Genetic analysis of the O-antigen of *Providencia alcalifaciens* O30 and biochemical characterization of a formyltransferase involved in the synthesis of a Qui4N derivative

Bin Liu2,4, Miao Chen2,4, Andrei V Perepelov5, Jianfeng Liu6, Olga G Ovchinnikova5, Dawei Zhou2,4, Lu Feng2,4, Antoni Rozalski7, Yuriy A Kniref2, and Lei Wang1,2,3,4

1To whom correspondence should be addressed: Tel: +86-22-66229588; Fax: +86-22-66229596; e-mail: wanglei@nankai.edu.cn

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

O-Antigen is a component of the outer membrane of Gram-negative bacteria and one of the most variable cell surface constituents, giving rise to major antigenic variability. The diversity of O-antigen is almost entirely attributed to genetic variations in O-antigen gene clusters. Bacteria of the genus *Providencia* are facultative pathogens, which can cause urinary tract infections, wound infections and enteric diseases. Recently, the O-antigen gene cluster of *Providencia* was localized between the *cpxA* and *yibK* genes in the genome. However, few genes involved in the synthesis of *Providencia* O-antigens have been functionally identified. In this study, the putative O-antigen gene cluster of *Providencia alcalifaciens* O30 was sequenced and analyzed. Almost all putative genes for the O-antigen synthesis were found, including a novel formyltransferase gene *vioF* that was proposed to be responsible for the conversion of dTDP-4-amino-4,6-dideoxy-4-formamido-D-glucose (dTDP-o-Qui4N) to dTDP-4,6-dideoxy-4-formamido-D-glucose (dTDP-o-Qui4NFo). *vioF* was cloned, and the enzyme product was expressed as a His-tagged fusion protein, purified and assayed for its activity. High-performance liquid chromatography was used to monitor the enzyme–substrate reaction, and the structure of the product dTDP-o-Qui4NFo was established by electrospray ionization tandem mass spectrometry and nuclear magnetic resonance spectroscopy. Kinetic parameters of *vioF* were determined, and effects of temperature and cations on its activity were also examined. Together, the functional analyses support the identification of the O-antigen gene cluster of *P. alcalifaciens* O30.

Keywords: formyltransferase / O-Antigen gene cluster / *Providencia alcalifaciens* / 4,6-Dideoxy-4-formamido-D-glucose biosynthesis

**Introduction**

O-Antigen (O-polysaccharide), containing many repeats of an oligosaccharide unit (O-unit), comprises the outermost domain of the lipopolysaccharide (LPS) molecule present in the outer cell wall membrane of Gram-negative bacteria. The O-units usually contain two to eight residues of a broad range of sugars, both common and rarely occurring, and their derivatives. The O-antigen exhibits variations in the types of sugars present, their arrangement within the O-unit and the linkages within and between O-units, making it one of the most variable cell constituents. The variability in the O-antigen thus provides a major basis of serotyping schemes for many Gram-negative bacteria.

O-Antigen is highly immunogenic and also used as a receptor by some bacteriophages, which may both contribute to maintenance of diversity by intermittent selection against specific O-antigen forms (Reeves and Wang 2002). The diversity of the O-antigen is also thought to be important in allowing the various clones to present variations in surface structures that may offer selective advantage in their specific niche. The O-antigen is also essential for the full function of bacteria and is related to bacterial virulence.

Bacteria of the genus *Providencia* from the *Enterobacteriaceae* family, a component of the normal intestinal flora, are facultative pathogens, which, under favorable conditions, may cause mainly urinary tract and wound infections as well as enteric diseases (O’Hara et al. 2000). These infections are frequently persistent, difficult to treat and may even result in fatal...
bacteremia. Currently, the genus *Providencia* consists of eight species (O’Hara et al. 2000; Somvanshi et al. 2006; Juneja and Lazzaro 2009), among which *P. alcalifaciens* has been described as a possible diarrhea-causing pathogen in travelers and children in developing countries (Yoh et al. 2005). The existing serological scheme of three *Providencia* species (*P. stuartii*, *P. rustigianii* and *P. alcalifaciens*) used in serotyping of clinical isolates is based on the O-antigens and flagella H-antigens and includes 63 O-serogroups and 30 H-serogroups (Ewing 1986). By now, more than 35 *Providencia* O-polysaccharide structures have been established (Knirel 2011).

The genes involved in the biosynthesis of O-antigens are generally found on the chromosome within an O-antigen gene cluster, and genetic variation in this area is the major basis for the diversity of O-antigen forms. Genes involved in O-antigen synthesis are classified into three main classes: (i) nucleotide sugar precursor synthesis genes, (ii) sugar transfer genes associated with O-unit formation and (iii) O-unit processing genes associated with the conversion of the O-unit to O-antigen, i.e. *wzx/wzy* genes for the Wzx/Wzy-dependent O-antigen synthesis and *wzm/wzt* genes for the ABC transporter-dependent O-antigen synthesis. Genetic analysis of the O-antigen gene cluster is useful for understanding the mechanisms of O-antigen diversification and provides a basis for the development of the DNA-based typing system (Liu et al. 2008).

