12-1-2003

Genetic Locus and Structural Characterization of the Biochemical Defect in the O-Antigenic Polysaccharide of the Symbiotically Deficient *Rhizobium etli* Mutant, CE166

Lennart Scott Forsberg
University of Georgia

K. Dale Noel
Marquette University, dale.noel@marquette.edu

Jodie M. Box
Marquette University

Russell W. Carlson
University of Georgia

Accepted version. *Journal of Biological Chemistry*, Vol. 278, No. 51 (December 2003): 51347-51359. DOI.

Genetic Locus and Structural Characterization of the Biochemical Defect in the O-Antigenic Polysaccharide of the Symbiotically Deficient *Rhizobium etli* Mutant, CE166

Replacement of *N*-Acetylquinovosamine with its Hexosyl-4-Ulose Precursor*

Authors: L. Scott Forsberg‡, K. Dale Noel§, Jodie Box§, and Russell W. Carlson‡

Abstract: *The O-antigen polysaccharide (OPS) of Rhizobium etli CE3 lipopolysaccharide (LPS) is linked to the core oligosaccharide via an N-acetylquinovosaminosyl (QuiNAc) residue. A mutant of CE3, CE166, produces LPS with reduced amounts of OPS, and a suppressed mutant, CE166α, produces LPS with nearly normal OPS levels. Both mutants are deficient in QuiNAc production. Characterization of OPS from CE166 and CE166α showed that QuiNAc was replaced by its 4-keto derivative, 2-acetamido-2,6-dideoxyhexosyl-4-ulose. The identity of this residue was determined by NMR and mass spectrometry, and by gas chromatography-mass spectrometry analysis of its 2-acetamido-4-deutero-2,6dideoxyhexosyl derivatives produced by reduction of the 4-keto group using borodeuteride. Mass spectrometric and methylation analyses showed that the 2-acetamido-2,6-dideoxyhexosyl-4-ulosyl residue was 3-linked and attached to the core-region external Kdo III residue of the LPS, the same position as that of QuiNAc in the CE3 LPS. DNA sequencing revealed that the transposon insertion in strain CE166 was located in an open reading frame whose predicted translation product, LpsQ, falls within a large family of predicted open reading frames, which includes biochemically characterized members that are sugar epimerases and/or reductases. A hypothesis to be tested in future work is that lpsQ encodes UDP-2-acetamido-2,6-dideoxyhexosyl-4-ulose reductase, the second step in the synthesis of UDP-QuiNAc from UDP-GlcNAc.*

The bacterial species *Rhizobium etli* forms a nitrogen-fixing symbiosis with bean (*Phaseolus vulgaris*). During the symbiotic interaction of *R. etli* with bean, the bacteria eventually penetrate the root nodule cells via a plant-derived infection thread and are, thus, surrounded by the plant plasma membrane upon entering the nodule cell. The *R. etli* lipopolysaccharide (LPS)[1] plays an essential role in the infection of the host cells. Mutants that lack the O-chain polysaccharide (OPS) of the LPS or have reduced amounts of OPS fail to infect the host cells, resulting in root nodules that are devoid of bacteria and are unable to fix nitrogen (1–3). Similar results have also been observed for LPS mutants of *R. leguminosarum* biovar viciae, and

Bradyrhizobium japonicum (for a review, see Ref. 4).

The complete glycosyl sequence of the *R. etli* parent strain OPS (5) and core region (6, 7) is shown in Fig. 1. The parent OPS consists of exactly five trisaccharide repeating units, composed of one residue each of fucose (Fuc), 3-*O*-methyl-6-deoxytalose (3Me6dTal), and methylglucuronate (GlcAMe). Considerable structural heterogeneity exists due to variability in the degree of endogenous *O*-methylation at fucosyl residues and due to partial loss of *O*-acetyl and methyl ester groups during mild acid hydrolysis or while in solution. As depicted in Fig. 1, the OPS repeating units are "capped" at the nonreducing end by a single residue of tri- or di-*O*-methylfucose (2,3,4-TOMeFuc or 2,3-DOMeFuc). The repeating unit located toward the proximal end of the OPS is attached to a nonrepeating glycosyl sequence, defined as the outer core region, consisting of fucose, mannose, and *N*-acetylquinovosamine (QuiNAc) (residues B–D); the QuiNAc residue of this outer core trisaccharide links the OPS repeating units to O-4 of the 3-deoxy-2-octulosonic acid (Kdo) III residue, itself uniquely located at the distal end of the inner core region.

A mutant of *R. etli* CE3 that produces an LPS with reduced amounts of OPS is CE166. This mutant also fails to invade the root nodule cells (8), and analysis of its LPS showed that it lacked QuiNAc (8). The amount of OPS could be restored to nearly normal levels by suppression with multiple copies of a plasmid containing the *lps*α region, which contains numerous genes encoding for the synthesis of the OPS but does not contain the gene mutated in CE166. The suppressed mutant, called CE166α, is able to form nitrogen-fixing nodules, but they form more slowly and are more widely dispersed on the host root compared with nodules formed by the parent, CE3 (8). The LPS from CE166α, while nearly normal in amount, also does not contain appreciable levels of QuiNAc (8). Thus, these results suggested that QuiNAc, or LPS structural features that require QuiNAc, are essential for normal nodulation of bean. An earlier report by Hrabak *et al.* (9) suggested that QuiNAc plays a role in the lectin-mediated interaction of *R. trifolii* with clover. A high level of OPS-containing LPS is essential for symbiotic infection; therefore, the synthesis of QuiNAc could play a role in regulating the level of OPS on the bacterial surface and, thereby, the symbiotic virulence of *R. etli* CE3.

Since QuiNAc is a key glycosyl residue linking the OPS to the core region of the LPS (see Fig. 1), it was not understood how mutant CE166 or CE166α could produce LPS that contained the OPS but lacked QuiNAc. Here we show that in the LPSs from these mutants, QuiNAc is replaced by 2-acetamido-2,6-dideoxyhexosyl-4-ulose, a likely biosynthetic precursor of QuiNAc synthesis.

Experimental Procedures

Growth of Bacteria

R. etli strain CE3, the Tn5-derived mutant CE166, and the suppressed mutant CE166α were grown in tryptone/yeast extract supplemented with Ca^{2+} (10, 11). Bacteria were harvested by centrifugation at late log/early stationary phase.

Isolation of Lipopolysaccharide and the O-specific Polysaccharide

Lipopolysaccharides were extracted using the hot phenol-water procedure, and purified by ribonuclease treatments and size exclusion chromatography (SEC) on Sepharose CL-4B as described previously (1, 12). The OPSs were isolated from the total LPS by mild hydrolysis in 1.0% acetic acid at 105 ℃ for 2 h. The hydrolysates we re subjected to freeze/thawing (three repetitions), and the precipitated lipid A was removed by ultracentrifugation at 160,000 X *g* for 1 h. The supernatants, containing the solubilized O-polysaccharide and the core region oligosaccharides, were passed through a column of polymyxin-agarose to remove remaining traces of free lipid A and intact LPS (6). The flow-through eluants were then chromatographed on Bio-Gel P-10 (45–90-μm fine, 1.5 X 90 cm) in 50 mM ammonium formate, pH 6.5, to separate the O-polysaccharides from the core oligosaccharides.

Analysis of Glycosyl Residues

Carbohydrate compositions of OPS and derived fractions were determined by gas chromatography-mass spectrometry (GC-MS) (electron impact) analysis of the trimethylsilyl methyl glycosides (13) using a 30-meter DB-5 fused silica capillary column (J&W Scientific). Carbohydrate identities of neutral and amino sugars and the location of endogenous *O*-methyl groups were determined by GC-MS analysis of the alditol acetates, using a 30-meter SP-2330 capillary column (Supelco) (13). Carboxylated glycosyl residues, and those containing keto functions, were also analyzed as the alditol acetates, requiring separate reduction steps in a specific sequence: 1) prereduction of OPS at the C-2 ketosidic position of terminal Kdo residues, using $NaBD_4$ in dilute NH_4OH, yielding the diastereomeric alditols; 2) normal acid hydrolysis (2 M trifluoroacetic acid, 105 ℃, 3 h) to cleave all glycosi dic linkages and cause formation of the Kdo 1,4-lactone (14), reduction at C-1 of the Kdo lactone using $NaBD_4$ in water (14), followed by reduction of all remaining glycosyl anomeric centers with excess $NaBD_4$; 4) methyl esterification of uronic acids (and other remaining carboxylated sugars) using 0.2 M methanolic HCl at 80 ℃ for 30 min; 5) reduction of methyl-esterified sugars with $NaBD_4$ in water; 6) peracetylation. Using this multistep procedure, reduction at C-4 of 4-hexulose residues was most efficient, presumably

due to the use of three separate reduction steps. Results described below indicate that prereduction (step 1), prior to acid hydrolysis and alditol acetate preparation, may be the major factor leading to efficient C-4 reduction of the acid-labile 4-hexulose residues. During the preparation of alditol acetates by the normal procedure, with only one reduction step following acid hydrolysis, the yield of C-4 deuterated derivatives was barely detectable. For glycosidic linkage analysis, OPS samples were prereduced with $NaBD_4$ prior to methylation in order to stabilize the reducing end residue. Subsequently, Kdo and uronic acid linkages were identified by sequential permethylation (Hakomori method), reduction of carboxymethyl groups with lithium triethylborodeuteride (Aldrich), remethylation (Hakomori), mild hydrolysis (0.1 M trifluoroacetic acid, 100 ℃, 30 min) to cleave any internal ketosidi c linkages, reduction ($NaBD_4$), normal hydrolysis (2 m trifluoroacetic acid, 121 ℃, 2 h), conv ersion to the partially methylated alditol acetates (13), and GC-MS analysis. OPS samples that were not prereduced prior to methylation failed to yield any Kdo derivatives (due to alkaline degradation of the reducing end Kdo residues) and also showed low yields of C-4 deuterium incorporation into the 4-hexulosyl residue.

Chemical Modifications

De-*O*-acetylation of O-polysaccharides was performed in 10% aqueous NH_4OH at 23 ℃ for 5 h. The progress of the reaction was followed by ^1H NMR and matrix-assisted laser desorption ionization-time-of-flight mass spectrometry. Total β-elimination (depolymerization) was performed on portions of the native or de-*O*-acetylated OPS as follows. The polysaccharide (8–30 mg) was reduced with $NaBD_4$ (10 mg/3 ml of 0.3 M NH_4OH) overnight to stabilize the reducing end glycosyl residue and then methyl-esterified by treating three times with methanol containing 1–2 drops of aqueous 1 M HCl and evaporating to dryness. The fully esterified polysaccharide was then treated with aqueous NaOH (0.25 M,38 ℃, 18 h), and the products were chromatographed by size exclusion chromatography (SEC) using Bio-Gel P-10 (Bio-Rad) (1.5 X 92 cm) in water, or on a Superdex-Peptide HR 10/30 column (Amersham Biosciences) in 50 mM NH_4HCO_3. Column eluants were monitored by refractive index and for the presence of $\Delta^{4,5}$ hexuronosyl residues by measuring the absorbance at 232 nm. Fractions containing oligosaccharides were combined and analyzed by electrospray ionization-mass spectrometry (ESI-MS) and GC-MS.

