Carbohydrate Research 361 (2012) 73–77

Contents lists available at SciVerse ScienceDirect

Carbohydrate Research

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

Note

Structure and gene cluster of the O-antigen of *Escherichia coli* O102

Andrei V. Perepelov [a,*], Quan Wang [b], Sof'ya N. Senchenkova [a], Yanping Gong [b], Alexander S. Shashkov [a], Lei Wang [b], Yuriy A. Knirel [a]

[a] N.D. Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, 119991 Moscow, Russian Federation
[b] TEDA School of Biological Sciences and Biotechnology, Nankai University, TEDA, Tianjin, China

ARTICLE INFO

Article history:
Received 8 June 2012
Received in revised form 18 July 2012
Accepted 31 July 2012
Available online 14 August 2012

Keywords:
Escherichia coli
Bacterial polysaccharide structure
O-Antigen
Lipopolysaccharide
O-Antigen gene cluster

ABSTRACT

The O-polysaccharide (O-antigen) of *Escherichia coli* O102 was studied by sugar analysis along with one- and two-dimensional ^{1}H and ^{13}C NMR spectroscopy. The following structure of the branched pentasaccharide repeating unit was established:

α-L-Rha*p*-(1→4)-β-D-Gal*p*NAc
1
↓
3
→4)-α-D-Gal*p*-(1→6)-β-D-Glc*p*-(1→3)-β-D-Gal*p*NAc-(1→

The O-antigen gene cluster of *E. coli* O102 was sequenced. The gene functions were tentatively assigned by comparison with sequences in the available databases and found to be in full agreement with the *E. coli* O102 O-polysaccharide structure.

© 2012 Elsevier Ltd. All rights reserved.

Escherichia coli clones including both commensal and pathogenic ones are normally identified by the combination of their somatic (O), flagellar (H), and sometimes capsular (K) antigens.[1] The O-antigen or O-polysaccharide, consisting of many repeats of an oligosaccharide (O-units), is an essential component of lipopolysaccharide on the surface of Gram-negative bacteria, and one of the most variable cell constituents. Till now, about 174 O-antigen forms have been recognized for *E. coli*.[2]

Different O-antigen forms are almost entirely due to genetic variations in the O-antigen gene cluster, which is located between *galF* and *gnd* in the chromosome in *E. coli*. The cluster usually contains genes involved in nucleotide-sugar synthesis, monosaccharide transfer, and O-antigen processing (*wzx* and *wzy* genes).[1]

Strains belonging to *E. coli* O102 have been isolated and identified as pathogens for avian colibacillosis[3] and also reported to be associated with Shiga toxin-producing *Escherichia coli* (STEC) infections in humans.[4] Recently, *E. coli* O102 strains have been reported as one of the major multidrug-resistant *E. coli* strains worldwide, which may cause infections in both hospital and community settings.[5–7] Therefore, studies of *E. coli* O102 O-antigen will be helpful for development of a diagnostic tool useful for prevention and control of diseases caused by *E. coli* O102 strains.

In this work, we established the structure of *E. coli* O102 O-antigen that has not been hitherto known and characterized the O-antigen gene cluster of these bacteria.

A high-molecular mass OPS was obtained by mild acid degradation of the lipopolysaccharide isolated from dried bacterial cells by the phenol–water procedure. Sugar analysis by GLC of the alditol acetates derived after full acid hydrolysis of the OPS revealed Glc, Gal, GalN, and Rha in the ratio ~1:1.5:1:1 (detector response). GLC analyses of the acetylated (*S*)-2-(+)-octyl glycosides demonstrated the D configuration of Glc, Gal, GalN, and the L configuration of Rha.

