

Structure and Molecular Characterization of *Streptococcus pneumoniae* Capsular Polysaccharide 10F by Carbohydrate Engineering in *Streptococcus oralis*^{*[5]}

Received for publication, March 16, 2010, and in revised form, May 11, 2010. Published, JBC Papers in Press, May 27, 2010, DOI 10.1074/jbc.M110.123562

Jinghua Yang[‡], Nirav Y. Shelat[§], C. Allen Bush[§], and John O. Cisar^{†1}

From the [‡]Oral Infection and Immunity Branch, NIDCR, National Institutes of Health, Bethesda, Maryland 20892 and the

[§]Department of Chemistry and Biochemistry, University of Maryland, Baltimore County, Baltimore, Maryland 21250

Although closely related at the molecular level, the capsular polysaccharide (CPS) of serotype 10F *Streptococcus pneumoniae* and coaggregation receptor polysaccharide (RPS) of *Streptococcus oralis* C104 have distinct ecological roles. CPS prevents phagocytosis of pathogenic *S. pneumoniae*, whereas RPS of commensal *S. oralis* functions as a receptor for lectin-like adhesins on other members of the dental plaque biofilm community. Results from high resolution NMR identified the recognition region of *S. oralis* RPS (*i.e.* Gal β 1–6GalNAc β 1–3Gal α) in the hexasaccharide repeat of *S. pneumoniae* CPS10F. The failure of this polysaccharide to support fimbriae-mediated adhesion of *Actinomyces naeslundii* was explained by the position of Gal f , which occurred as a branch in CPS10F rather than within the linear polysaccharide chain, as in RPS. Carbohydrate engineering of *S. oralis* RPS with *wzy* from *S. pneumoniae* attributed formation of the Gal f branch in CPS10F to the linkage of adjacent repeating units through sub terminal GalNAc in Gal β 1–6GalNAc β 1–3Gal α rather than through terminal Gal f , as in RPS. A gene (*wcrD*) from serotype 10A *S. pneumoniae* was then used to engineer a linear surface polysaccharide in *S. oralis* that was identical to RPS except for the presence of a β 1–3 linkage between Gal f and GalNAc β 1–3Gal α . This polysaccharide also failed to support adhesion of *A. naeslundii*, thereby establishing the essential role of β 1–6-linked Gal f in recognition of adjacent GalNAc β 1–3Gal α in wild-type RPS. These findings, which illustrate a molecular approach for relating bacterial polysaccharide structure to function, provide insight into the possible evolution of *S. oralis* RPS from *S. pneumoniae* CPS.

The Mitis group of viridans streptococci includes the important pathogen *Streptococcus pneumoniae* and 12 commensal species that inhabit the upper respiratory tract of man (1, 2). *Streptococcus mitis* and *Streptococcus oralis*, the two commensal species most closely related to *S. pneumoniae*, play an important role in colonization of tooth surfaces (3). Survival of these bacteria as pathogen or commensal depends on surface polysaccharides that have different ecological roles. Thus, cap-

sular polysaccharides (CPS)² of *S. pneumoniae* protect invading bacteria from phagocytic killing by the host, whereas the so called receptor polysaccharides (RPS) of *S. oralis* and related oral species function as receptors for lectin-like surface adhesins of other members of the dental plaque biofilm community, such as type 2 fimbriae-bearing *Actinomyces naeslundii* (4). Lectin-like recognition of RPS depends on the presence of an immunorecessive host-like motif; either GalNAc β 1–3Gal (Gn) or Gal β 1–3GalNAc (G), in the repeating units of different RPS structural types (5, 6). The presence of adjacent β 1–6-linked Gal f may also be important for exposing these motifs along linear polysaccharide chains (7, 8). Whereas the host-like features of RPS are critical for interbacterial adhesion, they contribute little to antigenicity, which instead depends on other features of these polysaccharides (6). RPS serotypes 1, 2, and 3, which occur in association with either Gn or G recognition motifs, are Glc and L-Rha-containing polysaccharides whereas RPS serotypes 4 and 5 lack Glc and L-Rha but instead contain ribitol phosphate in addition to Gal f , Gal p , and GalNAc. The latter polysaccharides contain Gn recognition motifs and thus, are designated RPS4Gn and RPS5Gn.