Identification of OPS Reducing End Residue by Reduction with Borodeuteride

The glycosyl residue located at the reducing end of the various OPS fractions was identified by reduction, using borodeuteride in NH_4OH as described above, followed by normal hydrolysis and peracetylation as described. Derivatives were analyzed using the SP-2330

column. This procedure yields the α/β peracetates of all internal residues and the deuterated alditol acetate of the reducing end residue.

Mass Spectrometry

ESI-MS was performed on a SCIEX API-III triple quadrapole mass analyzer (PE/SCIEX, Thornhill, Ontario, Canada) operated in the positive ion mode with an orifice potential of 35–50 V (5). Spectra are the accumulation of 10–60 scans collected over the *m*/*z* range 200–2400 with a mass step of 0.2–1.0 atomic mass units at 1 ms/step. OPS were desodiated by dialysis, followed by treatment with a strong cation exchange resin (H^+ form), and then filtered using 0.45-μm nylon spin filters and stored in polypropylene tubes. Underivatized samples were analyzed at a concentration of 2 μg/μl with a flow rate of 3–5 μl/min using a solution of 15% (v/v) methanol in deionized water containing 0.5% (v/v) acetic acid. The predicted molecular weights of the various OPSs and derived oligosaccharides were calculated using the following average incremental mass values, based on the atomic weights of the elements: hexose, 162.1424; hexuronic acid, 176.1259; 4-deoxyhex-4-enopyranosiduronic acid, 158.1106; Kdo, 220.1791; anhy-dro-Kdo, 202.1638; 6-deoxyhexose, 146.1430; mono-*O*-methyl-6-deoxyhexose, 160.1699; di-*O*-methyl-6-deoxyhexose, 174.1968; tri-*O*-methyl-6-deoxyhexose, 188.2236; 2-acetamido-2,6-dideoxyhexose, 187.1955; 2-acetamido-2,6-dideoxyhexosyl-4-ulose, 185.1796; *O*-acetyl, 42.0373; free reducing end, 18.0153.

NMR Spectroscopy

Proton NMR spectra of the native and de-*O*-acylated OPS samples were recorded at 35 ℃ in D_2O at 2 mg/ml. Spectra were recorded on a Varian 500-MHz spectrometer using the Varian software.

DNA Cloning and Sequence Determination

The 10.3-kb EcoRI fragment of CE166 DNA carrying the Tn5 insertion was cloned in vector pBlueScriptKS$^-$ (pBKS) (Stratagene). This plasmid (pET166) was digested with BamHI to allow separate subcloning of the inverted repeats at the termini of Tn5 and the rhizobial DNA connected to each Tn*5* terminus. p166LK is a religated (circularized) BamHI fragment from pET166 carrying most of the pBKS and 2.0 kb of rhizobial DNA connected to the Km resistance-conferring half of Tn*5*. Plasmid p166RS is pBKS in which the BamHI site carries 1.2 kb of rhizobial DNA connected to the other half of the Tn5 from pET166. Sequencing of these subcloned rhizobial DNA fragments was carried out by the Sanger method at the University of Wisconsin (Milwaukee, WI) and the Medical College of Wisconsin sequencing facilities. Primers included sequences near each end of pBKS, near the ends of Tn5, and nucleotide stretches

revealed in the rhizobial DNA. Only sequences that were verified by multiple overlapping sequencing runs are reported here.

Construction of Plasmids pJBQ2 and pJBQ3, Which Complement Allele *lps*Q166

A 1.7-kb stretch of DNA that included the *lps*Q ORF was amplified by PCR, using CE3 DNA as template. The forward primer was 5'-TTCATCAGGTCTCACATGCA-3', and the reverse primer was 5'-CTTCACTTCGGTACTGTTCG-3'. The resulting PCR product was ligated into plasmid pCR2.1 (Invitrogen), yielding plasmid pJBQ1. It was then subcloned into plasmid vectors pFAJ1700 (to give pJBQ2) and pFAJ1708 (to give pJBQ3). These vectors can be transferred to and replicate in *R. etli* (15). pJBQ2 was constructed by inserting an EcoRI fragment from pJBQ1 into the EcoRI site of pFAJ1700. The *lpsQ* PCR product cloned in pJBQ3 was inserted in pFAJ1708 following digestion of both pJBQ1 and pFAJ1708 by enzymes XbaI and SacI such that the *lpsQ* ORF would be transcribed downstream of the *nptII* promoter of pFAJ1708. Plasmids pJBQ2 and pJBQ3 were transferred from *E. coli* INVαF' cells into strain CE166 in mating mixtures containing *E. coli* MT616 and Tra^+ mobilizer plasmid pRK600 (16).

Results

NMR Analyses and Initial Comparison of the Parent and Mutant Strain O-polysaccharides

Evaluation of the glycosyl composition of the three different OPSs using standard derivatization procedures (trimethylsilyl methyl glycosides and alditol acetates of neutral and amino sugars) gave results consistent with previous analyses (5, 8), showing that the OPSs from mutants CE166 and CE166α were greatly diminished in QuiNAc (residue **B** in Fig. 1) compared with the parent strain CE3. The finding of detectable levels of QuiNAc in these preparations differed somewhat from results previously reported for CE166 OPS in which QuiNAc was essentially undetectable (accounting for <5% of parent strain levels) (6). NMR analysis of the OPSs (Fig. 2) showed that the ratio of *O*-acetyl to *N*-acetyl methyl protons of the CE166 OPS was approximately the same as that of the parent; 4.4 *versus* 4.0. A similar ratio was observed for the CE166α OPS (not shown). These results, together with the greatly decreased level of QuiNAc, suggested that both mutant OPSs contained an unidentified glycosyl residue that, like QuiNAc, carries an *N*-acetyl group.

Molecular Size of the Parent and Mutant Strain O-polysaccharides

ESI-MS analysis (Fig. 3) indicated that all three of the OPSs (as released from LPS by mild acid hydrolysis) have a similar molecular mass, each yielding several families of multiply charged ions. In Fig. 3*A*, the parent strain OPS yields two doubly charged ion families (I and II,

from high to lower mass, respectively), which differ by 202 mass units (Δ *m*/*z* = 101), reflecting partial loss of an anhydro-Kdo residue from the reducing terminus of the O-chain. As previously reported (5), this loss occurs in approximately one-third of the OPS molecules during the mild hydrolytic step and is a function of hydrolysis time. Based on the observed mass differences and the known structure of the parent OPS (5), family I ions should consist of molecules terminating in anhydro-Kdo, and family II ions should consist of molecules terminating in QuiNAc.

In Fig. 3*B*, these two ion families were again observed for the CE166 OPS, together with a third ion family (III) not observed in the parent OPS. This new, doubly charged ion family was also observed in the CE166α OPS spectrum (not shown). A predominant mass difference between representative ions from family II and family III is 187 mass units (Δ *m*/*z* = 93.5), suggesting the loss of a QuiNAc residue. Thus, family II ions are consistent with the structure R → Fuc → Man → QuiNAc (where R represents OPS repeating units plus capping residue; see Fig. 1), and family III ions are consistent with an R → Fuc → Man structure. For the mutant samples, the presence of QuiNAc-containing structures was somewhat unexpected, since these OPSs contained greatly reduced levels of QuiNAc. However, as stated above, the NMR spectrum indicated that the level of *N*-acetyl groups in the mutant OPS was very similar to that in the parent OPS, suggesting that the mutant contained an unidentified component, "X," which was *N*-acetylated. Aside from variations arising from *O*-acetyl and methyl group content, the similar *m*/*z* values of the mutant OPS ions with those of the parent imply that compound X would have a mass very similar to that of QuiNAc. If this were the case, then the mutant group I ions would consist of a mixture of R → Fuc → Man → QuiNAc → anhydro-Kdo and R → Fuc → Man → X → anhydro-Kdo components, and the group II ions would arise from the structures R → Fuc → Man → QuiNAc and R → Fuc → Man → X, assuming that X was a direct replacement for the QuiNAc residue.

One explanation for the existence of the group III ions in the mutant spectra is that the unidentified component is somewhat labile to the mild acid hydrolysis conditions and is partially removed during this procedure. A second possibility is that the mutant is able to form an R → Fuc → Man → Kdo sequence rather than the normal R → Fuc → Man → QuiNAc → Kdo due, for example, to the direct addition of Man II to Kdo III (residues **A** and **C** in Fig. 1) by the inner core LpcC mannosyl transferase. If this were the case, then, considering the specificity of LpcC, the Man II residue would be linked to O-5 of Kdo III. However, methylation analysis (not shown) showed that the mutant OPS contained only the expected 4-linked Kdo, making it unlikely that LpcC adds a Man to Kdo III.

Identification of 2-Acetamido-2,6-dideoxyhexosyl-4-ulose

The NMR and mass spectrometry data suggested that the mutant OPSs contained an unidentified, *N*-acetylated component, having a mass very similar to that of QuiNAc. As stated above, composition analyses using standard derivatization procedures failed to reveal any significant differences among the polysaccharides aside from the aforementioned reduction in QuiNAc in both mutant OPSs. Considering the possibility that the unidentified component might also contain a carboxyl or other carbonyl function, in addition to the proposed *N*-acetyl group, the "carboxyl-reduced" alditol acetates were prepared and analyzed. The results (Table I) confirmed that the OPSs from CE166 and CE166α are greatly diminished in QuiNAc, as expected. Interestingly, a series of new glycosyl residues were observed that were present only in the mutant OPSs. These residues were identified from their mass spectra as structural (configurational) isomers of QuiNAc. The GC-MS profiles and representative spectrum of these components are shown in Fig. 4 and are consistent with alditol acetates of various 2-acetamido-2,6-dideoxyhexosyl residues (epimers, having different retention times) each containing a single deuterium atom at C-4. Such residues would result from borodeuteride reduction of a 2-acetamido-2,6-dideoxyhexosyl-4-ulose residue. As shown by methylation analysis (see below), the deuterium was located exclusively at C-4 in all epimer derivatives. Thus, nonstereospecific reduction at C-4 would yield both the 4-deuterio-QuiNAc, (confirmed by its mass spectrum and co-chromatography with authentic QuiNAc) and its *galacto* isomer (*N*-acetyl-4-deuteriofucosamine). The two remaining C-4-deuterated derivatives are derived from reduction of the 4-ulose C-3 epimer, which most likely results from keto-enol isomerization of the 4-ulose residue prior to or during reduction.

The three OPS samples were prereduced (to stabilize the reducing end residue) and subjected to methylation analysis. The parent strain OPS yielded only the expected 3-linked QuiNAc derivative as reported previously (5). However, both mutant OPSs contained a series of new derivatives, representing the epimers of the 3-linked QuiNAc residue. The mass spectra of each derivative revealed that all epimers were deuterated at C-4, and all were 3-linked, each yielding a partially methylated alditol acetate, 2-*N*-methylacetamido-1,3,5-tri-*O*-acetyl-4-*O*-methyl-1,4-di-deuterio-2,6-dideoxyhexitol (having different configurations) as evidenced by primary fragment ions *m*/*z* 159, 289, 276, and 132 (data not shown). Each of the latter three ions is 1 mass unit higher than the corresponding ions derived from the nondeuterated parent strain compound and conclusively localizes deuterium attachment to C-4. Secondary ions *m*/*z* 202 (from 276) and *m*/*z* 90 (from 132) further prove the

location of deuterium to be C-4. Together these results show that a 3-linked 2-acetamido-2,6-dideoxyhexosyl-4-ulose residue occurs in the mutant OPS and is consistent with the conclusion that, in the mutant structure, it has replaced the 3-linked QuiNAc residue of the parental OPS.

