Introduction

The O antigen, consisting of numerous oligosaccharide repeats, is an essential component of the lipopolysaccharide (LPS) on the cell surface of Gram-negative bacteria. The O antigen is one of the most variable constituents on the cell surface due to the variation in the types of sugar present, their arrangement within the O unit and the linkages within and between O units. Consequently, the O antigen has become a major basis for serotyping schemes of Gram-negative bacteria (Reeves and Wang 2002). It is also an important virulence factor and there is direct evidence that O-antigen forms account for differences in the nature of pathogenicity (Pluschke et al. 1983; Achtman and Pluschke 1986).

Escherichia coli clones, including both commensal and pathogenic types, are normally identified by the combination of their O and H (and sometimes K) antigens (Nataro and Kaper 1998). To date, more than 180 O-antigen forms have been recognized for E. coli (Stenutz et al. 2006). Escherichia coli strain 86-381 was isolated from swine and its O antigen was first designated as O180 by Scheutz et al. (2004). The O180 isolate was shown to produce Verocytotoxin and was positive for the virulence genes vtx1 and astA. Some anomalous characteristics were also identified for the O180 strain, which was found to be adonitol positive and indole negative, in contrast to other E. coli strains (Scheutz et al. 2004).

Different O-antigen forms are almost entirely due to genetic variations in the O-antigen gene clusters (Reeves and Wang 2002). In E. coli, genes for O-antigen biosynthesis are normally clustered between galF and gnd and include a conserved 39-bp JUMPStart sequence required for the regulation of downstream genes (Hobbs and Reeves 1994), which is located in the intergenic region between galF and the O-antigen gene cluster (Marolda and Valvano 1998). Three groups of genes have been identified in the O-antigen gene

Structural and genetic characterization of the Escherichia coli O180 O antigen and identification of a UDP-GlcNAc 6-dehydrogenase

Quan Wang2,3,4†, Andrei V Perepelov5,†, Lothar Beutin6, Sof’ya N Senchenkova5, Yanli Xu2,3, Alexander S Shaskov5, Peng Ding2,3, Yuriy A Knirel5 and Lu Feng1,2,3,7,8

© The Author 2012. Published by Oxford University Press. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

Received on March 12, 2012; revised on May 19, 2012; accepted on June 5, 2012

The O antigen is an essential component of the lipopolysaccharides on the surface of Gram-negative bacteria and its variation provides a major basis for serotyping schemes. The Escherichia coli O-antigen form O180 was first designated in 2004, and O180 strains were found to contain virulence factors and cause diarrhoea. Different O-antigen forms are almost entirely due to genetic variations in the O-antigen gene clusters. In this study, the chemical structure and gene cluster of E. coli O180 O antigen were investigated. A tetrasaccharide repeating unit with the following structure: β-D-ManpNAc3NAcA-(1→2)-α-L-Rhap5-(1→3)-β-L-Rhap5-(1→4)-α-D-GlcNAc-(1→ was identified in the E. coli O180 O antigen, including the residue β-ManpNAc3NAcA (2,3-diacetamido-2,3-dideoxy-β-D-nannopyranuronic acid) that had not been hitherto identified in E. coli. Genes in the O-antigen gene cluster were assigned functions based on their similarities with those from available databases, and five genes involved in the synthesis of UDP-β-ManpNAc3NAcA (the nucleotide-activated form of β-ManpNAc3NAcA) were identified. The gnaA gene, encoding the enzyme involved in the initial step of the UDP-β-ManpNAc3NAcA biosynthetic pathway, was cloned and the enzyme product was expressed, purified and assayed for its activity. GnaA was characterized using capillary electrophoresis and electrospray ionization mass spectrometry and identified as a UDP-GlcNAc 6-dehydrogenase. The kinetic and physicochemical parameters of GnaA also were determined.

Keywords: biosynthesis / Escherichia coli O180 / O antigen / UDP-β-ManNAc3NAcA

Downloaded from https://academic.oup.com/glycob/article-abstract/22/10/1321/1988081 by guest on 02 January 2019

1To whom correspondence should be addressed: Tel: +86-22-66229592; Fax: +86-22-66229596; e-mail: fenglu63@nankai.edu.cn

2TEDA School of Biological Sciences and Biotecnology, Nankai University, 23 Hongda Street, TEDA; 3Key Laboratory of Molecular Microbiology and Technology, Ministry of Education; and 4The Engineering and Research Center for Microbial Functional Genomics and Detection Technology, Ministry of Education, Tianjin, People’s Republic of China;

3N. D. Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, Moscow 119991, Russian Federation; 5National Reference Laboratory for Glycobiology; and 6National Reference Laboratory for Glycobiology

†Advance Access publication on June 21, 2012
doi:10.1093/glycob/cws098

Advance Access publication on June 21, 2012
cluster including: (i) nucleotide sugar biosynthesis genes, (ii) sugar transferase genes and (iii) O-antigen processing genes for O-antigen flipase (wzx) and polymerase (wzy) (Reeves and Wang 2002).

In this study, the chemical structure and gene cluster of *E. coli* O180 O antigen were analyzed. The O unit was shown to contain a residue of 2,3-diacetamido-2,3-dideoxy-D-mannopyranuronic acid (D-Man at δ 61.5 (data of a distortionless enhancement of polarization δN)) and 61 oxygen-bearing sugar ring anomeric carbons at δ 1.80–5.16, two CH₃-C groups (H6 of Rha) at δ 1.23 and 1.29, three N-acetyl groups at δ 1.90–2.08 and other signals at δ 3.30–4.43. These data indicated that the repeating unit of the OPS is a tetrasaccharide containing two residues of Rha (RhaI and RhaII) and one residue of GlcNac, whereas the fourth monosaccharide is a diacetamido hexuronic acid, which was not detected in sugar analysis.

