Carbohydrate Polymers 388 (2026) 125547

Contents lists available at ScienceDirect

Carbohydrate Polymers

journal homepage: www.elsevier.com/locate/carbpol

Abequose-containing O-antigen polysaccharide from *Aeromonas sobria* K221 (serogroup PGO1): structure, O-antigen gene cluster correlation, and Abe-transferase prediction☆

Anna Turska-Szewczuk [a,*], Katarzyna Dworaczek [a], Arkadiusz Bomba [b], Maria Kubaczyńska [c], Dominika Drzewiecka [d], Ewelina Iwan [b], Ewelina Kamińska [b], Agnieszka Pękala-Safińska [e]

[a] Department of Genetics and Microbiology, Institute of Biological Sciences, M. Curie-Sklodowska University, Akademicka 19, 20-033, Lublin, Poland
[b] Department of Research Support, National Veterinary Research Institute, Partyzantow 57, 24-100, Pulawy, Poland
[c] Chair and Department of Medical Microbiology, Medical University of Lublin, Chodzki 1, 20-093, Lublin, Poland
[d] Department of Biology of Bacteria, Faculty of Biology and Environmental Protection, University of Lodz, Banacha 12/16, 90-237, Lodz, Poland
[e] Department of Preclinical Sciences and Infectious Diseases, Faculty of Veterinary Medicine and Animal Science, Poznan University of Life Sciences, Wolynska 35, 60-637, Poznan, Poland

ARTICLE INFO

Keywords:
Aeromonas
O-antigen polysaccharide structure
O-antigen gene cluster
Abequose
Abequosyltransferase
NMR

ABSTRACT

Aeromonas species are important opportunistic pathogens in freshwater aquaculture, including Polish fish farms, where outbreak-associated isolates have been assigned to local provisional serogroups. Among these, PGO1 is frequently detected, but the structural and genetic diversity of the corresponding O-specific polysaccharides remains poorly understood. Here, we characterized *Aeromonas sobria* strain K221, a PGO1 isolate recovered from common carp during an outbreak of motile *Aeromonas* infection/septicaemia. The O-specific polysaccharide (OPS), isolated from LPS, was analysed by chemical methods and $^{1}H/^{13}C$ NMR spectroscopy, and its O-repeating unit was identified as a linear pentasaccharide containing β-Glc*p*NAc, 2-*O*-acetylated α-Rha*p*, and α-Abe*p* residues. Bioinformatic analysis of the O-antigen gene cluster (OGC) revealed gene content consistent with the OPS structure and supported functional assignment of the biosynthesis locus, including a putative α-1,3-CDP-abequosyltransferase. Both the OPS and its OGC differed from those of other recently characterized PGO1 strains, indicating that K221 represents a distinct variant within this serogroup. The occurrence of 2-substituted abequose, not previously reported in *Aeromonas* O-polysaccharides or as a 2-substituted residue in bacterial O-polysaccharides, highlights an unusual structure–biosynthesis relationship. These findings reveal greater structural and genetic diversity within the *Aeromonas* PGO1 serogroup than previously recognized and expand current knowledge of bacterial O-specific carbohydrate polymers.

1. Introduction

Aeromonas bacteria, whose natural habitat is the aquatic environment, are primarily known as opportunistic pathogens of free-living and farmed fish. Mesophilic *Aeromonas* species are the etiological agents of Motile Aeromonas Infection/Motile Aeromonas Septicaemia (MAI/MAS) diseases in fish, including carp, trout, channel catfish, eel, and tilapia, whereas the psychrophilic *A. salmonicida* species causes furunculosis in salmonids (Janda & Abbott, 2010; Kozińska & Pękala, 2010; Tomas, 2012).

The virulence factors of *Aeromonas* that contribute to their high pathogenic potential and facilitate adhesion, colonization, and subsequent host-cell damage include outer membrane proteins (OMPs), capsular polysaccharide (CPS), lipopolysaccharide (LPS), S-layer, polar flagella and pili, iron-binding systems, exotoxins, and secretion systems (Garduno et al., 2000; Matys et al., 2020; Merino et al., 1996; Rabaan et al., 2001; Tomas, 2012). Among these, LPS plays a key role in virulence, serving as both a structural component of the outer membrane and a determinant of antigenic specificity. LPS is a negatively charged glycolipid composed of three distinct structural domains: lipid A, a non-repeating core oligosaccharide, and the O-antigen (O-specific polysaccharide, OPS). Lipid A, which anchors LPS in the outer membrane, is

☆ This article is part of a Special issue entitled: '10th BMMC' published in Carbohydrate Polymers.

* Corresponding author.
E-mail address: anna.turska-szewczuk@mail.umcs.pl (A. Turska-Szewczuk).

https://doi.org/10.1016/j.carbpol.2026.125547
Received 24 April 2026; Received in revised form 25 May 2026; Accepted 8 June 2026
Available online 9 June 2026

highly conserved among Gram-negative bacteria, including *Aeromonas* sp. By contrast, the O-antigen is the outermost and most variable part of the molecule and consists of linear or branched polysaccharides. Its variability arises from differences in sugar composition, sequence, linkage patterns, and non-carbohydrate substituents. Even minor changes in O-unit structure may generate new serotypes (Di Lorenzo et al., 2019; Whitfield et al., 2020).

In *Aeromonas*, O-antigen diversity is particularly high, with strains classified into nearly 100 established O-serogroups and additional provisional serogroups described in Europe and other regions (Cao et al., 2018; Tomas, 2012). In Poland, virulent isolates are most often assigned either to the serogroups of the National Institute of Health, Japan (NIH), scheme (O3, O6, O11, O16, O18, O33, O34) or to provisional serogroups, including PGO1, PGO2, PGO4, PGO6 (Sakazaki & Shimada, 1984; Kozińska, 2009; Kozińska & Pękala, 2010; unpublished veterinary data).

The structural diversity of O-antigens arises from variation in the composition and organization of the O-antigen gene cluster (OGC), which governs its biosynthesis. In enteric bacteria such as *Escherichia coli*, *Salmonella*, and *Shigella*, the OGC is typically located between the *galF* and *gnd* genes (Liu, Furevi, et al., 2020; Liu et al., 2008; Reeves et al., 2013), whereas in *Aeromonas* it is usually mapped between *acrB* and *oprM*, encoding the multidrug efflux pump subunit and an outer membrane protein, respectively (Cao et al., 2018, Kurzylewska, Bomba, et al., 2023; Kurzylewska, Turska-Szewczuk, et al., 2023).

The genes within OGC are categorized into three main functional groups: (1) nucleotide sugar synthesis genes; (2) glycosyltransferase (GT) genes; and (3) processing genes, responsible for translocating the O-unit across the inner membrane and polymerizing it into the complete O-polysaccharide chain (Whitfield et al., 2020). Variation in the O-antigen synthesis locus defines serogroups, but there are often other genes located elsewhere on the chromosome, sometimes within prophages, that can also be involved in O-antigen synthesis. These additional genes are generally not essential for synthesis of the main O-antigen chain, and carry out functions such as adding a side-branch sugar or *O*-acetyl groups that may not be present on all repeating units (Reeves et al., 2013; Van Puyvelde et al., 2022).

O-antigen diversity within bacterial species is driven primarily by horizontal gene transfer of OGC, resulting in the gain and loss of different structural types over evolutionary time. This process, rather than point mutations or recombination, is considered the major mechanism underlying the emergence of novel O-antigen structures (Reeves et al., 2013). O-antigens play key roles in host colonization and bacterial adaptation to specific ecological niches. At the same time, they high immunogenicity makes them major targets of the adaptive immune system. Consequently, the O-antigen is one of the most structurally diverse components of the bacterial cell surface, showing variation both between species and among strains within a species (Geue et al., 2017).

In the last decade, significant progress has been made in identifying O-antigen serogroups by molecular methods, especially for *E. coli* serogroups associated with diseases in humans and animals (DebRoy et al., 2016). The sequences of the O-unit processing genes, *wzx* encoding O-antigen flippase and *wzy* encoding O-antigen polymerase, are relatively unique for each individual O-type. Therefore, these two genes have been used as targeted in PCR- and microarray-based assays for the identification of *E. coli* O-serogroups (DebRoy et al., 2016). Such sequence data may also be useful for comparative studies of bacterial O-antigen evolution, including gene deletion, acquisition, or inactivation, as well as for investigating host adaptation, immune evasion, and the development of glycoconjugate vaccines.

Several *Aeromonas* strains belonging to the provisional PGO1 serogroup have recently been characterized; however, knowledge of the structural and genetic diversity within this serogroup remains limited. Previous studies demonstrated partial similarity among PGO1 O-antigens and their corresponding biosynthesis gene clusters (Kurzylewska, Turska-Szewczuk, et al., 2023), suggesting the presence of related but distinct serotypes. We therefore hypothesized that *Aeromonas* isolates from Polish aquaculture outbreaks may harbour additional variants within this serogroup. To test this hypothesis, we characterized the O-specific polysaccharide of *A. sobria* strain K221, isolated from carp during an outbreak of MAI/MAS disease, and analysed the genetic basis of its biosynthesis. By combining immunochemical, structural, and OGC analyses, we identified a distinct O-antigen variant within the PGO1 serogroup. Furthermore, this is the first report of abequose in an *Aeromonas* O-polysaccharide, where it occurs as a 2-substituted rather than terminal residue. In silico analyses also suggested the involvement of a putative α-1,3-CDP-abequosyltransferase in its biosynthesis. Together, these findings expand current knowledge of bacterial polysaccharide chemistry and biosynthesis and provide new insight into the diversity of O-specific carbohydrate polymers in aquatic *Aeromonas* isolates.

2. Materials and methods

2.1. Bacterial strain and cultivation

A. sobria strain K221 was isolated from the skin of a common carp during an outbreak of MAI/MAS disease on a Polish fish farm and classified into the provisional O-serogroup PGO1 after agglutination assays with antisera of the NIH scheme (Sakazaki & Shimada, 1984) and 20 additional antisera developed for Polish isolates representing new provisional serogroups (PGO1—PGO20) (Kozińska, 2009; Kozińska & Pękala, 2010).

For the LPS isolation, *A. sobria* K221 was cultivated in tryptic soy broth at 28 °C with aeration. After 72 h, the cells were harvested by centrifugation (8000 x*g*, 35 min).

2.2. Extraction of Lipopolysaccharide

Following enzymatic digestion as previously described by Dworaczek et al. (Dworaczek et al., 2021), *A. sobria* K221 biomass was extracted with 45% aqueous phenol according to the published method (Westphal & Jann, 1965). After phase separation by low-speed centrifugation (3000 x*g*, 1 h, 10 °C), LPS was recovered by dialysis, ultracentrifugation (105,000 ×*g*, 3 h), and freeze-drying. High-molecular-weight S-LPS was obtained from the phenol phase with a yield of approximately 1.7% of dry cell biomass.

2.3. SDS-PAGE analysis

LPS samples from *A. sobria* K221 and three O-related strains (*A. hydrophila* Pt679, *A. popoffii* A4 and *A. sobria* K928) (Kurzylewska, Bomba, et al., 2023; Kurzylewska, Turska-Szewczuk, et al., 2023) were separated by 12.5% SDS-Tricine PAGE and visualized by periodate oxidation followed by silver staining (Tsai & Frasch, 1982).

2.4. Serological studies

Western blotting was performed after transfer of LPSs onto Immobilon P (Millipore, USA), as previously described (Dworaczek et al., 2021), with rabbit polyclonal reference PGO1 O-antiserum kindly provided by Professor Alicja Kozińska (the National Veterinary Research Institute, Puławy, Poland) (Kozińska, 2009). Strain-specific polyclonal O-antisera against K221, Pt679, A4, and K928 were raised in New Zealand White rabbits as described previously (Kurzylewska, Turska-Szewczuk, et al., 2023). All animal experiments were approved by the Ninth Local Ethical Committee on Animal Testing in Lodz (permission number 51/ŁB 218/2021).

ELISA was performed as described previously (Kurzylewska, Turska-Szewczuk, et al., 2023). Flat-bottom Nunc-Immuno plates were coated with 0.1 μg of LPS per well, and bound antibodies were detected using peroxidase-conjugated goat anti-rabbit-IgG (Jackson ImmunoResearch, USA) with ABTS (Sigma-Aldrich) as the substrate. Absorbance was

measured at 405 nm using a Multiskan Go microplate reader (Thermo Fisher Scientific).

2.5. Degradation of LPS and isolation of the O-polysaccharide (OPS)

The phenol-soluble LPS sample (115 mg) was hydrolyzed with 2% aqueous acetic acid at 100 °C for 3 h. After removal of lipid A by centrifugation (13,000 xg, 30 min), the carbohydrate material was fractionated by GPC on a Sephadex G-50 Fine column using 1% acetic acid as the eluent and monitored with a differential refractometer (Knauer, Berlin, Germany). The OPS fraction was obtained in a yield of 14% of the hydrolyzed LPS sample.

2.6. Chemical analyses

For sugar analysis, an O-polysaccharide sample (0.3–0.4 mg) of *A. sobria* K221 was hydrolysed either under strong conditions with 10 M HCl at 80 °C for 30 min or under milder conditions with 1 M CF_3CO_2H at 100 °C for 1 h (Silipo et al., 2005). The products were reduced with $NaBD_4$, peracetylated with an Ac_2O-pyridine mixture, 1:1 (*v*/v), at 85 °C for 30 min, and analysed by GLC-MS as the corresponding alditol and aminoalditol acetates.

The O-polysaccharide (1.0 mg) was methylated with methyl iodide in dimethyl sulfoxide in the presence of powdered sodium hydroxide, according to the method of Ciucanu and Kerek (1984). The resulting methylated polysaccharide was hydrolysed with 2 M CF_3CO_2H at 100 °C for 4 h, or alternatively, with 1 M CF_3CO_2H at 100 °C for 1 h to facilitate the identification of a labile sugar derivative. The products were then *N*-acetylated, reduced with $NaBD_4$, and peracetylated, and analysed by GLC-MS as partially methylated alditol and aminoalditol acetates (PMAAs).

The absolute configuration of monosaccharides was determined by GLC of acetylated (*S*)-(+)-2-octyl- and racemic 2-octyl-glycoside derivatives using authentic sugar as standards according to a published method (Ovchinnikova et al., 2007). LPS of *E. coli* O111:B4 (Sigma-Aldrich) containing the authentic sample of Col*p* was a gift from Prof. Roman Paduch (M. Curie-Sklodowska University in Lublin). The authentic sample of Abe*p* was obtained from LPS of *Citrobacter freundii* O22 strain PCM 1555 (Polish Collection of Microorganisms, Institute of Immunology and Experimental Therapy in Wroclaw) (Katzenellenbogen et al., 2009).