The ^{13}C NMR spectrum of the OPS (Fig. 1) showed signals for five anomeric carbons in the region δ 99.6–105.8, four $HOCH_2$-C groups (C6 of hexoses) at δ 61.9, 62.2, 62.9, and 66.3 (data of the attached proton test), two nitrogen-bearing carbons (C2 of GalN) at δ 52.3 and 53.8, 18 oxygen-bearing non-anomeric sugar ring carbons in the region δ 68.8–82.7, and two *N*-acetyl group at δ 23.6 (2 CH_3), and 176.1 and 176.2 (CO). Accordingly, the ^{1}H NMR spectrum of the OPS contained signals for five anomeric protons at δ 4.59–5.19, other sugar protons in the region δ 3.30–4.42, one methyl group (H-6 of Rha) at δ 1.30, and two *N*-acetyl groups at δ 2.05 and 2.07. Therefore, the OPS has a pentasaccharide repeat (O-unit)

* Corresponding author. Tel.: +7 499 137 61 48; fax: +7 499 135 53 28.
E-mail address: perepel@ioc.ac.ru (A.V. Perepelov).

0008-6215/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.carres.2012.07.024

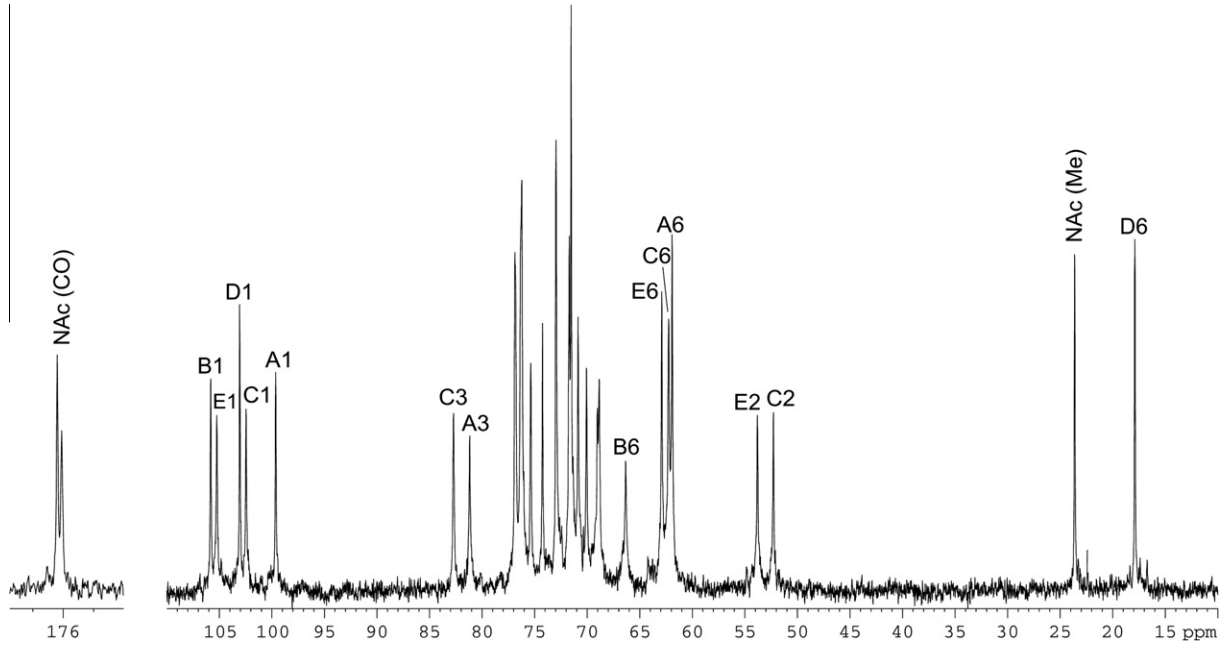


Figure 1. ^{13}C NMR spectrum of the OPS of *E. coli* O102. Numbers refer to carbons in sugar residues denoted as shown in Table 1.

containing two residues of D-GalNAc (denoted as units **C** and **E**), and one residues each of D-Gal, D-Glc, and L-Rha (units **A**, **B**, and **D**, respectively).