Labeling and Identification of the Reducing End Residue

In order to identify the glycosyl residues at the reducing end of the various OPS, samples were reduced with borodeuteride and then hydrolyzed and acetylated, yielding the α/β peracetates of all internal residues and the deuterated alditol of the reducing end residue. A comparison of the results obtained for the parent and mutant strain OPSs is shown in Fig. 5. The parent strain (Fig. 5*B*) yielded the α/β peracetates of QuiNAc (originating from molecules terminating in Kdo) and the alditol of QuiNAc, indicating that a portion of the parent OPS lacked Kdo, thereby terminating at the reducing end with QuiNAc. These results are consistent with the ESI-MS analyses (Fig. 3*A*) showing that the parent OPS preparation contains two families of ions consistent with R → Fuc → Man → QuiNAc → Kdo and R → Fuc → Man → QuiNAc structures. The mutant OPS (Fig. 5*A*) preparations yielded the α/β peracetates of QuiNAc (originating from molecules terminating in Kdo), the alditol of QuiNAc (originating from molecules terminating in QuiNAc), and substantial amounts of the alditol of Man (from molecules terminating in Man). In addition, the QuiNAc alditol detected in the mutant OPS co-eluted with a small percentage of its C-4 deuterated analogue (2-acetamido-4-deuterio-2,6dideoxyglucitol), as evident from the mass spectra (analogous to that shown in Fig. 5). The occurrence of the C-4-deuterated alditol supports the conclusion that the 2-acetamido-2,6dideoxyhexosyl-4-ulose residue is at the same location in the mutant OPS as is QuiNAc in the parent OPS. In the mutants, the presence of OPS molecules terminating in Man (Fig. 3*B*)or Man-ol (following reduction) (Fig. 4) is again attributed to acid lability of the 4-hexulose residue. The low yield of 4-deuterio-HexNAc derivatives observed in the mutant OPS using this protocol is probably due to incomplete reduction of the 4-keto group (due to only a single reduction step) and subsequent degradation, resulting in OPS with the reducing end Man residue. Thus, the unique structures observed in the mutant OPS are R → Fuc → Man, R → Fuc → Man → X, and R → Fuc → Man → X → Kdo.

Depolymerization of the OPSs and Analysis of Oligosaccharides

To further characterize the outer core glycosyl sequence at the reducing end of the mutant OPSs, the OPSs were subjected to alkali-catalyzed β-elimination. It was previously shown that treating the parent strain OPS with alkali causes β-elimination of the uronosyl residues, leading to depolymerization of the repeating units and leaving the nonrepeating

oligosaccharide regions of the OPS, lacking uronic acid, essentially intact (5). In the case of the parent OPS, these nonrepeating, alkali-stable regions consist mainly of glycosyl sequences that originate from the reducing end of the OPS (*e.g.* R → Fuc → Man → QuiNAc → Kdo, where R represents variable portions of a single OPS repeating unit, and Kdo represents the inner core Kdo III residue) (5). Further, partial β-elimination of the 3-linked QuiNAc residue may also occur during this procedure (5); nevertheless, there is enrichment of the outer core oligosaccharide region of the OPS due to degradation of the OPS repeating units.

This procedure was applied to both the CE166 and CE166α OPS, and the resulting alkali-resistant oligosaccharides were fractionated by SEC and then analyzed by ESI-MS, glycosyl composition, and linkage analyses. Two major SEC fractions were obtained for both the CE166 and CE166α mutants, and their ESI-MS spectra are shown in Fig. 6. Each SEC fraction consisted of a number of oligosaccharides, and their observed ions and proposed formulas are summarized in Table II. Both mutant OPS yielded several unique oligosaccharides, in addition to products identified previously in the parent OPS (5). One of these unique oligosaccharides (OS-1) has a mass (*m*/*z* 794.5) supporting a R → Fuc → Man-ol structure, consistent with the data described above. This oligosaccharide was abundant in the mutant spectra, particularly in the lower molecular weight SEC fraction (Fig. 6*B*), and was detected only at low levels in the parent strain spectrum (not shown). Other oligosaccharides, detected only in the mutant spectra, include the species at *m*/*z* 982.5 (OS-3, R → Fuc → Man → 6-deoxy-4-deuterio-HexNAc-ol), *m*/*z* 1181.8 (OS-7, R → Fuc → Man → 6-deoxy-4-deuterio-HexNAc → anhydro-Kdo), and *m*/*z* 1202.8 (OS-9, R → Fuc → Man → 6-deoxy-4-deuterio-HexNAc →Kdo-ol) (Fig. 6*A*). These ions are shifted 1 mass unit higher than their parent strain counterparts, consistent with the presence of a 4-C deuterio analogue of QuiNAc. The nondeuterated oligosaccharides, which contain QuiNAc, are also observed in the mutant spectra (compare OS-6 at *m*/*z* 1180.8 *versus* OS-7 at *m*/*z* 1181.8, and compare OS-8 at *m*/*z* 1201.8 *versus* OS-9 at *m*/*z* 1202.8). The presence of a 6-deoxy-4-deuterio-HexNAc residue in OS-7 and OS-9 is consistent with the results described above that show borodeuteride reduction at C-4 of the hex-4-ulosyl residue. In Fig. 6, the spectrum shows that the ion intensity of R → Fuc → Man → QuiNAc → anhydro-Kdo (*m*/*z* 1180.8) is greater than that of R → Fuc → Man → 6-deoxy-4-deuterio-HexNAc → anhydro-Kdo (*m*/*z* 1181.8). This result was not consistent with the composition data, which show that the mutant OPS contains only a low level of QuiNAc. However, as described above, a possible explanation of this apparent discrepancy is that the 4-keto group of 2-acetamido-2,6-dideoxyhexosyl-4-ulose is only partially reduced during the single prereduction

step prior to the β-elimination and ESI analysis, and the remaining nonreduced hex-4-ulose residue is subsequently degraded during the alkaline treatment. Since oligosaccharides containing the 2-acetamido-2,6-dideoxyhexosyl-4-ulose residue would show a decrement of 2 mass units compared with their normal counterparts (*i.e.* those containing QuiNAc), degradation of the 2-acetamido-2,6-dideoxyhexosyl-4-ulose residue would also account for the absence of any molecular species having a mass decrement of 2 mass units.

The Genetic Locus Mutated in Strain CE166

DNA carrying the Tn5 insertion in CE166 was cloned, and the sequence of a stretch of 2971 nucleotides was determined, as summarized in Fig. 7*A*. The Tn5 had inserted within an ORF designated as *lpsQ* that encodes a hypothetical polypeptide of 309 amino acids. The predicted LpsQ amino acid sequence falls within the large WcaG family of epimerases and reductases involved in deoxyhexose synthesis (www.ncbi.nlm.nih.gov/Structure/cdd/). Matched against ORF translation products whose functions have been studied by mutation analysis, LpsQ shows the greatest sequence alignment with WbpV and WbpK of *Pseudomonas aeruginosa* serotypes O6 and O5 (17). All three predicted proteins align throughout their sequences (Fig. 7*B*), the closest alignment being between WbpV and LpsQ (47% identity, 64% similarity). As noted in Fig. 7B, LpsQ and its homologs have the G*XX*G*XX*G motif for binding NAD(P) and a triad of amino acids (SX_nYX_3K) that are conserved among ORFs in the WcaG family. These latter three amino acids are at the active sites of WcaG (GDP-4-keto-6-deoxy-D-mannose epimerase/reductase) (18), GalE (UDP-galactose 4-epimerase) (19), and most of the studied examples within the larger family of short-chain dehydrogenases/reductases (20).

lpsQ is flanked by two ORFs homologous to highly conserved eubacterial genes, *murB* and *ddlB*, whose transcription is oriented oppositely to that of *lpsQ* (Fig. 7*A*). *murB* and *ddlB* are involved in peptidoglycan synthesis and are generally contiguous in the genomes of other α-proteobacteria (21–24); *murB* encodes UDP-N-acetylmuramate dehydrogenase, and *ddlB* encodes D-alanylalanine synthase. The 63 amino acids encoded by the deduced *ddlB* fragment were 76% identical and 87% similar to the N-terminal 63 amino acids encoded by the presumed *ddlB* of *Sinorhizobium meliloti* (25). The 279 amino acids encoded by the deduced *murB* fragment were 76% identical and 87% similar to the corresponding C-terminal amino acids encoded by the presumed *murB* of *S. meliloti* (25).

A PCR product from wild-type *R. etli* strain CE3 containing *lpsQ* that extends from nucleotide 260 to 1969 in Fig. 7*A*, was cloned into broad host vectors pFAJ1700 and FAJ1708. When either construct was transferred into strain CE166, normal LPS I abundance, normal

QuiNAc content, and normal symbiotic proficiency were restored (data not shown).

Discussion

The OPSs released by mild acid hydrolysis of the LPSs from the *R. etli* CE3 mutants, CE166 and CE166α, were shown to have the following structures: R → Fuc → Man → QuiNAc → Kdo (**1**), R → Fuc → Man → QuiNAc (**2**), R → Fuc → Man → X → Kdo (**3**), R → Fuc → Man → X(**4**), and R → Fuc → Man (**5**), where R represents the repeating unit oligosaccharide and capping residue as shown in Fig. 1 and where X has been identified as 2-acetamido-2,6-dideoxyhexosyl-4-ulose. The low levels of QuiNAc in these mutant LPSs imply that structures **1** and **2** are present in minor amounts. Structures **2** and **4** are due to the elimination of Kdo from a portion of the molecules during the mild acid hydrolytic release of OPS from the LPS; partial loss of this Kdo has also been reported as occurring during isolation of the parent strain OPS (5). Structure **5** is probably due to the elimination of the 2-acetamido-2,6-dideoxyhexosyl-4-ulose residue, also during the mild acid hydrolysis procedure. From these results it can be concluded that the *R. etli* CE3 mutants, CE166 and CE166α, produce LPSs in which the outer core oligosaccharide has two structures, **6** and **7**, shown in Fig. 8. Structure **6** is normally found in the OPS from the parent strain, whereas structure **7** is found in the mutant OPSs and is not present in the parent. In structure **7**, the QuiNAc is replaced by 2-acetamido-2,6-dideoxyhexosyl-4-ulose. Although rare in polysaccharides, 2-acetamido-2,6-dideoxy-D-*xylo*-hexos-4-ulose was first identified as a component of the pentasaccharide repeating unit in the capsular polysaccharide synthesized by *Streptococcus pneumoniae* type 5 (26). In LPS, two occurrences of this glycose have been reported, as a component of the tetrasaccharide repeating unit in the OPS from *Vibrio ordalii* serotype O:2 (27) and recently in the OPS repeating unit from *Flavobacterium columnare* ATCC 43622 (28). In all reported occurrences, the glycose is 3-linked and is part of a repeating unit that comprises a polysaccharide. In the case of the *R. etli* CE166 mutant, the glycose occurs as a single residue in the outer core portion and is not part of the OPS repeating unit. To our knowledge, this is the first example of a core region glycose residue being replaced by its biosynthetic precursor.

