Carbohydrate Research 344 (2009) 1533–1540

Contents lists available at ScienceDirect

Carbohydrate Research

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

The O-specific polysaccharide structure and biosynthetic gene cluster of *Yersinia pseudotuberculosis* serotype O:11

Monica M. Cunneen [a,†], Cristina De Castro [b,†], Johanna Kenyon [a], Michelangelo Parrilli [b], Peter R. Reeves [a], Antonio Molinaro [b], Otto Holst [c,*], Mikael Skurnik [d,e]

[a] *Division of Microbiology, School of Molecular and Microbial Biosciences, University of Sydney, Australia*
[b] *Department of Organic Chemistry and Biochemistry, University Federico II of Naples, Naples, Italy*
[c] *Division of Structural Biochemistry, Research Center Borstel, Leibniz-Center for Medicine and Biosciences, Parkallee 4a/c, D-23845 Borstel, Germany*
[d] *Department of Bacteriology and Immunology, Haartman Institute, University of Helsinki, Finland*
[e] *Helsinki University Central Hospital Laboratory Diagnostics, Helsinki, Finland*

ARTICLE INFO

Article history:
Received 6 March 2009
Received in revised form 24 April 2009
Accepted 28 April 2009
Available online 3 May 2009

This paper is dedicated to Professor Dr. Johannis P. Kamerling on the occasion of his 65th birthday

Keywords:
Yersinia pseudotuberculosis
O-Specific polysaccharide
Structure
Biosynthetic gene cluster
6-Deoxy-L-altrofuranose

ABSTRACT

In the *Yersinia pseudotuberculosis* serotyping scheme, 21 serotypes are present originating from about 30 different O-factors distributed within the species. With regard to the chemical structures of lipopolysaccharides (LPSs) and the genetic basis of their biosynthesis, a number, but not all, of *Y. pseudotuberculosis* strains representing different serotypes have been investigated. In order to present an overall picture of the relationship between genetics and structures, we have been working on the genetics and structures of various *Y. pseudotuberculosis* O-specific polysaccharides (OPSs). Here, we present a structural and genetic analysis of the *Y. pseudotuberculosis* serotype O:11 OPS. Our results showed that this OPS structure has the same backbone as that of *Y. pseudotuberculosis* O:1b, but with a 6d-L-Alt*f* side-branch instead of Par*f*. The 3′ end of the gene cluster is the same as that for O:1b and has the genes for synthesis of the backbone and for processing the completed repeat unit. The 5′ end has genes for synthesis of 6d-L-Alt*f* and its transfer to the repeating unit backbone. The pathway for the synthesis of the 6d-L-Alt*f* appears to be different from that for 6d-L-Alt*p* in *Y. enterocolitica* O:3. The chemical structure of the O:11 repeating unit is

2)-β-D-Man*p*-(1→4)-α-D-Man*p*-(1→3)-α-L-Fuc*p*-(1→3)-α-D-Glc*p*NAc-(1→
3)
↑
α-6d-L-Alt*f*-(1

1. Introduction

Most Gram-negative bacteria, including the genus *Yersinia*, possess in their cell envelope an outer membrane (OM), the outer leaflet of which comprises high amounts of lipopolysaccharide (LPS).[1] The smooth S-form LPSs are composed of the two biosynthetic entities, the lipid A–core region and the O-specific polysaccharide (OPS, O-antigen) that is covalently linked to the core. Rough-form LPSs lack the OPS. The OPS structures are highly variable and represent the basis for the O-serotyping of bacterial strains. In LPS biosynthesis,[2,3] enzymes are required that are specific for each LPS component and linkage; therefore, the number of LPS genes is high (⩾50) and are in most Gram-negative bacteria chromosomally located. Those genes encoding lipid A biosynthesis are scattered around the genome, whereas others involved in core and OPS biosynthesis each form a distinct gene cluster. Only a few bacteria are known in which the LPS genes are located in plasmids or in temperate bacteriophage genomes.

Twenty-one serotypes are presently included in the *Y. pseudotuberculosis* serotyping scheme that comprises 15 major O-serotypes,[4] the first five of which contain 2 or 3 O-subtypes, thus making up this number. The serotype classification is based on the distribution of about 30 different O-factors within the species. As identified by epidemiology, European isolates belong most often to serotypes O:1–O:3,[5] while in Japan serotypes O:3, O:4b, O:5a and O:5b are the common ones.[6] Some of the more recently recognised serotypes represent 'environmental' *Y. pseudotuberculosis*, which do not carry the virulence plasmid.[4,7]

The chemical structures of LPSs and the genetic basis of their biosynthesis have been determined for a number of *Yersinia* strains

* Corresponding author. Tel.: +49 4537 188472; fax: +49 4537 188745.
E-mail address: oholst@fz-borstel.de (O. Holst).
† These authors contributed equally.

0008-6215/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.carres.2009.04.030

representing different species and serotypes;[8–17]; thus, an overall picture of the relationship between genetics and structures is emerging. Here, all the studied gene clusters expressing the heteropolymeric OPS type are chromosomally arranged between the *hemH* and *gsk* genes. At the upstream of *hemH* is located *adk*, and downstream of *gsk* are the *rosB–rosA* and *ushA* genes. The latter are not directly involved in OPS biosynthesis; however, the *rosAB* operon is involved in regulation of the OPS expression.[18] In *Y. enterocolitica* serotypes O:3 and O:9 that express homopolymeric OPS, the OPS gene clusters are at an unknown position, that is, the locus between *hemH* and *gsk* is occupied by the outer core gene cluster.[19]

Several OPS structures of *Y. pseudotuberculosis* LPS have been identified.[20–23] As shown by sequencing, gene-specific PCR analysis and hybridisation studies, all serotypes except O:7, O:9 and O:10 carry the *ddhDABC* genes or their homologues, and are involved in the biosynthesis of CDP-3,6-dideoxyhexoses (CDP-DDH) in the beginning of their gene clusters downstream of the *hemH* gene.[15,24] The *ddhDABC* block-encoded enzymes produce a biosynthetic intermediate in the biosynthesis of the CDP-DDH, and completion of the specific CDP-DDH in each serogroup requires one or two additional gene(s) that determine(s) the serogroup specificity, for example, *abe* in serogroups O:2a, O:2b and O:2c (that carry abequose in their OPS); *prt* and *tyv* in serogroups O:4a and O:4b (that carry tyvelose in their OPS), *prt* in serogroups O:1a and O:1b (that carry paratose in their OPS) and *ascEF* in serotype O:5a (that carries ascarylose in its OPS).[25–28] The PCR analysis allowed the prediction of specific DDH for some serotypes for which no structural data were yet available, such as O:1c, O:11, O:14 and O:15.[17,29] In addition to the DDH pathway genes, the OPS clusters contain the genes encoding the biosynthetic enzymes of the precursors of other NDP-sugar components, the respective glycosyltransferases and the translocation and polymerisation proteins Wzx, Wzy and Wzz.[15] Interestingly, based on PCR analysis, all serotypes appear to use an almost identical Wzz-protein as a chain length determinant.[17]

In this work we have performed a structural and genetic analysis of the *Y. pseudotuberculosis* serotype O:11 OPS. Our results showed that this OPS structure has the same backbone as that of *Y. pseudotuberculosis* O:1b, but with a 6d-L-Alt*f* side-branch instead of Par*f*. The 3′ end of the gene cluster is the same as that for O:1b and has the genes for synthesis of the backbone and for processing the completed repeat unit. The 5′ end has genes for synthesis of 6d-L-Alt*f* and its transfer to the repeating unit backbone.

