Structural studies and biosynthetic aspects of the O-antigen polysaccharide from *Escherichia coli* O42

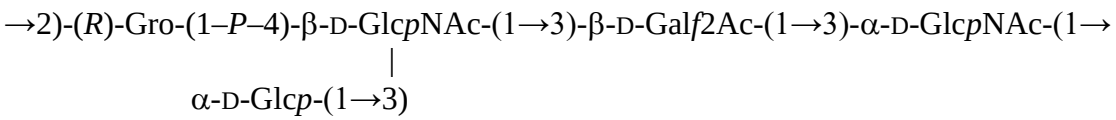
Carolina Fontana[a], Andrej Weintraub[b], Göran Widmalm[a,*]

[a] Department of Organic Chemistry, Arrhenius Laboratory, Stockholm University, S-106 91 Stockholm, Sweden.

[b] Karolinska Institute, Department of Laboratory Medicine, Division of Clinical Microbiology, Karolinska University Hospital, S-141 86 Stockholm, Sweden.

Abstract

The structure of the O-antigen polysaccharide (PS) from *Escherichia coli* O42 has been investigated by NMR spectroscopy as the main method, which was complemented with sugar analysis, mass spectrometry and analysis of biosynthetic information. The O-specific chain of the *O*-deacylated lipopolysaccharide (LPS-OH) consists of branched tetrasaccharide-glycerol repeating units joined by phosphodiester linkages. The lipid-free polysaccharide contains 0.8 equivalents of *O*-acetyl groups per repeating unit and has the following teichoic acid-like structure:



Based on biosynthetic aspects, this should also be the biological repeating unit. This O-antigen structure is remarkably similar to that of *E. coli* O28ac, differing only in the presence or absence, respectively, of a glucose residue at the branching point. The structural similarity explains the serological cross-reactivity observed between strains of these two serogroups, and also their almost identical O-antigen gene cluster sequences.

Keywords: *Escherichia coli*, NMR, *O*-acetylation, O-antigen

* Corresponding author. Tel.: +46 816 3742. E-mail address: gw@organ.su.se

1. Introduction

Within gram-negative bacteria, the *Enterobacteriaceae* family comprises some of the most common pathogens usually found in clinical specimens, such as those of the genera *Escherichia, Klebsiella, Salmonella* and *Shigella*. The *Escherichia coli* strains are rarely harmful and they usually coexist with their host sharing mutual benefits; however, some pathogenic strains have been implicated in a wide range of intestinal diseases, urinary tract infections, sepsis and meningitis.[1] In particular, the diarrheagenic *E. coli* strains that usually affect humans can be classified in six different pathotypes: enteropathogenic *E. coli* (EPEC), enterohaemorrhagic *E. coli* (EHEC), enterotoxigenic *E. coli* (ETEC), enteroaggregative *E. coli* (EAEC), enteroinvasive *E. coli* (EIEC) and diffusely adherent *E. coli* (DAEC).[2] The designation of *E. coli* strains in different serogroups is of particular importance in epidemiological studies, and three principal surface antigens have been employed to differentiate them: a) the lipopolysaccharide (LPS) O antigen, b) the capsular K antigen and c) the flagellar H antigen, and the O:K:H combination is then referred as a serotype.[3]

The O-antigen polysaccharide (PS) is the outermost part of the LPS and it usually consists of oligosaccharide repeating units. Since diverse monosaccharides can be found in different arrangements and linkages within and between repeating units, this is also considered the most variable portion of the LPS. The *E. coli* serogroups are presently numbered from O1 to O181 according to their serological properties but, since seven of them have been removed, only 174 are currently in use.[4,5] Moreover, slightly different O-antigen structures can be identified within the same serogroup giving rise to different subgroups.[6,7] The ECODAB (*E. coli* O-antigen database) contains information about the structure, NMR chemical shifts and cross-reactivity of many *E. coli* O-antigens[4,8] as well as information about glycosyltransferases (GTs) involved in the biosynthesis of O-antigens.[9]

The genes responsible for the biosynthesis of *E. coli* O-antigens are usually found in a single cluster in the bacterial chromosome between the *galF* and *gnd* housekeeping genes[10] with the exceptions being *E. coli* O8, O9, O55, O59 and O155 (in which some of the genes involved in the biosynthesis of the O-antigen can also be found between the *gnd* gene and *his* operon).[11–14] The genes found in the O-antigen gene cluster can be classified as follow: *i*) genes involved in the biosynthesis of sugar precursors (and additional non-sugar substituents when they are present), *ii*) genes involved in the O-unit formation (e.g. GTs, glycerophosphotransferases and *O*-acetyltransferases) and *iii*) O-antigen processing genes (e.g. genes encoding for the Wzx-flippase and Wzy-polymerase). The first group of genes is

highly conserved among species, and the major structural differences observed in the O-antigens are attributed to variations in genes encoding for glycosyltransferases and the polymerase. Furthermore, genes encoding for proteins that modify the O-specific chain after polymerization (such as *O*-acetyl- or glucosyltransferases) may be found outside the O-antigen gene cluster.[15–18] This is the case of *E. coli* O17, O44, O73, O77 and O106, which share the same O-antigen backbone (encoded by genes found in their almost identical O-antigen gene clusters), but they differ in the lateral glucosylation pattern (presumably caused by GTs genes located outside the respective O-antigen gene clusters).[19] In the case of *E. coli* O44, the gene that encodes for the GT responsible for the lateral glucosylation of the PS was found within a putative prophage.[19] Likewise, *E. coli* O13, O129 and O135 exhibit a common O-antigen backbone (formed by the action of enzymes encoded in their almost identical O-antigen gene cluster) that in the different serogroups is decorated with distinctive *O*-acetylation and glucosylation patterns (created by enzymes that are not encoded in their O-antigen gene clusters).[20]

E. coli O42 strains have been reported to cause diarrhea in both humans and animals, and they have been classified either as EIEC or ETEC, respectively.[21–24] Strains belonging to this serogroup were reported to give positive results in PCR assays and oligonucleotide-based DNA microarrays targeting the *wzx* and *wzy* genes of *E. coli* O28ac (and vice versa), and serological cross-reactivity was also observed between strains of these two serogroups.[25,26] The partial O-antigen gene cluster of *E. coli* O42 sequenced by Fratamico *et al.*[25] consisted of: *i*) the *wzx* and *wzy* O-antigen processing genes (encoding for the flippase and polymerase, respectively), ii) genes encoding for three GTs (*wbeS*, *wbeX* and *wbeY*) and an acetyltransferase (*wbeZ*) and iii) the partial sequence of a gene encoding for a UDP-galactopyranose mutase (*glf*), which is involved in the biosynthesis of a monosaccharide precursor (Gal*f*). This partial O-antigen gene cluster is almost identical to a region of the *E. coli* O28ac O-antigen gene cluster, differing only in the substitution of a single nucleotide in the case of *wbeX* gene and two nucleotides in the case of the *wbeY* gene. These mutations are responsible for the respective differences in the amino acid sequences of the proteins (that is, one and two amino acids are substituted in the WbeX and WbeY GTs, respectively). Since just a few amino acid substitutions may be enough to change the specificity of a glycosyltransferase, thereby altering the structure of the O-antigen,[27,28] a comparison of the structures of the O-antigen PSs of *E. coli* O42 and O28ac would help to reveal whether these mutations are responsible for the slightly different serological responses observed between these serogroups or not. Herein, we investigate the structure of the repeating unit of the O-

antigen PS of *E. coli* O42, and identify the structural difference with respect to that of the *E. coli* O28ac O-antigen PS.