The ¹H and ¹³C NMR spectra of the OPS were assigned using two-dimensional correlation spectroscopy (COSY), total correlation spectroscopy (TOCSY), rotating-frame nuclear overhauser effect correlation spectroscopy (ROESY), H-detected ¹H-¹³C heteronuclear single quantum coherence (HSQC) and heteronuclear multiple bond correlation (HMBC) experiments (Table I). Tracing connectivities in the COSY and TOCSY spectra within each sugar spin system starting from H1 and, for Rha residues, also from H6 enabled the unambiguous assignment of all proton signals. Correlations in the HSQC spectrum (Figure 2) of protons H2 and H3 with nitrogen-bearing carbons C2 and C3 at δ 4.43/52.8 and 4.28/54.6, respectively, showed that this monosaccharide is a 2,3-diamino sugar, and an H5/C6 cross-peak at δ 4.00/175.2 in the HMBC spectrum indicated that this sugar has the *manno* configuration and is thus 2,3-diacetamido-2,3-dideoxycellobiose (ManNAC3NAC).

The pyranose form of all monosaccharide residues was confirmed by the absence from the ¹³C NMR spectrum of any signals for non-anomeric sugar ring carbons at a lower field than δ 81. The ¹³C NMR chemical shifts δ 70.5 and 73.4 for C5 of RhaI and RhaII, when compared with the values δ 70.0 and 73.2 of α- and β-rhamnopyranose, respectively (Lipkind et al. 1988), demonstrated that RhaI is α-linked and RhaII is β-linked. The β-configuration of ManNAC3NAC was established and that of RhaI confirmed by strong H1/H5 cross-peaks in the ROESY spectrum (Figure 3). The absence of H1/H5 correlations and the presence of strong H1/H2 correlations defined the α-configuration of GlcNAc and confirmed the same configuration of RhaI.

**Results**

**Elucidation of the O-antigen structure of *E. coli* O180**

The O polysaccharide (OPS) was obtained by the mild acid degradation of the LPS, isolated from dried cells of *E. coli* O180 by the phenol–water procedure, followed by gel-permeation chromatography on Sephadex G-50. Sugar analysis by gas–liquid chromatography (GLC) of the alditol acetates derived after the full acid hydrolysis of the OPS revealed Rha and GlcN in the ratio of ~2:1 (detector response). GLC analysis of the acetylated (S)-2-octyl glycosides demonstrated the β-configuration of GlcNAc and the α-configuration of Rha.

The ¹³C NMR spectrum of the OPS (Figure 1) showed 4 anomic carbons at δ 98.1–102.6, 2 CH₃-C groups (C6 of Rha) at δ 18.1 (2 C), 1 HOCH₂-C group (C6 of GlcN) at δ 61.5 (data of a distortionless enhancement of polarization transfer NMR experiment), 13 oxygen-bearing sugar ring carbons in the region at δ 69.4–80.0, 3 nitrogen-bearing carbons at δ 52.8, 54.6 and 54.8, 1 CO₂H group at δ 175.2 and 3 N-acetyl groups at δ 23.25–23.34 (CH₃) and 175.2–176.8 (CO). Accordingly, the ¹H NMR spectrum of the OPS contained signals for four anomic protons at δ 4.83–5.16, two CH₃-C groups (H6 of Rha) at δ 1.23 and 1.29, three N-acetyl groups at δ 1.90–2.08 and other signals at δ 3.30–4.43. These data indicated that the repeating unit of the OPS is a tetrasaccharide containing two residues of Rha (RhaI and RhaII) and one residue of GlcNAc, whereas the fourth monosaccharide is a diacetamido hexuronic acid, which was not detected in sugar analysis.

The ¹H and ¹³C NMR spectra of the OPS were assigned using two-dimensional correlation spectroscopy (COSY), total correlation spectroscopy (TOCSY), rotating-frame nuclear overhauser effect correlation spectroscopy (ROESY), H-detected ¹H-¹³C heteronuclear single quantum coherence (HSQC) and heteronuclear multiple bond correlation (HMBC) experiments (Table I). Tracing connectivities in the COSY and TOCSY spectra within each sugar spin system starting from H1 and, for Rha residues, also from H6 enabled the unambiguous assignment of all proton signals. Correlations in the HSQC spectrum (Figure 2) of protons H2 and H3 with nitrogen-bearing carbons C2 and C3 at δ 4.43/52.8 and 4.28/54.6, respectively, showed that this monosaccharide is a 2,3-diamino sugar, and an H5/C6 cross-peak at δ 4.00/175.2 in the HMBC spectrum indicated that this sugar has the *manno* configuration and is thus 2,3-diacetamido-2,3-dideoxycellobiose (ManNAC3NAC).

The pyranose form of all monosaccharide residues was confirmed by the absence from the ¹³C NMR spectrum of any signals for non-anomeric sugar ring carbons at a lower field than δ 81. The ¹³C NMR chemical shifts δ 70.5 and 73.4 for C5 of RhaI and RhaII, when compared with the values δ 70.0 and 73.2 of α- and β-rhamnopyranose, respectively (Lipkind et al. 1988), demonstrated that RhaI is α-linked and RhaII is β-linked. The β-configuration of ManNAC3NAC was established and that of RhaI confirmed by strong H1/H5 cross-peaks in the ROESY spectrum (Figure 3). The absence of H1/H5 correlations and the presence of strong H1/H2 correlations defined the α-configuration of GlcNAc and confirmed the same configuration of RhaI.