Within a species the O-antigen gene clusters commonly map to the same site on the chromosome. For instance, the O-antigen gene clusters are located between *galF* and *gnd* in *Escherichia coli*, *Shigella* and *Salmonella enterica* strains. Recently, based on the structural and genetic data, we have found that the O-antigen gene clusters of *Providencia* strains are located between the *cpxA* and *yibK* genes (Ovchinnikova et al. 2012). However, few genes in the O-antigen gene clusters of *Providencia* have been functionally characterized apart from the GDP-cotilose synthesis genes in *P. alcalifaciens* O6 (Ovchinnikova et al. 2012).

In this study, we sequenced and analyzed a putative O-antigen gene cluster between *cpxA* and *yibK* of *P. alcalifaciens* O30. Almost all putative genes responsible for the O-antigen synthesis were found, including a novel formyltransferase gene *vioF*, which was further demonstrated biochemically to be involved in the synthesis of dTDP-4,6-dideoxy-4-formamido-D-glucose (dTDP-D-Qui4NFo). The data obtained confirmed the identity of the O-antigen gene cluster of *P. alcalifaciens* O30.

### Results

**Characterization of the O-antigen gene cluster of *P. alcalifaciens* O30**

A sequence of 21,256 bases was obtained between the *cpxA* and *yibK* genes from *P. alcalifaciens* O30 type strain 19,372. Seventeen open-reading frames (ORFs) were identified, all of which have the same transcriptional direction from *cpxA* to *yibK* (Figure 1). All of the genes were assigned functions based on their similarities to genes from available databases (Table I).

The O-unit of *P. alcalifaciens* O30 is a pentasaccharide containing two residues of α-glucuronic acid (α-GlcA) and one residue each of D-Qui4NFo, 2-acetamido-4-amino-2,4,6-trideoxy-D-galactose (D-GalNAc) and D-ribofuranose (D-Ribf) (Figure 1) (Kocharova et al. 2006). Orf2 shares a high-level identity to RmlA proteins from different bacteria that convert glucose 1-phosphate to dTDP-glucose. Orf4 shares 66% identity to VioA of E. coli O7, which has been identified biochemically as an aminotransferase responsible for the conversion of dTDP-6-deoxy-D-xilo-hexos-4-ulos into dTDP-D-Qui4N (Wang et al. 2007). Thus, *orf2* and *orf4* were identified as *rmlA* and *vioA*, respectively, and named accordingly. In addition to RmlA and VioA, 4,6-dehydratase RmlB is involved in the biosynthetic pathway of dTDP-D-Qui4N, catalyzing the conversion of dTDP-D-glucose to dTDP-6-deoxy-D-xilo-hexos-4-ulos (Wang et al. 2007). However, *rmlB* was not found in the O-antigen gene cluster of *P. alcalifaciens* O30, and we proposed that this gene is located outside the cluster. Orf1 shares 48% identity to a putative formyltransferase of Francisella tularensis. We suggested that *orf1* is responsible for the synthesis of dTDP-D-Qui4NFo from dTDP-D-Qui4N and named it *vioF*. Orf9 shares 95% identity to UDP-glucose 6-dehydrogenase Ugd of *P. rustigianii* DSM 4541, which converts the UDP-D-Glc to UDP-D-GlcA (Stevenson et al. 1996). Thus, *orf9* was proposed to be responsible for the synthesis of UDP-D-GlcA and named *ugd*. Orf11 and Orf13 share 65% identity to WbgX and WbgZ of *Shigella sonnei*, respectively, which are putative sugar synthetases involved in the synthesis of the O-antigen of *P. alcalifaciens* (Shepherd et al. 2000). As α-FucNAc4N is the only sugar shared by the O-antigens of *S. sonnei* and *P. alcalifaciens* O30, we proposed that *orf11* and *orf13* are responsible for the synthesis of UDP-D-FucNAc4N (Liu et al. 2008), and named them *wbgX* and *wbgZ*, respectively. ADP-D-Ribf is available from the NAD salvage pathway (Hillyard et al. 1981).

![Fig. 1. O-Antigen structure (top) and gene cluster organization (bottom) of *P. alcalifaciens* O30. Note: The sequences of *cpxA* and *yibK* are incomplete. The positions of each gene in the gene cluster: *vioF*, from positions 1707–2465; *rmlA*, from positions 2543–3418; *wzx*, from positions 3423–4853; *vioA*, from positions 4881–5990; *wpsh*, from positions 6055–7074; *wpal*, from positions 7064–8011; *wzy*, from positions 8008–9189; *wpja*, from positions 9158–10,003; *ugd*, from positions 10,082–11,248; *wpak*, from positions 11,251–12,018; *wbgX*, from positions 12,022–13,173; *wbgY*, from positions 13,173–13,784; *wbgZ*, from positions 13,795–15,657; *galE*, from positions 15,693–16,718; *wza*, from positions 17,287–18,186; *wzh*, from positions 18,176–18,637; *wzc*, from positions 18,687–20,768.](http://glycob.oxfordjournals.org/)

---

*O-antigen of *Providencia alcalifaciens* O30*

A sequence of 21,256 bases was obtained between the *cpxA* and *yibK* genes from *P. alcalifaciens* O30 type strain 19,372. Seventeen open-reading frames (ORFs) were identified, all of which have the same transcriptional direction from *cpxA* to *yibK* (Figure 1). All of the genes were assigned functions based on their similarities to genes from available databases (Table I).