2. Results

2.1. Isolation of the OPS and compositional analyses

The LPS was collected from the water phase of the hot phenol–water extraction in a yield of 80 mg (0.8% of the bacterial dry mass). Analysis by sodium dodecyl polyacrylamide gel electrophoresis (SDS–PAGE, data not shown) proved its *S*-form character. The OPS was isolated after mild acid hydrolysis and gel-permeation chromatography in a yield of 11.1 mg (17% of the LPS). Its sugar compositional and methylation analyses identified two D-Man*p* residues (one substituted at O-4 and the other at O-2 and O-3), one D-Glc*p*NAc (substituted at O-3), one L-Fuc*p* (linked at O-3) and one terminal 6-deoxy-L-*altro*-furanose (6d-L-Alt*f*) residue.

2.2. Structural analysis of the OPS

As visible from the proton and HSQC spectra (Figure 1), this was one of the cases in which the assignment of the anomeric region is very difficult. Only two signals (at δ 5.12 and 4.75) were present, the first of which comprised the overlapping chemical shifts of four different protons. This analytical problem was solved by applying the combined use of three heteronuclear NMR techniques, that is, gHSQC, gHMBC and gHSQC-TOCSY, which allowed a complete spectroscopic attribution (Table 1) and structure determination of the repeating unit of the OPS.

The NMR analysis started from the only straight anomeric signal, labelled as **E**. Its H-2 resonance (δ 4.32) correlated weakly with both H-1 and H-3, a feature characteristic for *manno*-configured monosaccharides. The signals of H-3 and H-4 were coincident, and H-5 and H-6 could be assigned through scalar correlation starting from the H-4 signal. This interpretation was in agreement with the heteronuclear 2D spectral data, for which assignment the correlations of H-2 were essential, that is, in the gHSQC-TOCSY experiment (Fig. 2) this signal correlated to all the six carbon atoms of the sugar ring, three of which were also displayed in the gHMBC spectrum (C-3, C-4 and C-1). Additionally, in the gHMBC experiment it was possible to identify C-4 at δ 66.7 by the correlation of the H-6 signal (δ 3.99), whereas the H-2 resonance gave an inter-residue long-range correlation with the anomeric carbon of **D**. C-2 and C-3 were both shifted downfield demonstrating their substitution. The anomeric configuration of **E** was deduced on the basis of the H-1/H-3 and H-1/H-5 NOE connectivities, which were diagnostic of a β-orientation of the glycosidic linkage and confirmed by the $^1J_{C\text{-}1,H\text{-}1}$ coupling (160.2 Hz), as observed for the anomeric signal in the HMBC spectrum. In conclusion, residue **E** was a 2,3-disubstituted β-Man*p*.

Analysis of residue **C** started from the methyl group resonance at δ 1.19, which correlated with H-5 (δ 4.35) in the COSY spectrum. This proton showed a heteronuclear long-range correlation with its anomeric carbon that was slightly separated from the other three carbon signals, whereas the gHSQC-TOCSY experiment disclosed the resonances of H-2, H-3 and H-4 at δ 3.94, 3.91 and 3.87, respectively; in particular, the H-4 resonance was deduced from the C-4 signal, which was in turn identified in the HMBC spectrum. Finally, proton H-2 was identified by its correlation with H-1 and H-3. The chemical shifts of the carbons of this residue, together with the scalar correlations linking H-1 up to H-3 and the lack of a H-4/H-5 scalar correlation, identified it as an α-configured Fuc*p* residue that was substituted at O-3, as indicated by the downfield shifted signal of its C-3 carbon. In conclusion, residue **C** was a 3-substituted α-Fuc*p*.

All ring proton coupling constants of residue **D** were around 10 Hz, and in the TOCSY spectrum, starting from the anomeric proton, all correlations up to the H-6 signals could be assigned. The $^1J_{C\text{-}1,H\text{-}1}$ coupling constant (175.3 Hz) established the α-configuration at the anomeric carbon. By gHSQC it was possible to identify C-2 as a nitrogen bearing carbon which was N-acetylated as proven by the downfield chemical shift of the H-2 signal and by the presence of one *N*-acetyl signal in the proton spectrum. Moreover, the C-3 signal was downfield shifted owing to its glycosylation. All of these data established residue **D** as a 3-substituted α-Glc*p*NAc.

The analysis of residue **A** started from the methyl group at δ 1.22, and the COSY experiment related this proton to H-5, whereas TOCSY densities indicated a close proximity of the other ring proton signals, that is, between δ 4.08 and 3.97. Disclosing information came from the HSQC-TOCSY spectrum in which H-6 correlated to four other carbon signals, two of which (δ 68.0 and 88.2) were also present in the gHMBC spectrum. These last two signals could easily be attributed to C-5 and C-4, respectively. The anomeric carbon of **A** was identified on the basis of its gHSQC-TOCSY correlations with the other ring protons, the chemical shifts of which were already established in the TOCSY experiment starting from H-6. The chemical shift of H-2 was identified in the COSY spectrum, in which the group of signals at δ 5.12 still contained two H-1/H-2 correlations, but only one was compatible with the values identified in the TOCSY experiment. The chemical shift of the C-2 signal was established