![Fig. 1. ¹³C NMR spectrum of the OPS of *E. coli* O180. Arabic numerals refer to carbons in sugar residues denoted by letters as shown in Table I.](https://academic.oup.com/glycob/article-abstract/22/10/1321/1988081)
The modes of sugar glycosylation were determined by relatively low-field positions of the signals for C2 of RhaI, C3 of RhaII, C4 of GlcNAc and ManNAc3NAcA at δ 80.3, 81.4, 77.9 and 71.6, respectively, in the $^{13}$C NMR spectrum of the OPS compared with their positions in the spectra of the corresponding non-substituted monosaccharides (Knirel et al. 1982; Lipkind et al. 1988).

The two-dimensional ROESY spectrum (Figure 3) showed strong inter-residue cross-peaks between the anomeric protons and protons at the linkage carbons at δ 5.01/4.12, 5.16/3.58, 4.83/3.66 and 5.12/3.93, which were assigned to ManNAc3NAcA H1/RhaI H2, RhaI H1/RhaII H3, RhaII H1/GlcNAc H4 and GlcNAc H1/ManNAc3NAcA H4 correlations, respectively. HMBC analysis showed the following inter-residue correlations between the anomeric protons and linkage carbons: ManNAc3NAcA H1/RhaI C2, RhaI H1/RhaII C3, RhaII H1/GlcNAc C4 and GlcNAc H1/ManNAc3NAcA C4 at δ 5.01/80.3, 5.16/81.4, 4.83/77.9 and 5.12/71.6, respectively. These data are consistent with the glycosylation pattern revealed by the $^{13}$C NMR chemical shift data and defined the sequence of the monosaccharide residues in the O unit.

The position at δ 54.6 of the signal for C3 of ManpNAc3NAcA showed that the constituent monosaccharides in the α-D-ManpNAc3NAcA disaccharide fragment have the same absolute configuration (a C3 chemical shift of δ ~ 53.0 would be observed in the case of their different absolute configurations; Knirel et al. 1982). Hence, ManNAc3NAcA has the $\delta$-configuration.

Based on the data obtained, it was concluded that the OPS of *E. coli* O180 has the following structure:

$$\rightarrow 4\beta-D-ManpNAc3NAcA-(1 \rightarrow 2)\alpha-L-Rhap^{I}(1 \rightarrow 3)\beta-L-Rhap^{II}(1 \rightarrow 4)\alpha-D-GlcpNAc-(1 \rightarrow \frac{\delta}{\delta})$$

### Table I. $^1$H and $^{13}$C NMR data of the OPS from *E. coli* O180 (δ, ppm)

<table>
<thead>
<tr>
<th>Sugar residue</th>
<th>Nucleus</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6 (6a,6b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A, $\rightarrow 4\beta-D-ManpNAc3NAcA-(1 \rightarrow$</td>
<td>$^1$H</td>
<td>5.01</td>
<td>4.43</td>
<td>4.28</td>
<td>3.93</td>
<td>4.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$^{13}$C</td>
<td>101.9</td>
<td>52.8</td>
<td>54.6</td>
<td>71.6</td>
<td>78.4</td>
<td>175.2</td>
</tr>
<tr>
<td>B, $\rightarrow 2\alpha-L-Rhap^{I}(1 \rightarrow$</td>
<td>$^1$H</td>
<td>5.16</td>
<td>4.12</td>
<td>3.88</td>
<td>3.30</td>
<td>3.79</td>
<td>1.23</td>
</tr>
<tr>
<td></td>
<td>$^{13}$C</td>
<td>102.6</td>
<td>80.3</td>
<td>71.1</td>
<td>73.6</td>
<td>70.5</td>
<td>18.1</td>
</tr>
<tr>
<td>C, $\rightarrow 3\beta-L-Rhap^{II}(1 \rightarrow$</td>
<td>$^1$H</td>
<td>4.83</td>
<td>4.11</td>
<td>3.58</td>
<td>3.41</td>
<td>3.39</td>
<td>1.29</td>
</tr>
<tr>
<td></td>
<td>$^{13}$C</td>
<td>102.0</td>
<td>71.8</td>
<td>81.4</td>
<td>72.8</td>
<td>73.4</td>
<td>18.1</td>
</tr>
<tr>
<td>D, $\rightarrow 4\alpha-D-GlcpNAc-(1 \rightarrow$</td>
<td>$^1$H</td>
<td>5.12</td>
<td>3.82</td>
<td>3.78</td>
<td>3.66</td>
<td>3.65</td>
<td>3.71, 3.91</td>
</tr>
<tr>
<td></td>
<td>$^{13}$C</td>
<td>98.1</td>
<td>54.8</td>
<td>71.5</td>
<td>77.9</td>
<td>72.3</td>
<td>61.5</td>
</tr>
</tbody>
</table>

The chemical shifts for the N-acetyl groups are δH 1.90–2.08; δC 23.25–23.34 (Me) and 175.2–176.8 (2 CO).

![Fig. 2. Parts of a two-dimensional $^1$H-$^1$C HSQC spectrum of the OPS of *E. coli* O180. The corresponding parts of the $^1$H and $^{13}$C NMR spectra are displayed along the horizontal and vertical axes, respectively. Arabic numerals refer to cross-peaks in sugars denoted by letters as shown in Table I.](https://academic.oup.com/glycob/article-abstract/22/10/1321/1988081)

![Fig. 3. Part of a two-dimensional $^1$H-$^1$H ROESY spectrum of the OPS of *E. coli* O180. The corresponding parts of $^1$H NMR spectrum are displayed along the axes. Arabic numerals refer to inter- (bold) and intra-residues (regular) proton pairs in sugar residues denoted by letters as shown in Table I.](https://academic.oup.com/glycob/article-abstract/22/10/1321/1988081)
Catalyzed by UDP-GlcNAc 6-dehydrogenase that catalyzes the conversion of UDP-D-GlcNAc to UDP-D-GlcNAcA (Miller et al.
1992; Lindqvist et al. 1993; Giraud et al.
1999). Among these enzymes, WbpA (GnaA in E. coli O180) is a UDP-GlcNAc 6-dehydrogenase that catalyzes the first step, conversion of UDP-D-GlcNAc to UDP-D-GlcNAcA (Miller et al.
2004).

