**Glycobiology** vol. 24 no. 5 pp. 450–457, 2014
doi:10.1093/glycob/cwu011
Advance Access publication on February 19, 2014

# Rapid structural elucidation of polysaccharides employing predicted functions of glycosyltransferases and NMR data: Application to the O-antigen of *Escherichia coli* O59

**Carolina Fontana[2], Magnus Lundborg[2,4], Andrej Weintraub[3], and Göran Widmalm[1,2]**

[2]Department of Organic Chemistry, Arrhenius Laboratory, Stockholm University, S-106 91 Stockholm, Sweden and [3]Department of Laboratory Medicine, Division of Clinical Microbiology, Karolinska Institute, Karolinska University Hospital, S-141 86 Stockholm, Sweden

*Received on January 20, 2014; revised on January 20, 2014; accepted on February 7, 2014*

**A computerized method that uses predicted functions of glycosyltransferases (GTs) in conjunction with unassigned NMR data has been developed for the structural elucidation of bacterial polysaccharides (PSs). In this approach, information about the action of GTs (consisting of possible sugar residues used as donors and/or acceptors, as well as the anomeric configuration and/or substitution position in the respective glycosidic linkages) is extracted from the *Escherichia coli* O-antigen database and is submitted, together with the unassigned NMR data, to the CASPER program. This time saving methodology, which alleviates the need for chemical analysis, was successfully implemented in the structural elucidation of the O-antigen PS of *E. coli* O59. The repeating unit of the O-specific chain was determined using the O-deacylated PS and has a branched structure, namely, →6)[α-D-Gal*p*A3Ac/4Ac-(1 → 3)]-α-D-Man*p*-(1 → 3)-α-D-Man*p*-(1 → 3)-β-D-Man*p*-(1 → 3)-α-D-Glc*p*NAc-(1→. The identification of the O-acetylation positions was efficiently performed by comparison of the $^{1}H$,$^{13}C$ HSQC NMR spectra of the O-deacylated lipopolysaccharide and the lipid-free PS in conjunction with chemical shift predictions made by the CASPER program. The side-chain D-Gal*p*A residue carries one equivalent of O-acetyl groups at the O-3 and O-4 positions distributed in the LPS in a 3:7 ratio, respectively. The presence of O-acetyl groups in the repeating unit of the *E. coli* O59 PS is consistent with the previously proposed acetyltransferase WclD in the O-antigen gene cluster.**

*Keywords:* CASPER / ECODAB / *Escherichia coli* / glycosyltransferases / O-antigen polysaccharide

## Introduction

The O-antigen polysaccharide (PS) is the most variable portion of the lipopolysaccharide (LPS), giving rise to a large number of serotypes among Gram-negative bacteria. It usually consists of oligosaccharide repeating units, and the genes encoding for the enzymes responsible for its biosynthesis are commonly located in a single cluster in the bacterial chromosome, between the *galF* and *gnd* genes; in some cases, additional genes can also be found in the region between the *gnd* gene and *his* operon. In this regard, three major classes of proteins can be recognized: (i) enzymes required for the biosynthesis of sugar precursors, (ii) glycosyltransferases (GTs) and (iii) O-antigen processing proteins (e.g. the flippase (Wzx), polymerase (Wzy) and copolymerase (Wzz)) (Raetz and Whitfield 2002; Samuel and Reeves 2003; Valvano et al. 2011). *Escherichia coli* strains can use either the Wzy/Wzx-dependent or the ABC-transporter pathway to assemble the O-specific chain, giving rise to hetero- or homopolysaccharides, respectively, with the former being the most widely distributed among these strains. In the Wzx/Wzy-dependent pathway the biosynthesis of the O-antigen PS initiates in the cytoplasmic side of the cell membrane, where the WecA enzyme catalyzes the formation of an undecaprenyl pyrophosphate (Und-*PP*) derivative of GlcNAc. As a consequence of the specificity of the WecA enzyme for GlcNAc, most of the PSs generated using this pathway have this monosaccharide residue located at the reducing end of their biological repeating units. However, in some cases, the GlcNAc residue can be epimerized to GalNAc by the action of a GlcNAc-*P*-*P*-Und epimerase, giving rise to PSs with biological repeating units having a GalNAc residue located at their reducing ends instead of a GlcNAc residue (Rush et al. 2010). In the subsequent steps, the oligosaccharide precursor of the biological repeating unit is built by the sequential action of a set of GTs, which are highly specific to sugar donors, sugar acceptors and the linkages between them. The oligosaccharide linked to Und-*PP* is then translocated to the periplasmic face of the cytoplasmic membrane by the Wzx flippase and further polymerized by the Wzy polymerase; the length of the polymer is regulated by the Wzz copolymerase (Raetz and Whitfield 2002; Caroff and Karibian 2003; Valvano et al. 2011).

The *Escherichia coli* O-antigen database (ECODAB) (Stenutz et al. 2006) has recently been expanded to include genetic information about the GTs involved in the biosynthesis of the *E. coli* O-antigen PSs (Lundborg et al. 2010). The functions of many of them were predicted using BLAST similarity

[1]To whom correspondence should be addressed: Tel: +46-8-16-37-42; Fax: +46-8-15-49-08; e-mail: gw@organ.su.se
[4]Present address: Department of Theoretical Physics, Royal Institute of Technology and Science for Life Laboratory, PO Box 1031, S-171 21 Solna, Sweden.

