian et al., 1985; Knirel et al., 2002; Katzenellenbogen et al., 2003). In all the strains, except for *Cit. braakii* PCM 1531 (O6), this monosaccharide was reported to be β-linked, as in *F. pulveris* G3872 studied in this work. In *Cit. braakii* PCM 1531, the α configuration of d-aradHex was assigned based solely on an H-1/H-2 correlation in the 1H,1H-ROESY spectrum (Katzenellenbogen et al., 2003). A comparison of the 1H- and 13C-NMR chemical shifts of the OPSs of *Cit. braakii* PCM 1531 (Katzenellenbogen et al., 2003) and *F. pulveris* G3872 (Table 1) showed that the anomeric configuration of d-aradHex in the former must be revised from α to β as shown in Fig. 3. Remarkably, the main chain of the OPS of *F. pulveris* G3872 shares the structure with the linear OPS of *Cit. braakii* PM 1561 (O28) (Kocharova et al., 1995; Knirel et al., 2002).

A gene cluster for *F. pulveris* G3872 O-antigen synthesis was found in the chromosomal locus between conserved JUMP-Start and gnd genes, which is the typical location for many enteric bacteria, including *S. enterica*, *E. coli*, *Shigella* spp., *Cronobacter* spp. and others. The occurrence in the cluster of the functional *wzm* and *wzt* genes, and the absence of *wzx* and *wzy* genes, suggested that the O antigen of *F. pulveris* G3872 is synthesized using the ABC-transporter export system. The *wzm* gene was demonstrated to be specific for

---

**Fig. 5.** Deletion analysis of the *wzm* gene of *F. pulveris* G3872 by silver stained SDS-PAGE of membrane extracts. Lane 1, *F. pulve- risis* G3872; lane 2, *wzm*-deficient G3872; lane 3, *wzm*-deficient G3872 complemented with pTrc99A plasmid carrying functional *wzm* from G3872.
F. pulveris G3872, which we propose to designate O1. Hence, wzm has a potential to be used for the identification and detection of this serogroup.

The OPS of F. pulveris G3872 includes three monosaccharide components, but genes for the synthesis of only dTDP-l-Rha, the nucleotide precursor of l-Rha, are present in the O-antigen gene cluster. These genes are usually found together at the 5’ end of the O-antigen gene cluster in the order rmlBDAC, but the order can vary, being for example rmlBADC in F. pulveris G3872. d-RlfB occurs in other structures generally present in bacteria, including RNA, and genes for the synthesis of its nucleotide precursor are located outside the O-antigen gene cluster. Although d-ara4dHex is a specific component of the OPS, genes for its synthesis were not found in the cluster.

In E. coli, Klebsiella pneumoniae and some other bacteria, synthesis of the O antigen is generally initiated by adding d-GlcNAc-P to UndP, which either is the first monosaccharide of the O unit in the Wzx/Wzy-dependent pathway [sometimes d-GlcNAc-PP-Und is further converted into d-GalNAc-PP-Und (Cunneen et al., 2013)] or functions as a primer in the ABC-transporter-dependent pathway (Greenfield & Whitfield, 2012). In both cases, the reaction is catalysed by d-GlcNAc-1-phosphate transferase WecA. The encoding gene, wecA, maps in the enterobacterial common antigen locus (Meier-Dieter et al., 1992) and is not duplicated in the O-antigen gene cluster. Presumably, F. pulveris G3872 also requires a d-GlcNAc primer for the O-antigen synthesis.

The first post-initiation step of the O-antigen synthesis by the ABC-transporter-dependent pathway is construction of a so called adaptor by transfer of one or two sugar residues to the d-GlcNAc primer. In a number of l-Rha-containing OPSs, synthesis of adaptor is catalysed by rhamnosyltransferase WbbL, which transfers a single l-Rha residue to d-GlcNAc-PP-Und (Rubires et al., 1997; Mills et al., 2004). Predicted rhamnosyltransferase Orf5 shares 21 % identity and 38 % similarity to WbbL of Acrononas hydrophila O1 (Merino et al., 2015) and, thus, may fulfil this function in F. pulvrius G3872. However, a role of Orf5 in the formation of a rhamnosidic linkage within the repetitive OPS domain rather than in the adaptor cannot be excluded.