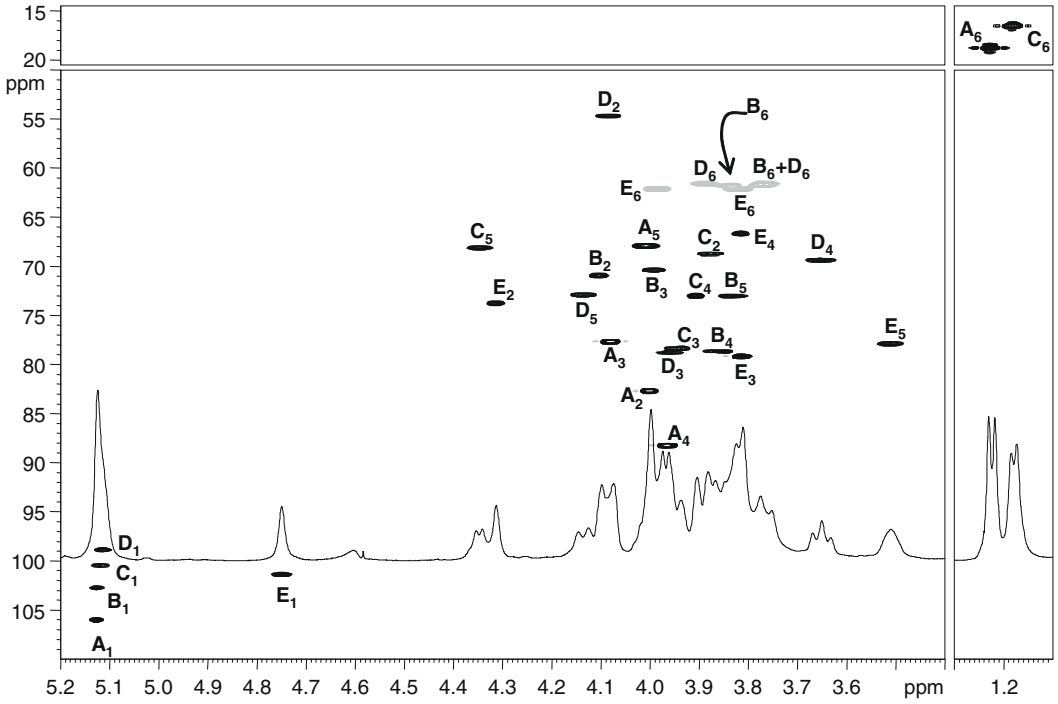

Figure 1. Assigned ^{1}H, ^{13}C gHSQC and proton spectra of the OPS from *Y. pseudotuberculosis* O:11 (D_2O, 318 K, 500 MHz). Hydroxymethyl signals are presented in antiphase (greytone).

Table 1
^{1}H and ^{13}C NMR data ($^{2}H_2O$, 500 MHz, 318 K, acetone as internal standard) of the OPS from *Y. pseudotuberculosis* O:11 LPS. Chemical shifts for the *N*-acetyl group were: ($\delta^{1}H$) 2.06, ($\delta^{13}C$) 22.4 and 175.6

	1	2	3	4	5	6
A	5.13	4.00	4.08	3.97	4.01	1.22
t-6d-α-L-Alt*f*	*106.4*	*82.6*	*77.7*	*88.2*	*68.0*	*18.8*
B	5.13	4.10	3.99	3.86	3.83	3.84–3.77
4-α-D-Man*p*	*102.8*	*70.9*	*70.3*	*78.7*	*73.0*	*61.8*
C	5.12	3.87	3.94	3.91	4.35	1.19
3-α-L-Fuc*p*	*100.5*	*68.7*	*78.4*	*73.0*	*68.0*	*16.6*
D	5.11	4.09	3.96	3.65	4.14	3.89-3.77
3-α-D-GlcpNAc	*98.9*	*54.7*	*78.8*	*69.3*	*72.9*	*61.6*
E	4.75	4.32	3.81	3.81	3.51	3.99-3.82
2,3-β-D-Man*p*	*101.4*	*73.8*	*79.2*	*66.7*	*77.8*	*62.1*

on the basis of the values assigned in the gHSQC-TOCSY correlation starting from H-6. Finally, the H-3/C-3 cross peak was the only correlation left. The stereochemistry of this residue was established by chemical analysis, and the configuration at the anomeric carbon was deduced from the chemical shifts of the reference compound.[36] In agreement with the methylation data, the chemical shifts of the carbon signals of all spin systems clearly indicated a downfield displacement for C-2 and C-4 signals owing to the furanose nature of the ring but not to glycosylation. Thus, residue **A** was identified as a terminal 6d-L-Alt*f*.

As for residue **B**, the shift of its H-2 was identified in the COSY experiment at δ 4.10, and the heteronuclear spectra helped in the identification of C-2 and the other ring carbons. In addition, the HMBC experiment identified both, C-3 and C-4; thus, the other correlations in the HSQC-TOCSY could be assigned to C-6 (δ 61.8) and to C-5 (δ 73.0). The discrimination between C-3 and C-4 cross peaks was possible through the HMBC correlation linking the H-6 signal (δ 3.77) to a carbon signal at δ 78.7 (C-4); thus, C-3 was the only correlation left. All chemical shifts of the proton signals were identified as a consequence of the gHSQC spectrum and definitely assigned by homonuclear spectroscopy. Thus, as proven by the α-glycosylation shift of C-4 and anomeric $^{1}J_{C\text{-}1,H\text{-}1}$ value (171.3 Hz from HMBC), residue **B** was identified as 4-substituted α-Man*p*.

The sequence of the five residues was inferred through the following key HMBC scalar correlations: H-2 of **E** with *C*-1 of **D**, H-4 of **B** with *C*-1 of **E**, H-3 of **C** with *C*-1 of **B**, H-3 of **D** with *C*-1 of **C** and H-3 of **E** with *C*-1 of **A**. This sequence was in agreement with the data obtained from methylation analysis.

In conclusion, *Y. pseudotuberculosis* O:11 LPS possessed an OPS with the pentasaccharide repeating unit shown in Figure 3.

2.3. O:11 O-antigen gene cluster sequencing

Previously, PCR screening had shown that much of the O:11 gene cluster was very similar to that of serotype O:1b,[17] and so primers based on the O:1b sequence were used to PCR amplify and sequence short segments of the gene cluster, with the remaining region between *ddhA* and *wbyH* being amplified by long-range PCR and sequenced. The complete gene cluster obtained was 19.4 kb in length, starting from the JUMPstart sequence, and contained 17 Orfs that are described in Table 2 with the proposed gene names, and presented in Figure 4 with the O:1b gene cluster, which is most closely related, included for comparison.

The *ddhD*, *ddhA*, *wbyH*, *wzx*, *wbyI*, *wbyJ*, *gmd*, *fcl*, *wzy*, *wbyK*, *manC*, *wbyL*, *manB* and *wzz* genes are present in the same order as in the O:1b gene cluster, with amino acid identity levels for the predicted gene products ranging from 93% to 99%. The functions of the O:1b gene products had been inferred by homology to previously characterised genes, and for the glycosyl transferase (GT) genes by their distribution among 10 related gene clusters.[16] All but *ddhD*, *ddhA* and *wbyH* would be required for synthesis of the 4-sugar backbone and subsequent processing, and the high level of similarity indicates that the two have identical backbone structures. However, the reported structure for O:1b backbone differs

