The O-antigen structure of bacterium *Comamonas aquatica* CJG

Xiqian Wang,† Anna N. Kondakova,‡ Yutong Zhu,³ Yuryi A. Knirel,²,* and Aidong Han¹,*

**Abstract**

Genus *Comamonas* is a group of bacteria that are able to degrade a variety of environmental waste. *Comamonas aquatica* CJG (*C. aquatica*) in this genus is able to absorb low-density lipoprotein but not high-density lipoprotein of human serum. Using ¹H and ¹³C NMR spectroscopy, we found that the O-polysaccharide (O-antigen) of this bacterium is comprised of a disaccharide repeat (O-unit) of d-glucose and 2-O-acetyl-l-rhamnose, which is shared by *Serratia marcescens* 06. The O-antigen gene cluster of *C. aquatica*, which is located between *coaX* and *tnp4* genes, contains rhamnose synthesis genes, glycosyl and acetyl transferase genes, and ATP-binding cassette transporter genes, and therefore is consistent with the O-antigen structure determined here.

**INTRODUCTION**

*Comamonas* belongs to *Comamonadaceae* of Proteobacteria. *Comamonas aquatica* CJG, isolated from Kunming Lake in 1963, is a Gram-negative bacterium, and can efficiently adsorb low-density lipoprotein (LDL), ox-LDL and very LDL (VLDL) rather than high-density lipoprotein (HDL) in human and mammal sera [1, 2].

O-polysaccharide, commonly called O-antigen as a major substance in responses of host immune systems, is a key component in the outer membrane of Gram-negative bacteria. It consists of repeats of an oligosaccharide (O-unit), which usually has two–eight residues of common and rare sugars and their derivatives. O-antigen contributes the major antigenic variability to the cell surface and is subject to intense selections by the host immune systems and other environmental factors.

Genes for O-antigen synthesis are normally located as a gene cluster. Typically, three different types of genes are present within the cluster: (i) nucleotide-sugar synthesis genes that encode enzymes involved in the synthesis of specific sugar components of an O-unit; (ii) sugar transferase genes that assemble the sugar substituents into the O-unit; and (iii) processing genes that encode enzymes involved in the O-unit’s translocation and polymerization [3].

Recently, we sequenced the whole *C. aquatica* CJG genome [4]. In this study, we analysed this genome and found an O-antigen gene cluster located between *coaX* and *tnp4*, which contains genes for dTDP-1-Rha synthesis, putative sugar and acetyl transferases, and the ATP-binding cassette (ABC)-2 transporter. We also determined the structure of the *C. aquatica* O-antigen, which is comprised of a disaccharide repeat of 2-O-acetyl-l-rhamnose and d-glucose. This work provides a molecular basis for further characterization of O-antigens of *Comamonas* for possible involvement in LDL/VLDL absorption.

**RESULT AND DISCUSSION**

With the genomic sequence of *C. aquatica* available, we wanted to assess whether this bacterium has an O-antigen gene cluster through a series of analyses of its whole genomic sequence. Because protein sequences in those well-known enterococci are extremely different from that in *C. aquatica*, it was not possible to identify those genes through linear protein domain similarities. We took a hidden Markov model (HMM) approach to analyse the essential domains of all functional proteins in *C. aquatica*. Prodigal [5] and HMMER v.3.0 [6] were used for gene analyses. The TMHMM Server v.2.0 was used to identify potential transmembrane segments [7]. Sequence alignments and
Comparisons were performed by the ClustalW program [8]. Characteristic rhamnose synthesis genes rmlA-D responsible for dTDP-L-rhamnose production helped us to localize the O-antigen cluster since they share high identities (75–88%) at the protein level to their homologues, which are usually located at the 5’ end of the O-antigen gene clusters in Salmonella and E. coli [9]. The rest of the genes were manually annotated using UniProt [10] and InterProScan [11] as the major protein and motif databases for possible gene functions.

A total of 15 proteins encoded by orf1-15, among which orf1-4 are rmlBDAC genes, were finally identified in this gene cluster (Fig. 1). Their functions were assigned based on the similarities of the amino acid sequences to those annotated in the available databases (Table S1, available in the online Supplementary Material). Notably, d-glucose synthesis genes are not in this cluster as d-glucose is a common sugar involved in many other cellular processes.