The relative amounts of these two structures present in these LPS preparations from CE166 and CE166α can be estimated from the composition data shown in Table II. This calculation is based on the assumption that the OPS from the parent exists exclusively as structure **6** and, therefore, that the level of QuiNAc in the parent OPS represents the total molar

fraction amount of QuiNAc present in an OPS preparation in which all molecules contain QuiNAc. Thus, in CE166 the percentage of structure **6** can be calculated from the mutant to parent molar QuiNAc ratio (see Table I), 0.13/0.62 or 21%, which implies that structure **7** comprises the remainder of the CE166 LPS, or 79%. Similarly, in the CE166α OPS, the percentages of structures **6** and **7** are 10% (0.06/0.62), and 90%, respectively.

The presence of small amounts of structure **6** in the mutant OPSs is a slightly different result from what would have been predicted from our previous report (8) in which the LPSs from CE166 and CE166α did not contain detectable levels of QuiNAc. One possible explanation for this difference in the two preparations is reversion. Initial studies with this mutant indicated that at low frequency it underwent an unusual type of reversion in which the Tn*5* ended up in another replicon of the bacterium (29). Later studies could not reproduce this result. A second related possibility is extragenic suppression. Suppressor mutations may allow detectable QuiNAc synthesis by occurring in a second, cryptic copy of *lpsQ* or a gene encoding a 4-reductase normally involved in synthesis of another deoxysugar. Both of these possibilities would be consistent with not observing QuiNAc in some cases (8) and observing a small amount in other cases (3). It should also be noted that the CE166α LPS, while containing almost parental levels of OPS, has much less QuiNAc than does the mutant CE166 LPS. Thus, whereas the current results are not identical to those previously reported with regard to the level of QuiNAc, it is clear that the *lps-166* mutation results in an LPS that is very deficient in the level of QuiNAc and, therefore, in the amount of the normal parental structure **6**.

In the major mutant structure **7**, the QuiNAc residue normally present in the parental LPS (*i.e.* structure **6**) has been replaced by its presumed 4-keto precursor, 2-acetamido-2,6-dideoxyhexosyl-4-ulose. Since QuiNAc is an essential sugar involved in joining the OPS to the core region of the LPS, the presence of the 4-keto precursor in place of QuiNAc suggests that the UDP-QuiNAc transferase is able to use UDP-2-acetamido-2,6-dideoxyhexosyl-4-ulose as a substrate. The lower level of OPS in CE166 compared with that in the parent CE3 would imply that this transferase uses UDP-*N*-acetyl-2-amino-6-deoxyhexosyl-4-ulose less efficiently than UDP-QuiNAc; *i.e.* it is likely not to recognize the 4-ulose as well as QuiNAc. The fact that CE166α, which contains multiple copies of the *lps α*-region, produces LPS with close to the normal level of OPS but containing *N*-acetyl-6-deoxyhexosyl-4-ulose in place of QuiNAc suggests that this transferase may be located on the *lps α*-region, and its overexpression allows more OPS to be made and subsequently added to the LPS core region.

The suggested biosynthetic pathway of D-QuiNAc is shown in Fig. 9. In this pathway, UDP-GlcNAc is converted into UDP-QuiNAc by two enzymatic activities: a 4,6-dehydratase, which forms UDP-*N*-acetyl-6-deoxyhexosyl-4-ulose from UDP-Glc-NAc, and a reductase, which stereospecifically reduces UDP-*N*-acetyl-6-deoxyhexosyl-4-ulose to UDP-QuiNAc. This work provides evidence for the assignment of LpsQ to the second step of QuiNAc synthesis (Fig. 9), since the *lpsQ* mutation allows the synthesis of the keto intermediate and highly abrogates the production of QuiNAc itself.

In vitro studies with two proteins have indicated that a single enzyme may catalyze both of the above steps in the synthesis of D-QuiNAc from D-GlcNAc (30–32). These two proteins are encoded by genes *wbpM* of *P. aeruginosa* (30) and *flaA1* of *Helicobacter pylori* (31). However, the predicted LpsQ ORF exhibits very little sequence similarity to these proteins. In paired comparisons by the Blast2 program (www.ncbi.nlm.nih.gov/blast/bl2seq/bl2.html), the FlaA1 sequence yields an alignment with LpsQ of low significance (“expect value” e-3), and this program finds no significant overall homology between LpsQ and WbpM. By contrast, FlaA1 and WbpM sequences align in a clearly significant manner (giving an expect value of 2e-25). They would appear to be in a different subfamily of the NDP-sugar dehydratase/epimerase/reductases within the larger family of short-chain dehydrogenase/reductases (20). There is also a recently proposed pathway for synthesis of L-QuiNAc, involving five steps catalyzed *in vitro* by three enzymes (33). In Blast2 and ClustalW analysis, LpsQ shows little or no significant sequence similarity to any of the proteins thus far assigned to this pathway.

Despite the existence of the apparently bifunctional enzymes WbpM and FlaA1, it still seems plausible that QuiNAc may be synthesized through the action of two separate enzymes, a dehydratase and a reductase (Fig. 9). However, if such is the case in *R. etli*, it is surprising that the gene for the first step (the hydratase) appears not to be contiguous with *lpsQ*. It was already surprising that *lpsQ* was not closely linked with other genes of *R. etli* CE3 that are required for its LPS core and O-antigen (3, 8, 34). However, it should be pointed out that many of the genes for LPS core synthesis in *R. etli* have not been identified. It also should be pointed out that *R. etli* appears to be especially prone to gene duplications and transpositions of small and long segments of DNA (29). There may be another (less active) copy of *lpsQ*, which is contiguous with an unduplicated gene that specifies the first step in Fig. 9. *lpsQ* appears to reside within a suite of cell wall synthesis genes. Generally, *murB* and *ddlB* are contiguous in other α-proteobacteria; however, there is precedent for an insertion between them. In *S. meliloti*, two hypothetical genes have been inserted between *ddlB* and *murB* (25). These two hypothetical

genes, coding for an aquaporin and a possible transport protein, as with *lpsQ*, are also not related to cell wall synthesis proteins. It is possible that this site may be a target for a type of transposition event that inserts various genes.

A number of subtle structural changes have been reported to occur to the OPS portion of *R. etli* CE3 LPS during symbiosis. These changes have been detected using monoclonal antibodies that are specific to the OPS portion of the LPS and involve methylation of the Fuc residues (8, 13, 35). During nodulation, the reactivity of the OPS for two of four mAbs greatly decreases (36). The binding of these mAbs involves, in part, the terminal capping TOMFuc residue (35) and a 2MeFuc residue[2] either in the repeat unit or in the outer core region (see the structure shown in Fig. 1). The CE166 and CE166α mutants do not differ from the parent with regard to their ability to bind these mAbs, and therefore, there apparently is no difference from the parent OPS in these fucosyl residues. Indeed, strain CE166α (8) produces parental levels of OPS, which, while deficient in QuiNAc, has a structure that is otherwise the same as the parent CE3 OPS as reflected by composition, mass spectrometric, NMR, and monoclonal antibody binding data (this report) (8). This very minor chemical change (*i.e.* replacement of QuiNAc with its 4-keto precursor) results in an altered nodulation phenotype (*i.e.* nodulation is delayed, and the distribution of nodules is altered) (8), suggesting that it produces a structural change that is significant enough to affect its role in symbiosis. In addition, this result suggests that structural features of this outer core region of CE3 LPS, in addition to changes in the methylation pattern of TOMFuc and Fuc residues, are involved in establishing symbiosis with the host legume.

Notes

- From the ‡Complex Carbohydrate Research Center, University of Georgia, Athens, Georgia 30602 and the §Department of Biological Sciences, Marquette University, Milwaukee, Wisconsin 53233

- This work was supported in part by National Institutes of Health Grant GM39583 (to R. W. C.), by Department of Energy (DOE) Grant DE-FG09-93ER20097 (to the Complex Carbohydrate Research Center), and by DOE Grant DE-FG02-98ER20307 (to K. D. N.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “*advertisement*” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.
- The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) AY391267.

- To whom correspondence should be addressed. Tel.: 706-542-4439; Fax: 706-542-4412;

E-mail: rcarlson@ccrc.uga.edu.

- [1]The abbreviations used are: LPS, lipopolysaccharide; OPS, O-specific polysaccharide; Kdo, 3-deoxy-D-*manno*-2-octulosonic acid; QuiNAc, 2-acetamido-2,6-dideoxyglucose (*N*-acetylquinovosamine); 3Me6dTal, 3-*O*-methyl-6-deoxytalose; SEC, size exclusion chromatography; GC-MS, gas-liquid chromatography-mass spectrometry; ESI-MS, electrospray ionization-mass spectrometry; Fuc, fucose; GlcAMe, methylglucuronate; ORF, open reading frame.
- [2] K. D. Noel, J. M. Box, and V. J. Bonne, manuscript in preparation.

Acknowledgments

- Former Marquette University student Kevin Barleben is acknowledged for isolating plasmid pET166. We also thank Matthew Hilley for technical assistance.

References

1. Carlson, R. W., Kalembasa, S., Turowski, D., Pachori, P., and Noel, K. D. (1987) *J. Bacteriol.* 169, 4923–4928
2. Noel, K. D., VandenBosch, K. A., and Kulpaca, B. (1986) *J. Bacteriol.* 168**,** 1392–1401
3. Cava, J. R., Elias, P. M., Turowski, D. A., and Noel, K. D. (1989) *J. Bacteriol.* 171, 8–15
4. Kannenberg, E. L., Reuhs, B. L., Forsberg, L. S., and Carlson, R. W. (1998) in *The Rhizobiaceae: Molecular Biology of Model Plant-associated Bacteria* (Spaink, H. P., Kondorosi, A., and Hooykaas, P. J. J., eds) Kluwer Academic Publishers, Dordrecht, The Netherlands
5. Forsberg, L. S., Bhat, U. R., and Carlson, R. W. (2000) *J. Biol. Chem.* 275, 18851–18863
6. Forsberg, L. S., and Carlson, R. W. (1998) *J. Biol. Chem.* 273, 2747–2757
7. Bhat, U. R., Bhagyalakshmi, S. K., and Carlson, R. W. (1991) *Carbohydr. Res.* 220, 219–227
8. Noel, K. D., Forsberg, L. S., and Carlson, R. W. (2000) *J. Bacteriol.* 182, 5317–5324
9. Hrabak, E. M., Urbano, M. R., and Dazzo, F. B. (1981) *J. Bacteriol.* 148, 697–711
10. Beringer, J. E. (1974) *J. Gen. Microbiol.* 84, 188–198
11. Sherwood, M. T. (1970) *J. Gen. Microbiol*. 71, 351–358
12. Carlson, R. W., Sanders, R. E., Napoli, C., and Albersheim, P. (1978) *Plant Physiol.* 62, 912–917
13. York, W. S., Darvill, A. G., McNeil, M., Stevenson, T. T., and Albersheim, P. (1985)