Characterization of GnaA in E. coli O180

It has been reported that the biosynthetic pathway of UDP-D-ManNAc3NAcA in P. aeruginosa includes five steps catalyzed by five enzymes (Westman et al.
2007, 2009). Among these enzymes, WbpA (GnaA in E. coli O180) is a UDP-GlcNAc 6-dehydrogenase that catalyzes the first step, conversion of UDP-D-GlcNAc to UDP-D-GlcNAcA (Miller et al.
2004).

(i) Overexpression and purification of GnaA. Plasmid pLW1476 containing gnaA was constructed. Expression of the gene in transfected E. coli BL21 was induced by isopropylthio-β-
-D-galactopyranoside (IPTG). The majority of the protein was detected in the soluble fraction as estimated by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS–PAGE) analysis (data not shown), and the protein was purified to near homogeneity by nickel ion affinity chromatography. The apparent molecular mass of 51.5 kDa was estimated by SDS–PAGE for His-tagged GnaA and corresponded well with the calculated mass (50.5 kDa; Supplementary data, Figure S1).

(ii) Characterization of GnaA activities by capillary electrophoresis (CE). The proposed reaction catalyzed by GnaA is shown in Figure 5. Following the addition of GnaA to a reaction mixture containing UDP-D-GlcNAc and nicotinamide adenine dinucleotide (NAD+) as substrates, two reaction products eluted at 14.8 and 19.0 min were detected by CE analysis (Figure 6). The product peak at 14.8 min was identified as nicotinamide adenine dinucleotide hydrogen (NADH) by comparison with the standard, and the peak at 19.0 min was subsequently identified as UDP-N-acetyl-D-glucosaminuronic acid (UDP-D-GlcNAcA) by mass spectrometry (MS) and MS/MS (MS2) analysis. No products were obtained when GnaA was heat-denatured before addition to the reaction mixture.

(iii) Identification of GnaA product by electrospray ionization (ESI) MS and tandem MS. The product of the GnaA reaction was purified by reversed phase high-performance liquid chromatography (RP HPLC; Supplementary data, Figure S2). The fraction containing the expected product and the substrate was collected and analyzed by negative-mode ESI MS (Supplementary data, Figure S3). An ion peak was obtained at m/z 620.03 (Supplementary data, Figure S3A), which was consistent with the expected mass for UDP-D-GlcNAcA (621.34 Da). MS2 analysis of the parent ion peak resulted in the detection of daughter ion peaks matching the fragments derived from UDP-D-GlcNAcA (Supplementary data, Figure S3C). The ion peak observed at m/z 606.06 (Supplementary data, Figure S3B) is consistent with the substrate UDP-D-GlcNAc (607.35 Da). On its MS2 analysis, daughter ion peaks were detected, which matched the fragments derived from UDP-D-GlcNAcA (Supplementary data, Figure S3D). The fragments corresponding to each peak are listed in Table III.

(iv) Kinetic parameters of GnaA. Kinetic parameters of GnaA for UDP-D-GlcNAc and NAD+ were measured. The initial velocities were measured and used for the kinetic parameter calculations. The kinetics of the reaction catalyzed by GnaA fitted well with the Michaelis–Menten model (Supplementary data, Figure S4). The K_m values of GnaA for UDP-GlcNAc and NAD+ were 0.84 and 4.16 mM, respectively. The detailed kinetic parameters for the enzyme are listed in Table IV.

(v) Determination of physicochemical parameters: optimal temperature and pH, effects of cations on GnaA activities. Activities of GnaA at temperatures ranging from 4 to 65°C are shown in Figure 7A. The highest conversion ratio

Sequencing and analysis of the E. coli O180 O-antigen gene cluster

A sequence of 14,733 bases between JUMPStart and gnd was obtained from E. coli O180, and 14 open reading frames (orfs) (excluding gnd) transcribed from JUMPStart to gnd were identified (Figure 4). All putative genes were assigned functions based on their similarities with those in available databases (Table II).

(i) Nucleotide sugar biosynthesis genes. Orfs 1–4 shared high-level identity (85–96%) with known RmlB, D, A, C from other E. coli and Shigella strains. The family of rml genes is responsible for the synthesis of dTDP-L-Rha (Marumo et al.
1992; Lindqvist et al. 1993; Giraud et al.
1999). orfs 1–4 were named rmbB, D, A, C accordingly. Orfs 5–8 and 11 shared 70, 17, 69, 36 and 56% identity with WbpA, WbpB, WbpD, WbpE and Wbpl of Pseudomonas aeruginosa PA01, respectively, which were identified to be responsible for the synthesis of UDP-D-ManNAc3NAcA (Wenzel et al.
2005; Westman et al.
2007, 2009). orfs 5–8 and 11 were named gnaA, mndA, mndC, mndB and mndD, respectively.

(ii) Sugar transferase genes. Four monosaccharides were identified in the O unit of E. coli O180. The wecA gene that is responsible for the transfer of GlcNAc-1-phosphate to the carrier lipid undecaprenol phosphate to initiate O-unit synthesis is located outside the O-antigen gene cluster (Amer and Valvano 2000); therefore, only three glycosyltransferase genes were expected to be present in the O180 O-antigen gene cluster. Orf12 and 14 belonged to the glycosyltransferase group 1 family (PF00534) and Orf13 to the glycosyltransferase group 2 family (PF00535). Orf12–14 also shared 32–56% identity with many of the glycosyltransferases of other species and were named wfcS, wfcT and wfcU, respectively.