2. Results

The LPS of *E. coli* O42 was obtained by hot phenol/water extraction of the bacteria, which were grown in Luria-Bertani (LB) medium. The LPS was delipidated under mild acid conditions to yield the lipid-free PS, which was purified by size exclusion chromatography. The ^{1}H and ^{13}C NMR spectra of this material suggested the presence of an *O*-acetyl group (resonances indicated with black filled triangles in Fig.1a and 1c, respectively). Subsequently, the *O*-deacylated LPS (LPS-OH) was obtained upon treatment of the LPS under alkaline conditions, and purified by size exclusion chromatography. The aforementioned resonances from the *O*-acetyl group were absent in the ^{1}H and ^{13}C NMR spectra of this material (Fig. 1b and 1d, respectively). The sugar analysis of the LPS-OH revealed glucose, galactose and 2-amino-2-deoxyglucose in a relative ratio 1.2:1.0:2.5, respectively. The determination of the absolute configuration of the monosaccharide components utilized authentic standards and showed that the aforementioned residues all have the D-configuration.

The ^{1}H NMR spectrum of the LPS-OH (Fig. 1b) revealed four resonances in the anomeric region that were denoted by **A** – **D** in order of decreasing ^{1}H chemical shifts (5.329, 5.076, 5.043 and 4.701, respectively). Resonances were present, inter alia, at δ_{H} 2.103 and 2.077 (corresponding to 3H each), indicating that the two 2-amino-2-deoxyglucoses detected in sugar analysis are *N*-acetylated. This is also consistent with the resonances in the ^{13}C NMR spectrum found at δ_{C} 22.98 (CH_3), 23.46 (CH_3), 175.05 (CO) and 175.15 (CO). Furthermore, the splitting of the ^{13}C resonances at 79.67, 78.12, 75.80, 74.92 and 65.53 ppm indicated the presence of phosphorous. The ^{31}P NMR spectrum of the LPS-OH contained a single signal at δ_{P} –0.2 ppm, consistent with a phosphodiester group.[29–31]

The LPS-OH was subjected to dephosphorylation with aqueous 48% hydrogen fluoride and, after purification by gel permeation chromatography, the resulting oligosaccharide material was analyzed by mass spectrometry (MS). The HR-MS spectrum in the positive mode of the underivatized oligosaccharide showed, inter alia, a peak at *m/z* 845.3004 corresponding to a compound of molecular formula $C_{31}H_{54}N_2NaO_{23}$ (calculated value 845.3010) that was attributed to the pseudomolecular ion $[M+Na]^+$. This information, combined with the above sugar analysis of the LPS-OH, revealed that the oligosaccharide is composed of one Glc, one Gal and two GlcNAc residues, as well as one residue of formula

$C_3H_8O_3$ that may be attributed to a glycerol moiety. Information about the monosaccharide sequence in the oligosaccharide was obtained by MS/MS in both the positive and negative mode from the precursor pseudomolecular ions *m/z* 845.3 (Fig. 2a) and 821.3 (Fig. 2b), respectively, which produced the corresponding daughter ions via A_1-, B- and C-type cleavages.[32] In the case of the MS/MS spectrum recorded in positive mode, the A_1-type cleavages are only observed at the glycosidic linkages involving GlcNAc residues, which reflects the higher stability of the HexNAc oxonium ions with respect to those originating from hexoses.[33] Additionally, the fragment *m/z* 591.2 in the MS/MS spectrum of Fig. 2a could be attributed to a double cleavage process (i.e. a B-type cleavage of the fragment *m/z* 753.3) sometimes observed for underivatized carbohydrates,[32] and the fragments *m/z* 641.3 and 438.2 in the MS/MS spectrum of Fig. 2b could be formed by loss of one molecule of water from the *m/z* 659.3 and 456.2 ions, respectively.[34]

The ^{13}C NMR spectra of the LPS-OH (Fig. 1d) revealed 31 resonances, in agreement with the number of carbons of the oligosaccharide material determined by MS. The multiplicity-edited ^{1}H,^{13}C-HSQC spectrum of the LPS-OH (Fig. 3a) showed in the region for the anomeric resonances three cross-peaks corresponding to hexopyranosyl residues (residues **A**, **C** and **D**) and one cross-peak corresponding to a hexofuranosyl residue (residue **B**). In addition, six hydroxymethyl groups were identified (shown in red color in the spectrum of Fig. 3a) consistent with the presence of four hexoses and one glycerol moiety, as anticipated from the MS analysis. The ^{1}H and ^{13}C resonances of the different residues were assigned using a combination of 2D NMR experiments such as ^{1}H,^{1}H-TOCSY, ^{1}H,^{1}H-NOESY, multiplicity-edited ^{1}H,^{13}C-HSQC, ^{1}H,^{13}C-H2BC and ^{1}H,^{13}C-HMBC; the ^{1}H and ^{13}C the chemical shifts for each residue are complied in Table 1. Residue **A** and **C** are α-linked since $^{3}J_{H1,H2}$ are 3.5-3.8 Hz and $^{1}J_{C1,H1}$ are 171-172 Hz, and residue **D** is β-linked since $^{3}J_{H1,H2}$ is 8.5 Hz and $^{1}J_{C1,H1}$ is 161 Hz.[35] The correlation patterns observed in the ^{1}H,^{1}H-TOCSY spectra indicate that residues **A**, **C** and **D** have the *gluco*-configuration, since all the protons in the spin system can be traced from H1. Moreover, residues **C** and **D** have C2 chemical shifts that correspond to *N*-acetylhexosamines (53.92 and 55.53 ppm, respectively) and thus are α-D-Glc*p*NAc and β-D-Glc*p*NAc; residue **A** is then α-D-Glc*p* and residue **B** is a D-galactofuranose. The ^{1}H and ^{13}C chemical shifts of the latter are almost identical to those reported for the β-D-Gal*f* residue of the *O*-deacetylated O-antigen PS of *E. coli* O28ac, thus indicating that residue **B** is also β-linked.[29] A regular aliased ^{1}H,^{13}C-HMBC spectrum was used to assign the resonances of the carbonyl groups via two- and three-bond proton-carbon

couplings. The resonance at 175.05 ppm has a correlation to the protons at 2.103 (*N*-acetyl methyl protons) and 4.081 ppm (H2 in residue **C**) whereas the resonance at 175.15 has a correlation to the protons at 2.077 (*N*-acetyl methyl protons) and 3.818 ppm (H2 in residue **D**), confirming the *N*-acetyl groups at the C2 positions in residues **C** and **D**, respectively.