Methods Enzymol. 118, 3–40
14. York, W. S., Darvill, A. G., McNeil, M., and Albersheim, P. (1985) *Carbohydr. Res.* 138, 109–126
15. Dombrecht, B., Vanderleyden, J., and Michiels, J. (2001) *Mol. Plant-Microbe Interact.* 14, 426–430
16. Finan, T. M., Kunkel, B., De Vos, G. F., and Signer, E. R. (1986) *J. Bacteriol.* 167, 66–72
17. Belanger, M. (1999) *Microbiology* 145, 3505–3521
18. Rocchetta, H. L., Burrows, L. L., and Lam, J. S. (1999) *Microbiol. Mol. Biol. Rev.* 63, 523–553
19. Thoden, J. B., Frey, P. A., and Holden, H. M. (1996) *Biochemistry* 35, 2557–2566
20. Jornvall, H., Persoon, B., Krook, M., Atrian, S., Gonzalez-Duarte, R., Jeffrey, J., and Ghosh, D. (1995) *Biochemistry* 34, 6003–6013
21. Andersson, S. G. E., Zomorodipour, A., Andersson, J. O., Sicheritz-Ponen, T., Alsmark, U. C. M., Podowski, R. M., Naslund, A. K., Eriksson, A.-S., Winkler, H. H., and Kurland, C. G. (1998) *Nature* 396, 133–143
22. Kaneko, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiumi, T., Sasamoto, S., Watanabe, A., Idesawa, K., Iriguchi, M., Kawashima, K., Kohara, M., Matsumoto, M., Shimpo, S., Tsuruoka, H., Wada, T., Yamada, M., and Tabata, S. (2002) *DNA Res.* 9, 189–197
23. Kaneko, T., Nakamura, Y., Asamizu, E., Kato, T., Sasamoto, S., Watamabe. A, Idesawa. K, Ishikawa. A, Kawashima, K., Kimura, Y., Kishida, Y., Kiyokawa, C., Kohara, M., Matsumoto, M., Matsuno, A., Mochizuki, Y., Nakayama, S., Nakazaki, N., Shimpo, S., Sugimoto, M., Takeuchi, C., Yamada, M., and Tabata, S. (2000) *DNA Res.* 7, 331–338
24. DelVecchio, V. G., Kapatral, V., Redkar, R. J., Patra, G., Mujer, C., Los, T., Ivanova, N., Anderson, I., Bhattacharyya, A., Lykidis, A., Reznik, G., Jablonski, L., Larsen, N., D'Souza, M., Bernal, A., Mazur, M., Goltsman, E., Selkov, E., Elzer, P. H., Hagius, S., O'Callaghan, D., Letesson, J. J., Haselkorn, R., Kyrpides, N., and Overbeek, R. (2002) *Proc. Natl. Acad. Sci. U. S. A.* 99, 443
25. Capela, D., Barloy-Hubler, F., Gouzy, J., Bothe, G., Ampe, F., Batut, J., Boistard, P., Becker, A., Boutry, M., Cadieu, E., Dreano, S., Gloux, S., Godrie, T., Goffeau, A., Kahn, D., Kiss, E., Lelaure, V., Masuy, D., Pohl, T., Portetelle, D., Puhler, A., Purnelle, B., Ramsperger, U., Renard, C., Thebault, P., Vandenbol, M., Weidner, S., and Galibert, F. (2001) *Proc. Natl. Acad. Sci. U. S. A.* 98, 9877–9882

26. Jansson, P.-E., Lindberg, B., and Lindquist, U. (1985) *Carbohydr. Res.* 101–110
27. Sadovskaya, I., Brisson, J. R., Khieu, N. H., Mutharia, L. M., and Altman, E. (1998) *Eur. J. Biochem.* 253, 319–327
28. MacLean, L. L., Perry, M. B., Crump, E. M., and Kay, W. W. (2003) *Eur. J. Biochem.* 270, 3440–3446
29. Cava, J. R. (1988) *Genetic Analysis of Lipopolysaccharide and Purine Biosynthesis and Their Requirement for Nodulation in Rhizobium leguminosarum Strain CFN42*. Ph.D. thesis, Marquette University, Milwaukee, WI
30. Creuzenet, C., and Lam, J. S. (2001) *Mol. Microbiol.* 41, 1295–1310
31. Creuzenet, C., Schur, M. J., Li, J., Wakarchuk, W. W., and Lam, J. S. (2000) *J. Biol. Chem.* 275, 34873–34880
32. Creuzenet, C., Urbanic, R. V., and Lam, J. S. (2002) *J. Biol. Chem.* 277, 26769–26778
33. Kneidinger, B., Larocque, S., Brisson, J.-R., Cadotte, N., and Lam, J. S. (2003) *Biochem. J.* 371, 989–995
34. Cava, J. R., Tao, H., and Noel, K. D. (1990) *Mol. Gen. Genet.* 221, 125–128
35. Duelli, D. M., Tobin, A., Box, J. M., Carlson, R. W., and Noel, K. D. (2001) *J. Bacteriol.* 183, 6054–6064
36. Tao, H., Brewin, N. J., and Noel, K. D. (1992) *J. Bacteriol.* 174, 2222–2229
37. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) *Nucleic Acids Res.* 22, 4673–4680
38. Rosano, C., Bisso, A., Izzo, G., Tonetti, M., Sturla, L., DeFlora, A., and Bolognesi, M. (2000) *J. Mol. Biol.* 303, 77–91

Appendix

Table 1

Glycosyl composition of R. etli parent and mutant strain O-polysaccharides as determined by GC-MS analysis of the derived, carboxyl-reduced, alditol acetates

Glycosyl reside	Molar ratio[a]		
	CE3	CE166	CE166-α
2,3,4-MeFuc	0.98	0.59	0.90
2,3-MeFuc	0.0	0.64	0.12
2-MeFuc	0.63	0.88	0.52
3-Me-6-deoxyTal[b]	4.30	4.10	3.81
Fuc	3.87	3.56	2.80
Man	1.00	1.00	1.00
GlcA[c]	3.84	3.98	3.08
Kdo[d]	0.71	0.93	1.04
QuiNAc	0.62	0.13	0.06
6-dexoxy-4-D-HexNAc[e]		0.38	0.50

[a]Ratios normalized to Man, determined from the GC-MS total ion current peak areas of the alditol acetates with response factor correction from authentic standards. Partially methylated standards of 6-deoxytalose were obtained from a *Streptococcal* type-specific polysaccharide, and the *R. etli* parent strain OPS served as a source of authentic QuiNAc as previously described (5).

[b]3-Me-6-deoxyTal, 3-*O*-methyl 6-deoxytalose.

[c]Determined as the carboxyl-reduced alditol acetate, containing two deuteriums at C-6.

[d]Determined as the carboxyl-reduced alditol acetate, containing two deuteriums at C-1.

[e]D-HexNAc, 2-acetamido-2,6-dideoxyhexose, determined as the keto-reduced alditol acetate, derived from the corresponding hex-4-ulose and containing one deuterium atom at C-4

Table 2

Ions derived from electrospray analysis and proposed formulas of the reducing end oligosaccharides produced from β-elimination of the R. etli mutant strain CE166 OPS

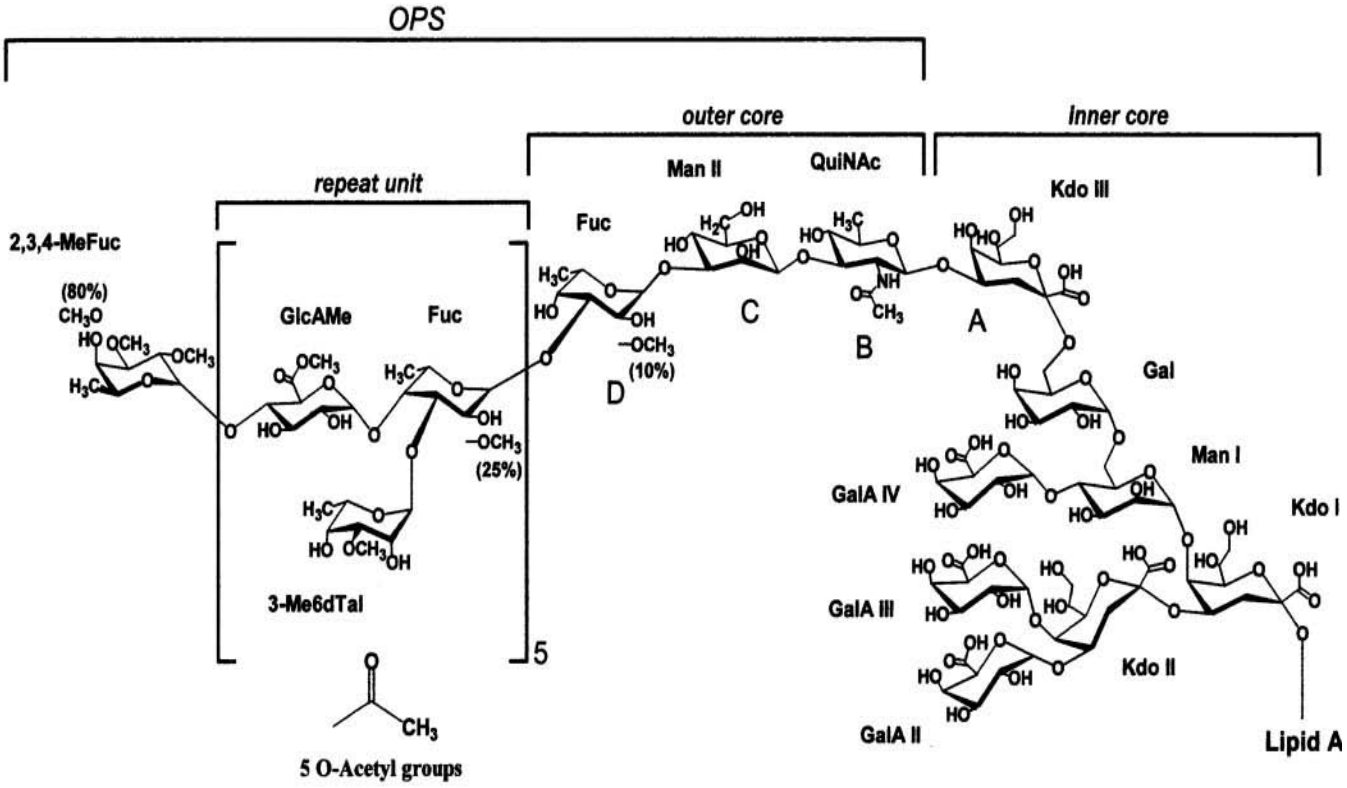
Observed ion[a]	Experimental mass		Calculated mass and proposed formula[b]	Predicted ion[c]
794.5	793.5	OS-1	$M + H^{+1}$	794.8
		793.8 + H^{+1}	[$\Delta^{4,5}$GlcA$_1$3Me6dTal$_1$Fuc$_2$Man-ol$_1$] + H^+	
981.5	980.5	OS-2	$M + H^{+1}$	982.0
		980.9 + H^{+1}	[$\Delta^{4,5}$GlcA$_1$3Me6dTal$_1$Fuc$_2$Man$_1$QuiNAc-ol$_1$] + H^+	
982.5	981.5	OS-3	$M + H^{+1}$	983.0
		981.9 + H^{+1}	[$\Delta^{4,5}$GlcA$_1$3Me6dTal$_1$Fuc$_2$Man$_1$4-D-HexNAc-ol$_1$] + H^+	
1020.5	1019.5	OS-4	$M + H^{+1}$	1020.9
		1019.9 + H^{+1}	[$\Delta^{4,5}$GlcA$_1$Fuc$_2$Man$_1$QuiNAc$_1$2,7anhydKdo$_1$] + H^+	
1041.5	1040.5	OS-5	$M + H^{+1}$	1042.0
		1040.9 + H^{+1}	[$\Delta^{4,5}$GlcA$_1$Fuc$_2$Man$_1$QuiNAc$_1$Kdo-ol$_1$] + H^+	
1180.8	1179.8	OS-6	$M + H^{+1}$	1181.1
		1180.1 + H^{+1}	[$\Delta^{4,5}$GlcA$_1$3Me6dTal$_1$Fuc$_2$Man$_1$QuiNAc$_1$2,7anhydKdo$_1$] + H^+	
1181.8	1180.8	OS-7	$M + H^{+1}$	1182.1
		1181.1 + H^{+1}	[$\Delta^{4,5}$GlcA$_1$3Me6dTal$_1$Fuc$_2$Man$_1$4-D-HexNAc$_1$2,7anhydKdo$_1$] + H^+	
1201.8	1200.8	OS-8	$M + H^{+1}$	1202.1
		1201.1 + H^{+1}	[$\Delta^{4,5}$GlcA$_1$3Me6dTal$_1$Fuc$_2$Man$_1$QuiNAc$_1$Kdo-ol$_1$] + H^+	
1202.8	1201.8	OS-9	$M + H^{+1}$	1203.1
		1202.1 + H^{+1}	[$\Delta^{4,5}$GlcA$_1$3Me6dTal$_1$Fuc$_2$Man$_1$4-D-HexNAc$_1$Kdo-ol$_1$] + H^+	

[a]A mass step of 0.2 atomic mass units was used during analysis.