The protein encoded by orf13 belongs to the ABC-2 transporter family (PF00005; E=6.0e-45), a permease component of ABC-type polysaccharide-polyol phosphate export systems. Orf13 exhibited 37% identity and 60% similarity to O-unit flippase (Wzm) in Methylobacterium morosus KoM1. Hydrophobicity analysis indicated that Orf13 is an integral membrane protein with six transmembrane segments, the average number for Wzm proteins (data not shown). Orf14 belongs to the ABC transporter family (PF00005; E=3.8e-90), an ATPase component of the ABC-type polysaccharide-polyol phosphate export system. Orf14 also shows 39% identity and 63% similarity to Wzt in Bacillus halodurans ATCC BAA-125. The presence of the sequence GLGAGKS, an ATP-binding protein motif (Walker box), as well as the ABC transporter family signature sequence YSSGMVQRLAFAVA [12], further indicates that orf14 encodes an ATP-binding protein. Therefore, we concluded that orf13 and orf14 are wzm- and wzt-type genes.

Glycosyltransferases are specific for sugar donors and acceptors and the linkage that they form. C. aquatica with an O-unit consisting of two monosaccharides (see below) requires an initial adaptor for the synthesis of O-antigen. Accordingly, three glycosyltransferase genes, orf9-11, were found in the O-antigen gene cluster and named rfbZ, wbuZ and adaZ, respectively. RfbZ belongs to the glycosyltransferase group 1 family (PF00534; E=2.1e-50) with 36% identity and 49% similarity to a glycosyltransferase of Ramlibacter tataouinensis ATCC BAA-407. It also shares 27% identity and 41% similarity to a dTDP-rhamnosyl transferase in Acidocella sp. MX-AZ02. WbuZ belongs to glycosyltransferase family 2 (PF00535; E=4.1e-127) with 49% identity and 65% similarity to a glycosyltransferase in Nitrosomonas sp. AL212.

AdaZ is a bifunctional enzyme, which possibly possesses N-acetyl-glucosaminyl-phosphatidylinositol deacetylase activity in addition to glycosyltransferase activity. This protein, which is bigger than most glycosyltransferases involved in the biosynthesis of bacterial O-antigens, has 597 amino acids. The C-terminal half has a reasonable similarity (49%) to a glycosyltransferase family 2 (PF00535; E=4.1e-127) protein in Candidatus magnetoturum sp. HK-1. The N-terminal half has a higher similarity (55%) to a de-N-acetylase (PF02585; E=2.4e-32) in Beggiatoa sp. PS. Here, AdaZ may possibly serve as an initial glycosyltransferase involved in the synthesis of an O-antigen adaptor in C. aquatica.

Gene orf12, named oatZ, encodes an acetyltransferase (PF00005; E=3.8e-90) in the C. aquatica O-antigen gene cluster, which is consistent with the 2-O-acetylation of the L-rhamnose residue (see below). Gene orf15, designated as wbdZ, encodes a methyl-transferase (PF08241; E=1.1e-44). In the ABC transporter-dependent pathway, O-antigen chain growth may be terminated by O-methylation of a sugar component [13]. In E. coli O8 and K. pneumonia O5, the O-polysaccharides are terminated with a 3-O-methylmannose residue [14]. The existence of orf15 suggests that the O-methylation could be a terminal modification of the O-polysaccharide in this strain.

Finally, the O-antigen cluster has four genes that may not be relevant to O-polysaccharide synthesis. Genes orf7 and orf8, named glyQ and glyS, encode glycine-tRNA ligase alpha and beta subunits responsible for the synthesis of glycine-tRNA ligase. The functions of genes orf5 and orf6 are unknown. Hydrophobicity analysis indicated that Orf5 is a membrane protein, called HypA, while Orf6, named HypB, contains a DNA-binding domain. Existence of these four genes may result from genomic recombination, and whether!

---

**Fig. 1.** Organization of the O-antigen gene cluster of C. aquatica CJG. The genes are coloured in groups and labelled on the top. The potential functions of each group are indicated below.
they play roles in the O-polysaccharide synthesis requires further investigation.

We further isolated the O-polysaccharide of *C. aquatica* and determined its structure using sugar analysis and NMR spectroscopy. The bacterium was grown in Luria–Bertani media at 37°C. The lipopolysaccharide was prepared from 2 g of dried bacterial cells and treated with 150 ml 45 % aqueous phenol by stirring at 70°C for 45 min [15]. The mixture was then dialysed against distilled water and acidified to pH 2 by adding 50 % aqueous trichloroacetic acid at 4°C. The supernatant was cleaned by centrifugation, dialysed and freeze dried. The O-polysaccharide prep was obtained after treating the lipopolysaccharide with 2 % acetic acid at 100°C for 3 h and cleaned by high-speed centrifugation. The supernatant was further purified by Sephadex G-50 gel chromatography in 0.05 M pyridinium acetate buffer, pH 4.5, to yield a high-molecular mass O-polysaccharide fraction.

Sugar analysis by gas–liquid chromatography of the acetylated alditols [16] and (S)-2-octyl glycosides [17] derived after hydrolysis with 2 M CF<sub>3</sub>CO<sub>2</sub>H (100°C, 3 h) showed that the O-polysaccharide of *C. aquatica* consisted of D-glucose and L-rhamnose in the ratio of ~1 : 1.