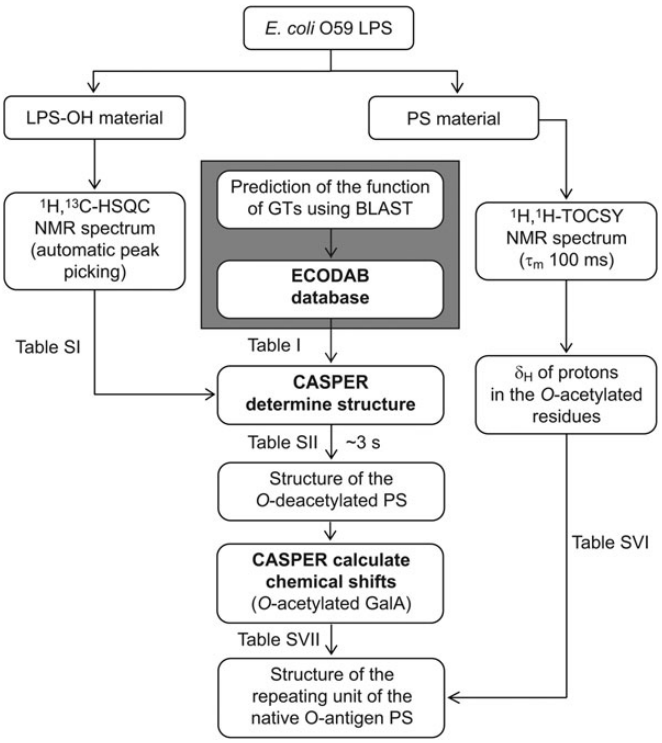


**Fig. 1.** Schematic representation of the methodology used to determine the structure of the O-antigen PS of *E. coli* O59.

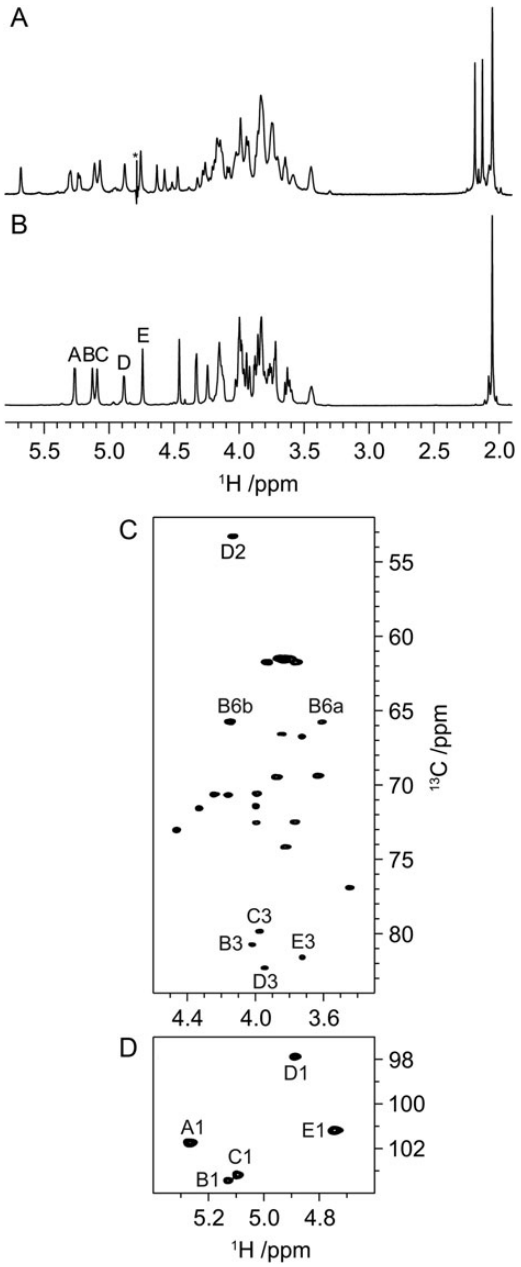


**Fig. 2.** Diffusion-filtered $^1$H NMR spectra of (**A**) the lipid-free native PS and (**B**) the LPS-OH of *E. coli* O59. The residual signal from the solvent is denoted by an asterisk and the anomeric protons are annotated by A–E in order of decreasing $^1$H chemical shifts. (**C** and **D**) Regions of the $^1$H,$^{13}$C HSQC spectrum of the LPS-OH showing the signals for the ring atoms and those of the hydroxymethyl groups (**C**) and the anomeric resonances (**D**). Resonances from anomeric and substitution positions, as well as a nitrogen-bearing carbon, are annotated.

searches (Altschul et al. 1990), and this information was used to suggest structural elements of the repeating units of the O-antigen PSs of unknown structure (Lundborg et al. 2010). These predictions have later been shown to be quite successful, once the structure of the O-antigen PS *E. coli* O177 was determined (Svensson et al. 2011a). Herein, we report a methodology for the structural elucidation of bacterial O-antigen PSs that employs the computer program CASPER (http://www.casper.organ.su.se/casper/) (Lundborg et al. 2011; Lundborg and Widmalm 2011) and combines biosynthetic information available in the ECODAB (http://www.casper.organ.su.se/ECODAB/) with unassigned experimental NMR data. The automated procedure consists of three steps: (i) identification of the corresponding serogroup in the ECODAB, (ii) submission of the possible structural elements as restrictions to the CASPER program (this is carried out automatically by clicking in a link available in the ECODAB) and (iii) input of unassigned experimental NMR data to the CASPER program. When using GT information to determine the structural elements in the "determine structure" module of the CASPER program, the information in the ECODAB is used to provide CASPER input, transparently to the user. The structural input can thus consist of possible carbohydrate residues, their linkage positions and their linkage orders, but in many cases there are several possible residues and linkage positions, related to each GT, that need to be tested by CASPER. The software evaluates only those structures that are consistent with the biosynthetic information and returns a list of possible candidates ranked according to the deviation between predicted and experimental $^1$H and $^{13}$C chemical shifts. This newly developed methodology has successfully been implemented in the elucidation of the hitherto unknown structure of the O-antigen PS of *E. coli* O59 (Ørskov and Ørskov 1984), and the procedure is outlined in Figure 1.