Signals in the ^{1}H and ^{13}C NMR spectra of the OPS were interpreted using two-dimensional homonuclear ^{1}H,^{1}H COSY, TOCSY, ROESY, and heteronuclear ^{1}H,^{13}C HSQC, HMQC-TOCSY, and HMBC experiments (Table 1). Based on intra-residue H,H and H,C correlations and coupling constant values, spin systems were assigned to residues of **A**–**E**, all being in the pyranose form. Relatively large $J_{1,2}$ coupling constant value of 7–8 Hz showed that units **B**, **C**, and **E** are β-linked, whereas a significantly smaller value of <2 Hz indicated the α-linkage of unit **A**. The position at δ 70.8 of the signal for C-5 indicated that unit **D** is α-linked (compare published data δ 70.0 and 73.2 for α- and β-Rha*p*, respectively).[8]

The spin systems for **C** and **E** were distinguished by correlation between protons at the nitrogen-bearing carbons (H2) and the corresponding carbons (C2) at δ 4.06/52.3 and 4.10/53.8. The signals for C3 and C4 of unit **A**, C6 of unit **B**, C3 of unit **C**, and C4 of unit **E** were shifted significantly downfield, as compared with their positions in the corresponding non-substituted monosaccharides.[8] The chemical shifts for the C2–C5 signals of unit **D** were close to those of unsubstituted α-Rha*p*. These data demonstrated the branching character of the OPS chain with unit **D** at the terminal position in the side chain and defined the glycosylation pattern in the O-unit.

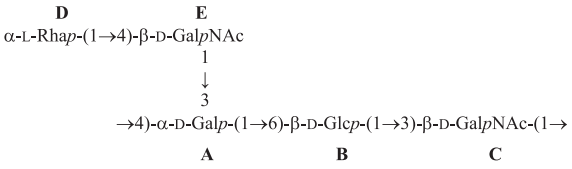
Table 1
^{1}H and ^{13}C NMR chemical shifts (δ, ppm) of the OPS of *E. coli* O102

Sugar unit	H1 C1	H2 C2	H3 C3	H4 C4	H5 C5	H6 (6a, 6b) C6
→3,4)-α-D-Gal*p*-(1→	4.95	3.72	3.95	4.42	3.96	3.69, 3.69
A	99.6	68.8	81.1	76.8	71.5	61.9
→6)-β-D-Glc*p*-(1→	4.63	3.30	3.50	3.62	3. 63	3.76, 4.02
B	105.8	74.2	76.8	70.0	75.3	66.3
→3)-β-D-Gal*p*NAc-(1→	5.01	4.06	3.73	4.18	3.58	3.79, 3.79
C	102.4	52.3	82.7	69.0	76.2	62.2
α-L-Rha*p*-(1→	5.19	4.16	3.78	3.47	3.75	1.30
D	103.0	71.5	71.7	72.9	70.8	17.9
→4)-β-D-Gal*p*NAc-(1→	4.59	4.10	3.87	4.06	3.75	3.79, 3.79
E	105.2	53.8	72.9	76.2	76.2	62.9

The chemical shifts for the *N*-acetyl groups are δ_H 2.05 and 2.07 (Me); δ_C 23.6 (2 Me) and 176.1, 176.2 (CO).

The ROESY spectrum of the OPS showed the following correlations between anomeric protons and protons at the linkage carbons: **A** H1,**B** H6b; **B** H1,**C** H3; **C** H1,**A** H4; **D** H1,**E** H4; and **E** H1,**A** H3 at δ 4.95/3.76, 4.63/3.73, 5.01/4.42, 5.19/4.06, and 4.59/3.95, respectively. The monosaccharide sequence thus determined was confirmed by a heteronuclear ^{1}H,^{13}C HMBC experiment, which showed correlations between anomeric protons and linkage carbons (Fig. 2) and vice versa.

Therefore, the OPS of *E. coli* O102 has the following structure:



Although the OPS is composed of rather common monosaccharides, to our knowledge, this structure is unique among the known bacterial polysaccharide structures.