All the sugar derivatives were analysed on an Agilent Technologies 7890A gas chromatograph (USA) connected to a 5975C MSD detector (inert XL EI/CI, Agilent Technologies, USA). The chromatograph was equipped with an HP-5MS capillary column (Agilent Technologies, 30 m × 0.25 mm) with helium as the carrier gas at a flow rate of 1 ml/min. The temperature program was as follows: 150 °C for 5 min, followed by an increase at 5 °C/min to 310 °C, and then held at 310 °C for 10 min.

2.7. NMR spectroscopy

Prior to measurement, an OPS sample (8 mg) was deuterium-exchanged by vacuum-drying twice from 99.9% D_2O and then examined as solution in 99.98% D_2O. For structural assignments, 1D and 2D NMR spectra for were recorded at 44 °C on a 500 MHz NMR Varian Unity Inova instrument and calibrated with external acetone (δ_H 2.225, δ_C 31.45). Standard Varian software (Vnmrj V. 4.2 rev.) was used to acquire and process the NMR data. Homonuclear and heteronuclear two-dimensional experiments: $^1H,^1H$ DQF-COSY, $^1H,^1H$ TOCSY, $^1H,^1H$ NOESY, $^1H,^{13}C$ HSQC, $^1H,^{13}C$ HSQC-TOCSY, and $^1H,^{13}C$ HMBC were carried out for signal assignments and determination of the sugar sequence. The mixing time of 100 and 200 ms was used in the TOCSY and NOESY experiments, respectively. The $^1H,^{13}C$ HSQC spectrum (band-selective gHSQCAD) measured without ^{13}C decoupling was used to determine the $^1J_{C1,H1}$ coupling constants for the anomeric carbons. The $^1H,^{13}C$ HMBC experiment was optimized for $J_{C,H} = 8$ Hz.

2.8. DNA isolation, whole-genome sequencing, and bioinformatic analyses

2.8.1. DNA extraction, quality assessment and library preparation

Genomic DNA was extracted from cultured cells using the automated extraction system Maxwell RSC 48 with the Maxwell RSC Cultured Cells DNA Kit (Promega), according to the manufacturer's instructions. DNA concentration was measured fluorometrically using a Qubit Fluorometer (Thermo Fisher Scientific), and purity was assessed spectrophotometrically using a NanoDrop instrument (Thermo Fisher Scientific). Sequencing libraries were prepared using the Nextera XT DNA Library Preparation Kit (Illumina), and library quality and fragment size distribution were evaluated using a Fragment Analyzer with the DNF-473 Standard Sensitivity NGS Kit (Agilent). Long-read sequencing libraries were prepared using the SQK-LSK109 Ligation Sequencing Kit with Native Barcoding Expansion (EXP-NBD114) (Oxford Nanopore Technologies).

2.8.2. DNA sequencing and analysis

Libraries were sequenced on the Illumina MiSeq platform using the MiSeq Reagent Kit v3 (Illumina), generating paired-end reads of 2 × 300 bp. Long-read sequencing was performed on the Oxford Nanopore MinION platform using FLO-MIN106D flow cells.

Short and long reads were trimmed using fastp v0.12.4 (Chen, 2025) and Chopper v0.5.0 (De Coster & Rademakers, 2023), respectively. Long-read assemblies were generated using Trycycler v0.5.3 following the workflow described by the author (Wick et al., 2021). Final assemblies were further refined using short reads with Polypolish v0.5.0 (Wick & Holt, 2022) and pypolca v0.3.1 (Bouras et al., 2024; Zimin & Salzberg, 2020). The quality of the assembly was evaluated using seqkit (Wei et al., 2024). Genome completeness and contamination were assessed with CheckM 1.2.2 (Parks et al., 2015). The genome was annotated using bakta 1.11 (Schwengers et al., 2021).

2.8.3. Species identification and bioinformatic analysis of the O-antigen gene cluster (OGC)

Species identification was performed by comparing the 16S rRNA gene sequence extracted from the whole-genome assembly against the SILVA 138 SSU database using BLAST 2.12.0 (https://www.arb-silva.de/documentation/release-138). In addition, average nucleotide identity (ANI) was calculated using fastANI v1.32 (Jain et al., 2018). MLPA (multilocus phylogenetic typing analysis) was done based on the selected housekeeping genes *gyrB*, *rpoD*, *recA*, *dnaJ*, *gyrA*, *dnaX*, and *atpD* (Martinez-Murcia et al., 2011). These seven genes from isolate K221 and all *Aeromonas* strains available in NCBI RefSeq and designated as reference genomes were concatenated and aligned using MAFFT (Katoh & Standley, 2013) (v7.490). An unrooted maximum likelihood phylogenetic tree was constructed using IQ-TREE v2.1.4 with 1000 bootstrap replicates and the '-m MFP' option to determine the best-fit substitution model.

The OGC gene cluster was inspected, extracted, and curated using UGENE 53.1 (Okonechnikov et al., 2012). Predicted protein sequences encoded within the OGC region were subsequently analysed for sequence identity and similarity using BLAST+ v2.15.0 (Camacho et al., 2009) against the NCBI BLAST RefSeq protein database (as of March 22, 2025).

Visualization was generated by Clinker v0.0.32 (Gilchrist & Chooi, 2021).

2.9. Structural prediction and molecular docking analysis of a putative Abe-transferase

The amino acid sequence of Orf12 was analysed to identify potential homologous proteins and predict its structural classification. Homology detection was performed using HHpred against the PDB_mmCIF70 database (probability >95%) (Söding et al., 2005). Structures of Orf12

and reference glycosyltransferases WbaR (WP_000038915.1) and WbyD (AKA20933.1) were predicted using ColabFold (AlphaFold2/MMseqs2) (Jumper et al., 2021; Mirdita et al., 2022; Steinegger & Söding, 2017). Structural similarity was evaluated using TM-align (TM-score, RMSD) and visualized in UCSF ChimeraX (Meng et al., 2023). Molecular docking with CDP-abequose (PubChem) (Kim et al., 2023) was performed using CB-Dock (AutoDock Vina) (Liu, Grimm, et al., 2020; Trott & Olson, 2010), and the top-ranked pose was selected based on binding energy and localization within the predicted catalytic pocket.

3. Results

3.1. Species confirmation

Aeromonas sp. strain K221, isolated from the skin of diseased carp and assigned to the provisional serogroup PGO1, was identified as *A. sobria*. This assignment was supported by ANI, which showed the highest similarity to *A. sobria* 25-AL00119 (96.98%) and the reference genome PRJEB7040 (96.72%), and further confirmed by MLPA, which placed strain K221 within the *A. sobria* clade (Supplementary Fig. S1).

3.2. SDS-PAGE analysis of A. sobria K221 lipopolysaccharide

SDS-PAGE of phenol-soluble LPS from *A. sobria* K221 revealed a ladder-like profile characteristic of smooth bacterial cells, with fast-migrating LMW species typical of R- and SR-LPS and slow-migrating S-LPS species differing in OPS chain lengths (Fig. 1a).

3.3. Serological characterization of A. sobria K221 within the PGO1 serogroup

Classification of *A. sobria* K221 within the PGO1 serogroup was supported by the reactivity of the reference antiserum with its phenol-soluble LPS in both Western blotting and ELISA. In Western blotting, the antiserum reacted with all LPS species, showing strongest binding to the long-chain S-LPS glycoforms, indicating that the cross-reactive epitope(s) are distributed throughout the LPS molecule (Fig. 1b). Similar positive reactions were observed for the O-related strains *A. hydrophila* Pt679, *A. popoffii* A4, and *A. sobria* K928 (Kurzylewska, Turska-Szewczuk, et al., 2023). In ELISA (Table 1), the reference

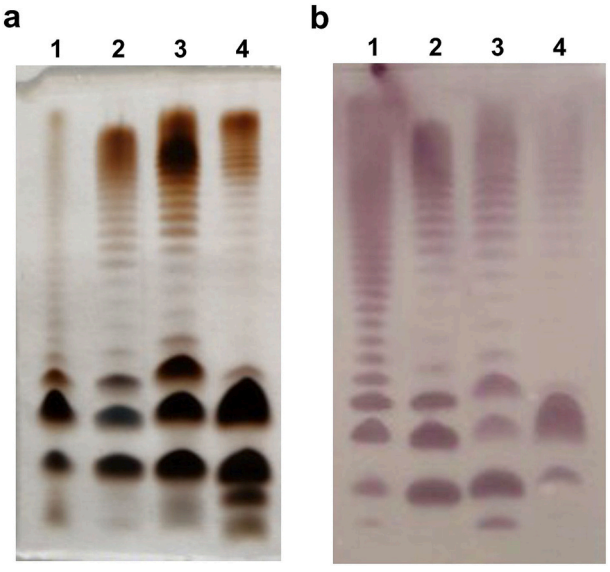

Fig. 1. Silver stained SDS-PAGE (**a**) of lipopolysaccharide samples and Western blot (**b**) with the reference PGO1 antiserum. Lanes: 1, *A. sobria* K221; 2, *A. hydrophila* Pt679; 3, *A. sobria* K928; and 4, *A. popoffii* A4. 3 μg of LPS were loaded per each lane.

Table 1
Reactivity (reciprocal titers) of PGO1 reference and anti-K221 sera with LPSs of *Aeromonas* strains K221, Pt679, A4, and K928 classified within the PGO1 serogroup[a].

Type of the antiserum	K221 LPS	Pt679 LPS	K928 LPS	A4 LPS
anti-PGO1[a] serum	64.000	64.000	16.000	8.000
anti-K221 serum	64.000	8.000	1.000	1.000

[a] the reactivity of the reference PGO1 antiserum with the biomass of the reference strain was at the level of 512.000 (Dworaczek et al., 2021).

antiserum reacted with *A. sobria* K221 LPS at a titer of 1: 64,000 and also with the LPS of Pt679, A4 and K928, although less strongly than in homologous systems (Dworaczek et al., 2021; Kurzylewska, Turska-Szewczuk, et al., 2023).

ELISA analysis of rabbit polyclonal antisera against the K221, Pt679, A4, and K928 O-serotypes showed limited cross-reactivity within the PGO1 serogroup. Clear reactivity was observed only between the anti-K221 serum and Pt679 LPS, indicating the presence of shared epitope (s) in the O-antigen and/or core region (Table 1), whereas the remaining antisera reacted only weakly with K221 LPS (titers $\leq$1:1000; data not shown). These findings suggest that the shared epitope contributes only marginally to the overall serological specificity of these O-serotypes.

3.4. Elucidation of the A. sobria K221 O-polysaccharide structure

GLC-MS analysis of the *A. sobria* K221 O-polysaccharide (OPS), performed after strong acid hydrolysis, reduction, and acetylation, showed the presence of 6-deoxymannose (Rha) and 2-amino-2-deoxy-glucose (GlcN) in comparable amounts. In contrast, hydrolysis under milder conditions, using 1 M TFA at 100 °C for 1 h, led to the detection of an additional compound (Silipo et al., 2005). Its chromatographic mobility was consistent with that of colitose (3,6-dideoxy-L-*xylo*-hexose) or abequose (3,6-dideoxy-D-*xylo*-hexose), which had previously been identified in the LPSs of *E. coli* O111 (Wang & Reeves, 2000) and *C. freundii* PCM 1555 (Katzenellenbogen et al., 2009), respectively. The presence of a 3,6-dideoxyhexose residue was further confirmed by NMR spectroscopy, as described below.

The determination of the absolute configuration of the monosaccharides by GLC of acetylated (*S*)-2-octyl glycosides and comparison with authentic standards (Ovchinnikova et al., 2007) indicated the presence of the L-configuration of Rha and the D-configuration of GlcN. The absolute configuration of 3,6-dideoxyhexose was found to have the D-configuration (3,6-dideoxy-D-*xylo*-hexose) and hence this enantiomer indicated abequose (Abe).

Methylation analysis of the OPS by GLC-MS completed the compositional data. After acid hydrolysis of the preparation with 1 M CF_3CO_2H (100 °C, 1 h), followed by reduction and acetylation, the following partially methylated alditol acetates (PMAA) were identified: 2,3,4-tri-*O*-methyl-6-deoxyhexitol-1-*d* (derived from terminal Rha*p*), 4-*O*-methyl-3,6-dideoxyhexitol-1-*d* (derived from 2-substituted Abe*p*), 2,4-di-*O*-methyl-6-deoxyhexitol-1-*d* (derived from 3-substituted Rha*p*), and 2-deoxy-4,6-di-*O*-methyl-2-(*N*-methyl)acetamidohexitol-1-*d* (derived from 3-substituted Glc*p*N), with peak area ratios of approximately 0.4: 1.0: 3.6: 7.7. Moreover, a 2-subsituted 3,6-dideoxyhexose derivative was also detected when stronger hydrolysis conditions were applied (2 M CF_3CO_2H, 100 °C, 4 h).