**Table I.** Information of the GTs involved in the biosynthesis of the O-antigen PS of *E. coli* O59 available in ECODAB (the *E. coli* O-antigen database)

| GT name | Relative position in the gene cluster | Linkage number | Donor[a] | Acceptor[a] | Similar to |
|---|---|---|---|---|---|
| WclA | 1 | 4 | α-D-Gal(1→<br>α-D-GalNAc(1→<br>α-D-GlcNAc(1→ | →3)-?-D-Man | O22 WbhP (ECODAB)<br>O148 WbbG (ECODAB)<br>O87 WfgO (ECODAB)<br>O146 WbwU (ECODAB) |
| WclB | 2 | 2, 3[b] | α-D-Man(1→ | →?)-?-D-Man | O9 MtfB (ECODAB)<br>O141 WfcC (ECODAB)<br>O9 MtfA (ECODAB)<br>O7:K1 WbbC (ECODAB)<br>O8 Orf469 (ECODAB) |
| WclC | 3 | 1 | β-D-Man(1→ | →3)-?-D-GlcNAc | O6 Orf7 (ECODAB)<br>O73 WbaD (ECODAB)<br>O77 WbaD (ECODAB)<br>O106 WbaD (ECODAB) |

The functions of the GTs involved in the assembly of the O-antigen PS of *E. coli* O59 were previously proposed using similarity searches on each of the three GTs identified in the O-antigen gene cluster (Guo et al. 2005; Lundborg et al. 2010).
[a]In order to contemplate the possibility of promiscuity of some GTs, the software will also consider the uronic acid of the corresponding monosaccharides listed as possible donor–acceptor.
[b]Note that this GT has been added twice, as it was observed that the PS was composed of pentasaccharide repeating units (five anomeric resonances were observed in the $^1$H NMR spectrum) and, on the other hand, the WclB GT shares a significant similarity with the MtfB and MtfA GTs of *E. coli* O9 that each have been proposed to catalyze the formation of two consecutive glycosidic linkages (Kido et al. 1995; Greenfield et al. 2012; Greenfield and Whitfield 2012).

## Results and discussion

The $^1$H NMR spectrum of the lipid-free PS (Figure 2A) revealed a complex material with two resonances corresponding to O-acetyl groups (at 2.128 and 2.185 ppm) that were absent in the $^1$H NMR spectrum of the O-deacylated lipopolysaccharide (LPS-OH) (Figure 2B). Five resonances corresponding to anomeric protons were observed between $\delta_H$ 5.264 and 4.742 in the $^1$H NMR spectrum of the LPS-OH preparation (denoted **A**–**E** in order of decreasing $^1$H chemical shifts), indicating that the PS is composed of pentasaccharide repeating units. On the other hand, three open reading frames (ORFs) encoding for putative GTs (WclA, WclB and WclC) were identified in the O-antigen gene cluster of *E. coli* O59 (Guo et al. 2005), and their functions were predicted using BLAST similarity searches (Table I) (Lundborg et al. 2010). The WclB GT was proposed to catalyze the formation of two consecutive glycosidic linkages due to its significant similarity with the MtfB mannosyltransferase of *E. coli* O9 (33% identity, *E*-value = $8 \times 10^{-36}$) (Kido et al. 1995; Greenfield et al. 2012; Greenfield and Whitfield 2012).

Unassigned experimental data from a decoupled $^1$H,$^{13}$C HSQC NMR spectrum of the LPS-OH (Figure 2C and D and Supplementary data, Table SI) was used as input information to the CASPER program, as well as the magnitude of the coupling constants obtained from the anomeric resonances in the $^1$H NMR spectrum (three small and two medium). The calculation (~3 s) returned a list of possible structures with a small score difference between the first and the second, but a larger difference with respect to that ranked in the third position (Figure 3 and Supplementary data, Table SII and Figure S1); as a consequence, only the two top-ranked structures were considered as possible candidates.

An alternative approach, employing a methodology previously developed in our laboratory (Lundborg et al. 2011), was used to validate these results and it is outlined in the Supplementary data, Figure S2. In this approach, the pre-hydrolyzed O-deacetylated PS of *E. coli* O59 was subjected to (+)-2-butanolysis, and the resulting mixture was analyzed by NMR spectroscopy without further purification. Both $^1$H,$^{13}$C HSQC and $^{13}$C NMR spectra were recorded, and the chemical shifts were extracted using automatic peak picking (cf. Supplementary data, Table SIII). The data were then analyzed using the CASPER program (calculation time ~1 s), which returned a list of three possible components: D-Man, D-GlcNAc and D-GalA (cf. Supplementary data, Table SIV). These results were used as input information to the "determine structure" module of the CASPER program, in addition to the NMR data from the LPS-OH. The two top-ranked structures suggested by the CASPER program using this alternative approach were identical to those obtained in the approach that utilized GT information (Supplementary data, Table SV and Figure S3).

In both approaches, the score difference between the structures ranked in the first and second positions could not be improved using the additional NMR data from the Supplementary data, Table SI, as input information, but the top-ranked structure (Figure 3, top) was confirmed as the correct one by full characterization of the spin system of the 3,6-disubstituted mannosyl residue (residue **B**) using intra-correlations from a 2D $^1$H,$^1$H TOCSY experiment (see Supplementary data, Figure S4A–E) and inter-residue correlations from an $^1$H,$^{13}$C HMBC (Supplementary data, Figure S4F and G). One should note that the three top-ranked structures (Figure 3) constitute positional isomers in which only one of the three mannosyl residues is 6-substituted. In the case of the $^1$H,$^1$H TOCSY spectra, the current version of the CASPER program only evaluates correlations from anomeric protons, and assignment of the H3–H6 resonances of mannosyl residues can be hampered when several of these residues are present in the repeating unit. However, the H6 resonances in the 6-substituted mannosyl residue were readily identified using correlations form the respective H2 proton (cf. Supplementary data,