Characterization of the O-antigen gene cluster of E. coli O102

The O-antigen gene cluster of *E. coli* O102 was found between the housekeeping genes *galF* and *gnd*. A DNA sequence of 12,783 bp was obtained, which contains 11 genes with transcription direction from *galF* to *gnd* (Fig. 3). Their functions were assigned based on similarities to genes from available databases (Table 2). The gene (*wecA*) responsible for the transfer of GlcNAc 1-phosphate/GalNAc 1-phosphate to the undecaprenol phosphate lipid carrier to initiate the O-unit synthesis, is located outside the O-antigen gene cluster.[9] Genes involved in the synthesis of common sugar nucleotide precursors, such as UDP-D-Glc, UDP-D-Gal, and UDP-D-GlcNAc (a precursor of UDP-D-GalNAc), are located outside the O-antigen gene cluster too.

Orf1–4 share high level identities to many known RmlB, D, A, and C proteins (78–96%). The *rmlBDAC* gene set responsible for

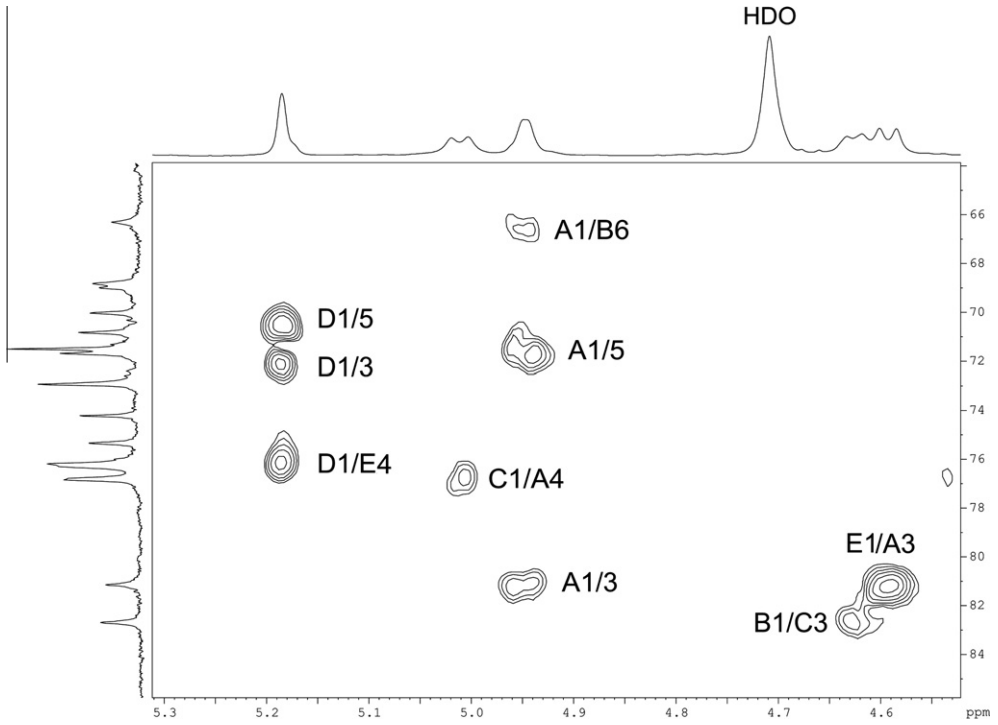


Figure 2. Part of a $^{1}H,^{13}C$ HMBC spectrum of the OPS from *E. coli* O102. The corresponding parts of the ^{1}H and ^{13}C NMR spectra are displayed along the horizontal and vertical axes, respectively. Arabic numerals before and after oblique stroke refer to protons and carbons, respectively, in the sugar residues denoted by letters as shown in Table 1.

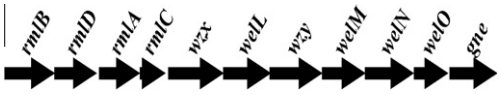


Figure 3. Organization of the *E. coli* O102 O-antigen gene cluster.