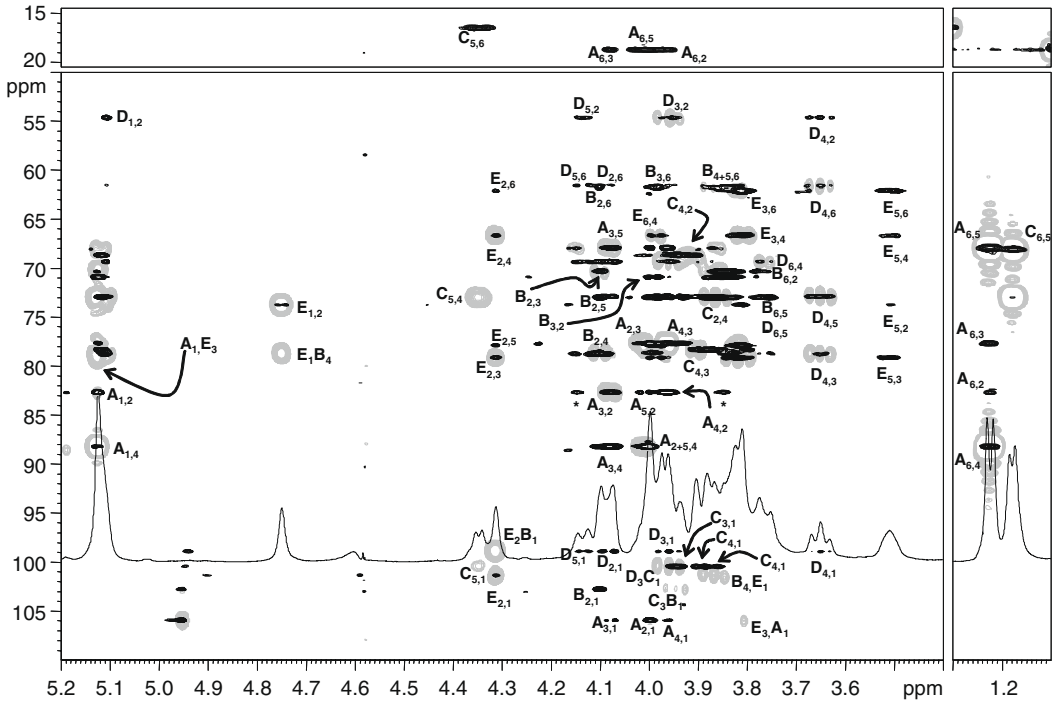

Figure 2. Assigned ^{1}H, ^{13}C gHMBC (grey), gHMQCTOCSY (black) and proton spectra of the OPS from *Y. pseudotuberculosis* O:11 (D_2O, 318 K, 500 MHz).

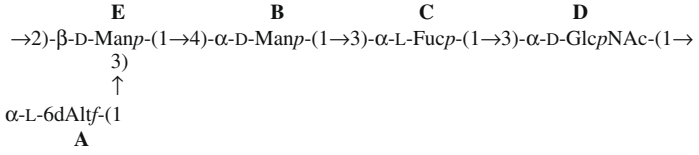

Figure 3. The pentasaccharide repeating unit of the OPS from *Y. pseudotuberculosis* O:11.

Table 2
Yersinia pseudotuberculosis O:11 O-antigen genes

Orf	Gene	Length (bp)	Length (AA)	BLAST Hit (accession number)	Protein description[a]	% Identity/similarity (AA hit/total)
1	*ddhD*	990	329	*Y. pestis* CO92 DdhD (NP_406594)	CDP-6-deoxy-Δ-3,4-glucoseen reductase	99 (327/329)/100 (329/329)
2	*ddhA*	786	261	*Y. pseudotuberculosis* IP 32953 DdhA (YP_069540)	Glucose-1-phosphate cytidylyltransferase	99 (259/261)/99 (260/261)
3	*ddhB*	1206	401	*Shewanella denitrificans* OS217 Sden_2656 (YP_563658)	NAD-dependent epimerase/dehydratase	66 (244/365)/80 (294/365)
4	*altA*	432	143	*Shewanella denitrificans* OS217 Sden_2655 (YP_563657)	Hypothetical protein	45 (60/132)/68 (91/132)
5	*altB*	867	288	*Clostridium beijerinckii* NCIMB 8052 Cbei_2579 (YP_001309691)	dTDP-4-dehydrorhamnose reductase	33 (92/303)/52 (158/303)
6	*wbyH*	1278	425	*Y. pseudotuberculosis* O:1b WbyH (CAB63294)	CDP-paratopyranose mutase[b]	93 (396/425)/96 (408/425)
7	*wzx*	1344	447	*Y. pestis* CO92 (O:1b[c]) Wzx (NP_406589)	O-Unit flippase[b]	90 (403/447)/94 (422/447)
8	*wbyI*	984	327	*Y. pseudotuberculosis* IP 31758 (O:1b) YpsIP31758_3045 (YP_001402005)	Paratofuranose β1,3 transferase[b]	94 (310/327)/98 (321/327)
9	*wbyJ*	1143	380	*Y. pestis* CO92 WbyJ (NP_406588)	Mannose (1,4) transferase[b]	96 (367/380)/98 (376/380)
10	*wzy*	1221	406	*Y. pseudotuberculosis* O:1b Wzy (CAB63298)	O-Unit polymerase[b]	96 (392/406)/98 (399/406)
11	*wbyK*	1014	337	*Y. pseudotuberculosis* O:1b WbyK (CAB63299)	Mannose α(1,3) transferase[b]	97 (330/337)/98 (332/337)
12	*gmd*	1122	373	*Y. pseudotuberculosis* O:1b Gmd (CAB63300)	GDP-D-mannose dehydratase	98 (368/373)/99 (371/373)
13	*fcl*	966	321	*Y. pseudotuberculosis* IP 32953 Fcl (YP_069551)	GDP-fucose synthetase	98 (316/321)/99 (320/321)
14	*manC*	1407	468	*Y. pestis* CO92 ManC (NP_406586)	Mannose-1-phosphate guanyltransferase	97 (456/468)/99 (465/468)
15	*wbyL*	744	247	*Y. pestis* CO92 WbyL (NP_406585)	Fucose α(1,3) transferase[b]	93 (231/247)/97 (240/247)
16	*manB*	1374	457	*Y. pestis* CO92 ManB (NP_406584)	Phosphomannomutase	97 (446/457)/98 (451/457)
17	*wzz*	1152	383	*Y. pseudotuberculosis* O:1b Wzz (CAB63305)	O-Antigen chain length determinant[b]	93 (357/383)/97 (372/383)

[a] From BLAST output unless otherwise stated.
[b] Function of homologue in *Y. pseudotuberculosis* O:1b (17).
[c] *Y. pestis* carries a non-functional O:1b gene cluster (17).