The O-unit of *P. alcalifaciens* O30 is a pentasaccharide containing two residues of α-glucuronic acid (α-GlcA) and one residue each of D-Qui4NFo, 2-acetamido-4-amino-2,4,6-trideoxy-D-galactose (D-GalNAc) and D-ribofuranose (D-Ribf) (Figure 1) (Kocharova et al. 2006). Orf2 shares a high-level identity to RmlA proteins from different bacteria that convert glucose 1-phosphate to dTDP-glucose. Orf4 shares 66% identity to VioA of E. coli O7, which has been identified biochemically as an aminotransferase responsible for the conversion of dTDP-6-deoxy-D-xilo-hexos-4-ulos into dTDP-D-Qui4N (Wang et al. 2007). Thus, *orf2* and *orf4* were identified as *rmlA* and *vioA*, respectively, and named accordingly. In addition to RmlA and VioA, 4,6-dehydratase RmlB is involved in the biosynthetic pathway of dTDP-D-Qui4N, catalyzing the conversion of dTDP-D-glucose to dTDP-6-deoxy-D-xilo-hexos-4-ulos (Wang et al. 2007). However, *rmlB* was not found in the O-antigen gene cluster of *P. alcalifaciens* O30, and we proposed that this gene is located outside the cluster. Orf1 shares 48% identity to a putative formyltransferase of Francisella tularensis. We suggested that *orf1* is responsible for the synthesis of dTDP-D-Qui4NFo from dTDP-D-Qui4N and named it *vioF*. Orf9 shares 95% identity to UDP-glucose 6-dehydrogenase Ugd of *P. rustigianii* DSM 4541, which converts the UDP-D-Glc to UDP-D-GlcA (Stevenson et al. 1996). Thus, *orf9* was proposed to be responsible for the synthesis of UDP-D-GlcA and named *ugd*. Orf11 and Orf13 share 65% identity to WbgX and WbgZ of *Shigella sonnei*, respectively, which are putative sugar synthetases involved in the synthesis of the O-antigen of *P. alcalifaciens* (Shepherd et al. 2000). As α-FucNAc4N is the only sugar shared by the O-antigens of *S. sonnei* and *P. alcalifaciens* O30, we proposed that *orf11* and *orf13* are responsible for the synthesis of UDP-D-FucNAc4N (Liu et al. 2008), and named them *wbgX* and *wbgZ*, respectively. ADP-D-Ribf is available from the NAD salvage pathway (Hillyard et al. 1981).
<table>
<thead>
<tr>
<th>Orf No.</th>
<th>Gene name</th>
<th>G + C content (%)</th>
<th>Conserved domain(s)</th>
<th>Similar protein(s), strain(s) (Genbank accession no.)</th>
<th>% Identical % similar (No. of aa overlap)</th>
<th>Putative function of protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>vioF</td>
<td>28.2</td>
<td>Formyltransferase (PF00551) $E$ value = $3.3 \times e^{-17}$</td>
<td>formyltransferase, <em>Francisella tularensis</em> (AAS60274)</td>
<td>48/65 (239)</td>
<td>Formyltransferase</td>
</tr>
<tr>
<td>2</td>
<td>rmlA</td>
<td>35.6</td>
<td>Nucleotidyl transferase (PF00483), $E$ value = $3.9 \times e^{-72}$</td>
<td>glucose 1-phosphate thymidylyltransferase, <em>Vibrio cholerae</em> (ADP80947)</td>
<td>75/89 (288)</td>
<td>Glucose 1-phosphate thymidylyltransferase</td>
</tr>
<tr>
<td>3</td>
<td>wzx</td>
<td>30.3</td>
<td>Polysaccharide biosynthesis protein (PF01943) $E$ value = $6 \times e^{-34}$</td>
<td>Wzx, <em>E. coli</em> O123 (ABG81783)</td>
<td>50/71 (417)</td>
<td>O-unit flippase</td>
</tr>
<tr>
<td>4</td>
<td>vioA</td>
<td>33.2</td>
<td>DegT/DnrJ/EryC1/StrS aminotransferase (PF01041) $E$ value = $3 \times e^{-34}$</td>
<td>VioA, <em>E. coli</em> O7 (AAS55721)</td>
<td>66/79 (367)</td>
<td>Aminotransferase</td>
</tr>
<tr>
<td>5</td>
<td>wpaH</td>
<td>25.9</td>
<td>Glycosyl transferase family 2 (PF00535) $E$ value = $4.7 \times e^{-19}$</td>
<td>glycosyltransferase, <em>Clostridium sp.</em> 7_2_43FAA (ZP_05129960)</td>
<td>35/52 (256)</td>
<td>Glycosyltransferase</td>
</tr>
<tr>
<td>6</td>
<td>wpaI</td>
<td>31.0</td>
<td>Phosphoribosyl transferase (PF00156), $E$ value = $5.7 \times e^{-11}$</td>
<td>phosphoribosyltransferase, <em>Mesorhizobium opportunistum</em> WSM2075 (ZP_05810989)</td>
<td>33/52 (322)</td>