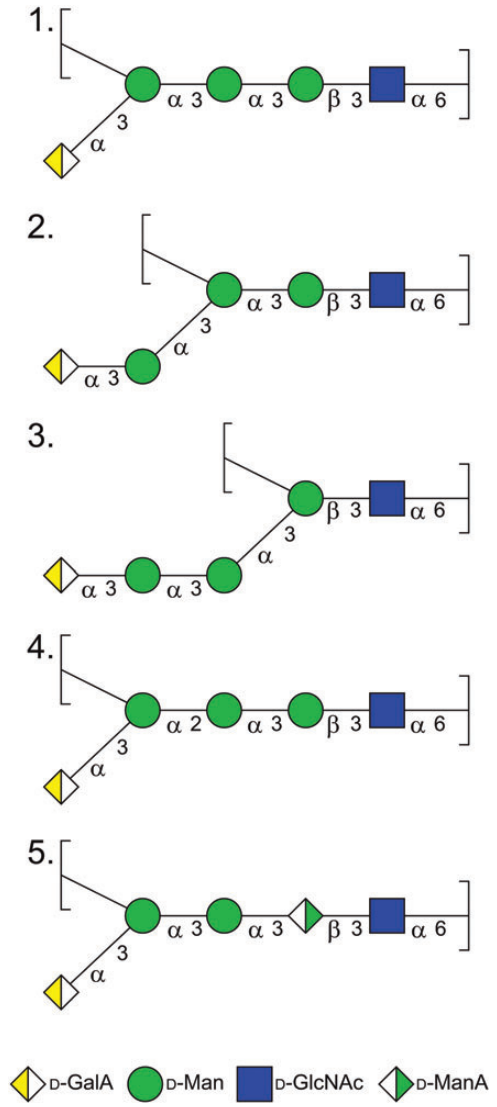


**Fig. 3.** CASPER output of the five top-ranked structures, in CFG format, for the O-deacetylated repeating unit of the O-antigen PS of *E. coli* O59. The relative deviations for structures 1–5 are 1.00, 1.02, 1.13 1.14 and 1.16, respectively. For standard carbohydrate listing of the ten top-ranked structures see Supplementary data, Table SII.

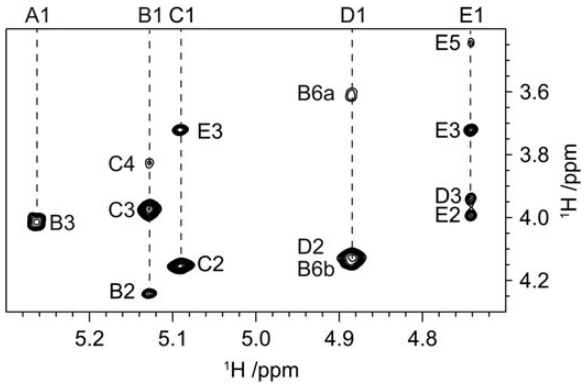


**Fig. 4.** Selected region of the $^{1}H,^{1}H$ NOESY spectrum ($\tau_m$ = 100 ms) of the LPS-OH of *E. coli* O59 showing inter- and intra-residue correlations from anomeric protons.

Figure S4E). An additional $^{1}H,^{1}H$ NOESY spectrum (Figure 4), which had not been used as input information in the structural analysis carried out by the CASPER program, was employed to confirm the inter-residue correlations and the sequence of the monosaccharides components in the repeating unit. The $^{1}H$ and $^{13}C$ chemical shifts of the biological repeating unit of the O-deacetylated O-antigen PS of *E. coli* O59 were straightforwardly assigned using the predictions made by the CASPER program for the top-ranked structure, and are compiled in Table II. The very good agreements between experimental and CASPER predicted $^{1}H$ and $^{13}C$ NMR chemical shifts are illustrated in the plots of Figure 5. One should note that the $^{1}H$ and $^{13}C$ resonances of carbohydrates are usually poorly dispersed, and these plots are only concentrated over a spectral region spanning 2.0 ppm in the case of $\delta_H$ and 60 ppm in the case of $\delta_C$.

$^{1}H$ chemical shifts of minor spin systems observed in the $^{1}H,^{1}H$ TOCSY spectra and attributed to the trisaccharide moiety located at the terminal non-reducing end of the O-antigen PS (i.e. this is an α-D-Gal*p*A-(1 → 3)-α-D-Man*p*-(1 → 3)-α-D-Man*p*-(1→ moiety) were assigned employing predictions made using the program CASPER and are compiled in Table III. Identification of the O-acetylated residues in the native PS was accomplished by comparison of the $^{1}H,^{13}C$ HSQC spectrum of the lipid-free PS with that of the LPS-OH. Only those correlations from the D-Gal*p*A residues represented significant chemical shift differences that could be attributed to perturbations due to O-acetylation (Jansson et al. 1987). The $^{1}H$ chemical shifts of the two populations of O-acetylated galacturonosyl residues were obtained from a $^{1}H,^{1}H$ TOCSY spectrum (Figure 6A and Supplemenaty data, Table SVI) and, based on CASPER simulations (Supplementary data, Table SVII), they were identified as the 3- and 4-O-acetylated derivatives. The $^{1}H$ and $^{13}C$ resonances of the O-acetylated PS could readily be assigned using the predictions made by CASPER and they are compiled in Table IV. The $^{13}C$ chemical shifts of the carbonyl resonances from N- and O-acetyl groups were extracted from a band-selective constant-time $^{1}H,^{13}C$ HMBC spectrum (Figure 6B). Integration of key resonances in the $^{1}H$ NMR spectrum of the native LPS showed that the ratio of D-Gal*p*A3Ac and D-Gal*p*A4Ac was 30 and 70%, respectively. This ratio was slightly different in the lipid-free PS preparation (50% each) and was attributed to migration of the O-acetyl groups during preparation and/or purification of the material (Roslund et al. 2008; Rönnols et al. 2013). The structure of the biological repeating unit of the O-antigen PS of *E. coli* O59 is shown in Figure 7, bottom, and the names of the GTs predicted to catalyze the formation of the different linkages are annotated in Figure 7, middle, in the proximity of the line representing the respective glycosidic linkages.