The ^{1}H NMR spectrum of the O-polysaccharide (Fig. 2) in the anomeric region showed seven proton signals, which included only five ones for H-1 protons at δ 5.09, 5.03, 4.88, 4.71, and 4.51, two other signals at δ 5.11 and 5.16 were resonances of H-2 that were shifted downfield due to 2-*O*-acetylation (see below). The high field region of the spectrum contained signals for three CH_3-CH groups (H-6 of Rha and Abe) at δ 1.15, 1.27 and 1.31, one C-CH_2-C group (H-3 of Abe) at δ 1.95 and 2.05, two signals for *N*-acetyl groups at δ 2.08 and 2.09, two signals for *O*-acetyl groups at δ 2.17 and 2.25, and ring proton signals in the

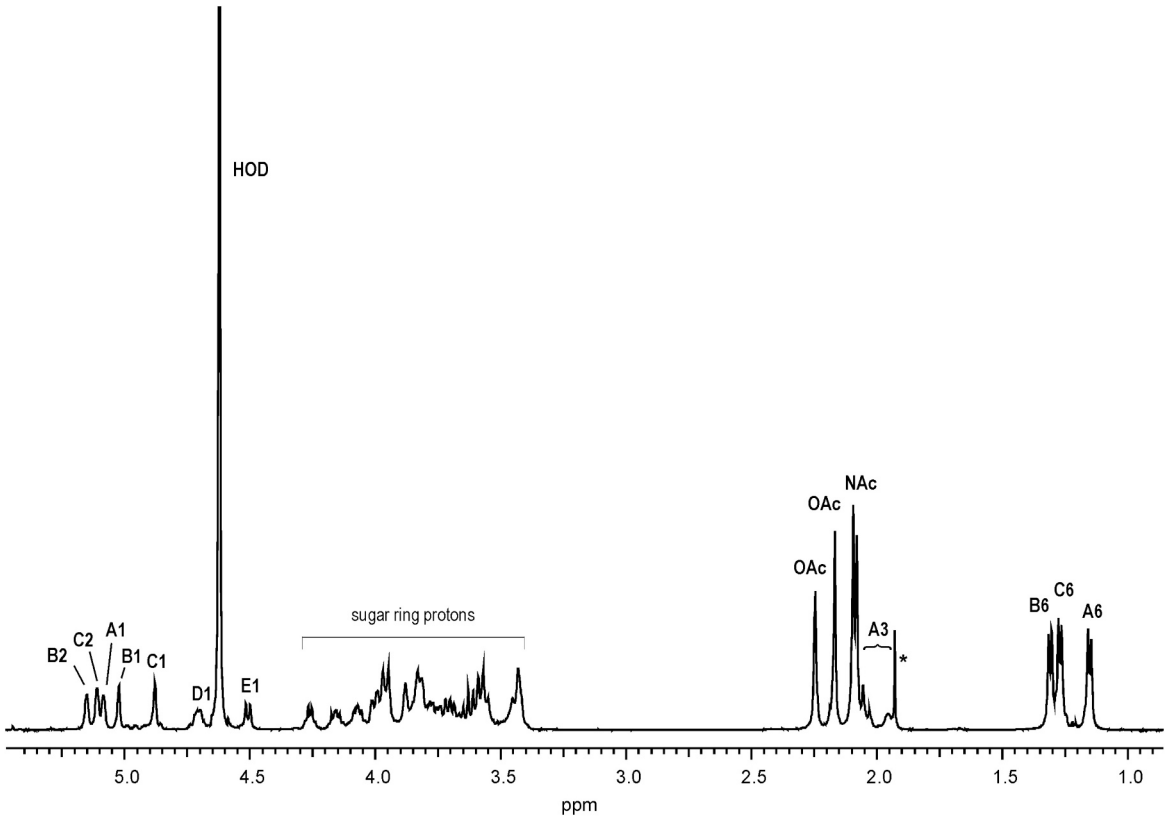

Fig. 2. ^{1}H NMR spectrum of the OPS of *A. sobria* K221. The spectrum was recorded at 44 °C in D_2O at 500 MHz. The capital letters and Arabic numerals refer to atoms in the sugar residues denoted as shown in Table 2. OAc - *O*-acetyl group, NAc - *N*-acetyl group, asterisk - free acetic acid. The ring proton signals in the *δ* 3.42–4.26 region are indicated by a bracket.

range of *δ* 3.42–4.26; some of the which overlapped.

The ^{1}H and ^{13}C resonances of the OPS from *A. sobria* strain K221 were assigned on the basis of 2D NMR spectroscopy. Homo- and heteronuclear experiments (^{1}H,^{1}H DQF-COSY, TOCSY, NOESY, ^{1}H,^{13}C HSQC, ^{1}H,^{13}C HSQC-TOCSY, and ^{1}H,^{13}C HMBC) were performed to assign all spin systems and to establish the sugar sequence of the O-antigen repeating unit. The ^{1}H and ^{13}C NMR chemical shifts are collected in Table 2.

The ^{1}H,^{13}C HSQC spectrum (Fig. 3) showed two signals at δ 3.72/56.6, and 3.83/56.9 of protons at the nitrogen-bearing carbons to the corresponding carbons, indicating that the OPS repeating unit contained N-acetamido sugars. In addition, two further correlation signals at δ 5.16/70.4, and 5.11/73.0 were assigned to the H-2/C-2 cross-peaks of *O*-acetylated sugar residues on the basis of their downfield chemical shifts.

The TOCSY and DQF-COSY spectra revealed five spin systems, labeled **A**, **B**, **C**, **D**, and **E**. Two of them, spin systems **B** and **C**, were characteristic of *manno*pyranoses, as indicated by relatively low vicinal coupling constants, $^{3}J_{1,2}$ (~2 Hz) and $^{3}J_{2,3}$ (3.5 Hz), and higher $^{3}J_{3,4}$ (8.5 Hz), $^{3}J_{4,5}$ (9 Hz), and $^{3}J_{5,6}$ (6 Hz) values. In the TOCSY spectrum, correlations of H-1 with H-2 and weakly with H-3, as well as correlations of H-2 with H-1, H-3, H-4, and H-5 and of H-6 with H-3, H-4, and H-5, were observed for both residues. These residues were assigned as 2-*O*-acetylated Rha*p* residues based on the high-field positions of their H-6 and C-6 resonances and the downfield displacement of the H-2/C-2 cross-peaks from δ 4.15/69.2 to 5.16/70.4 for residue **B** and 5.11/73.0 for residue **C** in the ^{1}H,^{13}C HSQC spectrum (Fig. 3), compared with published data (Shashkov et al., 2016). The *O*-acetylation of the sugars was further supported by the upfield shifts of the C-1 signals of residues **B** and **C** to δ 99.3 and 99.8, respectively, due to the β-effect of *O*-acetylation (Jansson et al., 1987), compared with the signal at δ 101.6 for the deacetylated Rha*p* (Shashkov et al., 2016). The α-anomeric configuration of the Rha*p* residues was inferred from the C-5 chemical shift at δ 70.2, consistent with published data for α- and β-Rha*p* (δ 70.0 and 73.0,

Table 2
^{1}H (500.43 MHz) and ^{13}C NMR (125 MHz) data (*δ*, ppm) for the OPS of *A. sobria* K221.

Sugar residue		H-1 C-1	H-2 C-2	H-3 C-3	H-4 C-4	H-5 C-5	H-6 C-6	NAc
→2)-α-Abe*p*-(1→	**A**	5.09	3.97	1.95; 2.05	3.88	4.26	1.15	
$^{1}J_{C,H}$ = 169 Hz; $^{3}J_{H1,H2}$ ~ 3 Hz		95.9	74.4	32.6	69.8	67.7	16.4	
→3)-α-L-Rha*p*2OAc-(1→	**B**	5.03	5.16	4.00	3.63	4.16	1.31	2.25
$^{1}J_{C,H}$ = 173 Hz; $^{3}J_{H1,H2}$ < 2 Hz		99.3	70.4	74.0	71.7	70.2	17.7	22.1; 174.4
→3)-α-L-Rha*p*2OAc-(1→	**C**	4.88	5.11	3.96	3.57	4.07	1.27	2.17
$^{1}J_{C,H}$ = 173 Hz; $^{3}J_{H1,H2}$ < 2 Hz		99.8	73.0	79.8	72.2	70.2	17.7	21.7; 174.6
→3)-β-D-Glc*p*NAc-(1→	**D**	4.71	3.83	3.69	3.59	3.45	3.82	2.08
$^{1}J_{C,H}$ = 162 Hz; $^{3}J_{H1,H2}$ ~ 8 Hz		103.1	56.9	82.0	69.3	77.1	61.9	23.8; 176.0
→3)-β-D-Glc*p*NAc-(1→	**E**	4.51	3.72	3.56	3.42	3.43	3.76;3.96	2.09
$^{1}J_{C,H}$ = 162 Hz; $^{3}J_{H1,H2}$ ~ 8.5 Hz		103.1	56.6	83.7	70.2	77.4	62.5	23.8; 176.2

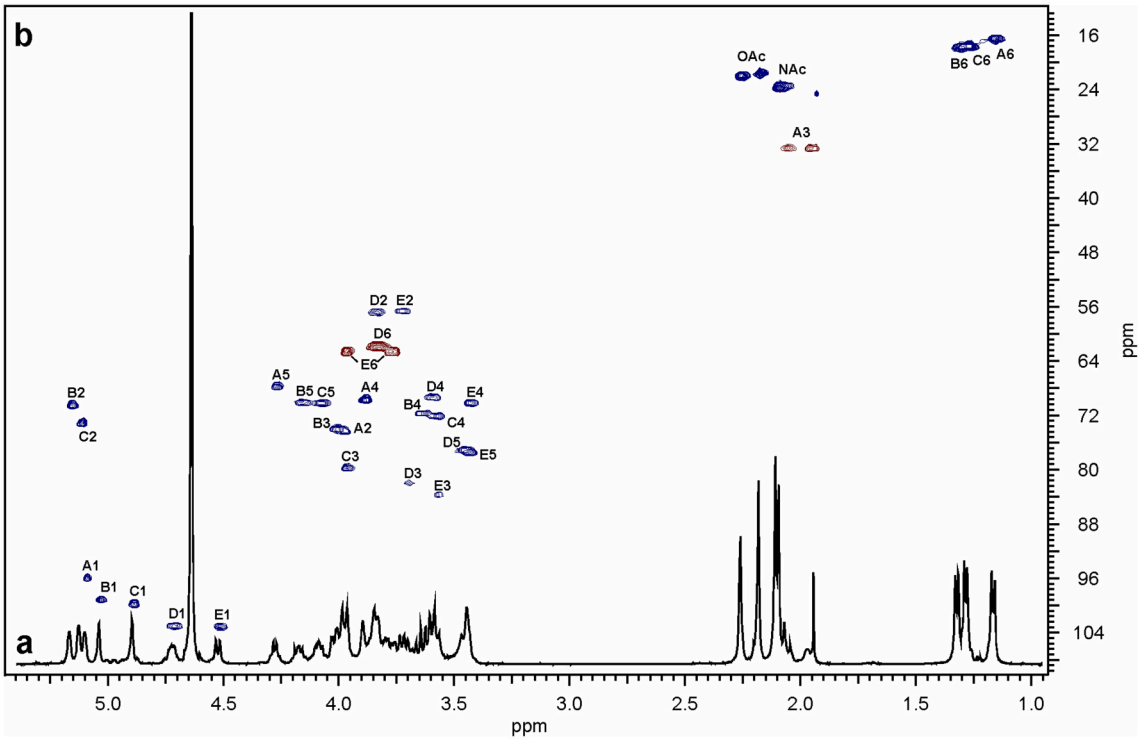

Fig. 3. (**a**) $^{1}H,^{13}C$ HSQC (500 × 125 MHz) and (**b**) ^{1}H NMR spectra of the OPS of *A. sobria* K221. Overlay of the anomeric region, the region for the ring atoms, and the region for the methyl groups of 6-deoxy sugars, *O*-acetyl, and *N*-acetyl groups. The spectrum was recorded at 44 °C in D_2O as a solvent. The capital letters and Arabic numerals refer to proton/carbon pairs in the respective sugar residues denoted as shown in Table 2.

respectively). In addition, intraresidue H-1/H-2 cross-peaks were observed for both Rha*p* residues in the NOESY spectrum (Fig. 4),

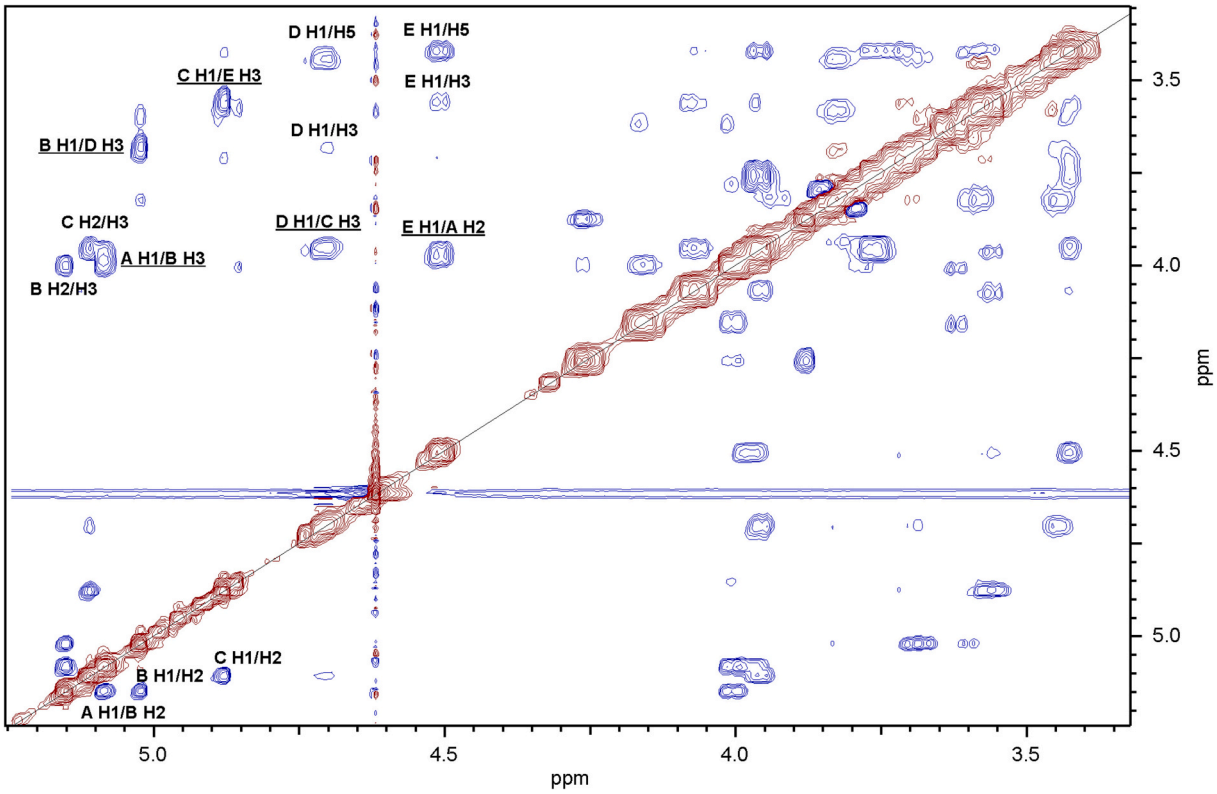

Fig. 4. $^{1}H,^{1}H$ NOESY spectrum of the OPS from *A. sobria* K221. The NOE correlations between anomeric protons and protons at the glycosidic linkages are underlined. Some other important H/H correlations are depicted as well. The capital letters and Arabic numerals refer to protons in the sugar residues denoted as shown in Table 2.

confirming their α-anomeric configuration.