the synthesis of dTDP-L-Rha has been well characterized.[10–13] Orf11 shares 58% identity to the *Yersinia enterocolitica* UDP-N-acetylglucosamine 4-epimerase (Gne), which catalyzes the conversion of D-GlcNAc to D-GalNAc (Bengoechea et al., 2002).[14] Orf5 and 7 are the only two proteins with predicted membrane segments. Orf5 has 8 predicted transmembrane segments, which is a typical topological character of Wzx proteins.[15] It also shares 49% identity to the Wzx protein of *Shigella flexneri* 2a. Orf7 has 10 predicted transmembrane segments and a periplasmic loop of 55 amino acid residues, which is a typical topological character of Wzy proteins.[16] Orf7 also shares 47% similarity to the polymerase of *Bacteroides fragilis*. Therefore, *orf5* and *orf7* are proposed to be the genes for O-antigen flippase and O-antigen polymerase, respectively, and named accordingly. Orf6, 8, 9, and 10 shared 34–46% identities to glycosyltransferases of other origins. Orf9 belonged to the glycosyltransferase family 1 (Pf00534), and orf8 and 10 belonged to the glycosyltransferase family 2 (PF00535). Therefore, *orf6*, *8*, *9*, and *10* were proposed to encode glycosyltransferases and named *welL, welM, welN*, and *welO*, respectively.

To sum up, the proposed functions of the genes in the O-antigen gene cluster of *E. coli* O102 correspond well to the O102 antigen structure established in this work.

1. Experimental

1.1. Cultivation of bacteria and isolation of lipopolysaccharide

E. coli O102 type strain (laboratory stock numbers G3100) was obtained from the Institute of Medical and Veterinary Science (Adelaide, Australia). Bacteria were grown to late log phase in 8 L of Luria-Bertani broth using a 10-L BIOSTAT C-10 fermentor (B. Braun Biotech Int., Germany) under constant aeration at 37 °C and pH 7.0. Bacterial cells were washed and dried as described.[17]

The lipopolysaccharide in a yield of 7.7 % was isolated from dried cells by the phenol–water method,[18] the crude extract was dialyzed without separation of the layers and freed from nucleic acids and proteins by treatment with 50% aqueous CCl_3CO_2H at pH 2. The supernatant was dialyzed and lyophilized.

1.2. Isolation of O-polysaccharide

Delipidation of the lipopolysaccharide (95 mg) was performed with 2% aq HOAc at 100 °C until precipitation of lipid A. The precipitate was removed by centrifugation (13,000*g*, 20 min), and the supernatant was fractionated by GPC on a column (56 × 2.6 cm) of Sephadex G-50 Superfine (Amersham Biosciences, Sweden) in 0.05 M pyridinium acetate buffer pH 5.5, monitored with a differential refractometer (Knauer, Germany). A high-molecular-mass O-polysaccharide (OPS) was obtained in a yield of 46% of the lipopolysaccharide mass.

1.3. Monosaccharide analysis

The O-polysaccharide was hydrolyzed with 2 M CF_3CO_2H (120 °C, 2 h). Monosaccharides were identified by GLC of the alditol acetates on an Agilent 7820A chromatograph (USA) equipped with an HP-5 column (0.32 mm × 30 m) using a temperature program of 160 (1 min) to 290 °C at 7 °C min^{-1}. The absolute configurations of the monosaccharides were determined by GLC of the acetylated (*S*)-2-octyl glycosides as described.[19]

1.4. Nuclear magnetic resonance spectroscopy

Sample was deuterium-exchanged by freeze-drying from D_2O and then examined as solution in 99.95% D_2O. Nuclear magnetic

Table 2
Characteristics of the ORFs located in the *E. coli* O102 O-antigen gene cluster