(iii) O-unit processing genes. Both Wzx and Wzy are typical inner membrane proteins. The only two orfs encoding predicted membrane proteins were orf9 and orf10. Orf9 and Orf10 shared 53 and 49% similarity with the O-antigen translocase of Methanosarcina barkeri str. fusa ro and O-antigen polymerase of E. coli O114, respectively. Therefore, orf9 and orf10 were named wzx and wzy.
<table>
<thead>
<tr>
<th>Gene</th>
<th>Position of gene</th>
<th>%G + C content</th>
<th>Conserved domain</th>
<th>Similar protein strain (GenBank accession number)</th>
<th>%Identity/%Similarity (number of amino acid overlap)</th>
<th>Putative function</th>
</tr>
</thead>
<tbody>
<tr>
<td>rmlB</td>
<td>107..1192</td>
<td>43.5</td>
<td>NAD-dependent epimerase/dehydratase family</td>
<td>dTDP-α-glucose 4,6-dehydratase <em>Shigella flexneri</em> (AAA53679)</td>
<td>96/98 (361)</td>
<td>dTDP-α-glucose 4,6-dehydratase</td>
</tr>
<tr>
<td>rmlD</td>
<td>1192..2091</td>
<td>47.7</td>
<td>RmlD substrate binding domain</td>
<td>dTDP-6-deoxy-α-glucose-3,5 epimerase <em>Shigella flexneri</em> (AAA53680)</td>
<td>96/98 (299)</td>
<td>dTDP-6-deoxy-α-glucose-3,5 epimerase</td>
</tr>
<tr>
<td>rmlA</td>
<td>2149..3027</td>
<td>42.6</td>
<td>Nucleotidyltransferase</td>
<td>Glucose-1-phosphate thymidylyltransferase <em>Shigella flexneri</em> (AAA53681)</td>
<td>96/99 (292)</td>
<td>Glucose-1-phosphate thymidylyltransferase</td>
</tr>
<tr>
<td>rmlC</td>
<td>3032..3571</td>
<td>38.3</td>
<td>dTDP-4-dehydrodhamnose 3,5-epimerase PF00922</td>
<td>dTDP-6-deoxy-α-mannose-dehydrogenase <em>Shigella flexneri</em> (AAA53682)</td>
<td>85/91 (169)</td>
<td>dTDP-6-deoxy-α-mannose-dehydrogenase</td>
</tr>
<tr>
<td>gnaA</td>
<td>3616..4920</td>
<td>41.3</td>
<td>UDP-glucose/GDP-mannose dehydrogenase PF03721</td>
<td>UDP glucose 6-dehydrogenase <em>Chromobacterium violaceum</em> (AAQ61790)</td>
<td>82/90 (430)</td>
<td>UDP glucose 6-dehydrogenase</td>
</tr>
<tr>
<td>mndA</td>
<td>4922..5980</td>
<td>41.1</td>
<td>Oxidoreductase family, NAD-binding Rossmann fold</td>
<td>Probable dehydrogenase <em>Chromobacterium violaceum</em> ATCC 12472 (AAQ61789)</td>
<td>75/86 (346)</td>
<td>Oxidoreductase</td>
</tr>
<tr>
<td>mndC</td>
<td>6007..6585</td>
<td>41.8</td>
<td>Bacterial transferase hexapeptide (three repeats)</td>
<td>wlb Bordetella pertussis (CAA62246)</td>
<td>70/84 (183)</td>
<td>Acetyltransferase</td>
</tr>
<tr>
<td>mndB</td>
<td>6588..7688</td>
<td>39.7</td>
<td>DegT/DnmI/EryC1/StrS aminotransferase PF01048</td>
<td>wlb Bordetella pertussis (CAA62247)</td>
<td>43/63 (366)</td>
<td>Aminotransferase</td>
</tr>
<tr>
<td>wex</td>
<td>7685..8944</td>
<td>34.4</td>
<td>Polysaccharide biosynthesis protein</td>
<td>O-antigen translocase <em>Chromobacterium violaceum</em> ATCC 12472 (AAQ61784)</td>
<td>31/53 (413)</td>
<td>O-unit flippase</td>
</tr>
<tr>
<td>wyx</td>
<td>9096..10310</td>
<td>32.8</td>
<td>None</td>
<td>Wyx Escherichia coli (AAT77177)</td>
<td>26/49 (366)</td>
<td>O-antigen polymerase</td>
</tr>
<tr>
<td>wxyD</td>
<td>10381..11472</td>
<td>36.1</td>
<td>UDP-N-acetylglucosamine 2-epimerase PF02350</td>
<td>wbd Bordetella pertussis (CAA62248)</td>
<td>50/66 (360)</td>
<td>UDP-N-acetylglucosamine 2-epimerase</td>
</tr>
<tr>
<td>wfcS</td>
<td>11525..12604</td>
<td>31.8</td>
<td>Glycosyltransferases group 1</td>
<td>LPS N-acetylglucosaminyltransferase <em>Thermus thermophilus</em> HB27 (AAAS0628)</td>
<td>32/49 (359)</td>
<td>Glycosyltransferase</td>
</tr>
<tr>
<td>wfeT</td>
<td>12626..13462</td>
<td>31.9</td>
<td>Glycosyltransferase group 2</td>
<td>Glycosyltransferase (rhamnosyltransferase) <em>Lactobacillus plantarum</em> WCFS1 (CAD63697)</td>
<td>37/59 (224)</td>
<td>Glycosyltransferase</td>
</tr>
<tr>
<td>wfeU</td>
<td>13500..14555</td>
<td>34.7</td>
<td>Glycosyltransferases group 1</td>
<td>WbuV <em>Shigella boydii</em> type 1 (AAV41074)</td>
<td>56/74 (353)</td>
<td>Glycosyltransferase</td>
</tr>
</tbody>
</table>
(75.02%) was detected at 37°C and was markedly reduced at 50°C (16.44%) or above. GnaA was active over a wide pH range with the highest conversion ratio at pH 8.0 (Figure 7B). The effects of cations, including NH$_4^+$, K$^+$, Na$^+$, Ca$^{2+}$, Mn$^{2+}$, Ni$^{2+}$, Co$^{2+}$, Zn$^{2+}$, Mg$^{2+}$, and Fe$^{3+}$ on the activities of GnaA are shown in Figure 7C. Compared with the reaction in the absence of cations, K$^+$ or NH$_4^+$ significantly enhanced the activity of GnaA and Mn$^{2+}$, Zn$^{2+}$, Co$^{2+}$, Ni$^{2+}$, and Fe$^{3+}$ completely inhibited it. The other cations, such as Ca$^{2+}$, Mg$^{2+}$, and Na$^+$, had no obvious effects on the activity of GnaA.