As judged from the ratio of integral intensities of the signals for H-1, H-2, and *O*-acetyl signals in the ^{1}H NMR spectrum, the *O*-acetylation of α-Rha*p* residues was stoichiometric. In the ^{1}H, ^{13}C HMBC spectrum, correlations were observed between the carbonyl carbons of the *O*-acetyl groups and H-2 of the Rha*p* residues at δ 174.4/5.16 (**B**) and 174.6/5.11 (**C**).

The ^{1}H resonances of spin systems **D** and **E** were assigned from the TOCSY, DQF-COSY, and NOESY spectra. In the TOCSY spectrum, correlations of H-1 with H-2, H-3, H-4, and H-5 were observed for both residues. The remaining proton resonances were assigned based on NOE correlations between H-5 and the H-6 protons and were confirmed by connectivities observed in the DQF-COSY spectrum. The ^{13}C resonances of spin systems **D** and **E** were assigned from the ^{1}H,^{13}C HSQC and ^{1}H,^{13}C HSQC-TOCSY spectra. The latter spectrum also showed H-4/C-6 and H-5/C-6 correlations. Based on the relatively large $^{3}J_{2,3}$, $^{3}J_{3,4}$, and $^{3}J_{4,5}$ coupling constants of 9–10 Hz, characteristic of the *gluco* configuration, together with the H-2/C-2 correlations at δ 3.83/56.9 and 3.72/56.6 in the ^{1}H,^{13}C HSQC spectrum (Fig. 3), spin systems **D** and **E** were assigned to Glc*p*N residues. The relatively large anomeric coupling constants ($^{3}J_{1,2}$–8–8.5 Hz) indicated the β-configuration of both residues. This assignment was confirmed by H-1/H-3 and H-1/H-5 correlations observed in the ^{1}H,^{1}H NOESY spectrum (Fig. 4).

In the TOCSY spectrum, the signals at δ 1.15 and δ 1.95 / 2.05 were assigned to H-6 and H-3a,b, respectively, of spin system **A**, which was identified as a 3,6-dideoxyhexose residue. In addition, correlations between H-1 and H-2, H-3, and H-4, between H-4 and H-1, H-2, H-3, and H-5, and between H-6 and H-5 were used to determine the ^{1}H chemical shifts of this residue; these assignments were confirmed by the connectivities observed in DQF-COSY and NOESY spectra. Based on the relatively low coupling constants, $^{3}J_{1,2} < 3$ Hz, $^{3}J_{2a,3e} \approx 4.5$ Hz, and $^{3}J_{3a,4e} \approx 3.5$ Hz, together with the relatively large $^{3}J_{2a,3a} \approx 12$ Hz value, the 3,6-dideoxyhexose was identified as the α-*xylo* derivative (Kochetkov, 1996; Senchenkova et al., 1997). Consistent with the *xylo* configuration, the sugar, which contains an axial proton H-2 and an equatorial proton H-4, showed H-3ax/H-5 and H-3eq/H-2 correlations in the NOESY spectrum at 2.05/4.26 and 1.95/3.97, respectively. Taking into account the absolute configuration, the monosaccharide was identified as 3,6-dideoxy-D-*xylo*-hexose, corresponding to abequose (Abe) (Kochetkov, 1996).

The ^{13}C chemical shifts of the 3,6-dideoxyhexose were assigned from the ^{1}H,^{13}C HSQC and ^{1}H,^{13}C HMBC spectra. The latter spectrum showed correlations involving the anomeric protons (H-1/C-3, H-1/C-5) and the H-6 (H-6/C-4 and H-6/C-5) protons.

Moreover, the absence of signals in the δ 85.0–88.0 region, characteristics of C-4 resonances of furanoses, together with the anomeric carbon resonances, confirmed that all sugar residues were in the pyranose form (Bock & Pedersen, 1983).

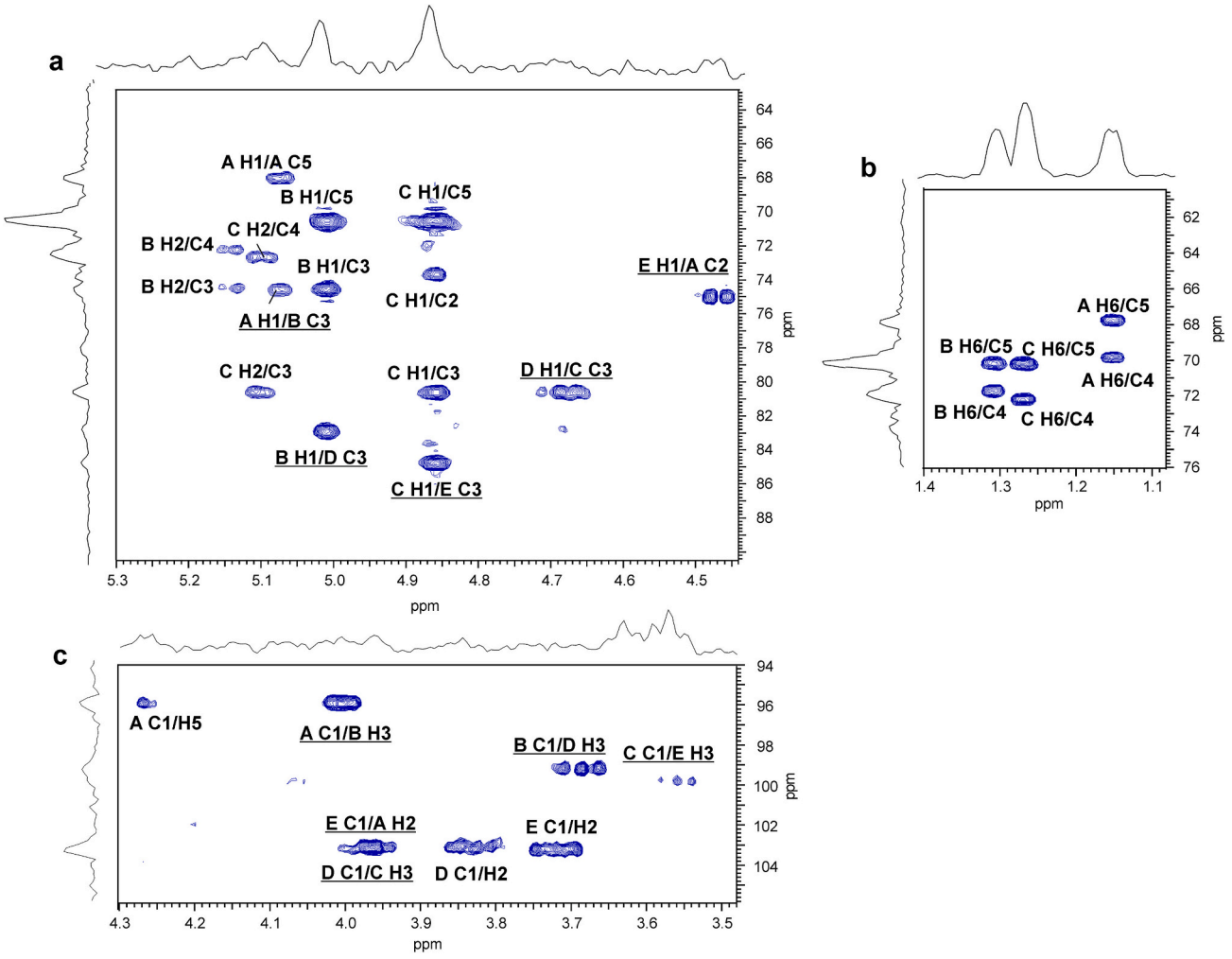

Fig. 5. Regions of the ^{1}H,^{13}C HMBC spectrum of the OPS of *A. sobria* K221. The corresponding parts of the ^{1}H and ^{13}C NMR spectra are displayed along the horizontal and vertical axis, respectively. The long-range heteronuclear correlations for (**a**) anomeric protons, (**b**) H-6 protons, and (**c**) anomeric carbons are marked. Inter-residue correlations important for establishing the sugar sequence in the O-antigen repeating unit are underlined. The capital letters and Arabic numerals refer to atoms in the sugar residues denoted as shown in Table 2.

The $^{1}J_{C1,H1}$ coupling constants, determined from the $^{1}H,^{13}C$ HSQC spectrum of the OPS recorded without ^{13}C decoupling, supported the anomeric assignments of the monosaccharide residues. The residues **A** – **C** with C-1/H-1 signals at δ 95.9/5.09, 99.3/5.03, and 99.8/4.88, respectively, showed $^{1}J_{C1,H1}$ values of 169–173 Hz, consistent with α-linkages. In contrast, residues **D** and **E**, with C-1/H-1 signals at δ 103.1/4.71 and 103.1/4.51, respectively, showed $^{1}J_{C1,H1}$ values of 162 Hz, consistent with β-linkages, as previously reported (Bock et al., 1973).

The downfield displacement of the C-3 carbon signals in spin systems **B** (δ 74.0), **C** (δ 79.8), **D** (δ 82.0), and **E** (δ 83.7), together with the downfield shift of C-2 to δ 74.4 in spin system **A**, relative to their corresponding positions in non-substituted monosaccharides, confirmed the glycosylation pattern of the monosaccharide residues in the OPS (Jansson et al., 1989; Shashkov et al., 1988; Torgov et al., 1990).

The sequence of the monosaccharide residues in the O-repeating unit was established by $^{1}H,^{13}C$ HMBC and $^{1}H,^{1}H$ NOESY experiments.

In the $^{1}H,^{13}C$ HMBC spectrum (Fig. 5), correlations were observed between anomeric protons and transglycosidic carbons: Abe*p* **A** H-1/ Rha*p* **B** C-3 at δ 5.09/74.0, Rha*p* **B** H-1/ Glc*p*N **D** C-3 at δ 5.03/82.0, Glc*p*N **D** H-1/ Rha*p* **C** C-3 at δ 4.71/79.8, Rha*p* **C** H-1/ Glc*p*N **E** C-3 at δ 4.88/83.7, and Glc*p*N **E** H-1/ Abe*p* **A** C-2 at δ 4.51/74.4. The spectrum also showed correlations between anomeric carbons and transglycosidic protons (Fig. 5 c).

The $^{1}H,^{1}H$ NOESY spectrum (Fig. 4) showed correlations between pairs of transglycosidic protons, namely **A** H-1,**B** H-3; **B** H-1,**D** H-3; **D** H-1,**C** H-3; **C** H-1,**E** H-3; **E** H-1,**A** H-2; at δ 5.09/4.00; 5.03/3.69; 4.71/3.96; 4.88/3.56, and 4.51/3.97, respectively. In addition, the spectrum showed intraresidual correlations of H-2 with H-1, H-3 for α-Rha*p*2OAc residues **B** and **C**, as well as interresidual correlations of H-2 of residues **B** and **C** with H-1 of the glycosylating residues α-Abe*p* (**A**) and β-Glc*p*N (**D**), respectively. The latter correlations further confirmed the linkages between the monosaccharides in the O-unit.

The absolute configuration of 3,6-dideoxy-D-*xylo*-hexose (Abe) in the OPS of *A. sobria* strain K221 was also confirmed by analysis of glycosylation effects on the ^{13}C NMR chemical shifts (Shashkov et al., 1988), using the known L configuration of Rha*p* as a reference.

Finally, on the basis of the results obtained, the novel structure of the O-polysaccharide from *A. sobria* strain K221 was elucidated (Fig. 6).

The linear pentasaccharide repeating unit of *A. sobria* strain K221 appears to be unique not only among *Aeromonas* O-polysaccharides but also, to the best of our knowledge, among all bacterial polysaccharides characterized to date (Bacterial Carbohydrate Structure Database: http://glyco.ac.ru/bcsdb) (Toukach & Egorova, 2016).

3.5. Organization and characterization of the A. sobria K221 O-antigen gene cluster

The draft genome sequence of *A. sobria* K221 consisted of three contigs, with a mean coverage of ~96×, a total genome size of 5,151,793 bp, and a GC content of 57.1%. Within this genome, the 30,458 bp O-antigen gene cluster (OGC) comprised 27 open reading frames (*orf1* – *orf27*) with tentatively assigned functions based on similarity to sequences in public databases (NCBI nr/nt, InterPro-related databases, KEGG) (Table 3; Fig. 7). The DNA sequence of the *A. sobria* K221 O-antigen gene cluster has been deposited in GenBank under accession number PZ284429.

Orf1 and *orf27* shared 99% identity with the housekeeping genes *acrB* and *oprM*, encoding a multidrug efflux pump subunit and an outer membrane protein, respectively. A similar flanking position of these genes around the OGC has also been reported in the O-antigen gene clusters of *A. hydrophila* and *A. sobria* K928 (Cao et al., 2018; Kurzylewska, Bomba, et al., 2023). In the *A. sobria* K221 OGC, four *rml* genes (*orf2-orf5*), arranged as *rmlBDAC*, were identified and proposed to direct dTDP-L-rhamnose biosynthesis based on their high identity to corresponding genes from *Aeromonas* spp. (96–99%) and *S. enterica* (79–94%).

Genes o*rf6-orf9* showed identity to *ddhA*, *ddhB*, ddhC (75–90%), and *ddhD* (54%) from *S. enterica* subsp. Enterica, consistent with a role in the first four steps of CDP-abequose biosynthesis. In contrast, *orf10* showed lower identity (47%) and similarity (67%) to the *S. enterica* CDP-abequose synthase, which catalyzes the final step of the pathway, namely the conversion of CDP-4-keto-3,6-dideoxy-D-glucose to CDP-abequose (Samuel & Reeves, 2003; Vogel et al., 2022). This lower similarity may reflect the rarity of abequose, which limits the availability of comparative genomic data for enzymes involved in the final step of its biosynthesis.