Gene name	Position of gene	G+C content (%)	Conserved domain(s)	Similar protein(s), strain(s) (Genbank accession No.)	%Identical aa /%similar aa (total No. of aa)	Putative function of protein
rmlB	707..1792	42.3	NAD dependent epimerase/ dehydratase family (PF01370) *E* value = $1.5 \times e^{-215}$	dTDP-glucose 4,6-dehydratase *Shigella flexneri 2a* str. 301 (AAN43643)	94/96(361)	dTDP-glucose 4,6-dehydratase
rmlD	1792..2694	46.1	RmlD substrate binding domain (PF04321) *E* value = $2.8 \times e^{-74}$	dTDP-6-deoxy-L-mannose dehydrogenase *Shigella flexneri 2a* str. 301 (AAN43642)	86/90(300)	dTDP-6-deoxy-L-mannose dehydrogenase
rmlA	2752..3630	44.0	Nucleotidyl transferase (PF00483) *E* value = $3.2 \times e^{-115}$	Glucose-1-phosphate thymidylyl transferase *Shigella flexneri 2a* str. 301 (AAN43641)	96/98(292)	Glucose-1-phosphate thymidylyl transferase
rmlC	3635..4183	35.2	dTDP-4-dehydrorhamnose 3,5-epimerase (PF00908) *E* value = $6.8 \times e^{-114}$	dTDP-4-dehydrorhamnose 3,5-epimerase *Shigella flexneri 2a* str. 301 (AAN43640)	78/85(182)	dTDP-4-dehydrorhamnose 3,5-epimerase
wzx	4252..5439	33.1	Polysaccharide biosynthesis protein (PF01943) *E* value = $7.6 \times e^{-32}$	Putative O-antigen transporter *Shigella flexneri 2a* str. 301 (AAN43639)	49/69(384)	Putative O-antigen transporter
welL	5433..6461	27.2		Eps9K *Streptococcus thermophilus* (AAN63754)	34/59(194)	Glycosyl transferase
wzy	6442..7572	28.3	O-antigen Polymerase (PF04932) *E* value = $9.0 \times e^{-2}$	Putative polymerase *Bacteroides fragilis* (AAG26475)	25/47(359)	O-Antigen polymerase
welM	7569..8501	31.2	Glycos_transf_2 (PF00535) *E* value = $8.1 \times e^{-44}$	Glycosyl transferase *Clostridium acetobutylicum* (AAK80132)	35/56(272)	Glycosyl transferase
welN	8498..9550	29.6	Glycos_transf_1 (PF00534) *E* value = $2.1 \times e^{-36}$	wbcM *Yersinia enterocolitica* (CAA87701)	34/53(355)	Glycosyl transferase
welO	9554..10291	29.5	Glycos_transf_2 (PF00535) *E* value = $7.0 \times e^{-37}$	Glycosyl transferase *Escherichia coli* O157:H7 EDL933 (AAG57100)	46/67(248)	Galactosyl transferase
gne	10330..11349	34.0	NAD dependent epimerase/ dehydratase family (PF01370) *E* value = 2.8 × e-74	Gne *Yersinia enterocolitica* (AAC60777)	58/75(336)	UDP-N-acetylglucosamine-4-epimerase

resonance (NMR) spectra were recorded on a Bruker Avance II 600 spectrometer (Germany) at 30 °C using internal TSP (δ_H 0) and acetone (δ_C 31.45) as reference. 2D NMR spectra were obtained using standard Bruker software, and Bruker TopSpin 2.1 program was used to acquire and process the NMR data. A mixing time of 100 and 150 ms was used in TOCSY and ROESY experiments, respectively.

1.5. Sequencing and analysis of genes

Chromosomal DNA was prepared as described previously.[20] The primers (#1523 and #1524) based on the housekeeping genes *galF* and *gnd*,[21] respectively, were used to amplify the O-antigen gene clusters of the *E. coli* strains. The PCR cycles used were as follows: denaturation at 94 °C for 10 s, annealing at 60 °C for 30 s, and extension at 68 °C for 15 min. Shotgun banks were constructed for each strain as described previously.[22] Sequencing was carried out using an ABI 3730 automated DNA sequencer (Applied Biosystems, Foster City, CA), and sequence data were analyzed using computer programs as described previously.[23] The DNA sequence of O-antigen gene clusters of *E. coli* O102 has been deposited in GenBank under the accession numbers JX087966.

Acknowledgements

This work was supported by the Russian Foundation for Basic Research (projects 11-04-91173-a and 11-04-01020-a), the Federal Targeted Program for Research and Development in Priority Areas of Russia's Science and Technology Complex for 2007–2013 (State Contract No. 16.552.11.7050), the National Natural Science Foundation of China (30900255 and 31111120026), and Tianjin Research Program of Application Foundation and Advanced Technology (10JCYBJC10100).