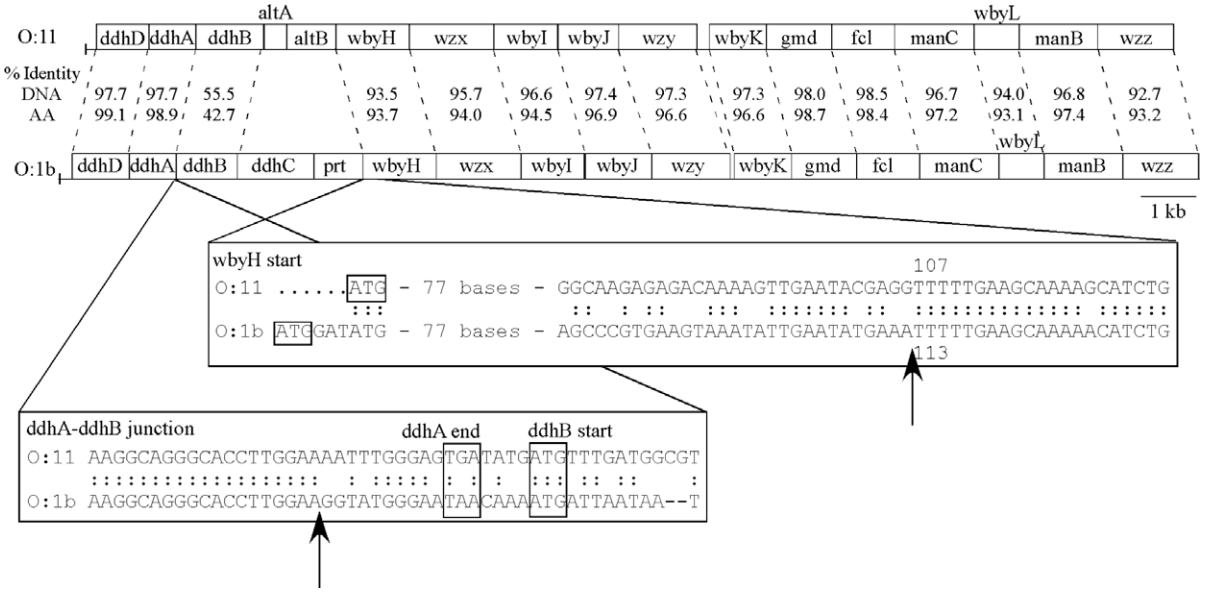

Figure 4. Comparison of the *Y. pseudotuberculosis* O:11 and O:1b O-antigen gene clusters. Clusters start at the JUMPstart sequence. Gene names, % DNA and amino acid identity are indicated. Panels of the *ddhA–ddhB* junction and within *wbyH* are also shown to indicate putative recombination sites (arrows). Gene start and end codons are shown (boxed). Base positions within *wbyH* are indicated.

from that for O:11 in that the terminal sugar is reported as α-Man, whereas O:11 has β-Man*p*.[30] Given that the O:1b structure was determined in 1981, and in the light of the genetic data, we suggest that the O:1b backbone structure is the same as that found now for O:11; however, this remains to be confirmed. On this basis we have assigned the same names and functions to the 14 common genes.

WbyH in O:1b was considered to be the expected CDP-Par*p* mutase on the basis of homology to the UDP-Glc*p* mutase Glf. It seems clear that WbyH in O:11 generates the furanose ring of the 6d-L-Alt*f* and that it acts at the nucleotide sugar level, suggesting that the other genes are involved in synthesis of NDP-6d-L-Alt*f*. We have retained the name *wbyH* but it is quite possible that the sequence differences confer substrate specificity, and if that were shown to be the case, the O:11 gene should be renamed.

We are left with *ddhD* and *ddhA* with 99% similarity to its O:1b counterpart, and three genes between *ddhA* and *wbyH* (*ddhB*, *altA and altB*) with relatively low-level similarity to their best hits in BLAST searches (see Table 2). We have accounted for the synthesis of the 4-sugar backbone and the processing of the repeat unit by Wzx, Wzy and Wzz, leaving only the synthesis of the 6d-L-Alt*f* precursor and the GT for 6d-L-Alt*f* transfer. The presence of *ddhA*, with only two codons different from the O:1b D-Glc-1-phosphate cytidyltransferase gene, indicates that the 6d-L-Alt*f* precursor must be CDP-6d-L-Alt*f*. However, although the *ddhB* gene is highly divergent with only 43% amino acid identity to *ddhB* of O:1b, those from both serotypes share the same Pfam-A domain (PF01370).[31] So for O:11, DdhB is most likely the CDP-sugar dehydratase, and it must act on the CDP-D-Glc to give 6-deoxy-D-*xylo*-hex-4-ulose, leaving Orf4 and Orf5 for conversion to CDP-6d-L-Alt. L-Altrose differs from D-Glc at C-4 and C-5, and the predicted steps would be epimerisation at C-5 followed by appropriate reduction at C-4. This proposed pathway is depicted in Figure 5. The most convincing hits are for Orf5, with the first hit to a gene with a predicted function to *wbbW* of *Y. enterocolitica* O3, the putative reductase discussed below. As both Orf5 and WbbW share the same Pfam-A domain (P404321), Orf5 is assumed to be the final reductase at C-4. Orf4 has its best hit to a *Shewanella* gene, with other hits to several putative dTDP-4-dehydrorhamnose 3,5 epimerases. Thus, although no Pfam domain was recognised, it is predicted to be the expected C-5 epimerase. Although the pathway details have to be confirmed, it seems clear that Orf4 and Orf5 are involved in the 6d-L-Alt*f* specific steps and are named *altA* and *altB*, respectively.

3. Discussion

The structural and genetic analysis of the *Y. pseudotuberculosis* serotype O:11 OPS has been performed. Our results showed that its structure possesses the same backbone as the OPS of *Y. pseudotuberculosis* O:1b; however, it contains a 6-deoxy-L-altrofuranose (6d-L-Alt*f*) side branch that replaces the paratofuranose (Par*f*) of the O:1b OPS. The 6d-L-Alt*f* gene is very rare in bacterial polysaccharides, as apart from *Y. pseudotuberculosis* there are reports only for *Campylobacter jejuni*[32] and two species of *Pectinatus*.[33] In the O-antigen of *Y. enterocolitica* O:3, 6d-L-Alt*p* is present.[34–36] Overall, the structure of the OPS of the *Y. pseudotuberculosis* serotype O:11 does not differ from the common architectural scheme of the OPS from other *Y. pseudotuberculosis,* in which to a tetrasaccharide repeating unit is attached a 6-deoxy or a 3,6-dideoxy monosaccharide residue as a single branching residue.[20–23]

The O:11 genes for the backbone structure are the same as in O:1b, and except for one linkage, the reported structures are the same, with the difference probably due to an error in the 1981 structure. Such errors were not uncommon with the techniques then available, and several predictions based on genetic similarity of *Y. pseudotuberculosis* gene clusters have since been confirmed, that is, serotypes 2a, 2b, 4a and 4b.[17,20–23]

Most interesting is the finding that the 6d-L-Alt*f* precursor must be CDP-6d-L-Alt*f*. This is again based on relating genes present to function, but in this case the synthesis of the CDP-based precursors of the 3,6-DDH is very well documented (see Ref. 37 for a review). As the O:11 DdhA protein differs at only two residues from that of the O:1b cluster, we feel confident that the nucleotide involved is CDP. However, for *Y. enterocolitica* O3, the only other occurrence of 6d-L-Alt for which we have any genetics, the genes involved are not closely related to those in *Y. pseudotuberculosis* O:11, and the predicted pathway involves dTDP. We have repeated the BLAST searches, and the evidence is still for a dTDP-based pathway for