The evolution of *S. mitis* from an ancestral *S. pneumoniae*-like pathogen was recently proposed from comparative taxonomic and genomic studies of these closely related species (2). Like the *S. mitis* genome, the nearly complete *S. oralis* genome³ is about 10% smaller than that of *S. pneumoniae*, thereby raising the possibility that *S. oralis* also evolved by genome reduction from ancestral *S. pneumoniae*, albeit at an earlier time than *S. mitis*. The possible evolution of pathogen to commensal is consistent with molecular similarities seen between the surface polysaccharides of different modern day species, such as *S. pneumoniae* CPS serotype 21 with *S. oralis* RPS1Gn and RPS2G (9) and *S. pneumoniae* CPS serogroup 10 with *S. oralis* RPS4Gn and RPS5Gn (10). The high shared synteny and homology seen across the chromosomal loci of CPS10F and RPS4Gn (*i.e.* *cps10F* and *rps4Gn*) is especially striking and

^{*} This work was supported, in whole or in part, by the Intramural Research Program of the NIDCR, National Institutes of Health.

^[5] The on-line version of this article (available at <http://www.jbc.org>) contains supplemental Table S1 and Figs. S1–S8.

¹ To whom correspondence should be addressed: Bldg. 30, Rm. 3A-301, 30 Convent Dr., NIDCR, National Institutes of Health, Bethesda, MD 20892-4352. Fax: 301-402-1064; E-mail: john.cisar@nih.gov.

² The abbreviations used are: CPS, capsular polysaccharide; *cps*, chromosomal locus for CPS biosynthesis; G-motif, Gal β 1–3GalNAc; Gn-motif, GalNAc β 1–3Gal; RPS, receptor polysaccharide; RPS[–], no cell surface RPS detected by dot immunoblotting; *rps*, chromosomal locus for RPS biosynthesis; COSY, double quantum-filtered homonuclear coherence spectroscopy; TOCSY, total correlation spectroscopy; TQF-COSY, triple quantum-filtered spectra; HSQC, heteronuclear single quantum coherence spectroscopy; NOESY, nuclear Overhauser spectroscopy; HMBC, heteronuclear multiple bond correlation spectroscopy.

³ S. Highlander, personal communication.

Molecular Basis of *S. pneumoniae* CPS10F Structure

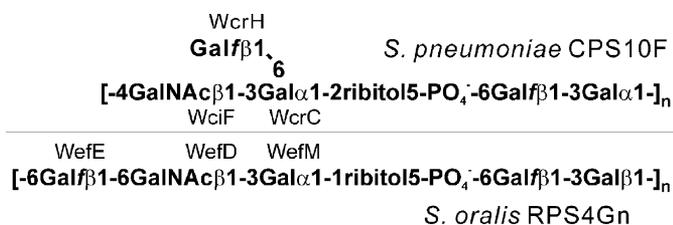


FIGURE 1. **Previously proposed molecular basis of polysaccharide structure for the host-like regions of *S. pneumoniae* CPS10F (11) and *S. oralis* RPS4Gn (10).**

includes the three genes for the recognition region in RPS. Two of these genes, *wefD* and *wefM* in *S. oralis* (10) and homologous *wciF* and *wcrC* in *S. pneumoniae* (11) are associated with synthesis of GalNAc β 1–3Gal in the repeating units of CPS10F and RPS4Gn (Fig. 1). The third gene, *wefE* of *S. oralis* and homologous *wcrH* of *S. pneumoniae*, encode putative GalF transferases that are 95% identical. However, based on the available structure of CPS10F (Fig. 1), WcrH was predicted (11) to link GalF to the Gal moiety of GalNAc β 1–3Gal, forming a branch in this polysaccharide (Fig. 1), whereas the action of WefE involved transfer of GalF to the terminal GalNAc β moiety, forming the linear recognition region in RPS4Gn of *S. oralis*. Whereas these findings suggest that WefE and WcrH differ in acceptor specificity, they also may indicate an error in the CPS10F structure (Fig. 1), which, as noted elsewhere (12), is not well established.

The present study was initiated to clarify the structural and corresponding molecular relationship that exists between CPS10F of *S. pneumoniae* and RPS4Gn of *S. oralis* and thereby, gain insight into the evolutionary history of these functionally distinct polysaccharides. The structure of CPS10F from *S. pneumoniae* 34355, the strain used to identify the *cps10F* locus (13), was determined by high resolution NMR and selected genes from this locus were characterized in *S. oralis* for their abilities to alter the structure and reactivity of RPS4Gn. The results associate the proposed evolution of RPS4Gn (and RPS5Gn) from a CPS10F-like ancestor with the different polymerases (Wzy) of *S. oralis* and *S. pneumoniae*. They also identify a previously unrecognized gene (*wcrF*) in the *cps10F* locus that appears to be critical for distinguishing different closely related members of CPS serogroup 10.

EXPERIMENTAL PROCEDURES

Bacterial Strains and Culture Conditions—Wild-type and mutant streptococci (Table 1) were cultured as previously described (10).