Five genes (*orf12, orf16, orf17, orf19*, and *orf20*) were assigned as putative glycosyltransferase genes based on high sequence identity to glycosyltransferases or glycosyltransferase family 2 proteins from *Aeromonas* spp., *S. enterica*, *Vibrio* spp., *Plesiomonas* spp., and *Yersinia* sp. In addition, *orf13* and *orf18* showed high identity to *O*-acetyltransferases, supporting their role in the 2-*O*-acetylation of rhamnose residues. The function of *orf19* is supported by its ~72% identity to an α-(1 → 3)-L-rhamnosyltransferase from *Aeromonas*, whereas the roles of *orf12*, *orf16*, and *orf17* remain tentative. Their functions were proposed based on the organization of genes within the *S. enterica* C2-C3 serogroup O-antigen cluster. Consistent with the reverse order of O-unit synthesis reflected by the gene arrangement, as described for the *S. enterica* C2-C3 locus (Reeves et al., 2013), *orf16*, located downstream of the *O*-acetyltransferase gene *orf13*, was proposed to encode a second rhamnosyltransferase, whereas *orf18* is positioned upstream of *orf19*, which also encodes a rhamnosyltransferase. Furthermore, *orf12* is located between *wzx* and *wzy*, upstream of the other genes required for O-unit assembly, and shares 36–39% identity with dideoxyhexosyltransferases from *S. enterica* C2–C3 and *Y. pseudotuberculosis* O2c, although with relatively low query coverage (15–22%) (Table 3). It was therefore proposed to encode a putative abequosyltransferase (see Section 3.6), whereas *orf17* was assigned as the remaining putative GlcNAc-transferase.

Based on sequence similarity, *orf20* was identified as *wecA*, which encodes the O-unit initiation enzyme. Its product showed 97% identity to the *Aeromonas* undecaprenyl-phosphate UDP-*N*-acetylglucosamine phosphotransferase WecA, suggesting its involvement in the transfer of β-D-GlcNAc from the corresponding nucleotide precursor to a membrane-linked undecaprenyl phosphate carrier. As in other *Aeromonas* O-antigen gene clusters, *wecA* was located upstream of *wzz*

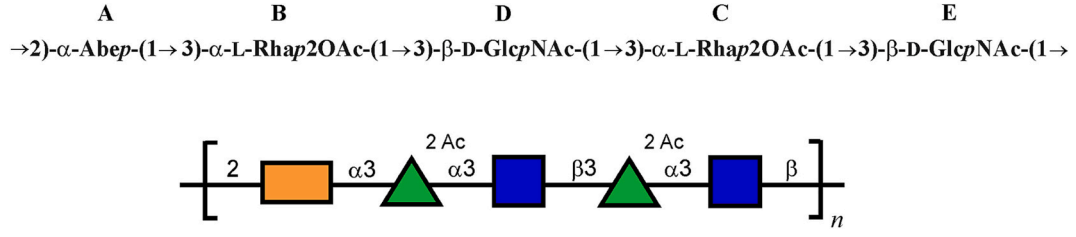

Fig. 6. Structure of the O-antigen repeating unit of *A. sobria* K221 and its schematic representation according to the Symbol Nomenclature For Glycans (SNFG) (Varki et al., 2015).

Table 3
Characteristics of *orfs* identified in the O-antigen gene cluster of *A. sobria* K221 (serogroup PGO1).

orf no.	Gene	Gene position in OGC sequence	No of aa	Similar protein(s) (strain, GenBank Accession No.)	% Identical/% similar (no. of aa overlap)	Putative function of protein
1	*acrB*	0–3150	1049	Multidrug efflux pump subunit AcrB efflux RND transporter permease subunit, *Aeromonas* sp. (WP_287974797.1)	99/99 (1049)	Multidrug efflux pump subunit AcrB
2	*rmlB*	3742–4827	361	dTDP-glucose 4,6-dehydratase, *Aeromonas veronii* (WP_167569593.1) *Salmonella enterica* subsp. *enterica* serovar Anatum (HCM4557973.1)	98/99 (357) 92/95 (357)	dTDP-D-glucose-4,6-dehydratase
3	*rmlD*	4827–5714	295	dTDP-4-dehydrorhamnose reductase, *Aeromonas veronii* (WP_335812626.1) *Salmonella enterica* subsp. *enterica* serovar Anatum (HCM4557972.1)	99/99 (295) 92/96 (295)	dTDP-4-dehydrorhamnose reductase
4	*rmlA*	5827–6705	292	glucose-1-phosphate thymidylyltransferase RfbA, *Aeromonas sobria* (WP_042019776.1) *Salmonella enterica* (EAP0475567.1)	98/99 (292) 94/97 (289)	Glucose-1-phosphate thymidylyltransferase
5	*rmlC*	6767–7315	182	dTDP-4-dehydrorhamnose 3,5-epimerase, *Aeromonas veronii* (WP_106842527.1) *Salmonella enterica* subsp. *enterica* serovar Anatum (HCM4557970.1)	96/98 (182) 79/84 (182)	dTDP-4-dehydrorhamnose 3,5-epimerase
6	*ddhD*	7319–8293	324	CDP-6-deoxy-delta-3,4-glucoseen reductase *Aeromonas jandaei* (RQM71106.1) CDP-6-deoxy-delta-3,4-glucoseen reductase *Salmonella enterica* (AMG27404.1)	89/94 (324) 54/70 (321)	CDP-6-deoxy-delta-3,4-glucoseen reductase
7	*ddhA*	8306–9079	257	glucose-1-phosphate cytidylyltransferase *Salmonella enterica* (EHQ2125277.1)	90/94 (257)	glucose-1-phosphate cytidylyltransferase
8	*ddhB*	9084–10,157	357	CDP-glucose 4,6-dehydratase *Salmonella enterica* HAK5849385.1 CDP-glucose 4,6-dehydratase (WP_000565902.1)	75/86 (353)	CDP-glucose 4,6-dehydratase
9	*ddhC*	10,277–11,590	437	lipopolysaccharide biosynthesis protein RfbH CDP-4-dehydro-6-deoxyglucose reductase *Salmonella enterica* (EII5742227.1)	89/95 (437)	CDP-4-dehydro-6-deoxyglucose reductase
10	*abe*	11,621–12,514	297	NAD-dependent epimerase/dehydratase family protein, CDP-abequose synthase *Salmonella enterica* (AMG27408.1)	47/67 (299)	putative CDP-abequose synthase
11	*wzx*	12,670–14,013	447	Hypothetical protein *Aeromonas veronii* (MGL4886952.1) Polysaccharide biosynthesis protein O156 family O-antigen flippase *Escherichia coli* (WP_165952102.1)	99/99 (447) 27/51 (448)	O-antigen flippase
12	*GT*	14,019–14,954	308	Glycosyltransferase family 2 *Aeromonas jandaei* (WP_124243661.1) CDP-abequose transferase WbaR of *S. enterica* (WP_000038915.1) CDP-abequose transferase WbyD of *Y. pseudotuberculosis* (AKA20933.1)	72/84 (302) 36 [a] (320) 39 [b] (339)	Glycosyltransferase family 2 putative abequose transferase
13	*OAcT*	14,966–15,601	211	*O*-acetyltransferase CatB-related *O*-acetyltransferase *Aeromonas jandaei* (WP_124243660.1)	84/92 (211)	*O*-acetyltransferase
14	*IS*	15,894–16,160	88	IS481 family transposase *Aeromonas veronii* (WP_335854406.1)	66/71 (97)	Transposase
15	*wzy*	16,409–17,683	424	Hypothetical protein *Plesiomonas* sp. (MGL4907140.1) hypotetical protein O132 family O-antigen polymerase *Escherichia coli* (EHH7693635.1)	97/98 (426) 27/51 (242)	O-antigen polymerase
16	*GT*	17,797–18,633	278	Glycosyltransferase *Vibrio parahaemolyticus* (WP_206974909.1)	44/59 (278)	Glycosyltransferase
17	*GT*	18,679–19,692	337	Glycosyltransferase family 2 *Plesiomonas* sp. (MGL4907142.1)	98/98 (337)	Glycosyltransferase family 2
18	*OAcT*	19,616–20,242	208	Acetyltransferase, putative lipopolysaccharide biosynthesis *O*-acetyl transferase WbbJ *Vibrio fluvialis* (MBY7919506.1)	62/72 (172)	*O*-acetyltransferase
19	*GT*	20,247–21,137	296	Alpha-L-Rha alpha-1,3-L-rhamnosyltransferase *Aeromonas salmonicida* (WP_290023938.1)	72/84 (296)	Rhamnosyltransferase
20	*wecA*	21,877–22,962	361	UDP-*N*-acetylglucosamine-undecaprenyl-phosphate *N*-acetylglucosamine-phosphotransferase *Aeromonas veronii* (WP_111901356.1)	97/98 (353)	UDP-GlcNAc-Undecaprenyl-phosphotransferase
21	*wzzE*	23,258–23,635	125	Wzz/FepE/Etk N-terminal domain-containing protein, *Aeromonas sobria* (MGL6150541.1)	66/74 (105)	Wzz/FepE/Etk N-terminal domain-containing protein
22	*wzz*	23,632–24,369	245	LPS O-antigen chain length determinant protein WzzB *Aeromonas veronii* (MFM5287211.1)	95/97 (390)	O-antigen chain length regulator
23	*waaL*	24,466–26,199	577	O-antigen ligase family protein *Aeromonas* sp. CU5 (WP_098969870.1)	97/98 (546)	O-antigen ligase

(*continued on next page*)

Table 3 (*continued*)

orf no.	Gene	Gene position in OGC sequence	No of aa	Similar protein(s) (strain, GenBank Accession No.)	% Identical/% similar (no. of aa overlap)	Putative function of protein
24	*IS3*	26,507–27,385	292	IS3 family ISAs17 transposase ORF B *Salmonella enterica* subsp. *enterica* serovar Enteritidis (EEE4881034.1)	99/99 (292)	IS3 family transposase
25	*IS3rep*	27,343–27,738	131	IS3 family ISAs17 transposase ORF A transposase *Salmonella enterica* subsp. *enterica* (AMW61116.1)	83/94 (131)	IS3 family transposase repressor
26	*hns*	28,159–28,569	136	H-NS histone family protein, *Aeromonas* sp. (WP_042018649.1)	100/100 (136)	DNA-binding protein H-NS
27	*oprM*	29,094–30,458	454	Outer membrane protein OprM, *Aeromonas* sp. (AXL04945.1)	98/98 (465)	Outer membrane protein OprM

[a] 36% amino acid identity (22% query coverage) with WbaR.
[b] 39% amino acid identity (15% query coverage) with WbyD.

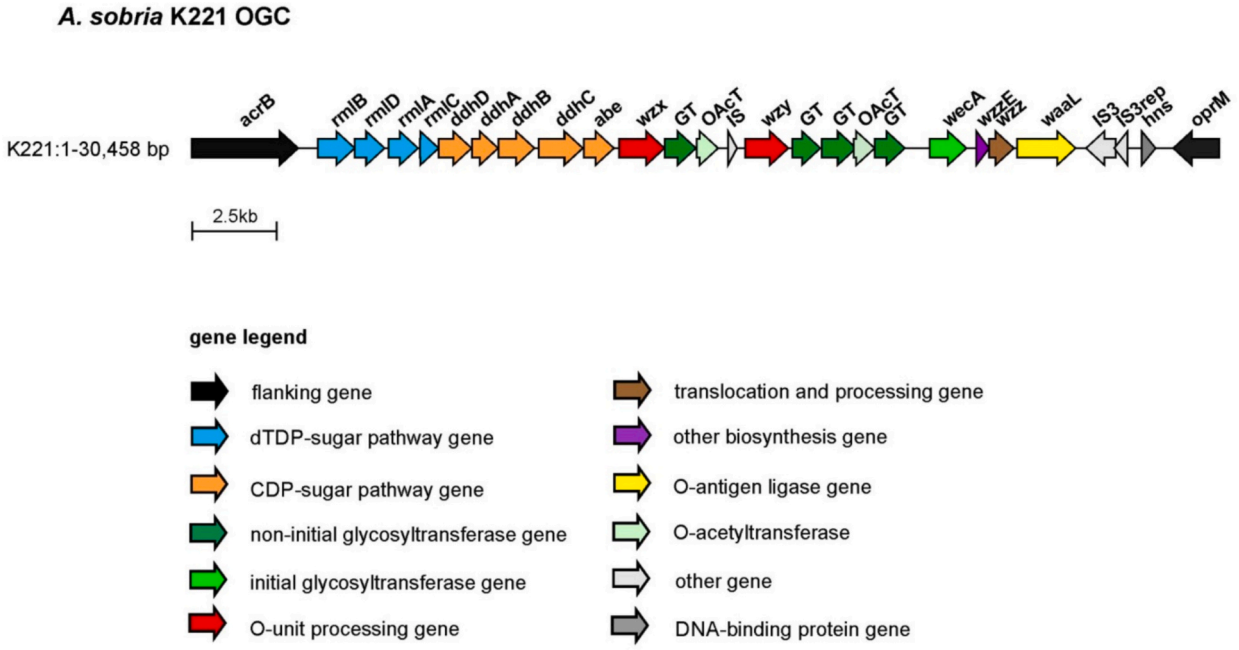

Fig. 7. Composition of the O-antigen gene cluster of *A. sobria* K221 classified to the PGO1 serogroup. Genes are represented as arrows indicating their transcription direction and are colored according to the predicted function of their products according to the legend shown below. The visualization of the OGC region was generated by Clinker (Gilchrist & Chooi, 2021).

(*orf22*) (Cao et al., 2018; Kurzylewska, Turska-Szewczuk, et al., 2023).

Orf11, orf15, orf22, and *orf23* were assigned to genes of the Wzx/Wzy-dependent pathway and identified as encoding Wzx, Wzy, Wzz, and WaaL, respectively, showing 27%, 27%, 95%, and 97% amino acid identity. These assignments were supported by the high sequence similarity of Wzz and WaaL and the lower, but still significant, identity of Wzx and Wzy, consistent with the sequence conservation patterns reported for O-antigen biosynthesis clusters in enteric bacteria, including *Shigella*, *Y. pseudotuberculosis*, and *E. coli* (De Castro et al., 2011; Liu et al., 2008; Liu, Furevi, et al., 2020). In addition, TMHMM 2.0 predicted 10, 10 and 13 transmembrane segments (TMS) for the proteins encoded by *orf11, orf15*, and *orf23*, respectively, further supporting their identification as the O-antigen flippase (Wzx), polymerase (Wzy), and ligase (WaaL).

The *orf14*, *orf24*, and *orf25* genes showed high identity to IS3-family transposases and their corresponding repressors. Whereas *orf14* is located upstream of *wzy* in the same orientation, *orf24* and *orf25* are positioned at the distal (3′) end of the O-antigen gene cluster in the opposite orientation. These elements are unlikely to have a substantial effect on OGC expression, given that the essential biosynthetic genes appear to function efficiently in O-antigen assembly.