In addition to the ORFs that encode for three putative GTs (*wclA*, *wclB* and *wclC*) and the polymerase (*wzy*), two genes that encode enzymes involved in the biosynthesis of the D-Man residue (*manC* and *manB*) were previously identified in the region between the *galF* and *gnd* genes (Figure 7, top). Furthermore, the region between the *gnd* gene and *hisI* operon

**Table II.** $^{1}$H and $^{13}$C NMR chemical shifts (ppm) at 60°C of the resonances of the LPS-OH from *E. coli* O59 and inter-residue correlations from $^{1}$H,$^{1}$H NOESY and $^{1}$H,$^{13}$C HMBC spectra

| Sugar residue | | $^{1}$H/$^{13}$C | | | | | | | | Correlation to atom (from anomeric atom) | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | Me | CO | NOE | HMBC |
| α-D-Gal*p*A-(1→ | **A** | 5.264 [3.7] | 3.870 | 3.987 | 4.327 | 4.460 | | | | H3, **B** | C3, **B** |
| | | (–0.04) | (0.04) | (0.07) | (0.04) | (0.07) | | | | | |
| | | 101.73 {172} | 69.46 | 70.56 | 71.58 | 73.06 | 176.19 | | | | H3, **B** |
| | | (8.66) | (0.43) | (0.30) | (–0.06) | (0.76) | (–0.24) | | | | |
| →3,6)-α-D-Man*p*-(1→ | **B** | 5.128 [n.r.][a] | 4.242 | 4.014 | 4.147 | 3.997 | 3.608, 4.145 | | | H3, **C** | C3, **C** |
| | | (–0.05) | (0.30) | (0.15) | (0.47) | (0.18) | | | | | |
| | | 103.47 {171} | 70.62 | 80.72 | 65.69 | 72.53 | 65.74 | | | | H3, **C** |
| | | (8.53) | (–1.07) | (9.47) | (–2.25) | (–0.81) | (3.75) | | | | |
| →3)-α-D-Man*p*-(1→ | **C** | 5.092 [n.r.][a] | 4.156 | 3.975 | 3.839 | 3.826 | ~3.848 | | | H3, **E** | C3, **E** |
| | | (–0.09) | (0.22) | (0.12) | (0.16) | (0.01) | | | | | |
| | | 103.19 {171} | 70.66 | 79.85 | 66.59 | 74.16 | 61.60 | | | | H3, **E** |
| | | (8.25) | (–1.03) | (8.60) | (–1.35) | (0.82) | (–0.39) | | | | |
| →3)-α-D-Glc*p*NAc-(1→ | **D** | 4.885 [3.4] | 4.129 | 3.943 | 3.628 | 3.765 | ~3.839 | 2.052 | | H6a, **B** | C6, **B** |
| | | (–0.33) | (0.25) | (0.19) | (0.14) | (–0.09) | | (–0.01) | | H6b, **B** | |
| | | 97.86 {171} | 53.27 | 82.33 | 69.38 | 72.53 | 61.45 | 22.90 | 174.97 | | H6b, **B** |
| | | (6.09) | (–1.73) | (10.59) | (–1.88) | (0.02) | (–0.33) | (0.03) | (–0.16) | | |
| →3)-β-D-Man*p*-(1→ | **E** | 4.742 [n.r.][a] | 3.994 | 3.720 | 3.721 | 3.443 | 3.757, 3.929 | | | H3, **D** | C3, **D** |
| | | (–0.15) | (0.04) | (0.06) | (0.12) | (0.06) | | | | | |
| | | 101.18 {160} | 71.44 | 81.61 | 66.78 | 76.94 | 61.76 | | | | H3, **D** |
| | | (6.63) | (–0.69) | (7.58) | (–0.91) | (–0.06) | | | | | |

$^{3}J_{H1,H2}$ values are given in Hertz in square brackets and $^{1}J_{H1,C1}$ values in braces. Chemical shift differences when compared with corresponding monosaccharides (Jansson et al. 1989) are given in parentheses.
[a]Peaks not resolved (n.r.), full width at half maximum are 7.8, 8.0 and 5.1 Hz for residues B, C and E, respectively.

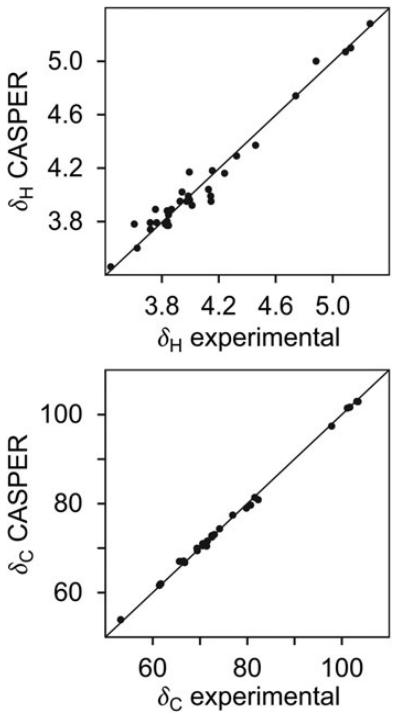


**Fig. 5.** Comparison between experimental and CASPER predicted $^{1}$H and $^{13}$C chemical shifts (top and bottom, respectively) for the anomeric, ring and hydroxymethyl groups of the O-deacetylated repeating unit of the O-antigen PS of *E. coli* O59.

**Table III.** $^{1}$H NMR chemical shifts (ppm) of the resonances of the minor spin systems from the LPS-OH of *E. coli* O59, attributed to residues located at the terminal non-reducing end of the O-antigen PS

| Sugar residue | | $^{1}$H/$^{13}$C | | | | | |
|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 |
| α-D-Gal*p*A-(1→ | **A′** | 5.28 | 3.87 | 3.99 | 4.32 | 4.42 | |
| | | (–0.02) | (0.04) | (0.07) | (0.03) | (0.03) | |
| | | [0.00] | [–0.02] | [0.00] | [0.03] | [0.05] | |
| →3)-α-D-Man*p*-(1→ | **B′** | 5.12 | 4.19 | 4.02 | 3.84 | 3.78 | 3.90, n.d. |
| | | (–0.06) | (0.25) | (0.16) | (0.16) | (–0.04) | (0.03) |
| | | [0.03] | [0.05] | [0.10] | [–0.01] | [–0.01] | |
| →3)-α-D-Man*p*-(1→ | **C′** | 5.10 | 4.22 | 3.99 | 3.83 | n.d. | n.d. |
| | | (–0.08) | (0.28) | (0.13) | (0.15) | | |
| | | [0.03] | [0.04] | [0.04] | [0.06] | | |