[b]Calculated from the average incremental mass of the component glycosyl residues as described under "Experimental Procedures." The proposed formulas of OS-6 and OS-8 are identical to those detected in the parent OPS (5). Glycosyl composition and linkage analysis of the mutant SEC fractions resulting from β-elimination confirmed the presence of 3-linked Man-ol, 3-QuiNAc-ol, and 3-(4-D-HexNAc) in the appropriate SEC fractions (data not shown). Oligosaccharides OS-4 and OS-5 appear to arise from OS-6 and OS-8, respectively, due to loss of a terminal 3Me6dTal residue; the 3Me6dTal is linked 1 → 3 to fucosyl residues in OS-6 and OS-8 (5). $\Delta^{4,5}$GlcA, 4-deoxyhex-4-enopyranosiduronate; 3Me6dTal, 3-*O*-methyl-6-deoxytalose; 2,7anhydKdo, 2,7-anhydro-Kdo*f*; Kdo-ol, Kdo reduced at C-2 with borodeuteride; Man-ol, Man reduced at C-1 with borodeuteride; QuiNAc-ol, QuiNAc reduced at C-1 with borodeuteride, etc.

[c]The predicted singly charged ion as determined from the proposed formula.

Figure 1
The complete glycosyl sequence of the carbohydrate portion of *R. etli* CE3 LPS



The OPS portion contains a nonrepeated trisaccharide at its reducing end defined as the outer core (residues *B–D*) that links the repeating units to the inner core region. Mild hydrolytic treatment of the LPS cleaves at ketosidic linkages, and the released OPS terminates in Kdo (residue A). The extent of *O*-methylation of the fucosyl residues is indicated (5).

Figure 2
The proton NMR spectra of the OPS from *R. etli* CE3 LPS (*A*) and the OPS from *R. etli* CE166 LPS (*B*)

The resonances for the anomeric, methoxyl, C-6 methyl, and *N*- and *O*-acetyl methyl protons are as indicated.

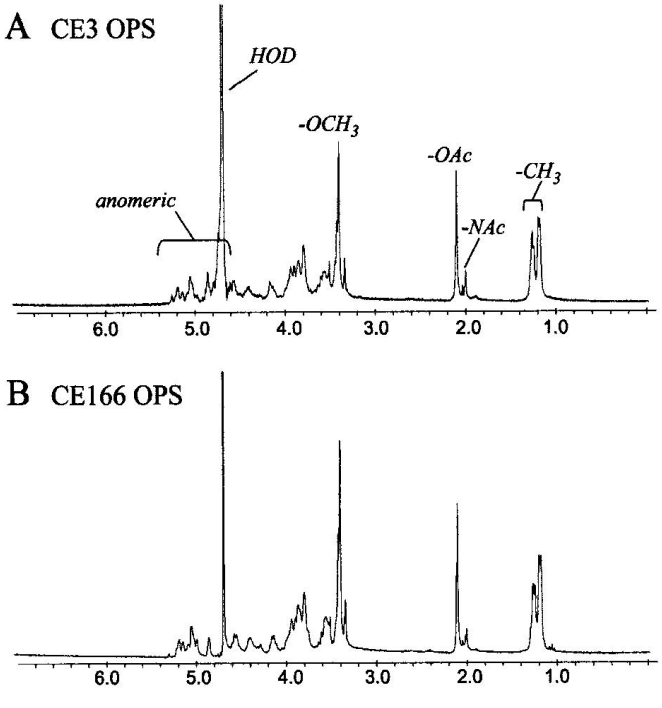
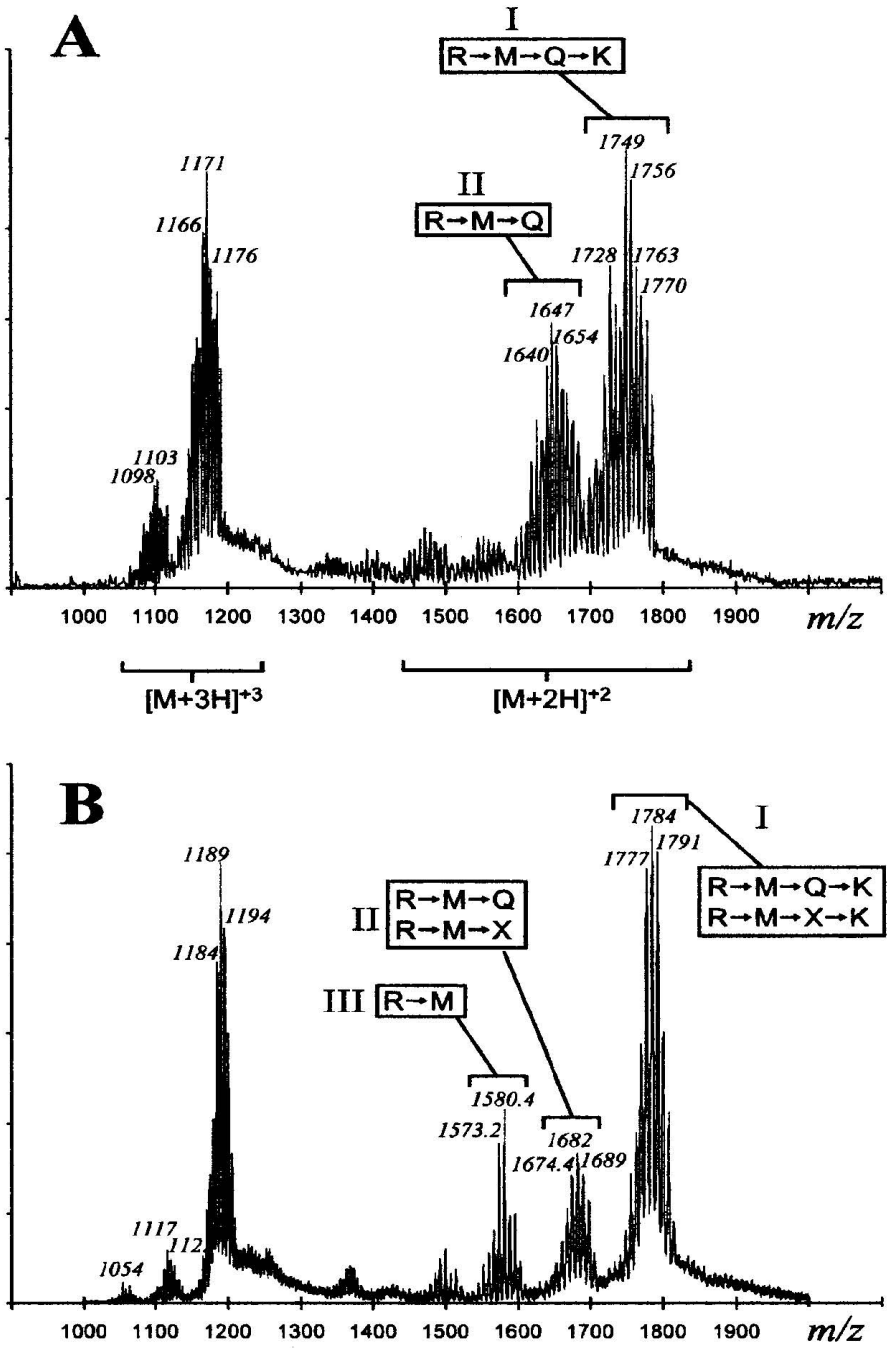


Figure 3

Positive ion ESI-MS spectra of the intact (native) OPS released by mild hydrolysis from the parent strain CE3 (*A*), and the CE166 mutant (*B*).



Protonated doubly and triply charged ions are observed in both spectra. Among doubly charged ions $(M + 2H)^{2+}$, ions within each family differ by ±7 *m*/*z* (14 mass units), reflecting heterogeneity in the degree of endogenous *O*-methylation. The triply charged ions differ by 4.7 *m*/*z* (rounded to Δ=5 *m*/*z* as shown). In *A*, the observed ion *m*/*z* 1749 (experimental mass 3496) corresponds to a molecular component (lactone form) having the following formula: $TOMeFuc_1GlcA_4(OAc)_4$-$GlcAMe_1$3M e6d$Tal_5Fuc_5$2$MeFuc_1Man_1$-$QuiNAc_1$ anhydro-Kdo1 (calculated mass 3494.3, predicted ion 1748.2). The electrospray technique tends to induce lactonization of carboxylated polysaccharides, resulting in a loss of a single H_2O from such species, presumably at the Kdo (or a GlcA) residue. Variation in *O*-acetyl content (± *m*/*z* 21; 42 mass units) also contributes to the multiplicity of ions within each family. *B*, the somewhat higher *m*/*z* values for the mutant, compared with the parent OPSs are consistent with an increase in *O*-acetyl and/or methyl groups (*e.g. m*/*z* 1784 compared with 1749 can be accounted for by one additional *O*-acetyl (+*m*/*z* 21) and two additional methyl groups (+*m*/*z* 14). Such differences could arise from variations in the mild acid hydrolysis conditions, storage time of the OPS in solution, or normal batch to batch variations in LPS. In *B*, the CE166 mutant OPS contains a third ion family, representing one aspect of the mutation (see "Results"). The spectrum of the CE166α OPS (not shown) is similar to that of CE166. *Q*, QuiNAc; *X*, analogue of QuiNAc; *M*, mannose; *K*, anhydro-Kdo; *R*, the repeating unit portion of the OPS. The anhydro forms of Kdo are generated during the mild hydrolytic release of OPS (5).

Figure 4
GC-MS analysis of the amino sugar components in parent and mutant strain OPSs

A, comparison of the total ion chromatograms of the carboxyl-reduced alditol acetates derived from the parent (*top*) and CE166 mutant (*bottom*) OPSs; the *arrows* mark the location of deuterated analogues of QuiNAc. Minor components eluting between 52 and 57 min are by-products of the Kdo derivatization. *B*, electron impact mass spectrum and fragmentation scheme of one of the deuterated amino sugar analogues detected in the CE166 and CE166α mutant OPSs. It is noteworthy that traces of the C-4 deuterated alditols were also detected during the preparation of the normal alditol acetates; however, their yields were increased during preparation of the carboxyl-reduced derivatives, a procedure that involves two additional treatments with borodeuteride (see "Experimental Procedures").