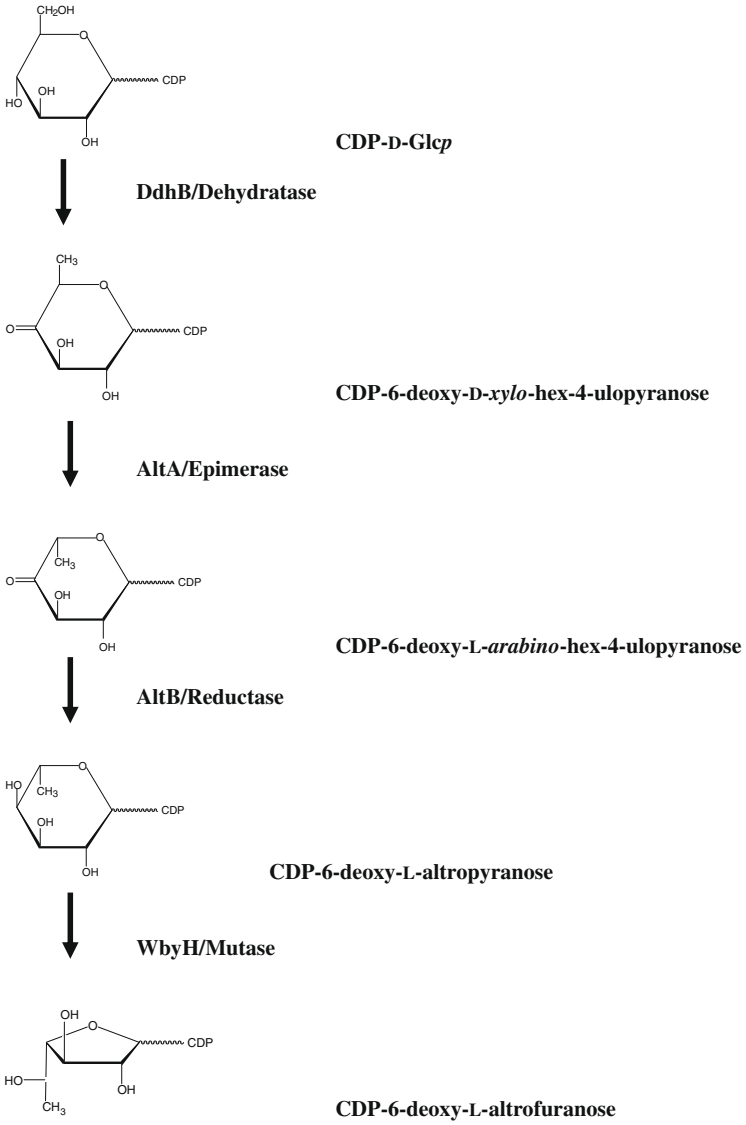

Figure 5. Proposed biosynthesis pathway of CDP-6-d-L-Alt*f*. See text for details.

6d-L-Alt*p* in *Y. enterocolitica* O:3; however, this again needs to be confirmed. Regardless of the NDP involved, the divergence between the genes is far greater than that which is normally seen for a pathway when comparing even unrelated species, and it is very striking for a variation within a genus. It seems clear that the two occurrences of 6d-L-Alt in *Yersina* had different origins, and the *Y. pseudotuberculosis* O:11 genes at least arose in a distantly related species.

Of the genes in common between O:11 and O:1b gene clusters, *ddhB* has the lowest level of identity. However, this extends upstream into the 3′ end of the *ddhA* gene, with the identity between the last 13 bases of *ddhA*, at 61.5%, being more comparable with the *ddhB* gene (55.5% identity), than to the whole of *ddhA* (97.7% identity). A similar region has been identified within the 5′ end of *wbyH* spanning ~100 bases that has 67.3% identity, and the much larger 3′ component having 96% identity (see Fig. 4 for putative recombination sites). The identification of such symmetrical putative recombination sites provides strong evidence for the acquisition of the *ddhB-altA-altB* gene block from a source outside *Yersinia*.

Although *ddhD* is present within the O:11 gene cluster, and is involved in the final C-4 reductase step of DDH synthesis in other serotypes, there is no assigned role for it in the 6d-L-Alt pathway proposed in this study, as AltB is assigned to the final C-4 reductase step. The substrate for the final reduction to CDP-6d-L-Alt and that for reduction to one of the CDP-3,6 dideoxyhexose (DDH) pathways would differ in the configuration at C-3 (with hydroxyl group or 3-deoxy) and at C-5 (L or D configuration), and it seems most unlikely that DdhD could function in the CDP-6d-L-Alt pathway.

The PCR profiles observed by Bogdanovich et al.[29] were identical for O:11 and O:14, but analysis by SDS–PAGE showed that O:14 lacked an O-antigen, and we therefore sequenced the gene cluster of O:14 strain CN3 in parallel with that for the O:11 strain R80. The same genes were present, and none of the sequence differences observed could account for the absence of the O-antigen in O:14. Very likely explanations are that there is either a mutation in the ligase gene in the LPS core gene cluster or a mutation in one of the core synthesis genes that changed the structure to prevent ligation of the OPS. The O:14 specific factor in the anti O:14 sera may be due to exposure of core sugars, but this would not justify recognition of a distinct serotype, and the O:14 form should be treated as a variant of O:11, or as an indicator of a rough strain if it reacts with rough derivatives of other serotypes.

4. Experimental

4.1. Bacterial growth and isolation of the LPS

Yersinia pseudotuberculosis serotype O:11 strains MW109-2 and R80 were obtained from Hiroshi Fukushima (Japan)[17,29] and routinely grown in Luria broth (LB) or on Luria agar. For fermentor cultivation of strain MW109-2, a 250-mL inoculum in LB was grown overnight at 25 °C on a rotary shaker at 200 rpm. Fermentor cultivation was done in a BIOSTAB B plus Santorius Fermentor with 4 L *Yersinia* fermentor medium as described elsewhere.[38] During fermentation, the pH was maintained at 7.4 by addition of 20% NaOH. The pO_2 of the culture was kept at 50% saturation by sterile air flow aeration and stirrer speed control. After 20–24 h fermentation at 25 °C, the bacteria were killed by a final concentration of 0.5–1% phenol, washed with PBS until the supernatant was clear and freeze-dried. The LPS was extracted from the dry bacteria by the hot phenol–water method.[39]

4.2. Isolation of the OPS

The LPS (65 mg) was hydrolysed with 1% HOAc (100 °C, 2 h), and the precipitating lipid A was sedimented by ultracentrifugation (100.000 *g*, 4 °C, 2 h). The supernatant that contained the OPS and smaller carbohydrates was lyophilised (yield: 40 mg, 61.5% of the LPS) and separated on a column (75 × 2.5 cm) of Sephadex G50 medium in pyridine–HOAc–H_2O (4:10:1000, by vol.). Two fractions were obtained, the first of which contained the OPS and was lyophilised (yield: 11.1 mg, 17% of the LPS).