**Discussion**

*Escherichia coli* O180 is a newly identified serogroup, and strains belonging to O180 have been found to contain virulence factors and cause diarrhea (Scheutz et al. 2004). In this work, the *E. coli* O180 O-antigen structure and gene cluster were identified, clarifying the genetic basis of this new O-antigen form. An unusual monosaccharide, D-ManpNAc3NAcA, was identified in the O180 O unit, which has been reported earlier as a component of the O antigens of *Pseudomonas* (Knirel and Kochetkov 1994) and *Bordetella* (Preston et al. 2006) but found for the first time in *E. coli* O antigens. The biosynthetic pathway of UDP-D-ManpNAc3NAcA has been elucidated in *P. aeruginosa*. This sugar nucleotide is synthesized from UDP-D-GlcNAc via a five-step pathway, with WbpA (a UDP-GlcNAc 6-dehydrogenase) catalyzing the initial step (Miller et al. 2004; Westman et al. 2007, 2009; Larkin and Imperiali 2009).

GnaA of *E. coli* O180 shared 70% identity with the homologous protein identified in *P. aeruginosa* and 70–83% identity with UDP-Glc 6-dehydrogenase, UDP-GlcNAC 6-dehydrogenase and UDP-ManNAC 6-dehydrogenase from a number of other species. The alignment of GnaA and other proteins sharing high identity (including the identified UDP-D-GlcNAC 6-dehydrogenase) revealed the presence of an NADH/nicotinamide adenine dinucleotide phosphate (NADPH)-binding motif GXGXXG (Wang et al. 2008) and residues involved in catalysis [Glu165, Lys218, Cys273 and Asp277 in the previously identified UDP-D-GlcNAC 6-dehydrogenase (NP_251849); Miller et al. 2004] in GnaA of O180 (Figure 8).

Analysis of the cation requirements showed that the activities of both O180 GnaA and known UDP-D-GlcNAC 6-dehydrogenase increased in the presence of K$^+$ or NH$_4^+$. The effects of Ca$^{2+}$, Mn$^{2+}$, Ni$^{2+}$, Co$^{2+}$, Zn$^{2+}$, Mg$^{2+}$, and Fe$^{3+}$ were not analyzed in the previously identified UDP-D-GlcNAC 6-dehydrogenase, and the inhibitory effects of these cations on the two enzymes cannot be compared.

The $K_m$ values of GnaA in *E. coli* O180 for UDP-GlcNAc and NAD$^+$ were 0.84 ± 0.11 and 4.16 ± 0.40, respectively, and those of the identified UDP-D-GlcNAC 6-dehydrogenase in *P. aeruginosa* were 0.094 and 0.22 mM, respectively, indicating a high affinity of the latter enzyme for the substrates.
Materials and methods

Bacterial strains and plasmids

The bacterial strains used in this study are listed in Table V. The bacteria were inoculated into Luria–Bertani (LB) medium and incubated overnight at 37°C.

Construction of shotgun bank, sequencing and sequence analysis

Chromosomal DNA was prepared as described previously (Bastin and Reeves 1995). The primer pair: 5′-ATTGGTACCTGTAAGCCAGGCGGTACTGC-3′ and 5′-CACTGGCC

Fig. 7. Effects of temperature (A), pH (B) and cations (C) on the conversion ratio of GnaA.
A TACCGACGACGCCGA TCTGTTGCTTGG-3′, based on the JUMPstart sequence and gnd, respectively (Wang and Reeves 1998), was used to amplify the O-antigen gene clusters of the E. coli O180 strain. Amplification by polymerase chain reaction (PCR) was performed as follows: denaturation at 94°C for 10 s, annealing at 60°C for 30 s and extension at 68°C for 15 min. Shotgun banks were constructed for each strain as described previously (Wang and Reeves 1998). Sequencing was carried out using an ABI 3730 automated DNA sequencer (Applied Biosystems, Foster City, CA), and sequence data were analyzed using computer programs as described previously (Feng et al. 2004).

### Isolation of LPS
Bacteria were grown to late log phase in 8 L of LB using a 10-L fermenter (BIOSTAT C-10, B. Braun Biotech International, Germany) under constant aeration at 37°C and pH 7.0. Bacterial cells were washed and dried as described (Robbins and Uchida 1962).

LPS (1.4 g) was isolated from dried cells (7.5 g) using the phenol–water method (Westphal and Jann 1965) and purified by the precipitation of nucleic acids and proteins with aqueous 50% trichloroacetic acid at 4°C.

### Degradation of LPSs
Delipidation of the LPS (100 mg) was performed with aqueous 2% HOAc (6 mL) at 100°C until the precipitation of lipid A. The precipitate was removed by centrifugation (13,000 × g, 20 min) and the supernatant was fractionated by gel-permeation chromatography on a Sephadex G-50 (S) column (56 × 2.6 cm) (Amersham Biosciences, Sweden) in 0.05 M pyridinium acetate buffer, pH 4.5, monitored by a differential refractometer (Knauer, Germany). A high-molecular-mass OPS was obtained in a yield of 42% of the LPS mass.