The substitution positions for the sugar residues were identified from ^{13}C NMR glycosylation shifts;[36] the ^{13}C chemical shifts of α-D-Glc*p*NAc, β-D-Glc*p*NAc, α-D-Glc*p*, β-D-Gal*f* and glycerol were used as references.[37–39] Residues **B** and **C** are both 3-substituted, with glycosylation shifts $\Delta\delta_C$ 8.42 and 6.85, respectively; thus these residues are →3)-β-D-Gal*f*-(1→ and →3)-α-D-Glc*p*NAc-(1→, respectively. Residue **D** has glycosylation shifts of 4.86 and 3.86 ppm at C3 and C4, respectively, showing that this residue is disubstituted and corresponds to →3,4)-β-D-Glc*p*NAc-(1→. Residue **A** does not show any significant ^{13}C glycosylation shift and thus it is a terminal α-D-Glc*p* (1→ group. Residue **E** (glycerol) is substituted at C2, with a ^{13}C glycosylation shift of 4.82 ppm. This residue is also 1-*O*-phosphorylated, since correlations were observed in the ^{1}H,^{31}P-HMBC spectrum (Fig. 4a) from the phosphorous atom to the H1 protons in the glycerol moiety (residue **E**) and to H4 in residue **D**. Consequently, residue **E** is →2)-Gro-(1-*P*– and it is connected to C4 of residue **D** via a phosphodiester linkage. This was corroborated by the splitting of the ^{13}C resonances of C1 and C2 in residue **E** and C3, C4 and C5 of residue **D** observed in the ^{1}H-decoupled ^{13}C NMR spectrum due to two- and three-bond phosphorous-carbon couplings ($^{2}J_{P,C1(E)}$ = 5.1 Hz and $^{3}J_{P,C2(E)}$ = 6.9 Hz, $^{3}J_{P,C3(D)}$ = 3.7 Hz, $^{2}J_{P,C4(D)}$ = 5.8 Hz and $^{3}J_{P,C5(D)}$ = 2.6 Hz). Additionally, the ^{1}H,^{31}P-hetero-TOCSY spectrum with a mixing time of 54 ms (Fig. 4b) allowed tracing of the full proton spin system of the glycerol residue (H1 to H3 in residue **E**) and the protons H1 to H5 in the spin system of residue **D**.

The sequence of the sugar residues in the O-antigen repeating unit was determined from ^{1}H,^{13}C-HMBC and ^{1}H,^{1}H-NOESY experiments. Three-bond heteronuclear correlations were observed from the anomeric carbon of each monosaccharide residue to the respective proton atoms at the substitution positions (Fig. 4c) as well as correlations from anomeric protons to the glycosyloxylated carbons (Table 1). Consequently, the structure of the repeating unit of the *O*-deacylated O-antigen PS of *E. coli* O42 is:

→2)-Gro-(1–*P*–4)[α-D-Glc*p*-(1→3)]-β-D-Glc*p*NAc-(1→3)-β-D-Gal*f*-(1→3)-α-D-Glc*p*NAc-(1→

The inter-residue correlations observed in the ^{1}H,^{1}H-NOESY spectrum (Table 1), and the sequence of monosaccharide residues determined by MS/MS (Fig. 2), are both consistent

with the proposed structure, which is remarkably similar to that of the *O*-deacetylated PS of *E. coli* O28ac:[29]

→2)-(*R*)-Gro-(1–*P*–4)-β-D-Glc*p*NAc-(1→3)-β-D-Gal*f*-(1→3)-α-D-Glc*p*NAc-(1→

Comparison of the ^{13}C chemical shifts of the →3)-α-D-Glc*p*NAc-(1→2)-Gro-(1– moiety in both polymers showed a maximum $\Delta\delta_{C(O42-O28ac)}$ deviation of 0.2 ppm for C1 in α-D-Glc*p*NAc and −0.2 ppm for C1 in glycerol that may be attributed to long range perturbations due to the presence/absence of the glucose residue in the *E. coli* O42/O28ac O-antigens, respectively. Having considered this, and the biosynthetic aspects that will be discussed below, the stereogenic center in the glycerol moiety can be anticipated to have the *R* configuration.

The ^{1}H and ^{13}C NMR spectra of the lipid-free PS showed resonances corresponding to an *O*-acetyl group: δ_C 21.06 (CH_3) and 173.41 (CO), and δ_H 2.147 (CH_3). The ^{1}H and ^{13}C chemical shifts assignments of each monosaccharide spin system were carried out using 2D ^{1}H,^{1}H-TOCSY, ^{1}H,^{1}H-NOESY, multiplicity-edited ^{1}H,^{13}C-HSQC and ^{1}H,^{13}C-HMBC experiments, and are compiled in Table 2. The *O*-acetylated residue could be readily identified by comparison of the multiplicity-edited ^{1}H,^{13}C-HSQC of the lipid-free PS with that of the LPS-OH (Fig. 3b and 3a, respectively), since only the chemical shifts of the D-Gal*f* residue (**B**) represented significant chemical shift differences that could be attributed to perturbations due to *O*-acetylation.[40,41] The exact location of the *O*-acetyl group was corroborated by an ^{1}H,^{13}C-HMBC correlation from the resonance at 173.41 ppm (CO) to the protons at 2.147 (*O*-acetyl methyl protons) and 4.861 ppm (H2 in residue **B**), confirming that the galactofuranosyl residue is 2-*O*-acetylated. Integration of the anomeric resonances of both D-Gal*f* and D-Gal*f*2Ac in the ^{1}H NMR spectrum of the lipid-free PS showed that the degree of *O*-acetylation was 80%. Thus, the structure of the teichoic acid like tetrasaccharide-glycerol repeating unit of the O-antigen PS of *E. coli* O42 is as shown in Fig. 5.