4.3. General and analytical methods

Monosaccharides were analysed as acetylated methyl glycosides.[40] Identification of the 6-deoxyhexose derivative was achieved by analysing the methyl glycoside derivatives of authentic 6-deoxyaltrose. The absolute configuration of the sugars was determined by GC–MS of either their chiral 2-octyl or their chiral 2-butyl derivatives.[41,42] Methylation analysis was performed according to Ciukanu and Kerek.[43] Briefly, the polysaccharide was methylated with iodomethane (CH_3I, 200 μL) in DMSO (1 mL) and powdered NaOH (100 mg), hydrolysed with 2 M trifluoroacetic acid (200 μL, 100 °C, 4 h), carbonyl-reduced with $NaBD_4$ (5 mg, 20–22 °C, 1 h), acetylated (Pyr 150 μL and Ac_2O 150 μL) and analysed by GC–MS. SDS–PAGE was performed as described.[44]

4.4. NMR Spectroscopy

1D and 2D NMR spectra were recorded with a Varian Inova 500 spectrometer of the Consortium INCA (L488/92, Cluster 11) operating at 318 K with an inverse *z*-gradient probe. The sample (11.1 mg) was solved in 0.6 mL ^2H_2O, and chemical shifts were referred relative to internal acetone [δ^1H 2.225, δ^{13}C 31.45]. For the homonuclear experiment, and the solvent-saturated DQF-COSY, TOCSY and NOESY spectra, 512 FIDs of 2048 complex data points were collected, with 48 scans per FID and using standard manufacturer software. The spectral width was set to 10 δ, the frequency carrier was placed at the residual HOD peak, and mixing times of 120 and 200 ms were used for TOCSY and NOESY, respectively. For the gHSQC and gHMBC gHSQC-TOCSY spectra, 512 FIDS of 2048 complex points were acquired with 50 scans per FID (100 for gHSQC-TOCSY), and the GARP sequence was used for ^{13}C decoupling during acquisition. Similar conditions were adopted for the gHSQC-TOCSY sequence. The latter was set to exclude the detection of the direct signals, and ^{13}C decoupling during acquisition was thus avoided. The TOCSY mixing time was set to 120 ms. Conversion of the Varian data and processing was performed with the standard Bruker TOPSPIN 1.3 program, and the spectra were assigned using the computer program PRONTO.[45]

4.5. DNA preparation and sequencing

Chromosomal DNA from a 10-mL overnight broth of strain R80 was extracted and purified as described.[46] Short-range PCR, using O:1b-based primers and high-fidelity VENT Taq Polymerase (NEB), was used to amplify the O:11 regions previously identified as being O:1b-like,[29] such that the gene cluster fragments overlapped and ranged between 1 and 2 kb in length. For PCR, the following cycling conditions were typically used: 94 °C for 2 min, followed by 30 cycles of 94 °C for 30 s, 54–60 °C for 30 s and 72 °C for 1 to 2 min, depending on the expected product size, and a final cycle of 72 °C for 5 min. Long-range PCR was used to amplify non-O:1b-like regions between *ddhA* and *wbyH*, using the same conditions as for short-range PCR but with a 3–4 min extension. PCR products were purified using the Ultraclean PCR cleanup kit (Mo Bio Laboratories) and then sequenced at the Sydney University and Prince Alfred Macromolecular Analysis Centre in Sydney, Australia, using Applied Biosciences 377 automated DNA sequencers. Primer sequences are available on request.

4.6. DNA Sequence analysis

From the initial sequence data obtained, primers were designed for gap-filling and sequence cleanup to complete the gene cluster and to ensure that both the forward and reverse strands were sequenced to minimise PCR and sequencing errors. Sequences were analysed and compiled using the PHRAP/PHRED package from the University of Washington Genome Centre and CONSED,[47] Sequences were submitted to ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html), and potential open reading frames (Orfs) were submitted to BLAST.[48] O-Antigen genes were named in accordance with the Bacterial Polysaccharide Genes Database.[49] Gene clusters were compared using the Artemis Comparison Tool.[50]

DNA sequences derived from this study have been deposited into the GenBank database under accession numbers FJ798742 and FJ798743.

Acknowledgements

We thank K. Jakob and S. Düpow (RCB) for their technical assistance and Dr. K. T. Duda for her help in fermentation. This work was supported by the funding from the Academy of Finland (Projects 104361 and 50441 to MS).

References

1. Holst, O.; Müller-Loennies, S. In *Comprehensive Glycosciences. From Chemistry to Systems Biology*; Kamerling, J. P., Boons, G.-J., Lee, Y. C., Suzuki, A., Taniguchi, N., Voragen, A. G. J., Eds.; Elsevier: Oxford, 2007; pp 1–67.
2. Raetz, C. R. H.; Whitfield, C. *Annu. Rev. Biochem.* **2002**, *71*, 635–700.
3. Valvano, M. A. *Front. Biosci.* **2003**, *8*, s452–s471.
4. Tsubokura, M.; Aleksic, S. *Contrib. Microb. Immunol.* **1995**, *13*, 99–105.
5. Aleksic, S.; Bockemühl, J.; Wuthe, H.-H. *Contrib. Microb. Immunol.* **1995**, *13*, 55–58.
6. Tsubokura, M.; Otsuki, K.; Sato, K.; Tanaka, M.; Hongo, T.; Fukusjhima, H.; Maruyama, T.; Inoue, M. *J. Clin. Microbiol.* **1989**, *27*, 790–791.
7. Nagano, T.; Kiyohara, T.; Suzuki, K.; Tsubokura, M.; Otsuki, K. *J. Vet. Med. Sci.* **1997**, *59*, 153–158.
8. Knirel, Y. A.; Kochetkov, N. K. *Biochemistry (Moscow)* **1994**, *59*, 1325–1383.
9. Jansson, P.-E. In *Endotoxin in Health and Disease*; Brade, H., Morrison, D. C., Opal, S., Vogel, S. N., Eds.; Marcel Dekker: New York, 1999; pp 155–178.
10. Holst, O. In *Endotoxin in Health and Disease*; Brade, H., Morrison, D. C., Opal, S., Vogel, S. N., Eds.; Marcel Dekker: New York, 1999; pp 115–154.
11. Holst, O. *Trends Glycosci. Glycotechnol.* **2002**, *14*, 87–103.
12. Holst, O. *FEMS Microbiol. Lett.* **2007**, *271*, 3–11.
13. Knirel, Y. A.; Dentovskaya, S. V.; Senchenkova, S. N.; Shaikhutdinova, R. Z.; Kocharova, N. A.; Anisimov, A. P. *J. Endotoxin Res.* **2006**, *12*, 3–9.