<td>Glycosyltransferase</td>
</tr>
<tr>
<td>7</td>
<td>wzy</td>
<td>27.1</td>
<td>Glycosyl transferase family 2 (PF00535) $E$ value = $1.4 \times e^{-21}$</td>
<td>Wzy, <em>E. coli</em> O79 (ACA24739.1)</td>
<td>25/50 (167)</td>
<td>O-Antigen polymerase</td>
</tr>
<tr>
<td>8</td>
<td>wpaJ</td>
<td>26.5</td>
<td>Glycosyl transferase family 2 (PF00535) $E$ value = $1.4 \times e^{-21}$</td>
<td>glycosyltransferase, <em>Vibrio cholerae</em>(ABK85714)</td>
<td>33/58 (261)</td>
<td>Glycosyltransferase</td>
</tr>
<tr>
<td>9</td>
<td>ugd</td>
<td>46.2</td>
<td>UDP-glucose/GDP-mannose dehydrogenase family, NAD binding domain (PF03721), $E$ value = $1.2 \times e^{-40}$</td>
<td>UDP-glucose 6-dehydrogenase, <em>Providencia rettgeri</em> DSM 4541 (ZP_05974404)</td>
<td>95/98 (388)</td>
<td>UDP-glucose 6-dehydrogenase</td>
</tr>
<tr>
<td>10</td>
<td>wpaK</td>
<td>31.1</td>
<td>Glycosyl transferase family 2 (PF00535), $E$ value = $7 \times e^{-38}$</td>
<td>glycosyl transferase family protein, <em>Shevanella pakeana</em>, ATCC 700345 (ABV86738)</td>
<td>44/66 (256)</td>
<td>Glycosyltransferase</td>
</tr>
<tr>
<td>11</td>
<td>wbgX</td>
<td>35.9</td>
<td>DegT/DnrJ/EryC1/StrS aminotransferase (PF01041) $E$ value = $2.8 \times e^{-27}$</td>
<td>WbgX, <em>S. sonnei</em> (AAK85171)</td>
<td>65/82 (384)</td>
<td>Aminotransferase</td>
</tr>
<tr>
<td>12</td>
<td>wbgY</td>
<td>31.7</td>
<td>Bacterial sugar transferase (PF02397), $E$ value = $1.3 \times e^{-22}$</td>
<td>WbgY, <em>S. sonnei</em> (AAK85172)</td>
<td>55/75 (192)</td>
<td>Glycosyltransferase</td>
</tr>
<tr>
<td>13</td>
<td>wbgZ</td>
<td>31.5</td>
<td>Polysaccharide biosynthesis protein (PF02719) $E$ value = $3 \times e^{-15}$</td>
<td>WbgZ, <em>S. sonnei</em> (AAK85173)</td>
<td>53/73 (613)</td>
<td>UDP-GlcNAc C6 dehydroxylase</td>
</tr>
<tr>
<td>14</td>
<td>galE</td>
<td>43.8</td>
<td>NAD-dependent epimerase/dehydroxylase family (PF01370), $E$ value = $5.3 \times e^{-50}$</td>
<td>UDP-glucose 4-epimerase, <em>Providencia rettgeri</em> DSM 1131 (ZP_06127560)</td>
<td>88/94 (339)</td>
<td>UDP-glucose 4-epimerase</td>
</tr>
<tr>
<td>15</td>
<td>wza</td>
<td>49.1</td>
<td>Polysaccharide biosynthesis/export protein (PF02563), $E$ value = $2.9 \times e^{-21}$</td>
<td>Putative polysaccharide export protein <em>Wza</em>, <em>Serratia odorifera</em> 4Rx13 (EFA17643)</td>
<td>63/78 (299)</td>
<td>Polysaccharide export protein</td>
</tr>
<tr>
<td>16</td>
<td>wzb</td>
<td>43.5</td>
<td>Low-molecular-weight phosphotyrosine protein phosphatase (PF01451), $E$ value = $9.1 \times e^{-37}$</td>
<td>Wzb, <em>Listonella anguillarum</em> (ABV89352)</td>
<td>47/68 (146)</td>
<td>Tyrosine phosphatase</td>
</tr>
<tr>
<td>17</td>
<td>wzc</td>
<td>38.5</td>
<td>Chain length determinant protein (PF02706), $E$ value = $9.1 \times e^{-37}$</td>
<td>Putative tyrosine-protein kinase <em>Wzc</em>, (EDK29043)</td>
<td>38/63 (700)</td>
<td>Tyrosine-protein kinase</td>
</tr>
</tbody>
</table>
Orf6 was found to belong to the phosphoribosyl transferase family (PF00156, $E = 5.7 \times 10^{-11}$), and is thus likely responsible for the transfer of $\beta$-Rib. Orf12 belongs to the bacterial sugar transferase family (PF02397, $E = 1.3 \times 10^{-57}$) and shares 55% identity to WbgY of *S. sonnei*, which is proposed to be responsible for the transfer of the first sugar ($\alpha$-FucNac4) of *S. sonnei* O-unit (Shepherd et al. 2000; Xu et al. 2002). We proposed that in *P. alcalifaciens* O30, t-FucNac4 is also the first sugar of the O-unit and that Orf12 has the same function as WbgY. Orf5, Orf8 and Orf10 were found to belong to the glycosyl transferase group 2 family (PF00535, $E = 5.7 \times 10^{-19}$), which is proposed to be WbgY. Orf5, Orf8, orf10 and orf12 were proposed to be glycosyltransferase genes and named wpaH, wpaI, wpaJ, and wbgY, respectively.