3. Discussion and conclusions

The O-antigen PSs of *E. coli* O28ac and O42 are remarkably similar, sharing the same trisaccharide-glycerol backbone (Fig. 6 top and bottom, respectively). The only difference between these two polysaccharides is that the β-D-Glc*p*NAc residue in the latter is 3-substituted with a glucose residue, which is absent in the O-antigen of *E. coli* O28ac. The

functions of the glycosyltransferases involved in the biosynthesis of these polymers were previously predicted using the respective O-antigen gene clusters[25] and BLAST similarity searches.[9] The names of the GTs anticipated to catalyze the formation of the different linkages are annotated in Figure 6, in the proximity of the respective glycosidic linkages. Since the polymerase (Wzy) was proposed to catalyze the formation of the α-D-Glc*p*NAc-(1→2)-Gro moiety, the biological repeating unit is defined by having a 3-substituted α-D-Glc*p*NAc residue at its reducing end. Furthermore, since the polymerase and the predicted glycerol phosphotransferase (WbeS)[25] of *E. coli* O28ac are identical to those of *E. coli* O42, the configuration of the stereogenic center in the glycerol residue is expected to be the same in both polysaccharides. The absolute configuration of glycerol moiety in the O-antigen PS of *E. coli* O28ac has been determined previously[29] and, thus, the same *R* configuration is expected in the O-antigen PS of *E. coli* O42. The very good agreement of the ^{13}C chemical shifts observed for the carbon resonances of the →3)-α-D-Glc*p*NAc-(1→2)-Gro-(1– moiety in both polysaccharides supported this hypothesis.

Furthermore, the WbeX and WbeY glycosyltransferases were predicted to catalyze the formation of the β-(1→3)-linkages between the D-Glc*p*NAc/D-Gal*f* and D-Gal*f*/D-Glc*p*NAc residues, respectively. The WbeX protein of *E. coli* O42 differs from that of *E. coli* O28ac just by the substitution of a single amino acid, whereas the WbeY protein differs by two amino acid substitutions. Contrary to what happened in the case or *E. coli* O9 and O9a,[27] and *E. coli* O107 and O117,[28] these differences seem not to affect the specificity of the enzymes for their substrates, catalyzing the formation of the same structural elements in both *E. coli* O42 and O28ac.

The genes responsible for the biosynthesis of 2-acetamido-2-deoxyglucopyranose, glucopyranose, galactopyranose and glycerol derivatives are considered housekeeping genes shared by other pathways and, thus, are located outside the O-antigen gene cluster.[15] The *glf* gene found in the O-antigen gene cluster of *E. coli* O28ac has been predicted to encode for an UDP-galactopyranose mutase using BLAST similarity searches; this is a protein composed of 381 amino acid residues that catalyzes the conversion of UDP-Gal*p* to UDP-Gal*f*.[25,42] A partial sequence of a *glf* gene, encoding for the first 302-amino acid fragment of the aforementioned protein, has also been identified in the O-antigen gene cluster of *E. coli* O42,[25] which is consistent with the presence of Gal*f* residues in the O-antigen polysaccharide. Based on the structural features of the O-antigen PS of *E. coli* O28ac, the WbeZ acetyltransferase has been anticipated to perform the 2-*O*-acetylation of the Gal*f* residue.[25]

This acetyltransferase is identical to that of *E. coli* O42, and a D-Gal*f*2Ac residue is indeed found in the repeating units of both of the O-antigen structures.

In the case of *E. coli* O42, the gene that encodes for the GT involved in the formation of the β-(1→3)-linkage between D-Glc*p* and D-Glc*p*NAc (branching point) was not present in the available O-antigen gene cluster reported by Fratamico *et al.*[25] One should note that whereas the whole O-antigen gene cluster between the *galF* and *gnd* genes has been sequenced for *E. coli* O28ac (GenBank: DQ462205, comprising the *wegS-wzx-wbeS-wzy-wbeX-wbeY-wbeZ-glf* genes), only a partial O-antigen gene cluster has been reported for *E. coli* O42 (GenBank: FJ539194, comprising the region from *wzx* and a truncated *glf* gene, in the same sequence order as in *E. coli* O28ac). Consequently, the gene that encodes for the *E. coli* O42 glucosyltransferase might be either in the remaining region of the O-antigen gene cluster that is not available, or located outside the O-antigen gene cluster. With the knowledge of the structure of both *E. coli* O28ac[29] and O42 O-antigens (elucidated herein), we have now identified another group of *E. coli* O-antigens sharing a common backbone structure and almost identical O-antigen gene clusters,[25] in addition to the previously identified *E. coli* serogroups O17/O44/O73/O77/O106 and O13/O129/O135.[19,20]

4. Experimental

4.1. Bacterial strain, conditions of growth and preparation of the lipopolysaccharide

E. coli O42:K-:H37, strain CCUG 11348, was obtained from Culture Collection, University of Göteborg, Sweden. The bacterium was grown in LB broth and the LPS isolated as previously described.[43]

4.2. Preparation of the *O*-deacylated lipopolysaccharide (LPS-OH)

The LPS (30.7 mg) was treated with 2 mL of NH_4OH 12.5% for 17 h at 37 °C. The solution was then neutralized with 4 M HCl, dialyzed against distilled water for 3 days and lyophilized. The crude material was purified by size exclusion chromatography on a HiLoad™ 16/60 Superdex™ 30 column (GE Healthcare), eluted with 1% 1-butanol in water at a flow rate of 1 mL·min^{-1} using an ÄKTA™ purifier system. RI and UV detection at 214 nm were used to monitor elution. The purified fractions were combined in two pools and lyophilized to yield a total of 18.5 mg of the LPS-OH.

4.3. Preparation of the lipid-free polysaccharide (PS)

The LPS (20 mg) was dissolved in 3 mL of 1% AcOH (pH 3) and stirred for 90 min at 100 °C. The lipid A was removed by centrifugation at 15000 × *g* for 20 min at 4 °C, and the supernatant was adjusted to pH 5 by adding 1 M NaOH very slowly at 0 °C and under continuous stirring. The solution was then dialyzed against distilled water and the crude material purified by size exclusion chromatography as described above to yield 5 mg of the purified PS.

4.4. Dephosphorylation of the LPS-OH

The LPS-OH (9.9 mg) was treated with 2 mL of aqueous 40% HF at –3 °C for 6 days. After evaporation of the solvent with a stream of dry air, 2 mL of H_2O were added and the solution neutralized with 1 M NH_4OH; all these steps were carried out at 0 °C. The solution was then concentrated to 0.5 mL and centrifuged; the supernatant was purified by size exclusion chromatography on a Superdex™ Peptide 10/300 GL (Tricorn™) column (GE Healthcare), using the same conditions as described above. Different fractions were collected and analyzed by mass spectrometry, and the fully dephosphorylated oligosaccharide was identified in fractions eluting at retention volumes between 14.6 and 17.6 mL.