3.6. Structural modeling and functional prediction of a putative Abe-transferase

To determine whether Orf12 adopts a GT-A or GT-B fold, sequence- and structure-based analyses were performed. Analysis of the amino acid sequence did not reveal conserved motifs characteristic of metal-dependent GT-A glycosyltransferases, such as the canonical H(X)H or DXD motifs involved in divalent metal coordination (Breton et al., 2012). HHpred analysis returned high-confidence matches (probability >99.8%) to glycosyltransferases representing both GT-A (5TZJ_C) and GT-B-like folds (e.g., 8BZ7_C and 3BCV_A; data not shown), precluding unambiguous classification based on sequence similarity alone. Structural evaluation of the AlphaFold-predicted model indicated a compact α/β/α architecture consistent with a single Rossmann-like domain and lacking a clearly defined two-domain organization typical of GT-B glycosyltransferases (Lairson et al., 2008). These features suggest similarity to a GT-A-like fold; however, the available data do not allow for definitive assignment, and further structural and functional analyses are required. Based on these results, Orf12 was assigned to the glycosyltransferase superfamily. In agreement with the sequence analysis, no conserved metal-binding motifs were identified (e.g., DXD or H(X)H) (Breton et al., 2006), and molecular docking simulations were therefore

performed without inclusion of divalent metal ions (see below).

Three-dimensional models of Orf12, WbaR, and WbyD were compared to assess structural relationships among O-antigen glycosyltransferases. Despite limited sequence identity (36–39%), structural superposition indicated conservation of the overall fold (Fig. 8). TM-align revealed 283 aligned residues (TM-score 0.80; RMSD 2.85 Å) for Orf12–WbaR and 285 residues (TM-score 0.78; RMSD 3.65 Å) for Orf12–WbyD (Fig. 8a,b). Together, these results show that Orf12 adopts the same overall fold as WbaR and WbyD, supporting a conserved glycosyltransferase architecture (Zhang & Skolnick, 2005).

Molecular docking simulations with CDP-abequose (PubChem) using CB-Dock and AutoDock Vina indicated a favorable binding mode within the catalytic pocket of Orf12, with a binding energy of −9.8 kcal/mol (Trott & Olson, 2010). The overall structure of Orf12 with the docked ligand is shown in Fig. 9. Several residues were predicted to interact with the ligand, including Arg13, Asp93, Tyr94, His177, and Lys206. Arg13 and Lys206 are likely involved in stabilizing the negatively charged diphosphate group through electrostatic interactions, while Asp93 may contribute to substrate positioning or catalytic stabilization (Kumar et al., 2018). Tyr94 may further stabilize the ligand through polar or π-stacking interactions, and His177 may assist in orienting the sugar moiety within the binding pocket (Fig. 9) (Liao et al., 2013). In addition to the predicted interactions, a potential hydrogen bond between the ligand (UNL1 N3) and Lys71 (O) was identified at a distance of 2.93 Å, suggesting a role for Lys71 in ligand stabilization, as shown in the close-up view of the binding pocket in Fig. 9. These results suggest that Orf12 can accommodate CDP-abequose within its catalytic pocket, consistent with its predicted role as a nucleotide-sugar–dependent glycosyltransferase involved in O-antigen biosynthesis.

4. Discussion

The O-antigen polysaccharide, which forms the outermost part of the cell-surface glycolipid, consists of repeating units containing one or more sugar residues and exhibits considerable structural diversity. As a major determinant of serogroup specificity and an important virulence factor recognized by both innate and adaptive immunity, it is highly relevant in both human and veterinary medicine (DebRoy et al., 2016).

In this study, we determined the structure of the O-specific polysaccharide from *A. sobria* strain K221, a representative of PGO1, one of the most prevalent serogroups in Polish aquaculture (Kozińska, 2009), and correlated this structure with the genetic organization of the O-antigen biosynthesis locus. The linear pentasaccharide repeating unit of the *A. sobria* K221 OPS differs from previously described *Aeromonas* O-polysaccharides in containing abequose, a sugar that has not previously been identified in the O-polysaccharides of this genus. This finding further highlights the distinctive structural features of the K221 O-polysaccharide within the PGO1 serogroup. Although abequose has previously been reported in the O-antigens of *S. enterica* serogroups B and C2-C3, *Y. pseudotuberculosis* O2a, O2b, and O2c (Kenyon et al., 2017), *C. freundii* PCM 1555 (O22) (Katzenellenbogen et al., 2009), and *Pectobacterium aquaticum* IFB5637 (Szulta et al., 2023), its occurrence in *A. sobria* K221 is unique. In particular, the presence of an α-1,2-linked Abe*p* residue in the main chain, rather than in a side branch, has not previously been reported in any bacterial polysaccharide.

Serological analysis using PGO1 antiserum indicated that the K221, A4, Pt679, and K928 O-antigens share structural features with the reference O-antigen, whose structure remains unknown. Among these strains, the K221 and Pt679 O-antigens appear to be more closely related to the reference O-antigen than those of A4 and K928. By contrast, the O-antigens of K928 and A4 are closely related to each other and share a terminal α-D-Fuc3N substituted with a 3-hydroxybutanoyl group, which seems to represent the major immunoreactive epitope (Kurzylewska, Bomba, et al., 2023; Kurzylewska, Turska-Szewczuk, et al., 2023). Our previous studies using polyclonal antisera (anti-Pt679, A4, and K928) demonstrated that the common linear L-Rha*p* trisaccharide present in the branched pentasaccharide repeating units contributes only partially to the cross-reactivity observed among these related O-antigens. Instead, serotype specificity is determined mainly by the terminal side-branch residue, namely α-D-Fuc3N carrying an amido-linked non-carbohydrate substituent, such as 3-hydroxybutanoyl or acetyl (Kurzylewska, Turska-Szewczuk, et al., 2023). By contrast, in the K221 serotype, structural modifications such as *O*-acetylation of rhamnose and the presence of 2-substituted abequose may constitute dominant epitopes responsible for serotype specificity. Moreover, the presence of *O*-acetylated sugar residues may account for the observed cross-reactivity with Pt679, whose OPS contains terminal α-D-Fuc3N residues carrying *N*-acetyl substituents.

In the *A. sobria* K221 OGC, 27 open reading frames (*orfs)* with putative function, including housekeeping genes, nucleotide sugar synthesis genes, glycosyltransferase genes, and O-antigen processing genes, were identified. In addition, the *A. sobria* K221 OGC contains the *O*-acetyltransferase genes within the locus itself, which determine the modification of two rhamnose residues in the O-unit. By contrast,

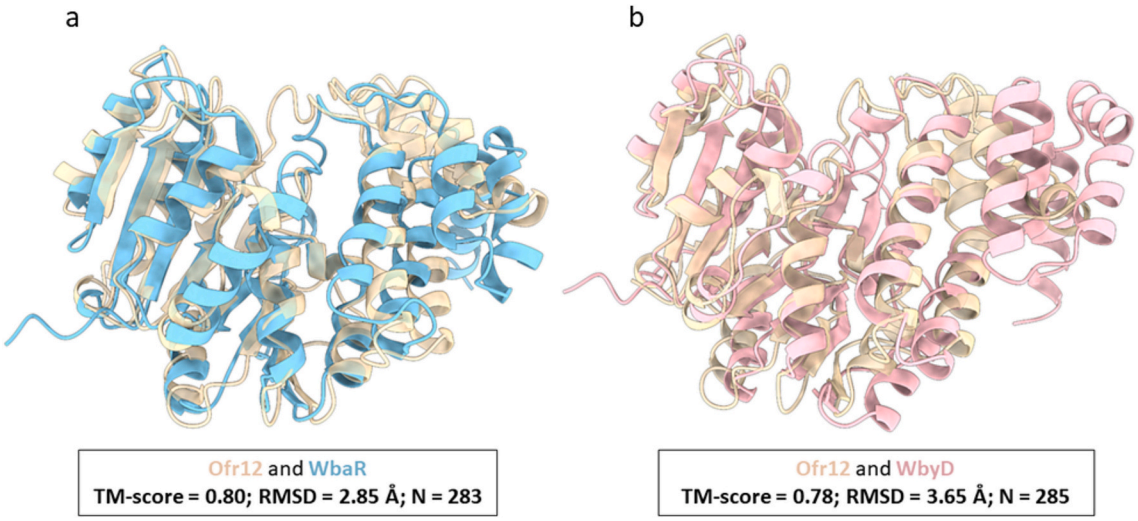

Fig. 8. Structural superposition of Orf12 with glycosyltransferases WbaR and WbyD. Superposition of Ofr12 and WbaR (TM-score = 0.80; RMSD = 2.85 Å; *N* = 283 aligned residues) (**a**). Superposition of Ofr12 and WbyD (TM-score = 0.78; RMSD = 3.65 Å; *N* = 285 aligned residues) (**b**). Ofr12 is shown in beige, WbaR in blue, and WbyD in pink.

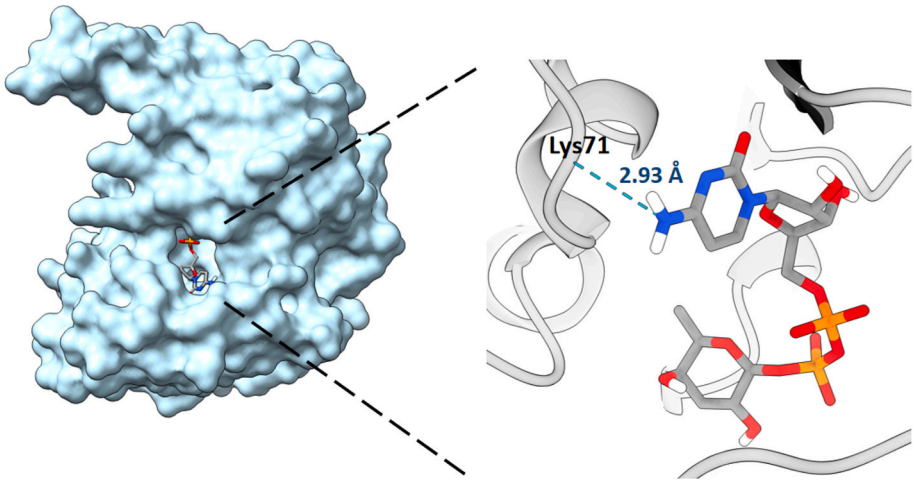

Fig. 9. Predicted binding mode of CDP-abequose in Orf12. The ligand is located within a putative binding pocket (left). A close-up view (right) shows a potential hydrogen bond between Lys71 and the ligand (UNL1 N3) at a distance of 2.93 Å.

although the O-unit of *S. boydii* O11 is characterized by four *O*-acetyl groups as side-chain modifications, none of the genes within its OGC shows sequence similarity to known *O*-acetyltransferases. This absence suggests that the corresponding *O*-acetyltransferase genes are located extrachromosomally or elsewhere in the genome, which is a common feature in *Shigella*, where *O*-acetylation is often mediated by prophages (Reeves & Wang, 2002; Tao et al., 2004).

An analysis of O-polysaccharide synthesis clusters containing 3,6-dideoxyhexoses revealed that the OGC of *A. sobria* K221 exhibits partial structural similarity to that of the *S. enterica* C2-C3 serogroup, particularly within the central region of the locus. Both clusters show a 'reverse-functional' gene arrangement: the 5′ end of the central region initiates with putative abequose biosynthesis gene (*abe*), followed by the flippase (*wzx*) and a putative dideoxyhexose (abequose) transferase. Other glycosyltransferase (GT) genes are situated downstream of the *wzy* gene and are organized in an inverse map order relative to their function (Reeves et al., 2013). This similarity extends to the positioning of the *O*-acetyltransferase gene, which may suggest a shared biosynthetic sequence in which rhamnose acetylation precedes abequose addition. Given that the C2-C3 group is regarded as an evolutionary outlier among galactose-initiated *S. enterica* serogroups, owing to its unique central region and the lack of variants shared with other groups (Reeves et al., 2013), the organizational similarities observed in *A. sobria* K221 may be consistent with either horizontal gene transfer or a distant common evolutionary origin, despite the significant divergence between these genera.

Sequence analysis revealed that, in contrast to *orf16* and *orf17*, *orf12* shows detectable similarity to the putative abequosyltransferases WbaR and WbyD from *S. enterica* C2–C3 and *Y. pseudotuberculosis* O:2c. Although the sequence similarity was low and largely confined to a conserved functional domain, the AlphaFold-predicted structures of these proteins aligned reasonably well, supporting possible functional similarity (Qin et al., 2024).

Furthermore, molecular docking with CDP-abequose supports the proposed assignment of Orf12 as an abequosyltransferase by revealing a favorable ligand orientation within the predicted catalytic site and interactions with residues involved in nucleotide-sugar recognition. These observations are consistent with Orf12 functioning as an abequosyltransferase; however, this assignment still requires experimental confirmation, as docking-based approaches alone cannot directly confirm enzymatic activity (Meng et al., 2011).

By contrast, despite the high nucleotide identity between the *ddhDABC* genes of *Y. pseudotuberculosis* and *Y. mollaretii* ATCC 43969 (~90%), and the significant homology of their *prt* and *tyv* genes (74–78%), the corresponding glycosyltransferases are unrelated (Kenyon et al., 2017). Furthermore, in the *Y. pseudotuberculosis* O2 antigen cluster, the *wzx*, abequosyltransferase, and *wzy* genes precede the other glycosyltransferase genes and occur in the reverse order relative to O-unit assembly. This arrangement is consistent with the gene organization observed in the OGCs of both *S. enterica* serogroup C2–C3 (Reeves et al., 2013) and *A. sobria* K221.

High amino acid similarity (95–97%) is characteristic for Wzz and WaaL, whereas significantly lower homology (~30%) for Wzx and Wzy is commonly observed among different O-serogroups of enteric bacteria and *Aeromonas* (Cao et al., 2018; Liu et al., 2008; Liu, Furevi, et al., 2020). Consistent with this, sequence comparison across 14 *A. hydrophila* O-serogroups studied to date showed that even the most closely related pairs of *wzx* and *wzy* shared only 58.1% and 69.3% identity, respectively (Cao et al., 2018). This divergence, particularly within Wzy, likely reflects its role in determining the specific glycosidic linkages formed during O-antigen polymerization. These findings highlight the variability of proteins involved in O-antigen specificity and support the use of *wzx* and *wzy* as molecular markers for *Aeromonas* serotyping and classification.