Both Wzx and Wzy are highly hydrophilic membrane proteins. Orf3 and Orf7 are the only two proteins with predicted multi-membrane segments. Orf3 was predicted to have 12 well-proportioned transmembrane segments, and shares 71% similarity to the putative O-antigen flipase of *E. coli* O123. Orf7 was found to have eight predicted transmembrane segments with a large periplasmic loop of 84 amino acid residues, which is a typical topological characteristic of Wzy proteins (Daniels et al. 1998). Orf7 shares 50% similarity to the Wzy protein of *E. coli* O79. Therefore, orf3 and orf7 were proposed to be the genes encoding O-antigen flipase (wzx) and O-antigen polymerase (wzy), respectively, and were named accordingly.

For four additional genes found in the O-antigen gene cluster of *P. alcalifaciens* O30, Orf14 shares 88% identity to GalE (UDP-glucose 4-epimerase) of *P. rettgeri* DSM 1131, and orf15-17 are homologues of wza, wzb and wzc genes, which are required for the surface expression of capsular polysaccharides (Rahn et al. 1999; Drummlsmith and Whitfield 2000; Wugeditsch et al. 2001). orf14–orf17 were named galE, wza, wzb and wzc, respectively. galE has been found in other O-antigen gene clusters of *P. alcalifaciens*, independently of the presence of galactose in the O-unit (Ovchinnikova et al. 2012). wza, wzb and wzc are also present upstream of yibK in all other *Providencia* strains studied, except for *P. rustigianii* DSM 454, and may play a role in the transport and regulation of *Providencia* O-antigens (Ovchinnikova et al. 2012).

Biochemical characterization of VioF

To determine the biosynthetic pathway of dTDP-$\alpha$-Qui4N and to confirm that the genes between cpxA and yibK are responsible for the synthesis of the O-antigen of *P. alcalifaciens* O30, the function of *vioF* was investigated. VioF was expressed as a His-tagged fusion protein in *E. coli* BL21 by isopropyl $\beta$-D-thiogalactopyranoside (IPTG) induction, and purified to near homogeneity by nickel ion affinity chromatography (Supplementary data, Figure S1). The molecular mass estimated by SDS–PAGE was 29.0 kDa, corresponding well to the calculated mass (29.3 kDa).

The activity of VioF was monitored by high-performance liquid chromatography (HPLC), and N-10-formyltetrahydrofolate was used as the formyl donor. VioF catalyzed conversion of dTDP-$\alpha$-Qui4N to a product that eluted at 14–17 min (Figure 2), which was purified by HPLC (data not shown) and identified by electrospray ionization mass spectrometry (ESI-MS). An [M − $\text{H}^+$] ion peak was observed at m/z 574.16 (Supplementary data, Figure S2A), which is in agreement with the expected molecular mass for dTDP-$\alpha$-Qui4NFo (575.36 Da). MS/MS (MS$^2$) analysis of this ion peak revealed ion peaks matching the fragments derived from dTDP-$\alpha$-Qui4NFo (Supplementary data, Figure S2B). Interpretation of each peak in the mass spectra is shown in Table II.

The identity of the VioF product as dTDP-$\alpha$-Qui4NFo was further confirmed by nuclear magnetic resonance (NMR) spectroscopy. The $^1$H NMR spectrum of the VioF product contained, *inter alia*, two characteristic signals for an $N$-formyl group at $\delta$ 8.03 and 8.21 for the E (minor) and Z (major) isomers, respectively (Kocharova et al. 2003). The spectrum was completely assigned using two-dimensional $^1$H,$^1$H correlation spectroscopy (COSY) and total correlation spectroscopy (TOCSY). In the TOCSY spectrum (Figure 3), cross-peaks were observed between all protons within each spin system.

### Table II. ESI-MS analysis data

<table>
<thead>
<tr>
<th>Composition of fragment</th>
<th>Molecular formula</th>
<th>Calculated molecular mass (Da)</th>
<th>[M − $\text{H}^+$] (m/z)</th>
</tr>
</thead>
<tbody>
<tr>
<td>dTDP-Qui4NFo (full scan)</td>
<td>C$<em>7$H$</em>{16}$P$_2$O$_5$N$_3$ Na</td>
<td>597.34</td>
<td>596.12</td>
</tr>
<tr>
<td>dTDP-Qui4NFo-Na</td>
<td>C$<em>7$H$</em>{16}$P$_2$O$_5$N$_3$</td>
<td>575.36</td>
<td>574.16</td>
</tr>
<tr>
<td>dTDP-Qui4NFo (MS$^2$, m/z 574.16)</td>
<td>C$<em>9$H$</em>{18}$P$_2$O$_8$N$_3$</td>
<td>401.18</td>
<td>401.13</td>
</tr>
<tr>
<td>dTDP</td>
<td>C$<em>{10}$H$</em>{18}$P$_2$O$_8$N$_3$</td>
<td>383.16</td>
<td>382.98</td>
</tr>
<tr>
<td>dTDP</td>
<td>C$<em>{10}$H$</em>{18}$P$_2$O$_8$N$_3$</td>
<td>332.12</td>
<td>332.11</td>
</tr>
<tr>
<td>dTDP-H$_2$O</td>
<td>C$<em>{10}$H$</em>{18}$P$_2$O$_8$N$_3$</td>
<td>321.20</td>
<td>321.03</td>
</tr>
<tr>
<td>Qui4NFo-PO$_3$H$_2$</td>
<td>C$<em>{10}$H$</em>{18}$P$_2$O$_8$N$_3$</td>
<td>252.14</td>
<td>252.15</td>
</tr>
<tr>
<td>dTDP-H$_2$O-thymine</td>
<td>C$_3$H$_5$PO$_4$</td>
<td>195.09</td>
<td>195.09</td>
</tr>
</tbody>
</table>
including those between H1 at δ 5.58 and H2−H6 of the Qui4NFo sugar moiety at δ 3.62 (H2), 3.81 (H3), 3.72 (H4), 4.18 (H5) and 1.20 (H6); between H1′ at δ 6.36 and H2′−H5′ of 2′-deoxyribose in the dTDP moiety at δ 2.38 (H2′), 4.64 (H3′) and 4.19 (H4′ + H5a′,5b′); and between H6 and CH3 of thymine at δ 7.75/1.94. The assigned chemical shifts were in good agreement with published data for dTDP and α-Qui4NAc (Wang et al. 2007), except for those of H3, H4 and H5 of Qui4NFo, which were higher by 0.04, 0.09 and 0.12 ppm, respectively, than those of Qui4NAc. These differences are evidently due to a higher deshielding effect of the N-formyl group when compared with the N-acetyl group.