4.5. Mass-spectrometry

The electrospray ionization high-resolution mass spectrum (ESI-HRMS) was recorded in positive mode using a MicrOTOF-Q™ mass spectrometer (Bruker Daltonics) using as solvent water:isopropanol (1:1) with 0.2% formic acid. An MS/MS spectrum in positive mode was obtained from the *m*/*z* 845.3 ion precursor (sodium adduct) and a MS/MS spectrum in negative mode was obtained from the *m*/*z* 821.3 ion precursor (M-H). Nitrogen was used as the collision gas.

4.6. Sugar analysis and absolute configuration

The LPS-OH (0.2 mg) was hydrolyzed with 2 M trifluoroacetic acid at 120 °C for 2 h. The sample was subsequently reduced with $NaBH_4$ and acetylated. The mixture of alditol acetates was analyzed by gas-liquid chromatography (GLC). The absolute configuration of D-Glc, D-Gal and D-GlcNAc were determined by GLC analysis of their acetylated (+)-2-butyl glycoside derivatives.[44]

4.7. GLC analyses

All the alditol acetates and the butyl glycoside derivatives of *N*-acetylglucosamine were separated on a PerkinElmer Elite-5 column with hydrogen as the carrier gas (25 psi) using a temperature program of 150 °C for 2 min, 3 °C min^{-1} up to 220 °C and then 10 min at 220 °C. The injector and detector temperatures were set to 220 and 250 °C, respectively. The acetylated butyl glycoside derivatives of glucose and galactose were separated on a PerkinElmer Elite-255 column with hydrogen as the carrier gas (25 psi) using a temperature program of 170 °C for 5 min, 2 °C min^{-1} up to 220 °C and then 1 min at 220 °C and, also, a temperature program of 180 °C for 5 min, 1 °C min^{-1} up to 220 °C and then 1 min at 220 °C. The injector and detector temperatures were set to 200 and 250 °C. The columns were fitted to a PerkinElmer Clarus 400 gas chromatograph equipped with flame ionization detectors. The retention times of the derivatives were compared with those of authentic reference compounds.

4.8. NMR Spectroscopy

The samples were deuterium-exchanged by freeze-drying from 99.9% D_2O and then examined by NMR spectroscopy as solutions in 99.96 % D_2O (0.55 mL) at 42 °C. Unless otherwise specified, the NMR spectra of the LPS-OH (8.6 mg) were recorded on Bruker Avance III 700 MHz spectrometer equipped with a 5 mm TCI ($^{1}H/^{13}C/^{15}N$) Z-Gradient (53.0 $G{\cdot}cm^{-1}$) CryoProbe™ and the NMR spectra of the lipid-free PS (5 mg, 75 mM buffer phosphate, pD 6) were carried out on a Bruker Avance III 600 MHz spectrometer equipped with a 5-mm inverse Z-gradient (55.7 $G{\cdot}cm^{-1}$) TXI probe ($^{1}H/^{13}C/^{31}P$). Chemical shifts are reported in ppm using internal sodium 3-trimethylsilyl-(2,2,3,3-$^{2}H_4$)-propanoate (TSP, δ_H 0.00), external 1,4-dioxane in D_2O (δ_C 67.40) or 2% H_3PO_4 in D_2O (δ_P 0.00) as references.

The 1D diffusion-filtered ^{1}H NMR spectra were recorded on a Bruker Avance III 600 MHz spectrometer using the 1D stimulated spin-echo pulse sequence with bipolar gradients and LED (ledbpgp2s1d).[45] Diffusion encoded sinusoidal gradients pulses (δ/2) of 1.8 ms and strength 50% of the maximum were used; the diffusion time was set to 100 ms. The $^{1}H,^{1}H$-TOCSY experiments[46] were obtained employing echo/antiecho gradient selection, an MLEV-17 spin-lock of 10 kHz and mixing times of 20, 40, 60 and 100 ms. The gradient selected $^{1}H,^{1}H$-NOESY experiments[47] were recorded with a mixing time of 100 ms. The multiplicity-edited $^{1}H,^{13}C$-HSQC experiments[48] were recorded employing the echo/antiecho method and the $^{1}H,^{13}C$-H2BC experiment[49] was recorded with a constant-time delay of 22 ms. The gradient selected $^{1}H,^{13}C$-HMBC experiments[50] were carried out with an evolution time of

~60 ms over a spectral region of 5.0-6.0 × 120 ppm in *F2* and *F1*, respectively. The ^{13}C offset was 60 ppm, and the correlations to the carbonyl carbons were determined from the aliased spectrum. The ^{31}P-based NMR spectra of the LPS-OH were obtained on a Bruker Avance III 600 MHz spectrometer. The $^{1}H,^{31}P$-HMBC spectrum[50,51] was recorded with an evolution time of 50 ms. The $^{1}H,^{31}P$-hetero-TOCSY experiments[52] were carried out with mixing times of 36 and 54 ms, using a DIPSI2 mixing sequence set at 3.2 kHz on both channels.

Acknowledgements

This work was supported by grants from the Swedish Research Council and the Knut and Alice Wallenberg Foundation. The research that has led to these results has received funding from the European Commission's Seventh Framework Programme FP7/2007-2013 under grant agreement no. 215536.

References

(1) Kaper, J. B.; Nataro, J. P.; Mobley, H. L. *Nat. Rev. Microbiol.* **2004**, *2*, 123–140.

(2) Nataro, J. P.; Kaper, J. B. *Clin. Microbiol. Rev.* **1998**, *11*, 142–201.

(3) Ørskov, F.; Ørskov, I. *Methods Microbiol.* **1984**, *14*, 43–112.

(4) Stenutz, R. R.; Weintraub, A.; Widmalm, G. *FEMS Microbiol. Rev.* **2006**, *30*, 382–403.

(5) Knirel, Y. A. In *Bacterial Lipopolysaccharides. Structure, Chemical Synthesis, Biogenesis and Interaction with Host Cells*; Knirel, Y. A.; Valvano, M. A., Eds.; Springer: Vienna, 2011; pp. 41–115.

(6) Zaccheus, M. V.; Broeker, N. K.; Lundborg, M.; Uetrecht, C.; Barbirz, S.; Widmalm, G. *Carbohydr. Res.* **2012**, *357*, 118–125.

(7) Urbina, F.; Nordmark, E.-L.; Yang, Z.; Weintraub, A.; Scheutz, F.; Widmalm, G. *Carbohydr. Res.* **2005**, *340*, 645–650.

(8) Rojas Macias, M. A.; Lütteke, T.; Widmalm, G. In *26 International Carbohydrate Symposium*; Madrid. poster no. P236, 2012.

(9) Lundborg, M.; Modhukur, V.; Widmalm, G. *Glycobiology* **2010**, *20*, 366–368.

(10) DebRoy, C.; Roberts, E.; Fratamico, P. M. *Anim. Heal. Res. Rev.* **2011**, *12*, 169–185.