Chemical shift differences when compared with the corresponding monosaccharides and the predictions made by the CASPER program are given in parentheses and in square brackets, respectively.
n.d., not determined.

contained not only the *wzx* and *wzz* genes but also those predicted to encode for two enzymes involved in the biosynthesis of the galacturonic acid precursor (*ugd* and *gla*) and an acetyltransferase (*wclD*) (Guo et al. 2005). According to the structure of the biological repeating unit of the O-antigen PS of *E. coli* O59 reported herein, the latter is anticipated to catalyze the O-acetylation of the galacturonosyl residue.

In conclusion, the developments presented in this work have led to a rapid automatic structural elucidation of the biological

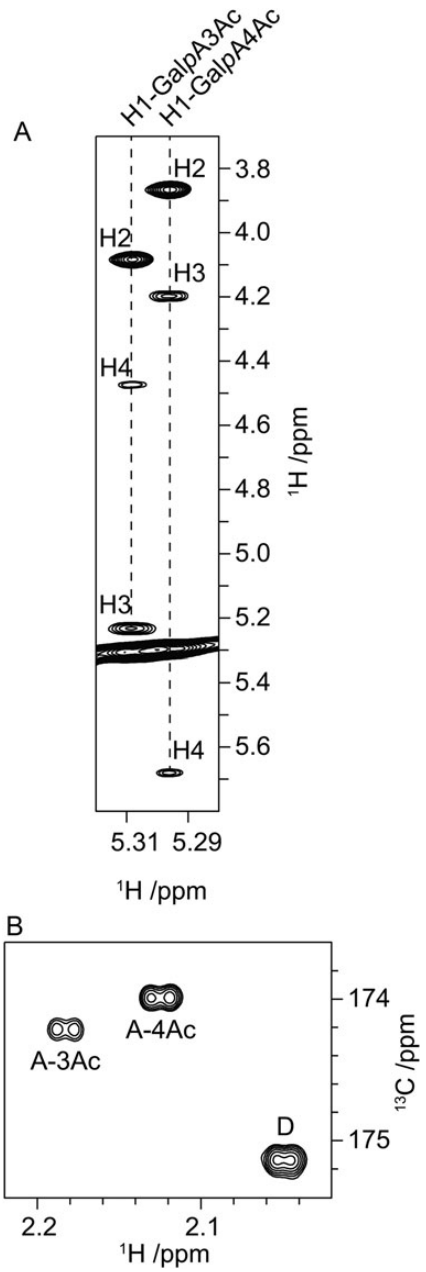


**Fig. 6.** (**A**) Selected region of the $^{1}H,^{1}H$ TOCSY NMR spectrum ($\tau_m$ = 100 ms) of the native PS of *E. coli* O59 showing correlations from the anomeric protons of the two O-acetylated galacturonosyl residues to H2–H4. (**B**) Region of the BS-CT-$^{1}H,^{13}C$ HMBC spectrum showing correlations from the carbonyl carbons of the O- and N-acetyl groups to their respective methyl protons.

repeating unit of a bacterial PS using predicted functions of GTs already available in the ECODAB. Even though different types of NMR data can be submitted to the CASPER program, in this particular case only chemical shifts correlations from a single 2D spectrum were sufficient to obtain the results in less than 3 s on a personal computer. This methodology does not require information about the identity and absolute configuration of the component monosaccharides (which is inferred from the GT predicted functions), avoiding the preparation and analysis of additional derivatized samples altogether.

## Materials and methods

### *Bacterial strain, conditions of growth and isolation of the LPS*

*E. coli* O59, strain CCUG11362, was obtained from the Culture Collection, University of Gothenburg. The bacterium was grown and the LPS isolated as previously described (Svensson et al. 2011b).

### *Preparation of the lipid-free PS*

The LPS (10.6 mg) was dissolved in 2 mL of 1% AcOH (pH 3) and stirred for 90 min at 100°C. The lipid A was removed by centrifugation at 15,000 × *g* for 20 min at 4°C, and the supernatant was adjusted to pH 5 by adding 1 M NaOH very slowly and under continuous stirring at 0°C. The solution was then dialyzed against distilled water for 3 days and purified by gel permeation chromatography using an ÄKTApurifier system equipped with a HiLoad™ 16/60 Superdex™ 30 column (GE Healthcare).

### *Preparation of the LPS-OH*

The LPS (13.4 mg) was treated with 1 mL of $NH_4OH$ 12.5% for 23 h at 37°C. The solution was then neutralized with 4 M HCl, dialyzed against distilled water for 2 days, lyophilized and purified by gel permeation chromatography as described above.

### *Preparation of the (+)-2-butyl glycosides*

The PS (3.0 mg, pH 7) was O-deacetylated by keeping the sample at 60°C until disappearance of the $^{1}H$ NMR resonances of the methyl protons of the O-acetyl groups (24 h). The O-deacetylated PS was then hydrolyzed with 2 M TFA (1.5 mL, 120°C) for 30 min and the (+)-2-butyl glycosides prepared as described before (Lundborg et al. 2011).