Relatively large $^{3}J_{2,3}$, $^{3}J_{3,4}$ and $^{3}J_{4,5}$ coupling constants of 9–10 Hz determined from the $^{1}$H NMR spectrum for the Qui4NFo residue are characteristic for all-axial orientation of the protons of the sugar pyranose ring, which has thus the gluco configuration. A $^{3}J_{1,2}$ coupling constant of 3.6 Hz showed that Qui4NFo has the α configuration. Therefore, the sugar moiety in the VioF product is 4,6-dideoxy-4-formamido-α-D-glucopyranose.

The $^{31}$P NMR spectrum of dTDP-α-Qui4NFo contained signals for a diphosphate group at δ −11.2 and −12.7. As expected, they showed intense cross-peaks with H5′ of dTDP at δ −11.2/4.19 and H1 of Qui4NFo at δ −12.7/5.58 in the $^{1}$H, $^{31}$P heteronuclear multiple-quantum coherence (HMQC) spectrum. These data confirmed finally that the VioF product is dTDP-α-Quip4NFo, whose structure is depicted in Figure 3.

Kinetic parameters of VioF were measured. The initial velocities were determined and used for calculating kinetic parameters for VioF. The kinetics of the reaction catalyzed by VioF fit well the Michaelis–Menten mode. The $K_{m}$ and $V_{max}$ values of VioF for dTDP-D-Qui4N were determined to be 547 ± 11 μM and 0.092 ± 0.015 μM s$^{-1}$, respectively.

Effects of temperature and divalent cations on the activity of VioF were examined as well. Activities of VioF at temperatures ranging from 4 to 75°C are shown in Supplementary data, Figure S3. VioF showed high activity over a wide temperature range from 15 to 50°C, which then began to decline at temperature above 50°C. The VioF activity was slightly enhanced by Fe$^{2+}$, inhibited by Fe$^{3+}$, Co$^{2+}$, Ni$^{2+}$ and relatively unaffected by Mg$^{2+}$, Ca$^{2+}$, Mn$^{2+}$ and EDTA (Supplementary data, Figure S3).

Discussion

In this study, the function of VioF of P. alcalifaciens O30 as a formyltransferase that converts dTDP-α-Qui4N to dTDP-α-Qui4NFo, using N-10-formyltetrahydrofolate as the formyl
donor, was fully characterized. d-Quin4NFo is a component of the O-antigen of *P. alcalifaciens* O30, and N-formyl-modified amino sugars have also been found in lipid A and O-antigens of several other Gram-negative bacteria (Knirel et al. 1985; Vinogradov et al. 1991; Katzenellenbogen et al. 1995; Muldoon et al. 2001; Breazeale et al. 2002; Kocharova et al. 2003). VioF is the second formyltransferase using an amino-sugar nucleotide substrate thus far to be characterized biochemically, and the other is the N-terminal domain of ArnA that catalyzes N-10-formyltetrahydrofuran-dependent formation of UDP-4-amino-4-deoxy-α-L-arabinose (UDP-4-ara4N) to give UDP-4-ara4NFo (Breazeale et al. 2005).

Formyltransferases utilizing other substrates have also been identified, including methionyl-tRNA N-formyltransferase (FMT) (Schmitt et al. 1996; Schmitt et al. 1998) and glycine-mide ribonucleotide formyltransferase (GARF) (Inglese et al. 1990). Like these formyltransferases, VioF requires N-10-formyltetrahydrofuran but does not recognize N-5-formyltetrahydrofuran or N-5,N-10-formyltetrahydrofuran. Multiple sequence alignment of VioF, N-terminal domain of ArnA, FMT and GARF revealed only a weak similarity of VioF to the others (Supplementary data, Figure S4). However, VioF contains three conserved amino acid residues (Asn-100, His-102 and Asp-137) known to function in GARF catalysis (Warren et al. 1996; Shim and Benkovic 1999). In FMT and GARF, the side chains of these three residues adopt the identical conformation (Schmitt et al. 1996). In FMT, they are in close vicinity to the docked 5-deaza-5,6,7,8-tetrahydrofuran molecule (Schmitt et al. 1996). The “SLLP” motif is conserved in many enzymes that bind tetrahydrofuran (Schmitt et al. 1998). The “GLNP” sequence was found at the corresponding site in VioF (Supplementary data, Figure S4), and we speculated that it may have a similar function to that of the “SLLP” motif.