(11) Sugiyama, T.; Kido, N.; Kato, Y.; Koide, N.; Yoshida, T.; Yokochi, T. *J. Bacteriol.* **1998**, *180*, 2775–2778.

(12) Kido, N.; Targov, V. I.; Sugiyama, T.; Uchiya, K.; Sugihara, H.; Kato, N.; Jann, K. *J. Bacteriol.* **1995**, *177*, 2178–2187.

(13) Wang, L.; Huskic, S.; Cisterne, A.; Rothemund, D.; Reeves, P. R. *J. Bacteriol.* **2002**, *184*, 2620–2625.

(14) Guo, H.; Kong, Q.; Cheng, J.; Wang, L.; Feng, L. *FEMS Microbiol. Lett.* **2005**, *248*, 153–161.

(15) Samuel, G.; Reeves, P. *Carbohydr. Res.* **2003**, *338*, 2503–2519.

(16) Raetz, C. R. H.; Whitfield, C. *Annu. Rev. Biochem.* **2002**, *71*, 635–700.

(17) Valvano, M. A.; Furlong, S. E.; Patel, K. B. In *Bacterial lipopolysaccharides. Structure, chemical synthesis, biogenesis and interaction with host cells*; Knirel, Y. A.; Valvano, M. A., Eds.; Springer Vienna, 2011; pp. 275–310.

(18) Wang, L.; Wang, Q.; Reeves, P. R. In *Endotoxins: structure, function and recognition*; Wang, X.; Quinn, P. J., Eds.; Springer: New York, 2010; pp. 123–152.

(19) Wang, W.; Perepelov, A. V.; Feng, L.; Shevelev, S. D.; Wang, Q.; Senchenkova, S. N.; Han, W.; Li, Y.; Shashkov, A. S.; Knirel, Y. A.; Reeves, P. R.; Wang, L. *Microbiology* **2007**, *153*, 2159–2167.

(20) Perepelov, A. V; Shevelev, S. D.; Liu, B.; Senchenkova, S. N.; Shashkov, A. S.; Feng, L.; Knirel, Y. A.; Wang, L. *Carbohydr. Res.* **2010**, *345*, 1594–1599.

(21) Constantiniu, S.; Buzdugan, I.; Romaniuc, A.; Cozma, L.; Filimon, R.; Dumbrava, M.; Nistor, A.; Teodoru, L.; Onu, P.; Danis, G. *J. Prev. Med.* **2001**, *9*, 67–73.

(22) Wasteson, Y.; Olsvik, Ø.; Skancke, E.; Bopp, C. A.; Fossum, K. *J. Clin. Microbiol.* **1988**, *26*, 2564–2566.

(23) Prada, J.; Baljer, G.; De Rycke, J.; Steinrück, H.; Zimmermann, S.; Stephan, R.; Beutin, L. *Vet. Microbiol.* **1991**, *29*, 59–73.

(24) Holland, R. E.; Walker, R. D.; Sriranganathan, N.; Wilson, R. A.; Ruhl, D. C. *Vet. Microbiol.* **1999**, *70*, 261–268.

(25) Fratamico, P. M.; Yan, X.; Liu, Y.; Debroy, C.; Byrne, B.; Monaghan, A.; Fanning, S.; Bolton, D. *Can. J. Microbiol.* **2010**, *56*, 308–316.

(26) Geue, L.; Monecke, S.; Engelmann, I.; Braun, S.; Slickers, P.; Ehricht, R. *Microbiol. Immunol.* **2014**, *58*, 77–86.

(27) Kido, N.; Kobayashi, H. *J. Bacteriol.* **2000**, *182*, 2567–2573.

(28) Wang, Q.; Perepelov, A. V.; Feng, L.; Knirel, Y. A.; Li, Y.; Wang, L. *FEMS Immunol. Med. Microbiol.* **2009**, *55*, 47–54.

(29) Rundlöf, T.; Weintraub, A.; Widmalm, G. *Carbohydr. Res.* **1996**, *291*, 127–139.

(30) Linnerborg, M.; Weintraub, A.; Widmalm, G. *Carbohydr. Res.* **1999**, *320*, 200–208.

(31) Landersjö, C.; Weintraub, A.; Widmalm, G. *Eur. J. Biochem.* **2001**, *268*, 2239–2245.

(32) Dell, A. In *Advances in Carbohydrate Chemistry and Biochemistry*; Tipson, R. S.; Horton, D., Eds.; Academic Press, 1987; pp. 19–72.

(33) Dell, A.; Jang-Lee, J.; Pang, P.; Parry, S.; Sutton-Smith, M.; Tissot, B.; Morris, H. R.; Panico, M.; Haslam, S. M. In *Glycoscience*; Fraser-Reid, B.; Tatsuta, K.; Thiem, J., Eds.; Springer-Verlag: Berlin, 2008; pp. 2193–2216.

(34) Pfenninger, A.; Karas, M.; Finke, B.; Stahl, B. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 1331–1340.

(35) Bundle, D. R.; Lemieux, R. U. *Methods Carbohydr. Chem.* **1976**, *7*, 79–86.

(36) Söderman, P.; Jansson, P.-E.; Widmalm, G. *J. Chem. Soc. Perkin Trans. 2* **1998**, 639–648.

(37) Jansson, P.-E.; Kenne, L.; Widmalm, G. *Carbohydr. Res.* **1989**, *188*, 169–191.

(38) Baumann, H.; Tzianabos, A. O.; Brisson, J. R.; Kasper, D. L.; Jennings, H. J. *Biochemistry* **1992**, *31*, 4081–4089.

(39) De Boer, W. R.; Kruyssen, F. J.; Wouters, J. T. M.; Kruk, C. *Eur. J. Biochem.* **1976**, *62*, 1–6.

(40) Jansson, P.-E.; Kenne, L.; Schweda, E. *J. Chem. Soc. Perkin Trans. 1* **1987**, 377–383.

(41) Rönnols, J.; Pendrill, R.; Fontana, C.; Hamark, C.; Angles d'Ortoli, T.; Engström, O.; Ståhle, J.; Zaccheus, M. V.; Säwén, E.; Hahn, L. E.; Iqbal, S.; Widmalm, G. *Carbohydr. Res.* **2013**, *380*, 156–166.

(42) Nassau, P. M.; Martin, S. L.; Brown, R. E.; Weston, A.; Monsey, D.; McNeil, M. R.; Duncan, K. *J. Bacteriol.* **1996**, *178*, 1047–1052.

(43) Svensson, M. V.; Weintraub, A.; Widmalm, G. *Carbohydr. Res.* **2011**, *346*, 449–453.

(44) Leontein, K.; Lönngren, J. *Methods Carbohydr. Chem.* **1993**, *9*, 87–89.

(45) Wu, D.; Chen, A.; Johnson, C. S. *J. Magn. Reson. Ser. A* **1995**, *115*, 260–264.