### *NMR spectroscopy*

The samples were deuterium exchanged by freeze-drying from 99.9% $D_2O$ and then examined as solutions in 99.96% $D_2O$. The NMR spectra were recorded on Bruker Avance 500 MHz and Bruker Avance III 700 spectrometers equipped with 5 mm TCI ($^{1}H/^{13}C/^{15}N$) Z-Gradient (53.0 G cm$^{-1}$) CryoProbes and the chemical shifts are reported in ppm using internal sodium 3-trimethylsilyl-(2,2,3,3-$^{2}H_4$)-propanoate ($\delta_H$ 0.00) or external 1,4-dioxane in $D_2O$ ($\delta_C$ 67.40) as references. The NMR spectra of the LPS-OH (7.1 mg) were recorded in $D_2O$ solution (0.55 mL) at 60°C at a $^{1}H$ frequency of 500 MHz, with exception of the $^{13}C$ NMR spectrum, which was recorded at the higher magnetic field. The NMR spectra of the PS (2.7 mg) were recorded in $D_2O$ solution (0.55 mL, pD ∼5) at 25°C on a 700 MHz spectrometer, whereas the (+)-2-butyl glycosides were analyzed as a $D_2O$ solution (0.18 mL, in a 3 mm NMR tube) at 70°C using a 500 MHz spectrometer.

### *Outline of the procedure used by the CASPER program*

The script to generate CASPER input based on GT data in ECODAB is written in Python. The procedure is:

1. CASPER input from the web interface:
   - Chemical shifts (of any combination of experiments)
   - Chemical shift corrections

**Table IV.** $^{1}H$ and $^{13}C$ NMR chemical shifts (ppm) at 60°C of the resonances of the O-acetylated galacturonosyl residues from the native O-antigen PS of *E. coli* O59

| Sugar residue | | $^{1}H/^{13}C$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | Me | CO |
| α-D-Gal*p*A3Ac-(1→ | **A-3Ac** | 5.307 | 4.081 | 5.232 | 4.473 | 4.574 | | 2.185 | |
| | | (0.01) | (0.25) | (1.31) | (0.18) | (0.18) | | | |
| | | {–0.03} | {–0.04} | {0.02} | {0.02} | {0.11} | | {0.02} | |
| | | 101.86 | 67.22 | 73.59 | 69.49 | 72.70 | n.d. | 21.26 | 174.20 |
| | | (8.79) | (–1.81) | (3.33) | (–2.15) | (0.40) | | | |
| | | {0.51} | {0.21} | {–0.34} | {0.12} | {0.04} | | {–0.14} | {–0.10} |
| α-D-Gal*p*A4Ac-(1→ | **A-4Ac** | 5.295 | 3.864 | 4.196 | 5.680 | 4.632 | | 2.128 | |
| | | (0.00) | (0.03) | (0.28) | (1.39) | (0.24) | | | |
| | | {–0.03} | {–0.05} | {0.01} | {0.05} | {0.09} | | {–0.03} | |
| | | 102.06 | 69.49 | 68.92 | 73.77 | 71.64 | n.d. | 21.12 | 173.99 |
| | | (8.99) | (0.46) | (–1.34) | (2.13) | (–0.66) | | | |
| | | {0.51} | {0.08} | {–0.01} | {–0.10} | {–0.22} | | {0.02} | {–0.41} |

Chemical shift differences when compared with corresponding monosaccharide (Jansson et al. 1989) and the CASPER predictions are given in parentheses and braces, respectively. The ratio of D-Gal*p*A3Ac and D-Gal*p*A4Ac in the lipid-free PS material was 1:1, respectively, whereas in the native LPS the ratio was 3:7. This difference was attributed to migration of the O-acetyl groups during preparation and purifications of the lipid-free PS material.
n.d., not determined.

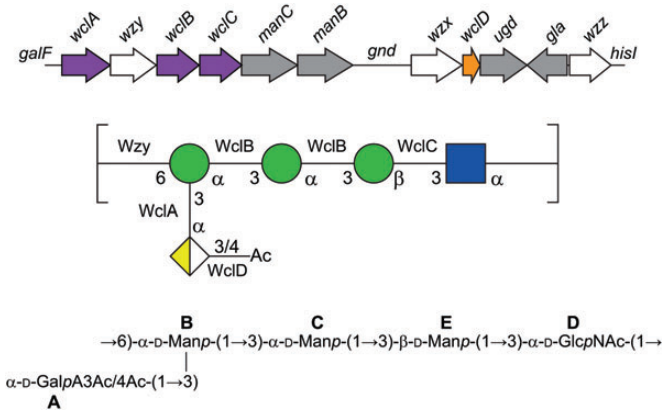


**Fig. 7.** (Top) Representation of the O-antigen gene cluster of *E. coli* O59 (Guo et al. 2005), where the genes are predicted to encode for the O-unit processing proteins, GTs, an O-acetyltransferase and enzymes involved in the biosynthesis of monosaccharides. (Middle) The structure of the biological repeating unit of the O-antigen PS of *E. coli* O59 in CFG notation, where the names of the GTs predicted to be involved in the biosynthesis of each respective linkage are annotated. (Bottom) The structure of the biological repeating unit of the *E. coli* O59 PS in standard nomenclature. The ratio of D-Gal*p*A3Ac and D-Gal*p*A4Ac in the native LPS was 3:7, respectively (but 1:1 in the lipid-free PS material).

- Number of minimal $J_{H1,H2}$ and $J_{C1,H1}$ coupling constants in respective ranges
- WecA rule

2. Read ECODAB entry
3. Sort residues in order of linkages (if specified)
4. If the reducing end of the repeating unit is not identified from the linkage order:
   - Set it to α-D-GlcNAc, β-D-GlcNAc, α-D-GalNAc or β-D-GalNAc
   - Otherwise add the acceptor of the first linkage as the first unit.
5. Iterate through the donors of the linkages:
   - Add the residues specified. If there are multiple possible donor residues for one linkage add all of them as possible residues in the structure determination.
   - If the possible residues are Glc, Gal or Man also add the corresponding uronic acid as a possible alternative (e.g. if α-D-Glc is listed also add α-D-GlcA as an alternative).
   - If the linkage order is specified:
     Restrict the CASPER structure determination by requiring that the residues are connected.
     If the glycosylation position of the acceptor is specified, then add the respective "linkage position restriction" to the CASPER structure determination.
   - Otherwise:
     Try to determine which residues are linked together by comparing the acceptor of a specific GT with the available donors for the other GTs. If there is a unique match of the acceptor in the list of donors then mark the residues as linked together, as well as the linkage position (if this information is available in the GT information).
     If the linkages can be deduced, restrict the possible linkages in the CASPER structure determination as mentioned above.
6. Submit input to CASPER.