References

1. Reeves, P. P.; Wang, L. *Curr. Top. Microbiol. Immunol.* **2002**, *264*, 109–135.
2. Wang, L.; Wang, Q.; Reeves, P. R. *Subcell. Biochem.* **2010**, *53*, 123–152.
3. Blanco, J. E.; Blanco, M.; Mora, A.; Jansen, W. H.; Garcia, V.; Vazquez, M. L.; Blanco, J. *Vet. Microbiol.* **1998**, *61*, 229–235.
4. Eklund, M. F.; Scheutz, F.; Siitonen, A. J. *Clin. Microbiol.* **2001**, *39*, 2829–2834.
5. Wolk, M.; Valinsky, L.; Sompolinsky, D.; Sechter, I.; Schmidt, H.; Tetry, S.; Agmon, V. *FEMS Microbiol. Lett.* **2004**, *239*, 249–254.
6. Naseer, U.; Haldorsen, B.; Tofteland, S.; Hegstad, K.; Scheutz, F.; Simonsen, G. S.; Sundsfjord, A. *APMIS* **2009**, *117*, 526–536.
7. Billard-Pomares, T.; Tenaillon, O.; Le Nagard, H.; Rouy, Z.; Cruveiller, S.; Medigue, C.; Arlet, G.; Denamur, E.; Branger, C. *Antimicrob. Agents Chemother.* **2011**, *55*, 1270–1273.
8. Lipkind, G. M.; Shashkov, A. S.; Knirel, Y. A.; Vinogradov, E. V.; Kochetkov, N. K. *Carbohydr. Res.* **1988**, *175*, 59–75.
9. Alexander, D. C.; Valvano, M. A. *J. Bacteriol.* **1994**, *176*, 7079–7084.
10. Samuel, G.; Reeves, P. *Carbohydr. Res.* **2003**, *338*, 2503–2519.
11. Lindqvist, L.; Kaiser, R.; Reeves, P. R.; Lindberg, A. A. *Eur. J. Biochem.* **1993**, *211*, 763–770.
12. Marumo, K.; Lindqvist, L.; Verma, N.; Weintraub, A.; Reeves, P. R.; Lindberg, A. A. *Eur. J. Biochem.* **1992**, *204*, 539–545.
13. Giraud, M. F.; McMiken, H. J.; Leonard, G. A.; Messner, P.; Whitfield, C.; Naismith, J. H. *Acta Crystallogr. D Biol. Crystallogr.* **1999**, *55*, 2043–2046.

14. Bengoechea, J. A.; Pinta, E.; Salminen, T.; Oertelt, C.; Holst, O.; Radziejewska-Lebrecht, J.; Piotrowska-Seget, Z.; Venho, R.; Skurnik, M. *J. Bacteriol.* **2002**, *184*, 4277–4287.
15. Liu, D.; Cole, R. A.; Reeves, P. R. *J. Bacteriol.* **1996**, *178*, 2102–2107.
16. Daniels, C.; Vindurampulle, C.; Morona, R. *Mol. Microbiol.* **1998**, *28*, 1211–1222.
17. Robbins, P. W.; Uchida, T. *Biochemistry* **1962**, *1*, 323–335.
18. Westphal, O.; Jann, K. *Methods Carbohydr. Chem.* **1965**, *5*, 83–91.
19. Leontein, K.; Lönngren, J. *Methods Carbohydr. Chem.* **1993**, *9*, 87–89.
20. Bastin, D. A.; Reeves, P. R. *Gene* **1995**, *164*, 17–23.
21. Feng, L.; Wang, W.; Tao, J.; Guo, H.; Krause, G.; Beutin, L.; Wang, L. *J. Clin. Microbiol.* **2004**, *42*, 3799–3804.
22. Wang, L.; Reeves, P. R. *Infect. Immun.* **1998**, *66*, 3545–3551.
23. Feng, L.; Senchenkova, S. N.; Yang, J.; Shashkov, A. S.; Tao, J.; Guo, H.; Zhao, G.; Knirel, Y. A.; Reeves, P.; Wang, L. *J. Bacteriol.* **2004**, *186*, 383–392.