### Chemical analyses
The OPS was hydrolyzed with 2 M CF₂CO₂H (120°C, 2 h). Monosaccharides were identified by GLC of the alditol acetates on a Hewlett-Packard 5890 chromatograph (USA) equipped with an Ultra-1 column using a temperature gradient of 150–290°C at 5°C/min. The absolute configurations of the monosaccharides were determined by GLC of the acetylated (S)-2-octyl glycosides as described (Leontein et al. 1978; Gerwig et al. 1979).

---

**Table V.** Strains, plasmids and primers used in this study

<table>
<thead>
<tr>
<th>Strain or plasmid</th>
<th>Description</th>
<th>Source a</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strains</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G2258</td>
<td>Escherichia coli O180 strain 86–381</td>
<td>A</td>
</tr>
<tr>
<td>Escherichia coli BL21 (DE3)</td>
<td>F’ ompT hsdS₂ (rK₂ mK₂) gal dcm (DE3)</td>
<td>Novagen</td>
</tr>
<tr>
<td>Escherichia coli DH5α</td>
<td>F’ supE44 thi-1 gyrA96 relA1 (lacZYA-argF)</td>
<td>U169</td>
</tr>
<tr>
<td><strong>Plasmids</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pET-28a(+)</td>
<td>T7 Express vector, Kan’</td>
<td>Novagen</td>
</tr>
<tr>
<td>pET28a-6 containing N-terminally 6-histidine-tagged E. coli O180 gnaA at the NdeI/XhoI site</td>
<td>This study</td>
<td></td>
</tr>
</tbody>
</table>

aA, the International Escherichia and Klebsiella Centre, Statens Serum Institut, Copenhagen, Denmark; B, Tianjin Biochip Corporation, Tianjin, China.

---

**Fig. 8.** Alignment of GnaA and similar proteins. The GenBank accession numbers of these proteins are shown, and GnaA is from E. coli O180. The region of the NADH/NADPH binding motif is marked by a box and the residues involved in catalysis are shown by triangles.
NMR spectroscopy

An OPS sample was deuterium-exchanged by freeze-drying twice from 99.5% D2O and then examined as a solution in 99.95 D2O at 30°C. NMR spectra were recorded on a Bruker DRX-500 spectrometer (Germany) using acetone (δ1 2.225, δc 31.45) as internal reference. Two-dimensional NMR spectra were obtained using standard Bruker software, and Bruker XWINNMR 2.6 software was used to acquire and process the NMR data. Mixing times of 200 and 100 ms were used in TOCSY and ROESY experiments, respectively.

Plasmid construction, protein expression and purification

The gnaA gene was amplified by PCR using the following primers: 5′-GGGAATTCCATATGTTACACTGCCTAT TAATAAAA-3′ and 5′-CCGCTCGAGTTAAGCCCTACAA CATTGTTTT-3′. A total of 25 cycles were performed using the following conditions: denaturation at 95°C for 60 s, annealing at 48°C for 30 s and extension at 72°C for 90 s, in a final volume of 50 µL. The amplified genes were cloned into pET28a+ to construct pLW1238 and the presence of the insert was confirmed by sequencing using an ABI 3730 Sequencer.

*Escherichia coli* BL21 carrying the recombinant plasmid was grown in LB medium containing 50 µg/mL kanamycin overnight at 37°C. The overnight culture (5 mL) was inoculated into 500 mL of fresh LB and grown at 37°C until A600 reached 0.6. The expression of gnaA was induced with 0.1 mM IPTG for 4 h at 25°C. Cells were subsequently harvested by centrifugation, washed with binding buffer (50 mM Tris–HCl, pH 8.0, 300 mM NaCl and 10 mM imidazole), resuspended into 5 mL of the same buffer containing 1 M of phenylmethanesulfonyl fluoride and 1 mg/mL lysozyme and sonicated. Cell debris was removed by centrifugation and total soluble proteins in the supernatant were collected. The His6-tagged fusion proteins in the supernatant were purified by nickel ion affinity chromatography using a Chelating Sepharose Fast Flow column (GE Healthcare, Madison, WI) according to the instructions provided by the manufacturer. Unbound proteins were washed out with 100 mL of wash buffer (50 mM Tris–HCl, pH 8.0, 300 mM NaCl and 20 mM imidazole). The fusion proteins were eluted with 3 mL of elution buffer (50 mM Tris–HCl, pH 8.0, 300 mM NaCl and 250 mM imidazole) and dialyzed overnight against 50 mM Tris–HCl buffer (pH 8.0) at 4°C. Protein concentration was determined by the Bradford method.

Enzyme activity assays and CE analysis

GnaA activity was assayed in reaction mixture containing 0.5 mM UDP-α-D-GlcNAc, 1.5 mM NAD+, 5 mM KCl, 0.5 mM K2HPO4–KH2PO4 (pH 8.0) and 12.3 µM purified GnaA protein in a total volume of 20 µL and the reaction was carried out at 37°C for 1 h. Products generated from the reaction were analyzed by CE. Enzyme activities were indicated by the conversion of substrates into products.

CE was performed using a Beckman Coulter P/ACE MDQ Capillary Electrophoresis System with a PDA detector (Beckman Coulter, CA). The capillary was bare silica 75 µm (internal diameter, i.d.) × 57 cm, with the detector positioned at 50 cm. The capillary was conditioned before each run by washing with 0.1 M NaOH, followed by deionized water and finally with 25 mM borate-sodium hydroxide (pH 9.6) (used as the mobile phase; 2 min/wash). Samples were loaded by pressure injection at 0.5 psi for 10 s and separation was carried out at 20 kV and monitored by measuring UV absorbance at 260 nm. Peak integration and trace alignments were performed with the Beckman P/ACE station software (32 Karat™ version 5.0). The conversion ratio was calculated by comparing the peak areas of substrate and product.