(46) Bax, A.; Davis, D. G. *J. Magn. Reson.* **1985**, *65*, 355–360.

(47) Kumar, A.; Ernst, R. R.; Wüthrich, K. *Biochem. Biophys. Res. Commun.* **1980**, *95*, 1–6.

(48) Schleucher, J.; Schwendinger, M.; Sattler, M.; Schmidt, P.; Schedletzky, O.; Glaser, S. J.; Sørensen, O. W.; Griesinger, C. *J. Biomol. NMR* **1994**, *4*, 301–306.

(49) Nyberg, N. T.; Duus, J. Ø.; Sørensen, O. W. *J. Am. Chem. Soc.* **2005**, *127*, 6154–6155.

(50) Bax, A.; Summers, M. F. *J. Am. Chem. Soc.* **1986**, *108*, 2093–2094.

(51) Zartler, E. R.; Martin, G. E. *J. Biomol. NMR* **2011**, *51*, 357–367.

(52) Kellogg, G. W. *J. Magn. Reson.* **1992**, *98*, 176–182.

Figures

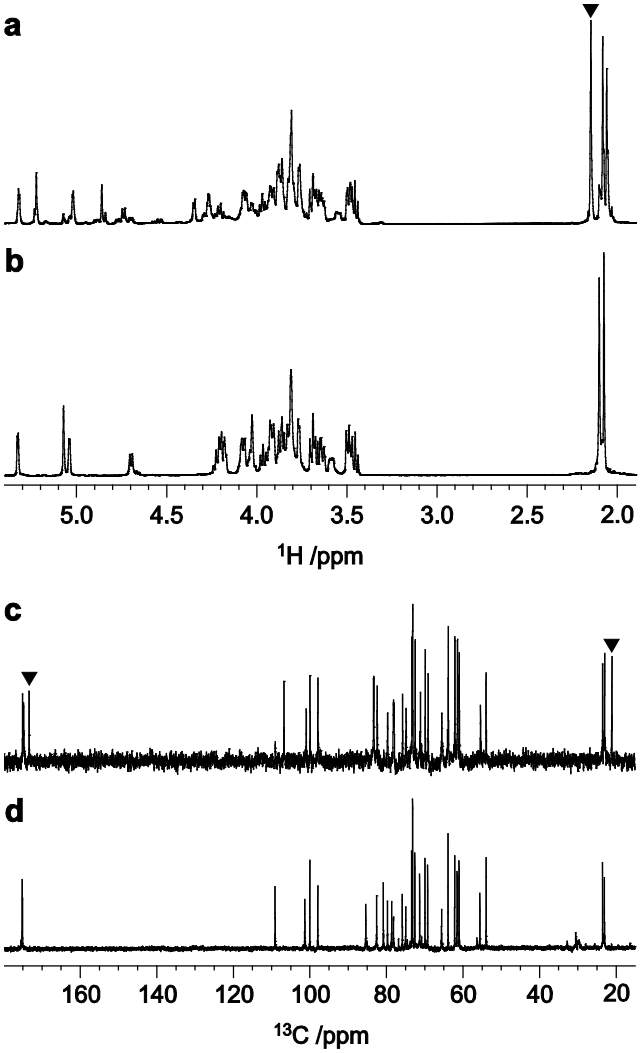


Figure 1. Diffusion filtered ^{1}H NMR and ^{13}C NMR spectra of the lipid-free PS (**a** and **c**, respectively) and *O*-deacylated LPS (**b** and **d**, respectively) from *Escherichia coli* O42. Resonances from the *O*-acetyl groups are indicated with black filled triangles.

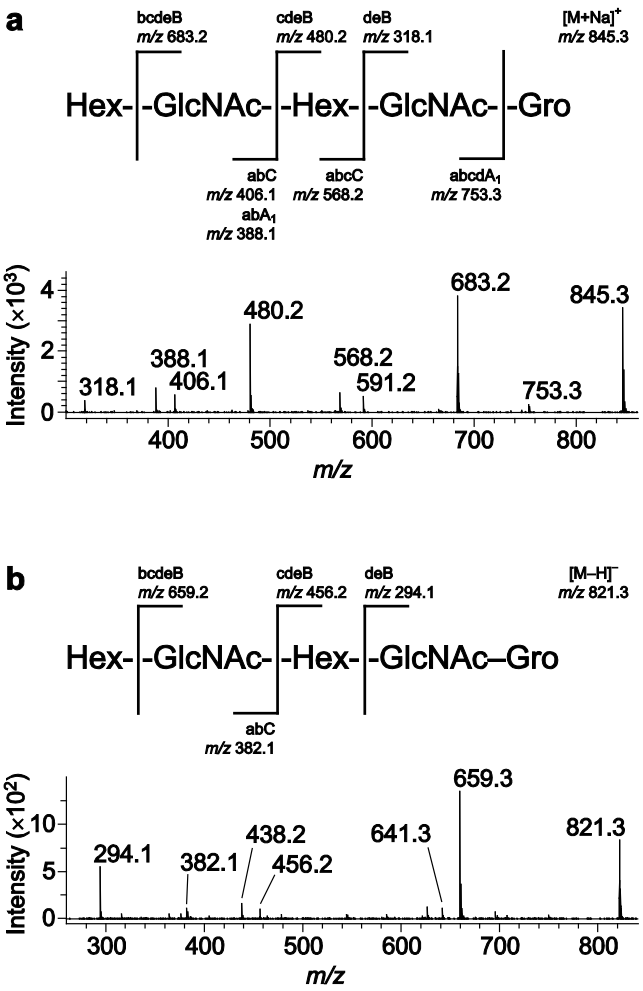


Figure 2. The MS/MS spectra of the: **a)** pseudomolecular $[M+Na]^+$ ion *m/z* 845.3 and **b)** pseudomolecular $[M–H]^–$ ion *m/z* 821.3 recorded in positive and negative mode, respectively. The detected ions and the resulting sugar sequence of the tetrasaccharide-glycerol compound, obtained after cleavage of the phosphodiester group in the LPS-OH, are shown on the top of each spectrum.