IS and transposable elements are well-documented drivers of bacterial genome evolution (Bennett, 2004). In O-antigen clusters they are typically adjacent to recombination sites. A notable example occurs in *Y. pseudotuberculosis*, where a remnant IS630 is located near a recombination breakpoint at which the *abe* gene has replaced the *prt* and *tyv* genes, leaving behind a residual *tyv* fragment (Wang et al., 2002). The presence of IS elements in the *A. sobria* K221 cluster may likewise indicate ancestral recombination events that facilitated the acquisition of its unique O-antigen biosynthetic genes. Insertion events can lead to gene inactivation, while the clustering of multiple IS elements may facilitate the mobilization of large genomic segments (Liu et al., 2008). For instance, in the *Shigella* B12 O-antigen cluster, an IS630 element is located between *rmlD* and *rmlA* in an opposite orientation; however, its presence does not impede the expression of downstream genes, as evidenced by the successful synthesis of L-Rha.

In contrast, in *Shigella* B6, an IS629 element disrupts *wbaM* (the ribosyltransferase gene), resulting in the loss of the D-Rib*f* side branch. Although this is the only difference between the B6 and B10 O antigen gene clusters, the D-Rib*f* residue constitutes an immunodominant epitope, as no cross-reactivity is observed between these serogroups (Liu et al., 2008).

Comprehensive phenotypic and genotypic studies undertaken to characterize and compare the O-antigen polysaccharide of *A. sobria* K221 and the related O-antigens of *A. hydrophila* Pt679, *A. popoffii* A4, and *A. sobria* K928 (Kurzylewska, Turska-Szewczuk, et al., 2023), all classified within the PGO1 serogroup, appear to be the most appropriate

approach. These four representatives of PGO1 are the first strains whose O-serotypes have been fully elucidated at both genetic and structural levels.

The present study demonstrated that the composition of the genetic region of *A. sobria* K221 was consistent with the chemical structure of the corresponding O-antigen. Although *Aeromonas* strains possessing closely related O-antigens share a high level of OGC sequence identity, specific structural elements determine the uniqueness of individual O-serotypes (Kurzylewska, Turska-Szewczuk, et al., 2023). In *A. sobria* K221, structural modifications of the O-polysaccharide, including *O*-acetylation of rhamnose residues and the presence of 2-substituted abequose in the main chain, may represent relevant epitopes underlying the serotype specificity of this strain within serogroup PGO1. Moreover, the presence of *O*-acetylated Rha*p* residues may account for the observed cross-reactivity with Pt679, whose O-polysaccharide contains α-D-Fuc3N residues with *N*-acetyl substituents in the side branches.

O-acetyl groups are important immune determinants in several bacterial polysaccharide vaccines, including those against meningococcal serogroup A and *S. typhi* Vi. These groups are often considered critical to the structural identity of polysaccharides because they may constitute an important part of the immunodominant epitopes and therefore play a major role in eliciting a functional immune response (Berti et al., 2018). Consistent with this, Van Puyvelde et al. (Van Puyvelde et al., 2022) demonstrated that variation in O-antigen *O*-acetylation patterns among *S. typhi*murium clinical isolates from the Democratic Republic of the Congo may influence the optimal design of O-antigen-based vaccines. The *S. typhi*murium O-antigen is composed of repeating units containing mannose, rhamnose and galactose, which confer O12 specificity.

Additional structural diversity arises from the side-branch sugar abequose, which confers O4 specificity, and from modifications such as glucosylation and *O*-acetylation (Wang et al., 2002).

Notably, O5 specificity results from 2-*O*-acetylation of abequose, a modification that influences the binding of multiple antibody types, and may therefore be relevant to vaccine design (Van Puyvelde et al., 2022).

Taken together, integrated chemical and genetic characterization of O-antigen polysaccharides within a given serogroup may support the development of PCR-based methods (DebRoy et al., 2016) for rapid serotyping of *Aeromonas* spp. associated with severe MAI/MAS diseases in aquaculture and improve the selection of vaccine candidates for immunoprophylaxis.

5. Conclusions

The combined structural and serological characterization of the *A. sobria* K221 O-antigen confirmed our hypothesis that the PGO1 serogroup comprises strains exhibiting greater O-polysaccharide diversity than previously recognized. Analysis of the O-antigen gene cluster also enabled the assignment of putative functions to genes within the O-antigen biosynthesis locus. The O-antigen repeating unit was identified as a linear pentasaccharide composed of two β-Glc*p*NAc residues, two 2-*O*-acetylated α-Rha*p* residues, and one α-Abe*p* residue. This is the first report of abequose in an *Aeromonas* O-polysaccharide, where it was identified as a 2-substituted residue rather than as a terminal sugar, highlighting the unusual structure of the K221 O-antigen. Bioinformatic analyses of Orf12 using UCSF ChimeraX and TM-align, together with molecular docking indicated that this protein may represent a putative α-1,3 CDP-abequosyltransferase. Analysis of O-polysaccharide synthesis clusters containing 3,6-dideoxyhexoses revealed that the *A. sobria* K221 OGC exhibits partial structural similarity to that of the *S. enterica* C2-C3 serogroup, particularly within the central region of the locus, suggesting the possibility of interspecies recombination. In the *A. sobria* K221 serotype, structural modifications such as rhamnose *O*-acetylation and the presence of 2-substituted abequose may constitute dominant epitopes underlying serotype specificity. Given the serological and structural heterogeneity of the *Aeromonas* sp. PGO1 serogroup, which is prevalent in Polish aquaculture, the establishment of sub-serogroups appears justified and warrants further investigation.

CRediT authorship contribution statement

Anna Turska-Szewczuk: Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Katarzyna Dworaczek:** Writing – original draft, Visualization, Software, Methodology, Investigation, Conceptualization. **Arkadiusz Bomba:** Writing – review & editing, Visualization, Software, Methodology, Data curation. **Maria Kubaczyńska:** Writing – review & editing, Validation, Investigation. **Dominika Drzewiecka:** Writing – review & editing, Validation, Investigation, Data curation. **Ewelina Iwan:** Writing – review & editing, Data curation. **Agnieszka Pękala-Safińska:** Writing – review & editing, Validation, Investigation.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT (OpenAI) in order to improve language, grammar, and readability. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors thank Dr. Paweł Sowiński (Intercollegiate NMR Laboratory, Department of Chemistry, Technical University of Gdansk, Poland) for recording the NMR spectra, MSc Hubert Pietras for technical support in LPS isolation, and Prof. Alicja Kozińska for kindly providing the reference PGO1 antiserum.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.carbpol.2026.125547.

Data availability

Data will be made available on request.

References

Bennett, P. M. (2004). Genome plasticity: insertion sequence elements, transposons and integrons, and DNA rearrangement. *Methods in Molecular Biology, 266*, 71–113. https://doi.org/10.1385/1-59259-763-7:071

Berti, F., De Ricco, R., & Rappuoli, R. (2018). Role of O-Acetylation in the Immunogenicity of Bacterial Polysaccharide Vaccines. *Molecules, 23*, 1340. https://doi.org/10.3390/molecules23061340

Bock, K., Lundt, I., & Pedersen, C. (1973). Assignment of anomeric structure to carbohydrates through germinal 13C-H coupling constants. *Tetrahedron Letters, 14* (13), 1037–1040. https://doi.org/10.1016/S0040-4039(01)95898-8

Bock, K., & Pedersen, C. (1983). Carbon-13 nuclear magnetic resonance spectroscopy of monosaccharides. *Advances in Carbohydrate Chemistry and Biochemistry, 41*, 27–66. https://doi.org/10.1016/S0065-2318(08)60055-4

Bouras, G., Judd, L. M., Edwards, R. A., Vreugde, S., Stinear, T. P., & Wick, R. R. (2024). How low can you go? Short-read polishing of Oxford Nanopore bacterial genome assemblies. *Microbial Genomics*. https://doi.org/10.1099/mgen.0.001254

Breton, C., Fournel-Gigleux, S., & Palcic, M. M. (2012). Recent structures, evolution and mechanisms of glycosyltransferases. *Current Opinion in Structural Biology, 22*(5), 540–549. https://doi.org/10.1016/j.sbi.2012.06.007. Oct.

Breton, C., Snajdrová, L., Jeanneau, C., Koca, J., & Imberty, A. (2006). Structures and mechanisms of glycosyltransferases. *Glycobiology, 16*(2), 29R–37R. https://doi.org/10.1093/glycob/cwj016. Feb.

Camacho, C., Coulouris, G., Avagyan, V., et al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics, 10*, Article 421. https://doi.org/10.1186/1471-2105-10-421

Cao, H., Wang, M., Wang, Q., Xu, T., Du, Y., Li, H., … Liu, B. (2018). Identifying genetic diversity of O antigens in Aeromonas hydrophila for molecular serotype detection. *PLoS One, 13*, Article e0203445. https://doi.org/10.1371/journal.pone.0203445

Chen, S. (2025). fastp 1.0: An ultra-fast all-round tool for FASTQ data quality control and preprocessing. *iMeta, 4*(5). https://doi.org/10.1002/imt2.70078. e70078. Sep 9.

Ciucanu, I., & Kerek, F. (1984). A simple and rapid method for the permethylation of carbohydrates. *Carbohydrate Research, 131*, 209–217. https://doi.org/10.1016/0008-6215(84)85242-8

De Castro, C., Kenyon, J. J., Cunneen, M. M., Reeves, P. R., Molinaro, A., Holst, O., & Skurnik, M. (2011). Genetic characterisation and structural analysis of the O-specific polysaccharide of Yersinia pseudotuberculosis serotype O:1c. *Innate Immunity, 17*(2), 183–190. https://doi.org/10.1177/1753425910364425. Apr.

De Coster, W., & Rademakers, R. (2023). NanoPack2: population-scale evaluation of long-read sequencing data. *Bioinformatics, 39*(5), Article btad311. https://doi.org/10.1093/bioinformatics/btad311

DebRoy, C., Fratamico, P. M., Yan, X., Baranzoni, G., Liu, Y., et al. (2016). Comparison of O-Antigen Gene Clusters of All O-Serogroups of Escherichia coli and Proposal for Adopting a New Nomenclature for O-Typing. *PLoS One, 11*(1), Article e0147434. https://doi.org/10.1371/journal.pone.0147434

Di Lorenzo, F., De Castro, C., Silipo, A., & Molinaro, A. (2019). Lipopolysaccharide structures of Gram-negative populations in the gut microbiota and effects on host interactions. *FEMS Microbiology Reviews, 43*, 257–272. https://doi.org/10.1093/femsre/fuz002

Dworaczek, K., Kurzylewska, M., Laban, M., Drzewiecka, D., Pękala-Safińska, A., & Turska-Szewczuk, A. (2021). Structural studies of the lipopolysaccharide of *Aeromonas veronii* bv. *sobria* strain K133 which represents new provisional serogroup PGO1 prevailing among mesophilic aeromonads on Polish fish farms. *International Journal of Molecular Sciences, 22*, Article 4272. https://doi.org/10.3390/ijms22084272

Garduno, R. A., Moore, A. R., Oliver, G., Lizama, A. L., Garduno, E., & Kay, W. W. (2000). Host cell invasion and intracellular residence by Aeromonas salmonicida: Role of the S-layer. *Journal of Clinical Microbiology, 46*, 660–668. https://doi.org/10.1139/w00-034

Geue, L., Menge, C., Eichhorn, I., Semmler, T., Wieler, L. H., Pickard, D., … Barth, S. A. (2017). Evidence for Contemporary Switching of the O-Antigen Gene Cluster between Shiga Toxin-Producing Escherichia coli Strains Colonizing Cattle. *Frontiers in Microbiology, 8*, 424. https://doi.org/10.3389/fmicb.2017.00424

Gilchrist, C. L. M., & Chooi, Y. H. (2021). Clinker & clustermap.js: Automatic generation of gene cluster comparison figures. *Bioinformatics, 37*, 2473–2475. https://doi.org/10.1093/bioinformatics/btab007

Jain, C., Rodriguez-R, L. M., Phillippy, A. M., et al. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications, 9*, 5114. https://doi.org/10.1038/s41467-018-07641-9

Janda, J. M., & Abbott, S. (2010). The genus Aeromonas: taxonomy, pathogenicity, and infection. *Clinical Microbiology Reviews, 23*, 35–73. https://doi.org/10.1128/CMR.00039-09

Jansson, P.-E., Kenne, L., & Schweda, E. (1987). Nuclear magnetic resonance and conformational studies on monoacetylated methyl α- and β-D-gluco- and D-galacto-pyranosides. *Journal of the Chemical Society, Perkin Transactions, 1*, 377–383. https://doi.org/10.1039/P19870000377

Jansson, P.-E., Kenne, L., & Widmalm, G. (1989). Computer-assisted structural analysis of polysaccharides with an extended version of casper using 1H- and 13C-n.m.r. data. *Carbohydrate Research, 188*, 169–191. https://doi.org/10.1016/0008-6215(89)84069-8

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature, 596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution, 30*, 772–780. https://doi.org/10.1093/molbev/mst010

Katzenellenbogen, E., Kocharova, N. A., Toukach, P. V., Górska, S., Korzeniowska-Kowal, A., Bogulska, M., … Knirel, Y. A. (2009). Structure of an abequose-containing O-polysaccharide from Citrobacter freundii O22 strain PCM 1555. *Carbohydrate Research, 344*(13), 1724–1728. https://doi.org/10.1016/j.carres.2009.06.005. Sep 8.

Kenyon, J. J., Cunneen, M. M., & Reeves, P. R. (2017). Genetics and evolution of Yersinia pseudotuberculosis O-specific polysaccharides: a novel pattern of O-antigen diversity. *FEMS Microbiology Reviews, 41*(2), 200–217. https://doi.org/10.1093/femsre/fux002

Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B. A., Thiessen, P. A., Yu, B., Zaslavsky, L., Zhang, J., & Bolton, E. E. (2023). PubChem 2023 update. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkac956

Kochetkov, N. K. (1996). Unusual monosaccharides: Components of O-antigenic polysaccharides of microorganisms. *Russian Chemical Reviews, 65*(9), 735–768. https://www.russchemrev.org/RCR229pdf.

Kozińska, A. (2009). *Genotypic and serological analysis of domestic mesophilic isolates Aeromonas sp. in terms of pathogenicity and the type of disease symptoms caused by them in fish. Habilitation thesis*. Puławy, Poland: The National Veterinary Institute—The National Research Institute.