This biochemical study of the VioF enzyme involved in the synthesis of the O-antigen in the genus *Providencia* provides strong experimental support for the identified O-antigen gene cluster. The putative genes accounting for almost all functions expected for synthesis of the O-antigen of *P. alcalifaciens* O30 were found between the cxpA and yibK genes. The only exception is rmlB, which may be located outside the O-antigen gene cluster. Similarly, in *P. alcalifaciens* O36 and O19, RmlB is involved in the biosynthetic pathways of unique monosaccharides present in their O-antigens, but no rmlB has been found in their O-antigen gene clusters (Ovchinnikova et al. 2012). Analysis of the available *Providencia* genomes revealed homologs of rmlB within the enterobacterial common antigen gene clusters (Ovchinnikova et al. 2012).

Another unusual component of the O-antigen of *P. alcalifaciens* O30, d-FucNAc4N, is shared by the O-antigen of *S. sonnei*. Two pathways have been proposed for the biosynthesis of UDP-d-FucNAc4N (Kowal and Wang 2002; Liu et al. 2008). One includes WbgV and WbgX (Shepherd et al. 2000) and the other includes WbgZ and WbgX (Xu et al. 2002). A homolog of wbgZ was found in the O-antigen gene cluster of *P. alcalifaciens* O30, suggesting that this bacterium utilizes the latter pathway. It is likely although not proven that WbgZ catalyzes the conversion of UDP-d-GlcNAc to UDP-2-acetamido-2,6-dideoxy-d-xylono-hexos-4-ulse, and then WbgX converts the WbgZ product to UDP-d-FucNAc4N. Biochemical characterization of WbgZ and WbgX in vitro will be a subject of a further study.

In most *E. coli*, *Shigella* and *S. enterica* strains, WecA catalyzes the transfer of GlcNAc 1-phosphate or GalNAc 1-phosphate to an undecaprenol phosphate carrier, thus triggering the O-unit assembly, and the wecA gene is located outside the O-antigen gene cluster (Alexander and Valvano 1994). WecA has also been suggested to initiate the O-antigen synthesis in *P. alcalifaciens* O6, O19, O36, O44 and *P. stuartii* O47 (Ovchinnikova et al. 2012). However, no GlcNAc or GalNAc residue has been found in the O-antigen of *P. alcalifaciens* O30. In *S. sonnei*, a d-FucNAc4N residue is attached to the lipopolysaccharide core as the first sugar of the O-unit (Gamian and Romanowska 1982). WbgY has been suggested to be an enzyme that transfers d-FucNAc4N to the lipid carrier and, therefore, to play the same role as WecA (Xu et al. 2002). As a homolog of WbgY is present in the O-antigen gene cluster of *P. alcalifaciens* O30, we speculated that d-FucNAc4N is also the first sugar of the O-unit of this bacterium. Consequently, the O-units would be linked to each other by the d-FucNAc4N linkage, which is formed by Wzy-catalyzed polymerization. While further functional studies are needed to fully characterize the O-antigen gene cluster of *Providencia*, the findings of this study contribute overall to the cumulative knowledge of the diverse repertoire of O-antigens.
GenBank, Clusters of Orthologous Groups (COG) and Pfam protein motif databases (Tatusov et al. 2001; Bateman et al. 2002). The program TMHMM 2.0 (http://www.cbs.dtu.dk/services/TMHMM-2.0/) was used to identify potential transmembrane segments.

Cloning and plasmid construction

vioF from P. alcalifaciens O30 was amplified by PCR using the primers wl-19862 (5-GGAATTCGAATTGCAATGAT GAG-3) and wl-19864 (5-CCGCTCGAGCTCTTGACTTA AAGATAATTC-3). A total of 30 cycles were performed using the following conditions: denaturation at 95°C for 30 s, annealing at 50°C for 30 s and extension at 72°C for 1 min, in a final volume of 20 μL. The amplified genes were cloned into pET21a+ to construct pLW1048, and the presence of the inserts was confirmed by sequencing using an ABI 3730 sequencer.

Protein expression and purification

Escherichia coli BL21 (DE3) carrying pLW1048 was grown in LB medium containing 100 μg mL−1 ampicillin overnight at 37°C with shaking. The overnight culture was inoculated into 500 mL of fresh medium and grown to OD600 0.6. Expression of VioF was induced by 0.1 mM IPTG at 30°C for 4 h. After the IPTG induction, cells were harvested by centrifugation, washed with binding buffer (50 mM Tris–HCl, pH 8.0, 300 mM NaCl and 10 mM imidazole), resuspended in the same buffer containing 1 mM phenylmethanesulfonyl fluoride and 1 mg mL−1 lysozyme and sonicated (Ultraschallprozessor UP200S). The cell debris was removed by centrifugation, and the supernatant containing soluble proteins was collected. The His6-tagged fusion proteins were purified by nickel ion affinity chromatography with a Chelating Sepharose Fast Flow column (GE Healthcare) according to the manufacturer’s instructions. Unbound proteins were removed with 100 mL of wash buffer (50 mM Tris–HCl, pH 8.0, 300 mM NaCl and 25 mM imidazole). 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 containing 20% glycerol (pH 7.4) at 4°C. The protein concentration was determined by the Bradford method. Purified proteins were stored at −80°C.