## Supplementary data

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

## Funding

This work was supported by grants from the Swedish Research Council and The Knut and Alice Wallenberg Foundation. The research leading to these results has received funding from the European Commission's Seventh Framework Programme FP7/2007-2013 under grant agreement n° 215536.

## Abbreviations

BLAST, basic local alignment search tool; CASPER, computer-assisted spectrum evaluation of regular polysaccharides; CFG, consortium for functional glycomics; GalA, galacturonic acid; GlcNAc, 2-acetamido-2-deoxy-glucose; ManA, mannuronic acid; ECODAB, *Escherichia coli* O-antigen database; GT, glycosyltransferase; H2BC, heteronuclear 2-bond correlation; HMBC, heteronuclear multiple-bond correlation; HSQC, heteronuclear single-quantum correlation; LPS, lipopolysaccharide; LPS-OH, O-deacylated lipopolysaccharide; NMR, nuclear magnetic resonance; NOESY, nuclear Overhauser effect spectroscopy; ORF, open reading frame; PS, polysaccharide; TFA, trifluoroacetic acid; TOCSY, total correlation spectroscopy.

## References


Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol*. 215:403–410.

Caroff M, Karibian D. 2003. Structure of bacterial lipopolysaccharides. *Carbohydr Res*. 338:2431–2447.

Greenfield LK, Richards MR, Vinogradov E, Wakarchuk WW, Lowary TL, Whitfield C. 2012. Domain organization of the polymerizing mannosyltransferases involved in synthesis of the *Escherichia coli* O8 and O9a lipopolysaccharide O-antigens. *J Biol Chem*. 287:38135–38149.

Greenfield LK, Whitfield C. 2012. Synthesis of lipopolysaccharide O-antigens by ABC transporter-dependent pathways. *Carbohydr Res*. 356:12–24.

Guo H, Kong Q, Cheng J, Wang L, Feng L. 2005. Characterization of the *Escherichia coli* O59 and O155 O-antigen gene clusters: The atypical *wzx* genes are evolutionary related. *FEMS Microbiol Lett*. 248:153–161.

Jansson P-E, Kenne L, Schweda E. 1987. Nuclear magnetic resonance and conformational studies on monoacetylated methyl D-*gluco*- and D-*galacto*-pyranosides. *J Chem Soc Perkin Trans 1*. 377–383.

Jansson P-E, Kenne L, Widmalm G. 1989. Computer-assisted structural analysis of polysaccharides with an extended version of CASPER using $^{1}$H- and $^{13}$C-N.M.R. data. *Carbohydr Res*. 188:169–191.

Kido N, Targov VI, Sugiyama T, Uchiya K, Sugihara H, Kato N, Jann K. 1995. Expression of the O9 polysaccharide of *Escherichia coli*: Sequencing of the *E. coli* O9 rfb gene cluster, characterization of mannosyl transferases, and evidence for an ATP-binding cassette transport system. *J Bacteriol*. 177: 2178–2187.

Lundborg M, Fontana C, Widmalm G. 2011. Automatic structure determination of regular polysaccharides based solely on NMR spectroscopy. *Biomacromolecules*. 12:3851–3855.

Lundborg M, Modhukur V, Widmalm G. 2010. Glycosyltransferase functions of *E. coli* O-antigens. *Glycobiology*. 20:366–368.

Lundborg M, Widmalm G. 2011. Structural analysis of glycans by NMR chemical shift prediction. *Anal Chem*. 83:10–13.

Ørskov F, Ørskov I. 1984. Serotyping of *Esherichia coli*. *Methods Microbiol*. 14:43–112.

Raetz CRH, Whitfield C. 2002. Lipopolysaccharide endotoxins. *Annu Rev Biochem*. 71:635–700.

Rönnols J, Pendrill R, Fontana C, Hamark C, Angles d'Ortoli T, Engström O, Ståhle J, Zaccheus MV, Säwén E, Hahn LE, et al. 2013. Complete $^{1}$H and $^{13}$C NMR chemical shift assignments of mono- to tetrasaccharides as basis for NMR chemical shift predictions of oligosaccharides using the computer program CASPER. *Carbohydr Res*. 380:156–166.

Roslund MU, Aitio O, Wärnå J, Maaheimo H, Murzin DY, Leino R. 2008. Acyl group migration and cleavage in selectively protected β-D-galactopyranosides as studied by NMR spectroscopy and kinetic calculations. *J Am Chem Soc*. 130: 8769–8772.

Rush JS, Alaimo C, Robbiani R, Wacker M, Waechter CJ. 2010. A novel epimerase that converts GlcNAc-P-P-undecaprenol to GalNAc-P-P-undecaprenol in *Escherichia coli* O157. *J Biol Chem*. 285:1671–1680.

Samuel G, Reeves P. 2003. Biosynthesis of O-antigens: Genes and pathways involved in nucleotide sugar precursor synthesis and O-antigen assembly. *Carbohydr Res*. 338:2503–2519.

Stenutz R, Weintraub A, Widmalm G. 2006. The structures of *Escherichia coli* O-polysaccharide antigens. *FEMS Microbiol Rev*. 30:382–403.

Svensson MV, Weintraub A, Widmalm G. 2011a. Structural studies of the O-antigenic polysaccharide from *Escherichia coli* O177. *Carbohydr Res*. 346:2300–2303.

Svensson MV, Weintraub A, Widmalm G. 2011b. Structural elucidation of the O-antigenic polysaccharide from *Escherichia coli* O175. *Carbohydr Res*. 346:449–453.

Valvano MA, Furlong SE, Patel KB. 2011. Genetics, biosynthesis and assembly of O-antigen. In: Knirel YA, Valvano MA, editors. Bacterial Lipopolysaccharides. Structure, Chemical Synthesis, Biogenesis and Interaction with Host Cells. Vienna: Springer. p. 275–310.