14. Skurnik, M.; Zhang, L. *APMIS* **1996**, *104*, 849–872.
15. Skurnik, M. In *Yersinia Molecular and Cellular Biology*; Carniel, E., Hinnebusch, B. J., Eds.; Horizon Bioscience: Wymondham, 2004; pp 215–241.
16. Reeves, P. R.; Pacinelli, E.; Wang, L. *Adv. Exp. Med. Biol.* **2003**, *529*, 199–206.
17. Skurnik, M.; Peippo, A.; Ervelä, E. *Mol. Microbiol.* **2000**, *37*, 316–330.
18. Bengoechea, J. A.; Zhang, L.; Toivanen, P.; Skurnik, M. *Mol. Microbiol.* **2002**, *44*, 1045–1062.
19. Skurnik, M.; Venho, R.; Toivanen, P.; Al-Hendy, A. *Mol. Microbiol.* **1995**, *17*, 575–594.
20. Kondakova, A. N.; Ho, N.; Bystrova, O. V.; Shashkov, A. S.; Lindner, B.; Creuzenet, C.; Knirel, Y. A. *Carbohydr. Res.* **2008**, *343*, 1383–1389.
21. Kondakova, A. N.; Bystrova, O. V.; Shaikhutdinova, R. Z.; Ivanov, S. A.; Dentovskaya, S. V.; Shashkov, A. S.; Knirel, Y. A.; Anisimov, A. P. *Carbohydr. Res.* **2009**, *344*, 152–154.
22. Kondakova, A. N.; Bystrova, O. V.; Shaikhutdinova, R. Z.; Ivanov, S. A.; Dentovskaya, S. V.; Shashkov, A. S.; Knirel, Y. A.; Anisimov, A. P. *Carbohydr. Res.* **2009**, *344*, 531–534.
23. Kondakova, A. N.; Bystrova, O. V.; Shaikhutdinova, R. Z.; Ivanov, S. A.; Dentovskaya, S. V.; Shashkov, A. S.; Knirel, Y. A.; Anisimov, A. P. *Carbohydr. Res.* **2009**, *344*, 405–407.
24. Skurnik, M.; Bengoechea, J. A. *Carbohydr. Res.* **2003**, *338*, 2521–2529.
25. Thorson, J. S.; Lo, S. F.; Liu, H. W. *J. Am. Chem. Soc.* **1993**, *115*, 5827–5828.
26. Thorson, J. S.; Lo, S. F.; Liu, H. W.; Hutchinson, C. R. *J. Am. Chem. Soc.* **1993**, *115*, 6993–6994.
27. Thorson, J. S.; Liu, H. *J. Am. Chem. Soc.* **1993**, *115*, 7539–7540.
28. Thorson, J. S.; Lo, S. F.; Ploux, O.; He, X. M.; Liu, H. W. *J. Bacteriol.* **1994**, *176*, 5483–5493.
29. Bogdanovich, T.; Carniel, E.; Fukushima, H.; Skurnik, M. *J. Clin. Microbiol.* **2003**, *41*, 5103–5112.
30. Isakov, V. V.; Gorshkova, R. P.; Tomshich, S. V.; Ovodov, Y. S.; Shashkov, A. S. *Bioorg. Khim.* **1981**, *7*, 559–562.
31. Finn, R. D.; Tate, J.; Mistry, J.; Coggill, P. C.; Sammut, J. S.; Hotz, H. R.; Ceric, G.; Forslund, K.; Eddy, S. R.; Sonnhammer, E. L.; Bateman, A. *Nucleic Acids Res.* **2008**, *36*, D281–D288.
32. Hanniffy, O. M.; Shashkov, A. S.; Moran, A. P.; Prendergast, M. M.; Senchenkova, S. N.; Knirel, Y. A.; Savage, A. V. *Carbohydr. Res.* **1999**, *319*, 124–132.
33. Senchenkova, S. N.; Shashkov, A. S.; Moran, A. P.; Helander, I. M.; Knirel, Y. A. *Eur. J. Biochem.* **1995**, *232*, 552–557.
34. Ellwood, D. C.; Kirk, G. R. *Biochem. J.* **1971**, *122*, 14P.
35. Hoffman, J.; Lindberg, B.; Brubaker, R. R. *Carbohydr. Res.* **1980**, *78*, 212–214.
36. Gorshkova, R. P.; Kalmykova, E. N.; Isakov, V. V.; Ovodov, Y. S. *Eur. J. Biochem.* **1985**, *150*, 527–531.
37. He, X. M.; Liu, H. *Annu. Rev. Biochem.* **2002**, *71*, 701–754.
38. De Castro, C.; Skurnik, M.; Molinaro, A.; Holst, O. *Innate Immun*., accepted for publication.
39. Westphal, O.; Jann, K. *Methods Carbohydr. Chem.* **1965**, *5*, 83–91.
40. Kamerling, J. P.; Gerwig, G. J. In *Comprehensive Glycoscience: From Chemistry to Systems Biology*; Kamerling, J. P., Boons, G.-J., Lee, Y. C., Suzuki, A., Taniguchi, N., Voragen, A. G. J., Eds.; Elsevier: Oxford, 2007; pp 1–67.
41. Leontein, K.; Lindberg, B.; Lönngren, J. *Carbohydr. Res.* **1978**, *62*, 359–362.
42. Gerwig, G. J.; Kamerling, J. P.; Vliegenthart, J. F. G. *Carbohydr. Res.* **1978**, *62*, 349–357.
43. Ciucanu, I.; Kerek, F. *Carbohydr. Res.* **1984**, *131*, 209–217.
44. Oertelt, C.; Lindner, B.; Skurnik, M.; Holst, O. *Eur. J. Biochem.* **2001**, *268*, 554–564.
45. Kjaer, M.; Andersen, K. V.; Poulsen, F. M. *Methods Enzymol.* **1994**, *239*, 288–308.
46. Maloy, S. R. *Experimental Techniques in Bacterial Genetics*; Jones and Bartlett: Boston, 1990. p 63.
47. Gordon, D.; Abajian, C.; Green, P. *Genome Res.* **1998**, *8*, 195–202.
48. Altschul, A. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. *J. Mol. Biol.* **1990**, *215*, 403–410.
49. Reeves, P. R.; Hobbs, M.; Valvano, M. A.; Skurnik, M.; Whitfield, C.; Coplin, D.; Kido, N.; Klena, J.; Maskell, D.; Raetz, C. R.; Rick, P. D. *Trends Microbiol.* **1996**, *4*, 495–503.
50. Carver, T. J.; Rutherford, K. M.; Berriman, M.; Rajandream, M. A.; Barrell, B. G.; Parkhill, J. *Bioinformatics* **2005**, *21*, 3422–3423.