Enzyme activity assays

dTDP-n-Qu4N was prepared using RmlA, RmlB and VioA from E. coli O7 as described earlier (Wang et al. 2007). N-5,N-10-formyltetrahydrofolate was prepared from N-5-formyltetrahydrofolate (Promega) as described previously (Breazeale et al. 2002). The reaction mixture contained 1 mM dTDP-n-Qu4N, 7 mM N-5,N-10-formyltetrahydrofolate with N-5,N-10-formyltetrahydrofolate as the precursor, a 30-min preincubation at 30°C before adding purified VioF, was necessary for the conversion to the actual formyl donor, N-10-formyltetrahydrofolate (Breazeale et al. 2002), 50 mM HEPEs buffer (pH 7.5), 2 mM DTT, 2.7 μM purified VioF in a total volume of 50 μL. The reaction was then carried out at 37°C for 30 min before termination by adding an equal volume of chloroform.

The product was analyzed by HPLC and structurally characterized by ESI-MS and NMR spectroscopy. The enzymatic activity was demonstrated by the conversion of the substrate into the product.

HPLC and ESI-MS analysis

The product was monitored by HPLC using a Shimadzu LC-20A (Kyoto, Japan) with a Venusil MP-C18 column (5 μm particle, 4.6 × 250 mm) (Agela Technologies). The large-scale purification as required for NMR analysis was performed by HPLC using a 1:19 acetonitrile and 50 mM triethylammonium acetate mixture (pH 6.0) as the mobile phase at a flow rate of 0.6 mL min−1 and monitored by UV detection at 260 nm. The conversion ratio was calculated by comparing the peak areas of the substrate and the product. The fraction containing the expected product was lyophilized and re-dissolved in methanol before injecting into a Finnigan LCQ Advantage MAX ion trap mass spectrometer (Thermo Electron, CA) operated at negative mode (4.5 kV, 250°C) for ESI-MS analysis. For MS² analysis, nitrogen was used as the collision gas and helium was used as the auxiliary gas; collision energies used were typically 20–30 eV.

NMR spectroscopy

A sample of dTDP-D-Qu4NFo (~0.2 mg) was deuterium-exchanged by freeze drying from 99.9% D₂O, dissolved in 99.95% D₂O (150 μL) and examined using a Shigemi microtube (Japan). NMR spectra were recorded on a Bruker Avance II 600-MHz spectrometer (Germany) at 30°C using internal sodium 3-trimethylsilylpropionate-2,2,3,3-d₄ (δH 0) and external aqueous 85% H₃PO₄ (δp 0) as references. NMR spectra were obtained using standard Bruker software, and Bruker TopSpin 2.1 program was used to acquire and process the NMR data. A mixing time of 200 ms was used in the TOCSY experiment.

Measurement of kinetic parameters

To measure the Kₘ and Vₘₐₓ values of VioF for dTDP-n-Qu4N, reactions were carried out with various concentrations of dTDP-n-Qu4N (0.29–1.0 mM) and a fixed concentration of VioF (0.14 μM). All reactions were performed at 37°C for 5 min in a final volume of 50 μL. Conversion of dTDP-D-Qu4N to dTDP-n-Qu4NFo was monitored by HPLC. Kₘ and Vₘₐₓ values were calculated based on the Michaelis–Menten equation. The reported data were the average of three independent experiments.

Determination of temperature and divalent cations requirements

To determine the temperature optima for VioF, reactions were carried out at different temperatures (4, 15, 25, 30, 37, 50, 55, 65, 75°C), pH 8.0, for 30 min. To test the effects of divalent cations on the enzyme activity, reactions were carried out in the presence of 10 mM MgCl₂, MnCl₂, FeCl₂, CaCl₂, CoCl₂, NiCl₂, NiSO₄ or EDTA at 37°C, pH 8.0, for 30 min.
Nucleotide sequence accession number

The DNA sequence of *P. alcalifaciens* O30 O-antigen gene cluster has been deposited in GenBank under the accession number JQ801294.

Supplementary Data

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

Funding

This work was supported by the National 973 program of China (2012CB721101, 2012CB721001, 2009CB522603, 2011CB504900), National Natural Science Foundation of China (NSFC) Key Program (31030002), NSFC General Program (30900041 and 81171524), Tianjin Research Program of Application Foundation and Advanced Technology (10JCYBJC10000), the Fundamental Research Funds for the Central Universities (65020121 and 65020061), the Russian Foundation for Basic Research (Grants 08-04-92221), the Federal Targeted Program of Higher Education of China (20090031120023), the Program of Higher Education of China (20090031120023), the Russian Foundation for Basic Research, the Russian Foundation of Application Foundation and Advanced Technology Program (30900041 and 81171524), Tianjin Research Program of Higher Education of China (Grant 803, core funding for statutory research activity of the Department of Immunobiology of Bacteria).

Conflict of interest

None declared.

Abbreviations

d-FucNAc4N, 2-acetamido-4-amino-2,4,6-trideoxygalactose; d-GlcA, d-glucuronic acid; d-Qu4N, 4-amino-4,6-dideoxy-d-glucose; d-Qu4NFo, 4,6-dideoxy-4-formamido-d-glucose; d-Ribf, d-ribofuranose; COSY, correlation spectroscopy; ESI-MS, electrospray ionization mass spectrometry; FMT, methionyl-tRNA N-formyltransferase; GARF, glucamine ribonucleotide formyltransferase; HMQC, heteronuclear multi-quantum coherence; HPLC, high-performance liquid chromatography; IPTG, isopropyl β-D-thiogalactopyranoside; MS2, tandem mass spectrometry; NMR, nuclear magnetic resonance; TOCSY, total correlation spectroscopy.

References