Kozińska, A., & Pękala, A. (2010). Serotyping of Aeromonas species isolated from Polish fish farms in relation to species and virulence phenotype of the bacteria. *Bulletin of the Veterinary Institute in Pulawy, 54*, 315–320.

Kumar, K., Woo, S. M., Siu, T., Cortopassi, W. A., Duarte, F., & Paton, R. S. (2018). Cation-π interactions in protein-ligand binding: theory and data-mining reveal different roles for lysine and arginine. *Chemical Science, 9*(10), 2655–2665. https://doi.org/10.1039/c7sc04905f. Jan 31.

Kurzylewska, M., Bomba, A., Dworaczek, K., Pękala-Safińska, A., & Turska-Szewczuk, A. (2023). Structure and gene cluster annotation of the O-antigen of Aeromonas sobria strain K928 isolated from common carp and classified into the new Aeromonas PGO1 serogroup. *Carbohydrate Research*. , Article 108809. https://doi.org/10.1016/j.carres.2023.108809

Kurzylewska, M., Turska-Szewczuk, A., Dworaczek, K., Bomba, A., Drzewiecka, D., & Pękala-Safińska, A. (2023). Immunochemical studies and gene cluster relationships of closely related O-antigens of Aeromonas hydrophila Pt679, Aeromonas popoffii A4, and Aeromonas sobria K928 strains classified into the PGO1 serogroup dominant in Polish aquaculture of carp and rainbow trout. *Carbohydrate Research, 531*. https://doi.org/10.1016/j.carres.2023.108896, 108896.)Sep.

Lairson, L. L., Henrissat, B., Davies, G. J., & Withers, S. G. (2008). Glycosyltransferases: structures, functions, and mechanisms. *Annual Review of Biochemistry, 77*, 521–555. https://doi.org/10.1146/annurev.biochem.76.061005.092322

Liao, S. M., Du, Q. S., Meng, J. Z., Pang, Z. W., & Huang, R. B. (2013). The multiple roles of histidine in protein interactions. *Chemistry Central Journal, 7*(1), 44. https://doi.org/10.1186/1752-153X-7-44. Mar 1.

Liu, B., Furevi, A., Perepelov, A. V., Guo, X., Cao, H., Wang, Q., … Widmalm, G. (2020). Structure and genetics of Escherichia coli O antigens. *FEMS Microbiology Reviews, 44*, 655–683. https://doi.org/10.1093/femsre/fuz028

Liu, B., Knirel, Y. A., Feng, L., Perepelov, A. V., Senchenkova, S. N., Wang, Q., … Wang, L. (2008). Structure and genetics of Shigella O antigens. *FEMS Microbiology Reviews, 32*(4), 627–653. https://doi.org/10.1111/j.1574-6976.2008.00114.x. Jul.

Liu, Y., Grimm, M., Dai, W. T., Hou, M. C., Xiao, Z. X., & Cao, Y. (2020). CB-Dock: a web server for cavity detection-guided protein–ligand docking. *Acta Pharmacologica Sinica, 41*, 138–144. https://doi.org/10.1038/s41401-019-0228-6

Martinez-Murcia, A. J., Monera, A., Saavedra, M. J., Oncina, R., Lopez-Alvarez, M., Lara, E., & Figueras, M. J. (2011). Multilocus phylogenetic analysis of the genus Aeromonas. *Systematic and Applied Microbiology, 34*, 189–199. https://doi.org/10.1016/j.syapm.2010.11.014

Matys, J., Turska-Szewczuk, A., & Sroka-Bartnicka, A. (2020). Role of bacterial secretion systems and effector proteins - insights into Aeromonas pathogenicity mechanisms. *Acta Biochimica Polonica, 67*, 283–293. https://doi.org/10.18388/abp.2020_5410

Meng, E. C., Goddard, T. D., Pettersen, E. F., Couch, G. S., Pearson, Z. J., Morris, J. H., & Ferrin, T. E. (2023). UCSF ChimeraX: Tools for structure building and analysis. *Protein Science, 32*, Article e4792. https://doi.org/10.1002/pro.4792

Meng, X. Y., Zhang, H. X., Mezei, M., & Cui, M. (2011). Molecular docking: a powerful approach for structure-based drug discovery. *Current Computer-Aided Drug Design, 7*(2), 146–157. https://doi.org/10.2174/157340911795677602. Jun.

Merino, S., Rubires, X., Aguillar, A., Guillot, J. F., & Tomas, J. M. (1996). The role of the O-antigen lipopolysaccharide on the colonization *in vivo* of the germfree chicken gut by*Aeromonas hydrophila*serogroup O:34. *Microbial Pathogenesis, 20*, 325–333. https://doi.org/10.1006/mpat.1996.0031

Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., & Steinegger, M. (2022). ColabFold: making protein folding accessible to all. *Nature Methods, 19*, 679–682. https://doi.org/10.1038/s41592-022-01488-1

Okonechnikov, K., Golosova, O., Fursov, M., & UGENE team. (2012). Unipro UGENE: A unified bioinformatics toolkit. *Bioinformatics, 28*(8), 1166–1167. https://doi.org/10.1093/bioinformatics/bts091. Apr 15.

Ovchinnikova, O. G., Kocharova, N. A., Wykrota, M., Shashkov, A. S., Knirel, Y. A., & Rozalski, A. (2007). Structure of a colitose-containing O-polysaccharide from the lipopolysaccharide of Providencia alcalifaciens O6. *Carbohydrate Research, 342*(14), 2144–2148. https://doi.org/10.1016/j.carres.2007.05.010

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., & Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research, 25*(7), 1043–1055. https://doi.org/10.1101/gr.186072.114. Jul.

Qin, J., Hong, Y., & Totsika, M. (2024). Determining glycosyltransferase functional order via lethality due to accumulated O-antigen intermediates, exemplified with Shigella flexneri O-antigen biosynthesis. *Applied and Environmental Microbiology, 90*(6). https://doi.org/10.1128/aem.02203-23. e0220323. Jun 18.

Rabaan, A. A., Gryllos, I., Tomas, J. M., Shaw, J. G., Infect. Immun, Rabaan, A. A., … Shaw, J. G. (2001). Motility and the polar flagellum are required for *Aeromonas caviae* adherence to HEp-2 cells. *Infection and Immunity, 69*, 4257–4267. https://doi.org/10.1128/iai.69.7.4257-4267.2001, 2001.

Reeves, P. P., & Wang, L. (2002). Genomic organization of LPS-specific loci. *Current Topics in Microbiology and Immunology, 264*(1), 109–135.

Reeves, P. R., Cunneen, M. M., Liu, B., & Wang, L. (2013). Genetics and Evolution of the Salmonella Galactose-Initiated Set of O Antigens. *PLoS One, 8*(7), Article e69306. https://doi.org/10.1371/journal.pone.0069306

Sakazaki, R., & Shimada, T. (1984). O-serogrouping scheme for mesophilic Aeromonas strains. *Japanese J. Med. Sci. Biol., 37*, 247–255. https://doi.org/10.1016/S0923-2508(98)80323-9

Samuel, G., & Reeves, P. (2003). Biosynthesis of O-antigens: genes and pathways involved in nucleotide sugar precursor synthesis and O-antigen assembly. *Carbohydrate Research, 338*(23), 2503–2519. https://doi.org/10.1016/j.carres.2003.07.009. Nov 14.

Schwengers, O., Jelonek, L., Dieckmann, M. A., Beyvers, S., Blom, J., & Goesmann, A. (2021). Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. *Microbial Genomics, 7*(11). https://doi.org/10.1099/mgen.0.000685, 000685. Nov.

Senchenkova, S. N., Shashkov, A. S., Knirel, Y. A., Schwarzmüller, E., & Mayer, H. (1997). Structure of the O-specific polysaccharide of Salmonella enterica ssp. arizonae 050 (Arizona 9a,9b). *Carbohydrate Research, 301*(1–2), 61–67.

Shashkov, A. S., Lipkind, G. M., Knirel, Y. A., & Kochetkov, N. K. (1988). Stereochemical factors determining the effects of glycosylation on the 13C chemical shifts in carbohydrates. *Magnetic Resonance in Chemistry, 26*, 735–747. https://doi.org/10.1002/mrc.1260260904

Shashkov, A. S., Zhang, W., Perepelov, A. V., Weintraub, A., Liu, B., Widmalm, G., & Knirel, Y. A. (2016). Structure of the O-polysaccharide of Escherichia coli O132. *Carbohydrate Research, 427*, 44–47. https://doi.org/10.1016/j.carres.2016.03.016

Silipo, A., Molinaro, A., Nazarenko, E. L., Gorshkova, R. P., Ivanova, E. P., Lanzetta, R., & Parrilli, M. (2005). The O-chain structure from the LPS of marine halophilic bacterium *Pseudoalteromonas carrageenovora*-type strain IAM 12662T. *Carbohydrate Research, 340*(17), 2693–2697. https://doi.org/10.1016/j.carres.2005.09.007

Söding, J., Biegert, A., & Lupas, A. N. (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Research, 33*, W244–W248. https://doi.org/10.1093/nar/gki408

Steinegger, M., & Söding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotechnology, 35*, 1026–1028. https://doi.org/10.1038/nbt.3988

Szulta, S., Kowalczyk, A., Czerwicka-Pach, M., Motyka-Pomagruk, A., Śledź, W., Kaczyńska, N., Łojkowska, E., & Kaczyński, Z. (2023). The chemical structure of the O-polysaccharide isolated from the lipopolysaccharide of Pectobacterium brasiliense IFB5527, a phytopathogenic bacterium of high economic importance. *Carbohydrate Research, 527*, Article 108806. https://doi.org/10.1016/j.carres.2023.108806

Tao, J., Feng, L., Guo, H., Li, Y., & Wang, L. (2004). The O-antigen gene cluster of Shigella boydii O11 and functional identification of its wzy gene. *FEMS Microbiology Letters, 234*(1), 125–132. https://doi.org/10.1016/j.femsle.2004.03.021. May 1.

Tomas, J. M. (2012). *ISRN Microbiology*, 1–22. https://doi.org/10.5402/2012/256261

Torgov, V. I., Shibaev, V. N., Shashkov, A. A., & Rozhnova, S. S. (1990). Structural studies of the O-specific polysaccharide from *Salmonella kentucky* strain 98/39 (O:8, H:i, Z6). *Carbohydrate Research, 208*, 293–300. ISSN 0008-6215 https://doi.org/10.1016/0008-6215(90)80115-J.

Toukach, P. V., & Egorova, K. S. (2016). Carbohydrate structure database merged from bacterial, archaeal, plant and fungal parts. *Nucleic Acids Research, 44*, D1229–D1236. https://doi.org/10.1093/nar/gkv840

Trott, O., & Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry, 31*, 455–461. https://doi.org/10.1002/jcc.21334

Tsai, C. M., & Frasch, C. E. (1982). A sensitive silver stain for detecting lipopolysaccharides in polyacrylamide gels. *Analytical Biochemistry, 119*, 115–119. https://doi.org/10.1016/0003-2697(82)90673-X

Van Puyvelde, S., Gasperini, G., Biggel, M., Phoba, M. F., Raso, M. M., de Block, T., … Micoli, F. (2022). Genetic and structural variation in the O-Antigen of *Salmonella enterica* Serovar Typhimurium isolates causing bloodstream infections in the Democratic Republic of the Congo. *mBio, 13*(4). https://doi.org/10.1128/mbio.00374-22. e0037422. Aug.

Varki, A., Cummings, R. D., Aebi, M., Packer, N. H., Seeberger, P. H., Esko, J. D., … Kornfeld, S. (2015). Symbol Nomenclature for Graphical Representations of Glycans. *Glycobiology, 25*(12), 1323–1324. https://doi.org/10.1093/glycob/cwv091. Dec.

Vogel, U., Beerens, K., & Desmet, T. (2022). Nucleotide sugar dehydratases: Structure, mechanism, substrate specificity, and application potential. *Journal of Biological Chemistry, 298*, Article 101809. https://doi.org/10.1016/j.jbc.2022.101809

Wang, L., Andrianopoulos, K., Liu, D., Popoff, M. Y., & Reeves, P. R. (2002). Extensive Variation in the O-Antigen Gene Cluster within One Salmonella enterica Serogroup Reveals an Unexpected Complex History. *Journal of Bacteriology, 184*. https://doi.org/10.1128/jb.184.6.1669-1677.2002

Wang, L., & Reeves, P. R. (2000). The *Escherichia coli* O111 and *Salmonella enterica* O35 Gene Clusters: Gene Clusters Encoding the Same Colitose-Containing O Antigen Are Highly Conserved. *Journal of Bacteriology*, 182. https://doi.org/10.1128/jb.182.18.5256-5261.2000

Wei, S., Sipos, B., & Zhao, L. (2024). SeqKit2: A Swiss Army Knife for sequence and alignment processing. *iMeta*. , Article e191. https://doi.org/10.1002/imt2.191

Westphal, O., & Jann, K. (1965). Bacterial lipopolysaccharides. Extraction with phenol water and further applications of the procedure. In R. L. Whistler (Ed.), *Methods Carbohydr. Chem.* (pp. 83–91). New York: Academic Press.

Whitfield, C., Williams, D. M., & Kelly, S. D. (2020). Lipopolysaccharide O-antigens bacterial glycans made to measure. *The Journal of Biological Chemistry, 295*, 10593–10609. https://doi.org/10.1074/jbc.REV120.009402

Wick, R. R., & Holt, K. E. (2022). Polypolish: Short-read polishing of long-read bacterial genome assemblies. *PLoS Computational Biology, 18*(1), Article e1009802. https://doi.org/10.1371/journal.pcbi.1009802

Wick, R. R., Judd, L. M., Cerdeira, L. T., et al. (2021). Trycycler: Consensus long-read assemblies for bacterial genomes. *Genome Biology, 22*, Article 266. https://doi.org/10.1186/s13059-021-02483-z

Zhang, Y., & Skolnick, J. (2005). TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Research, 33*, 2302–2309. https://doi.org/10.1093/nar/gki524

Zimin, A. V., & Salzberg, S. L. (2020). The genome polishing tool POLCA makes fast and accurate corrections in genome assemblies. *PLoS Computational Biology, 16*(6), Article e1007981. https://doi.org/10.1371/journal.pcbi.1007981