RP HPLC and ESI-MS analysis

The GnaA reaction mixture was separated by RP HPLC using an LC-20AT HPLC (Shimadzu, Japan) with a Venusil MP-C18 column (5 µm particle, 4.6 by 250 mm) (Agela Technologies, Inc., Willmington, DE). The mobile phase was composed of 5% acetonitrile and 95% 50 mM triethylamine-acetic acid (pH 6.5), and the flow rate was 0.6 mL/min. Fractions containing the expected products were collected, lyophilized and redissolved in water before being injected into a Finnigan LCQ Advantage MAX ion trap mass spectrometer (Thermo Fisher, CA) at a negative mode (4.5 kV, 250°C) for ESI tandem MS analysis.

Determination of temperature and pH optima and cofactor requirements of GnaA

For parameter characterization, all reactions contained 1 mM UDP-α-D-GlcNAc and 3 mM NAD+, 5 mM KCl and 1.23 µM GnaA in a buffer of 50 mM K2HPO4–KH2PO4. To determine the optimum temperature for GnaA, reactions were carried out at 4, 15, 25, 37, 50 and 65°C for 40 min. For investigation of pH effects, reactions were performed at pH 6.5, 7.0, 7.5, 8.0, 8.5, 9.0, 9.5 and 10.0 at 37°C for 40 min. To analyze the effects of different cations on GnaA activity, all reactions contained 12.3 µM of the enzyme, 1 mM UDP-α-D-GlcNAc, 3 mM NAD+ and 50 mM Tris–HCl in a total volume of 20 µL. Salts (5 mM), such as KCl, NH4Cl, NaCl, MgCl2, MnCl2, NiCl2, CaCl2, CoCl2, ZnCl2 and FeCl3, were added. Ethylene diamine tetraacetic acid was also tested and results were compared with those obtained in the absence of cations as the control. Reactions were incubated at 37°C. All enzyme activities were determined by CE.

Measurement of kinetic parameters

All reactions contained 50 mM K2HPO4–KH2PO4 (pH 8.0), 5 mM KCl and 1.23 µM of enzyme in a final volume of 20 µL. To measure the Km and Vmax values of GnaA for UDP-α-D-GlcNAc, reactions were carried out over a range of UDP-α-D-GlcNAc concentrations (0.125–2.5 mM) and a constant concentration of NAD+ (1.5 mM). To measure the Km and Vmax values of GnaA for NAD+, reactions were carried out over a range of NAD+ concentrations (0.25–5 mM) and a constant concentration of UDP-α-D-GlcNAc (0.5 mM). All reactions were carried out at 37°C for 3 min in a total volume of 20 µL. The reactions were terminated by adding an equal volume of chloroform (20 µL). Conversion of UDP-α-D-GlcNAc to UDP-α-D-GlcNAAc and NAD+ to NADH was monitored by CE. Km and Vmax values were calculated based on the
Michaelis–Menten equation. Results are the average of three independent experiments.

**Nucleotide sequence accession numbers**
The DNA sequence of the O-antigen gene cluster of *E. coli* O180 has been deposited in the GenBank database under the accession number JQ751058.

**Supplementary data**
Supplementary data for this article is available online at http://glycob.oxfordjournals.org/.

**Funding**
This work was supported by grants from the National Basic Research Program of China (973 Program) (2012CB721001, 2012CB721101, 2011CB504900), the General Program of the National Natural Science Foundation of China (30900255, 81071392, 31111120026), the Key Program of the National Natural Science Foundation of China (31030002), Tianjin Research Program of Application Foundation and Advanced Technology (10JCZDJC10100), the Research Fund for the Doctoral Program of Higher Education of China (2011003110022), the Fundamental Research Funds for the Central Universities (65020121, 65020061), the Federal Targeted Program for Research and Development in Priority Areas of Russia’s Science and Technology Complex for 2007–2013 (State contract No. 16.552.11.7050) and the Russian Foundation for Basic Research (11-04-91173-NSF, 11-04-01020).

**Acknowledgements**
We thank Dr Flemming Scheutz (The International Escherichia and Klebsiella Centre, Statens Serum Institut, Copenhagen, Denmark) for providing *E. coli* O180 strain 86-381 used in this study.

**Conflict of interest**
None declared.

**Abbreviations**
CE, capillary electrophoresis; COSY, correlation spectroscopy; d-ManpN3NAcA, 2,3–diametamido-2,3-dideoxy-6-manno-pyranuronic acid; ESI, electrospray ionization; GLC, gas-liquid chromatography; Glc, glucose; GlcNAcA, N-acetylglucosaminuronic acid; GlcPNAc or GlcNAc, N-acetylglucosamine; HMBC, heteronuclear multiple bond correlation; HSQC, heteronuclear single quantum coherence; IPTG, isopropylthio-β-D-galactopyranoside; LB, Luria–Bertani; LPS, lipopolysaccharide; ManNAc, N-acetylmannosamine; ManpN3NAcA, 2,3–diametamido-2,3-dideoxy-manno-pyranuronic acid; MS, mass spectrometry; NAD+, nicotinamide adenine dinucleotide; NADH, nicotinamide adenine dinucleotide phosphate; NADPH, nicotinamide adenine dinucleotide phosphate hydrogen; OPS, O polysaccharide; PCR, polymerase chain reaction; Rhap or Rha, rhamnopyranose or rhamnose; ROESY, rotating-frame nuclear overhauser effect correlation spectroscopy; RP HPLC, reversed-phase high-performance liquid chromatography; SD–PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; TOCSY, total correlation spectroscopy; UDP-N-acetylglucosaminuronic acid.

**References**


E. coli O180 O antigen and UDP-D-Man3NAcA biosynthesis