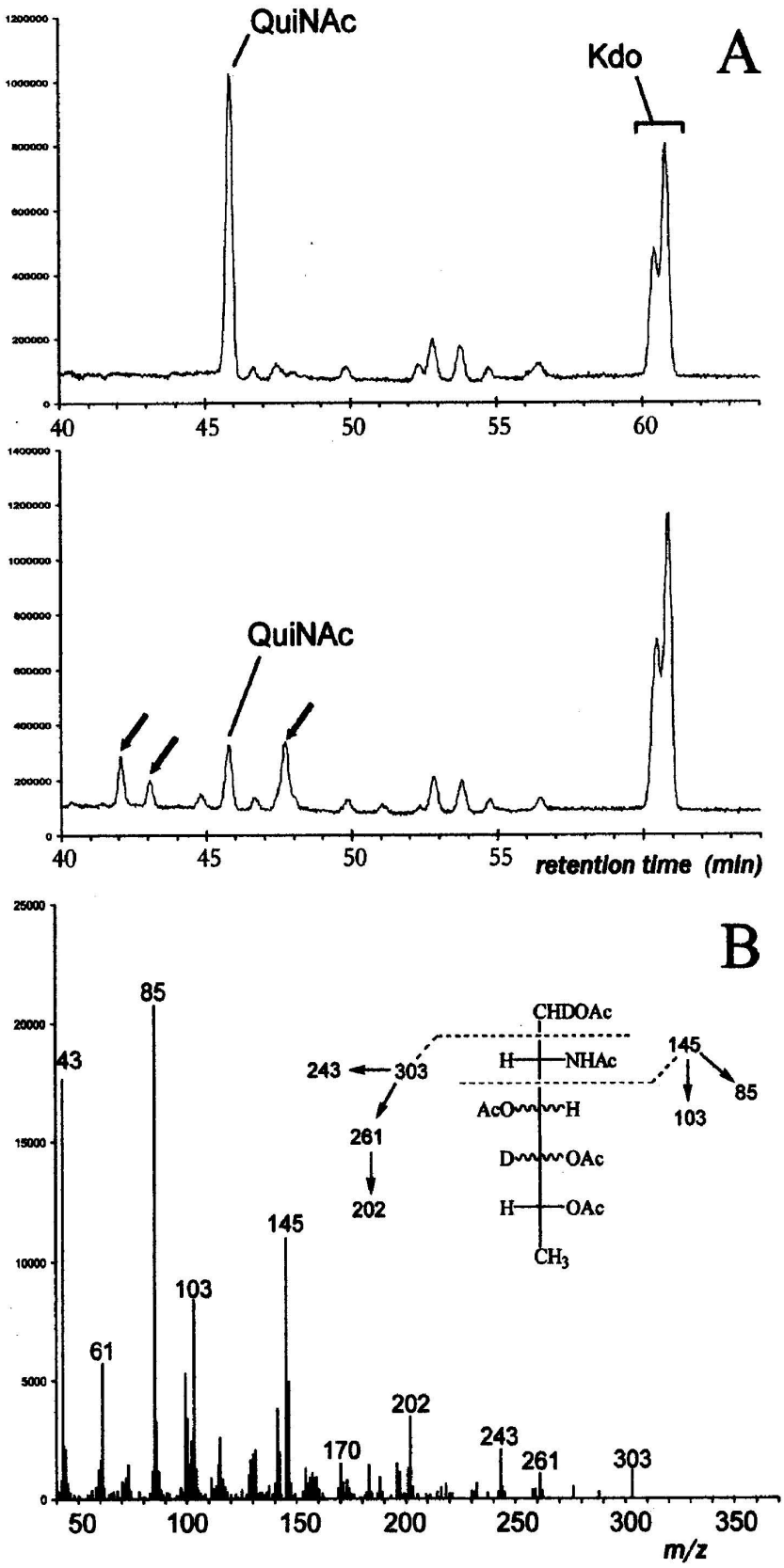
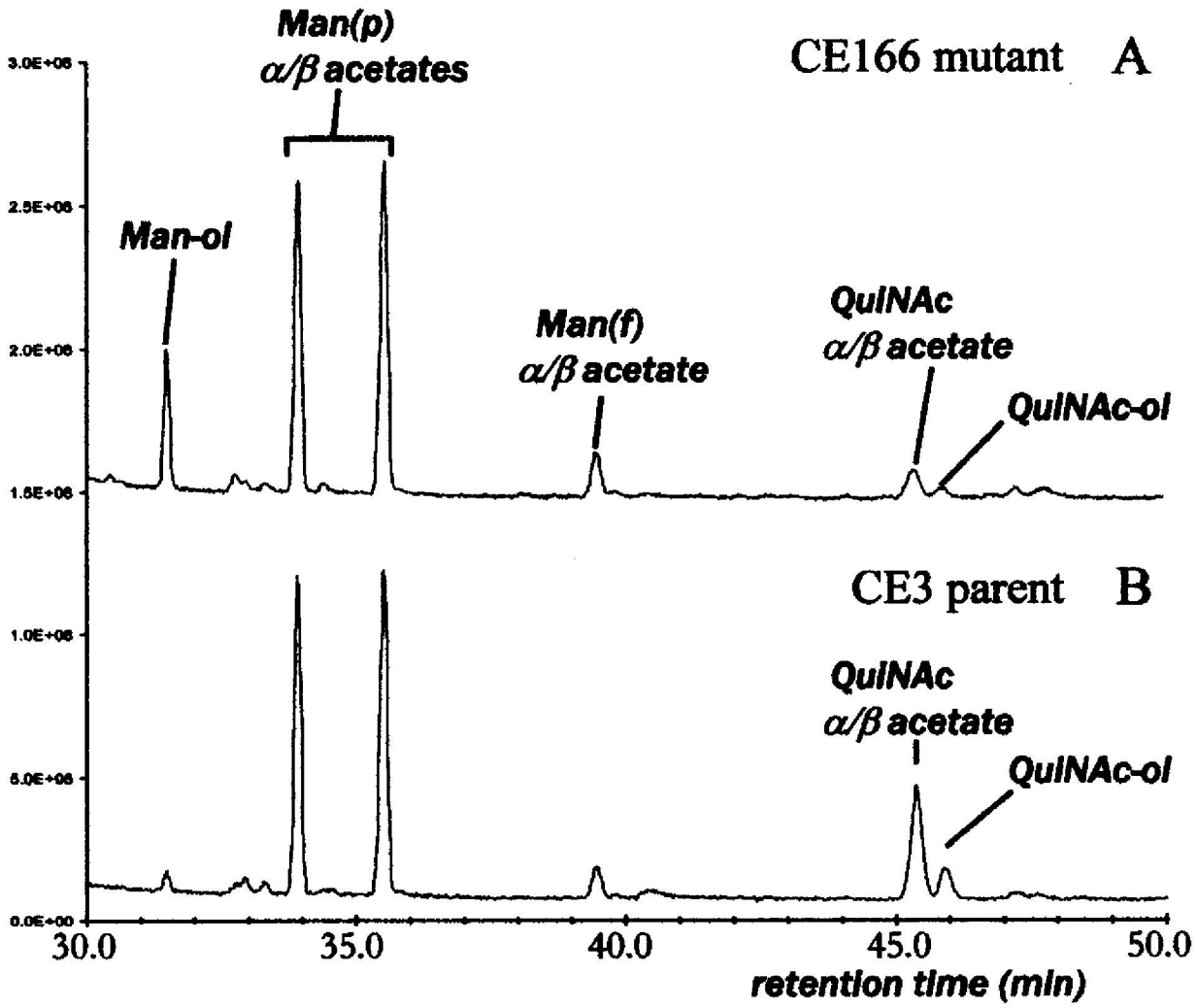


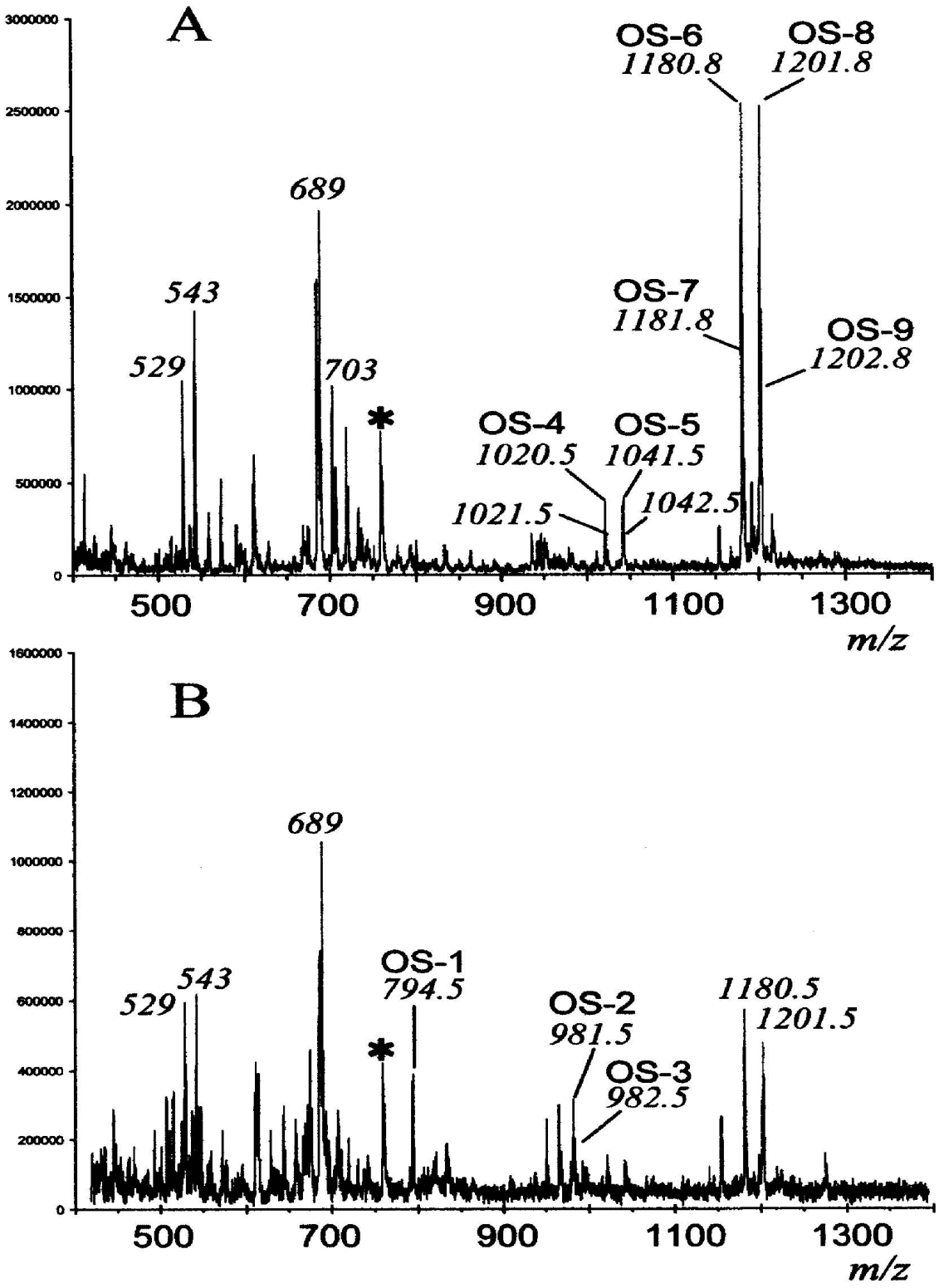
Figure 5
Labeling and comparison of the reducing end residues of the OPSs by GC-MS analysis



The native OPSs were reduced with borodeuteride and then hydrolyzed and acetylated, yielding the peracetates of all internal residues and the deuterated alditols of the reducing end residues. *A*, the CE166 mutant OPS yields a small amount of QuiNAc-ol, co-eluting with 4-deuterio-QuiNAc-ol (identified from the mass spectrum of its alditol acetate), and Man-ol. The α/β peracetates of QuiNAc, and 4-deuterio-QuiNAc (which co-elute) were also observed; *B*, the parent strain OPS yielded primarily the QuiNAc α/β peracetates and lesser amounts of QuiNAc-ol, confirming the structural assignments for the ion families obtained by ESI-MS (see Fig. 3). The Kdo alditol derivative, visible in Fig. 4, is not produced under these conditions (see "Experimental Procedures"). All three OPS samples contained a percentage of molecules terminating in Kdo, revealed by the "carboxyl-reduced alditols" (Fig. 4) and by the ESI-MS analysis (Fig. 3).