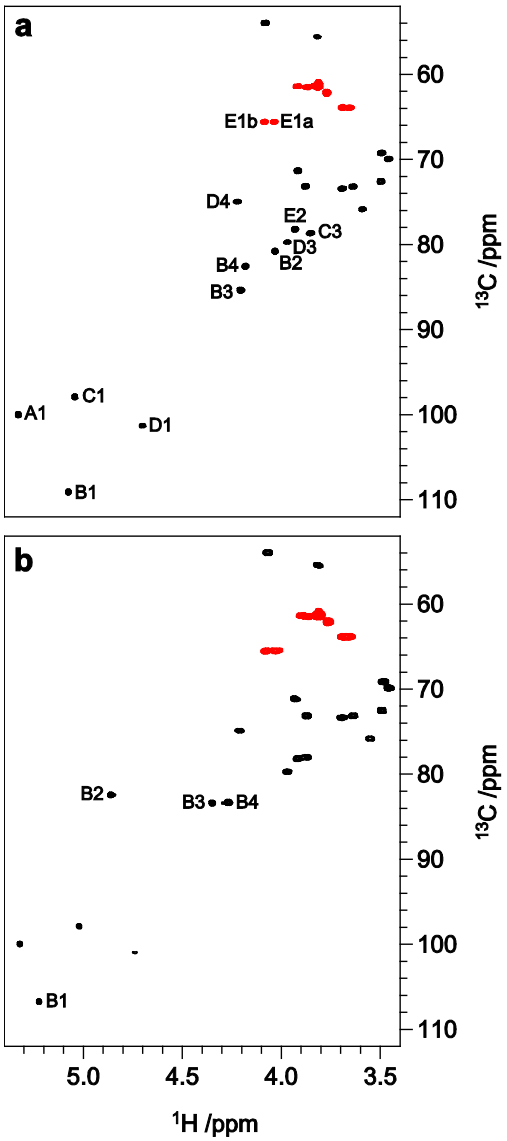


Figure 3. Selected region of the multiplicity-edited ^{1}H,^{13}C-HSQC spectrum of *O*-deacylated LPS (**a**) and lipid free PS (**b**) from *Escherichia coli* O42, showing the anomeric and ring atom regions, as well as the hydroxymethyl groups (in which the cross-peaks from the latter appear in red). Substitution positions and cross-peaks with significant chemical shifts changes upon *O*-deacetylation (B1, B2, B3 and B4) are annotated.

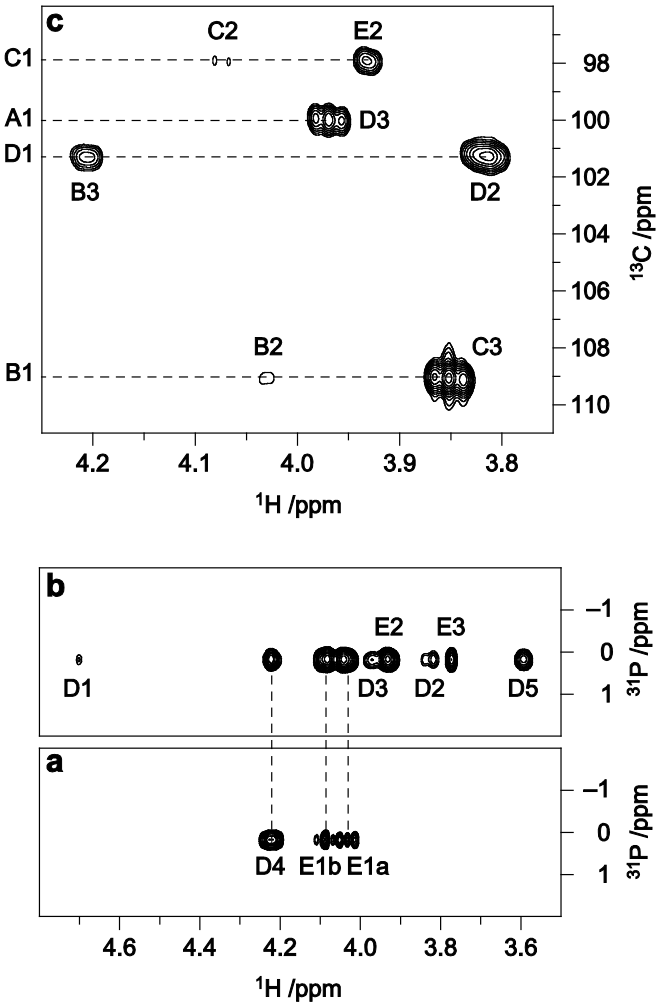


Figure 4. The $^{1}H,^{31}P$-HMBC (**a**) and $^{1}H,^{31}P$-hetero-TOCSY (τ_{mix} = 54 ms) (**b**) spectra of the *O*-deacylated LPS from *Escherichia coli* O42 showing correlations from the phosphorous atom. Selected region of the $^{1}H,^{13}C$-HMBC spectrum (**c**) showing intra- and inter- residue correlations from anomeric carbons.

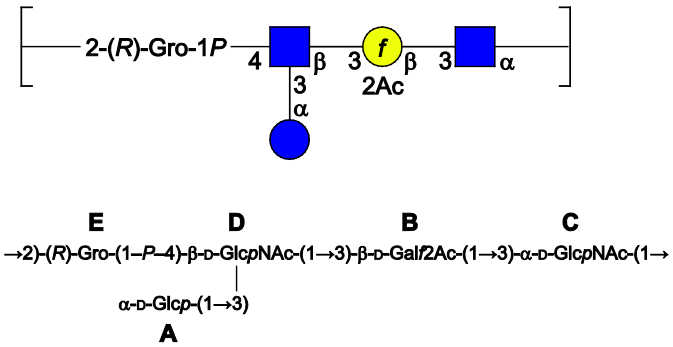


Figure 5. Structure of the biological repeating unit of the O-antigen LPS from *Escherichia coli* O42 in CFG-notation (*top*) and standard nomenclature (*bottom*). The 2-*O*-acetylation of residue **B** in the lipid-free PS was ~80%.

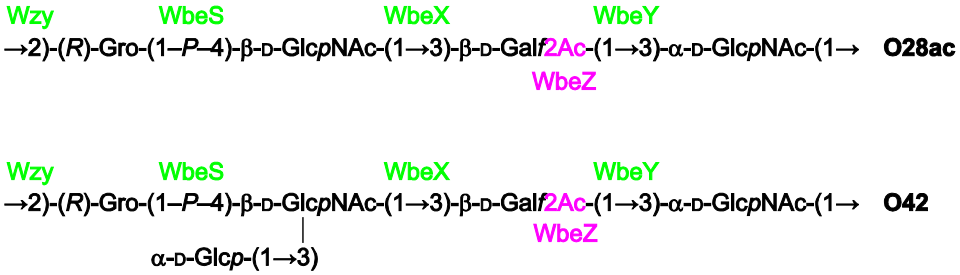


Figure 6. Comparison of the structures of the biological repeating unit of the O-antigen PSs of *E. coli* O28ac and O42. The glycosyltransferases (green color) involved in the biosynthesis of the different structural elements, and the *O*-acetyltransferase (magenta color), are annotated in the proximity of the respective linkages. With exception of WbeY and WbeX, all the enzymes are identical for both *E. coli* O42 and O28ac serogroups (the WbeY proteins differ by two amino acids whereas the WbeX proteins differ by just a single amino acid).