Another predicted glycosyltransferase, Orf8, which is significantly larger than most monofunctional glycosyltransferases, contains two glycosyltransferase domains and an N-terminal methyltransferase domain. Large bifunctional glycosyltransferases involved in the synthesis of heteropolysaccharide O antigens employing the ABC-transporter system have been reported in various bacteria (Izquierdo et al., 2003; Saigi et al., 1999; Feng et al., 2004). Most likely, the methyltransferase domain is responsible for terminating the O-antigen chain elongation by adding a single methyl group to the non-reducing end of the OPS, which is essential for coupling the polysaccharide synthesis to export (Greenfield & Whitfield, 2012). Accordingly, a putative carbohydrate-binding module appended to the C-terminus of the nucleotide-binding domain of the ABC transporter (Wzt) presumably recognizes a methyl-modified monosaccharide residue (Mann et al., 2016). An O-methyl group was found in the G3872 OPS, but its location could not be determined owing to the content level being too low. Neither Orf8 nor Orf9 showed homology to any known ribosyltransferase or rhamnosyltransferase and could be assigned to specific glycosidic linkages.

Remarkably, the O-antigen gene cluster of F. pulveris G3872 lacks genes for the synthesis and, most likely, also for the transfer of d-ara4dHex, and, therefore, proteins responsible for the attachment of d-ara4dHex residues are encoded elsewhere in the genome. This suggests that decoration of the O antigen with the side-chain d-ara4dHex groups likely occurs late in the overall biosynthesis pathway. Such modifications, which in many instances are associated with lysogenic seroconverting bacteriophages, have long been known in the Wzy/Wzx-dependent pathway (Guan et al., 1999; Allison & Verma, 2000; Mann & Whitfield, 2016). They extend the O-antigen diversity and may impact other properties of bacteria in a species-specific manner. The best studied is the side-chain glucosylation, which typically occurs in the periplasm and requires a three-component module. Two enzymes are responsible for the synthesis of d-Glc-P-Und from UDP-d-Glc and its export from the cytoplasm to the periplasm, where it provides the donor for the modifying glucosylation of an O-antigen polymerization product catalysed by a third enzyme, an integral membrane serotype-specific glucosyltransferase.

Recently, it has been demonstrated that there are no mechanistic barriers for bacteriophage-mediated O-antigen glycosylation in the ABC-transporter-dependent pathway (Mann et al., 2015). Bioinformatics data suggest that decoration of the galactan O antigens of K. pneumoniae O2 and O9 and related Raoullella terrigena by side-chain galactose residues could be natural examples of bacteriophage-mediated galactosylation (Mann & Whitfield, 2016). Another example could be E. coli O99, in which side-chain glucosylation of the rhamnan O-antigen synthesized by the ABC-transporter-dependent pathway is catalysed by enzymes encoded outside the O-antigen gene cluster (Pereplov et al., 2009).

To date, four genomes of F. pulvrius have been sequenced. Analysis of the sequences between the JumpStart and gnd gene revealed two distinct O-antigen gene cluster types. In both of them, the wzx and wzy genes are present, and hence, unlike in F. pulvrius G3872, the O antigens of the corresponding strains are synthesized by the Wzx/Wzy-dependent pathway.

The discovery in this work of an O-antigen modification by a different, rarely occurring monosaccharide in the ABC-transporter-dependent pathway raises a number of issues to be elucidated, including the currently unknown biosynthesis pathway of d-ara4dHex, the origin and genomic location of genes for synthesis, export and transfer of this monosaccharide, genetic and biochemical mechanisms of
glycosylation, as well as a possible biological role of the O-antigen modification by 3-ara4dHex.

ACKNOWLEDGEMENTS

The authors thank Dr Olga G. Ovchinnikova for valuable discussions and critical reading of the manuscript. This work was supported by the Russian Foundation for Basic Research (14-04-00663), the Chinese National Science Fund for Distinguished Young Scholars (30788001), NSFC General Program Grants 31000044, 30670038, 30771175 and 30900041, and the Fundamental Research Funds for the Central Universities (65010721).

REFERENCES


Edited by: D. Hood