Figure 6

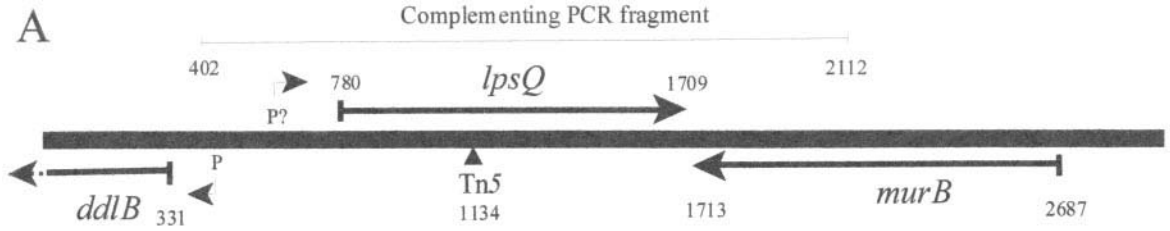
Positive ion ESI-MS spectra of oligosaccharides derived from base-catalyzed β -elimination of the CE166 mutant OPS.



The OPS was prereduced with borodeuteride to stabilize the reducing end to alkali, subjected to β-elimination, and fractionated by SEC. The two major SEC fractions were analyzed: the higher molecular weight fraction that is enriched in oligosaccharides originating from the reducing end of the OPS and terminating in anhydro-Kdo or Kdo-ol (*A*) and the lower molecular weight fraction that is enriched in reducing end oligosaccharides terminating in Man-ol, QuiNAc-ol, or 2-*N*-acetamido-4-deuterio-2,6-dideoxyhexitol (*B*). These oligosaccharides represent different forms of the reducing end of the OPS; all reducing end residues have been reduced with borodeuteride except for those terminating in 2,7-anhydro-Kdo residues, which arise during the mild hydrolytic release of the OPS (5) and resist reduction. All of these oligosaccharides terminate at the nonreducing end with a 4-deoxyhex-4-enopyranosyluronic acid residue, resulting from elimination of the 4-*O*-substituent from the glucuronosyl residues (5). All molecular ions are of the formula $(M + H)^{1+}$, and their structures are summarized in Table II. Oligosaccharides originating from the nonreducing end of the OPS co-eluted during SEC and are also visible in the spectra; their masses and structures are identical to those described previously for the parent OPS (5) and include species at *m*/*z* 703, *m*/*z* 689, *m*/*z* 543, and *m*/*z* 529. The ion (*) is a contaminant in the solvent.

Figure 7

***A*, the DNA locus mutated in strain CE166**



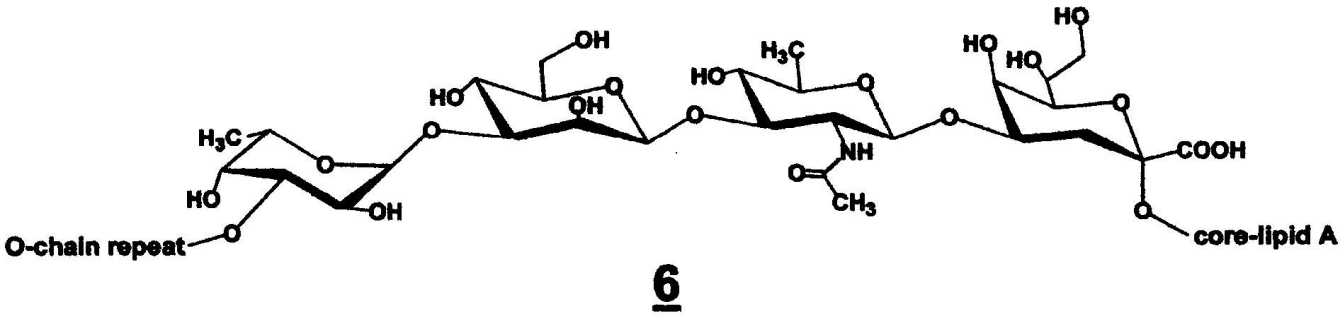
B

```
LPSQ          ---MRCLVTGAAGFVGSPLVKRLHAEKIYDLVATTRSQTPAFPPEVAHFPIEITGGTDWT 57
WBPV          MTRHNVLVTGATGFIGAALVNSLCSSGQYKVWAGCRRRGGAWPRGVTPLLLGELGSSVVW 60
WBPK          ------MVTGASGFVGSALCCELARTG-YAVIAVVRRVVERIP-SVTYIEADLTDPATFA 52
                    :****:**:*:.*   *     * : *  *      *  *: :     . :

LPSQ          AALEGVDVIVHLAARVHIMNDRAADPLAEFRRTNTAAALNLAEQAASAGVKRFVFVSTIK 117
WBPV          DAESAIDTVVHCAARVHVMSETASDPLVEFRKANVQGTLDLAREAVSRGVRRFIFISSIK 120
WBPK          GEFPTVDCIIHLAGRAHILTDKVADPLAAFREVNRDATVRLATRALEAGVKRFVFVSSIG 112
                   :* ::* *.*.*::.: .:***. **..*  .:: ** .* . **:**:*:*:*

LPSQ          VN--GEENDRPFRHDDRPKPIDPYGISKLECEIGLREIAARTGMEVVIIRPPLVYGPGAR 175
WBPV          VNGEGTEPGRPYTADSPPNPVDPYGVSKREAEQALLDLAEETGLEVVIIRPVLVYGPGVK 180
WBPK          VNG-NSTRQQAFNEDSPAGPHAPYAISKYEAEQELGTLLRGKGMELVVVRPPLIYANDAP 171
              **  .    :.:  *. . *  **.:** *.*  *  :   .*:*:*::** *:*. ..

LPSQ          GNFALLVNLVRKKLPLPFASLKNHRTLVAVQNLVDLIIACATHPAAPGEIFLAGDGEDLS 235
WBPV          ANVQTMMRWLKRGVPLPLGAIHNRRSLVSLDNLVDLIITCIEHPAAVGQVFLVSDGEDLS 240
WBPK          GNFGRLLKLVASGLPLPLDGVRNARSLVSRRNIVGFLSLCAEHPDAAGELFLVADGEDVS 231
              .*.  ::. :   :***: .::* *:**:  *:*.::  *  ** * *::**..****:*

LPSQ          TPALIRGIAAGLGVKPMLVPFPPALLQMAAKALGKEAVYQRLCGSLQVDITRARDVLGWS 295
WBPV          TTELLRRMGRALGAPARLLPVPASWIGAAAKVLNRQAFARRLCGSLQVDIMKTRQVLGWT 300
WBPK          IAQMIEALSRGMGRRPALFTFPAVLLKLVMCLLGKASMHEQLCGSLQVDASKARRLLGWV 291
               . ::. :. .:*  . *...*.  :  .   *.: :. .:********  ::* :***

LPSQ          PVVTPREGLKLAVE---------- 309
WBPV          PPVGVDQALEKTARSFLDRQ----- 320
WBPK          PVETIGAGLQAAGREYILRQRERRK 316
              *:::   :*  :
```

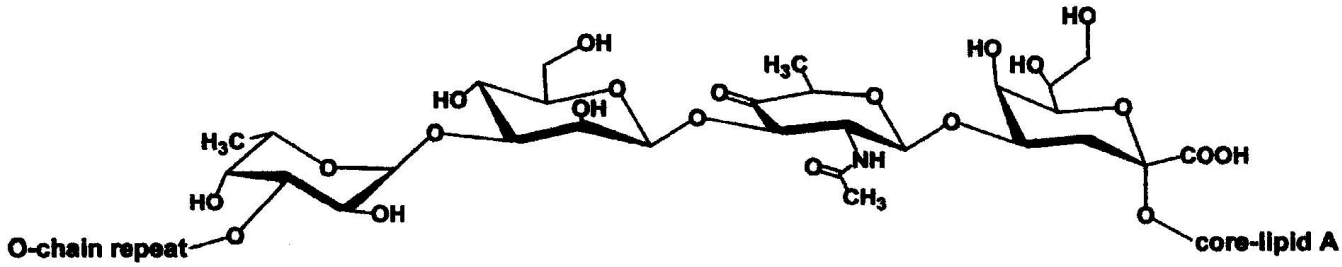
The *numbers* refer to nucleotides, with position 1 on the leftmost end being the first of the 2971 nucleotides whose sequence was determined at this locus. A stretch of 34 nucleotide pairs upstream of *ddlB* (nucleotides 431–465) are almost exactly conserved between *S. meliloti*, *Agrobacterium tumefaciens*, and this *R. etli* sequence. This stretch is probably the promoter for *ddlB;* it has a very strong match with consensus σ70-type promoters. A weaker match with σ70-type promoters, oriented toward *lpsQ,* is indicated by *P?*. *Above* the predicted ORFs is shown the expanse of DNA that was amplified from wild-type strain CE3 by PCR, which restores CE166 to LpsQ+ phenotypes when cloned in vectors pFAJ1700 and pFAJ1708. *B*, alignment of the predicted LpsQ amino acid sequence with the sequences of two homologs from *P. aeruginosa* by use of the ClustalW program (37) (available on the World Wide Web at www.ebi.ac.uk/clustalw/index.html). *Boldface letters* within *shaded areas* indicate the G*XX*G*XX*G motif for NAD(P) binding and the SX_nYX_3K residues commonly found at the active sites of enzymes in the SDR superfamily (*e.g.* residues 114, 139, and 143 of LpsQ). The *boxed string* of identical residues near the C terminus may participate in substrate specificity. It differs greatly from a stretch of residues that are conserved among fucose synthetases at the same place in alignments of the polypeptide sequences (38).

Figure 8

The outer core structures observed in the LPSs from R. etli CE3, and mutants CE166 and CE166α



6



7

R represents the capping and repeating unit portion of the OPS (refer to Fig. 1 for the complete structure of the OPS).

Figure 9
The suggested biosynthetic scheme for the synthesis of UDP-QuiNAc