An O-polysaccharide sample was dissolved in 99.9 % D<sub>2</sub>O for the NMR spectroscopic experiments on an Avance II 600 MHz instrument (Bruker, Germany) at 30°C using sodium 3-(trimethylsilyl) propanoate-2,2,3,3-d<sub>4</sub> (δ<sub>H</sub> 0, δ<sub>C</sub> 1.6) as the internal reference for calibration. Two-dimensional ¹H, ¹H COSY, ¹H, ¹H ROESY and ¹H, ¹³C HSQC spectra were acquired and processed using standard Bruker software TopSpin 2.1. A mixing time of 150 ms was used in the ROESY experiment.

Assignments of the ¹H and ¹³C NMR chemical shifts are presented in Fig. 2 and listed in Table 1. The presence of signals for two anomic atoms at δ<sub>H</sub> 5.04 and 5.05, δ<sub>C</sub> 95.8 and 100.4 showed that the O-polysaccharide has a disaccharide O-unit. ³J<sub>HH</sub> coupling constants estimated from the two-dimensional NMR spectra and a comparison of the ¹³C NMR chemical shifts with published data [18] enabled identification of two monosaccharide components as α-glucopyranose (α-GlcP) and β-rhamnopyranose (β-Rhap).

Low-field positions at δ 78.4 and 76.0 of the signals for C-4 of Glc and C-3 of Rha, respectively, indicated that the O-polysaccharide is linear with four-substituted Glc and three-substituted Rha residues. The structure of the O-polysaccharide was confirmed by a ¹H, ¹H ROESY experiment, which showed Glc H-1/Rha H-3 and Rha H-1/Glc H-4 correlations at δ 5.04/3.85 and 5.05/3.65, respectively. The

![Fig. 2. Two-dimensional ¹H, ¹³C HSQC spectrum of the O-polysaccharide of *C. aquatica* CJG. One-dimensional ¹H and ¹³C NMR spectra are shown along the horizontal and vertical axes, respectively. Numbers are referred to the H/C pairs in sugar residues denoted by G for Glc and R for Rha.](image-url)
chemical shift of δ 95.8 for C-1 of Glc showed that the constituent monosaccharides have different absolute configurations [19]. As in other natural carbohydrates, Glc is assumed to be D, and hence, Rha has the L configuration. A low-field position at δ 5.63 of the signal for H-2 of Rha was evidently due to a deshielding effect, indicating O-acetylation of this monosaccharide at position 2.

Based on these data, it was concluded that the O-polysaccharide of *C. aquatica* has the following structure: →3)-β-1-L-Rhap2Ac-(1→4)-α-D-GlcP-(1→. The only species that is currently known to have this O-polysaccharide structure is *Serratia marcescens* O6 [20]. The biological implication of such a similarity remains to be explored in the future.

In summary, we located and characterized the O-antigen gene cluster in *C. aquatica* genome. Five genes, adoZ, rpfZ, wbuZ, wzm and wzt, are specific for this bacterial strain. The region from rmlB to glyS is conserved in two other stains of *C. aquatica* (data not shown). The predicted functions of most genes in this cluster are consistent with the O-antigen structure determined by the NMR spectroscopy.

### Funding information
This work was supported by National Science Foundation of China (31370723 and 31570752 to A. H.), National Key Basic Research Program (2013CB910603 to A. H.).

### Acknowledgements
We thank Dr A.S. Shashkov for providing access to the NMR spectrometer and Dr Zhiliang Ji for useful discussions.

### Conflicts of interest
The authors declare that there are no conflicts of interest.

### Ethical statement
The funders had no role in experimental design, data collection and analysis, decision to publish, or preparation of the manuscript.

### References

**Table 1.** $^1$H and $^{13}$C NMR chemical shifts (δ, ppm) of the O-polysaccharide of *C. aquatica* CJG.

<table>
<thead>
<tr>
<th>Residue</th>
<th>Nucleus</th>
<th>Monosaccharide</th>
<th>O-Acetyl</th>
</tr>
</thead>
<tbody>
<tr>
<td>→3)-β-1-Rhap2Ac-(1→</td>
<td>$^1$H</td>
<td>5.05</td>
<td>3.85</td>
</tr>
<tr>
<td>→4)-α-D-GlcP-(1→</td>
<td>$^1$H</td>
<td>5.04</td>
<td>3.80</td>
</tr>
<tr>
<td>→3)-β-1-Rhap2Ac-(1→</td>
<td>$^1$C</td>
<td>100.4</td>
<td>76.0</td>
</tr>
<tr>
<td>→4)-α-D-GlcP-(1→</td>
<td>$^1$C</td>
<td>95.8</td>
<td>73.9</td>
</tr>
</tbody>
</table>