Antibodies and Immunochemical Methods—Previously described methods (6, 14) were followed to prepare rabbit antiserum against *S. pneumoniae* 34355 by repeated intravenous injections of heat-inactivated whole bacteria and also to affinity purify anti-CPS10F IgG by 4 M MgCl₂ elution of bound antibody from partially oxidized CPS10F coupled to Affi-Gel Hz. The cross-reactive fraction of antibody referred to as anti-RPS2Gn/4Gn IgG was prepared from rabbit antiserum (R102) against RPS2Gn-producing *S. gordonii* 38 (6) by 4 M MgCl₂ elution from partially oxidized RPS4Gn coupled to Affi-Gel Hz. This antibody was absorbed with RPS[−] *S. gordonii* XC2 (14) to ensure RPS specific immunoreactivity. Similarly, anti-RPS4Gn

IgG was absorbed with RPS[−] *S. oralis* YC3 (10). Absorptions were performed by incubating 25 μ g of IgG with $\sim 2 \times 10^9$ freshly washed bacteria in a total volume of 0.5 ml of phosphate-buffered saline containing 2 mg/ml bovine serum albumin for 2 h at 4 °C prior to centrifugation of absorption mixtures to remove added bacteria and membrane filtration (0.22 μ m pore size) of antibody-containing supernatants. Dot immunoblotting (10) was performed to compare binding of each antibody at 50 ng/ml to decreasing numbers of streptococci spotted on nitrocellulose membranes.

Isolation of Polysaccharides—*S. pneumoniae* 34355 was cultured to late stationary phase in 18 liters Todd Hewett Broth that was passed through a Millipore (Billerica, MA) PBTk Ultrafiltration Membrane (30-kDa pore size) to remove macromolecules prior to autoclaving. Following inoculation, cultures were incubated 3 days at 37 °C; liquid phenol was then added to a final concentration of 1% to kill virulent bacteria (15), which were removed by centrifugation of the culture medium followed by membrane filtration (0.22- μ m pore size) of the supernatant. Cell-free culture supernatant containing high molecular weight CPS10F was concentrated above a PBHK Ultrafiltration Membrane (100-kDa pore size) along with several volumes of water and buffer (pH 7.4, 10 mM TrisCl containing 15 mM NaCl, 2.5 mM MgCl₂, 0.5 mM CaCl₂, and 0.1% azide), which were added to wash residual media components through the membrane. The retained high molecular weight fraction was harvested from the concentrator and incubated with 15 mg of DNase I and 75 mg RNase (both from Sigma) in a total volume of 650 ml for several hours prior to the addition of 500 mg of protease (Sigma) for overnight digestion at 37 °C. The digest was chilled on ice prior to precipitation of protein in the presence of trichloroacetic acid, which was added to a final concentration of 5% (w/v). Precipitate was removed by centrifugation and membrane filtration (0.22- μ m pore size) prior to neutralization of the cold filtrate by dropwise addition of concentrated Tris. Soluble material was further digested with 20,000 units of mutanolysin (Sigma) in 400 ml of 20 mM sodium/potassium phosphate buffer (pH 6.7) containing 0.5 mM MgCl₂, 0.5 mM CaCl₂, 0.5% azide to cleave possible peptidoglycan links between CPS10F and C-polysaccharide (16). Following digestion, added protein was again removed by precipitation in the presence of 5% trichloroacetic acid, as described above. The soluble fraction was dialyzed against water followed by 10 mM TrisCl buffer (pH 8.0) containing 100 mM NaCl and applied to a DEAE Sephacel (GE Healthcare) anion exchange column equilibrated with this buffer. The column was rinsed with starting buffer prior to elution with a linear gradient of NaCl (100–200 mM) in 10 mM Tris buffer. Column fractions were monitored by the phenol sulfuric acid reaction (17) and by immunodiffusion performed with rabbit anti-serogroup 10 serum (Statens Serum Institute). CPS10F emerged from the column as a symmetrical peak in fractions containing from 120 to 150 mM NaCl. Material from the region between 128 and 140 mM NaCl was used for structural studies. The central region of the ¹H-¹³C spectrum of this material was fully assigned to CPS10F (Fig. 2); contaminating C-polysaccharide was less than 5% based on comparisons of one-dimensional ¹H (not shown) and two-dimensional ¹H-¹³C NMR spectra with the data of Karlsson *et al.* (18). Comparable

spectra of CPS10F samples that were not digested with mutanolysin prior to chromatography contained additional peaks that could be assigned to the form of C-polysaccharide that has two phosphocholine residues per repeating subunit (results not shown). RPS-like cell wall polysaccharides produced by plasmid-bearing mutant constructs of *S. oralis* C104 were isolated from mutanolysin digests of protease-treated cell walls and purified by DEAE Sephacel column chromatography as described previously (6, 10).

Chemical Methods for Carbohydrate Composition and Linkage Analysis—Glycosyl composition analysis was done by gas chromatography-mass spectrometry (GC-MS) of the monosaccharide TMS derivatives at the Complex Carbohydrate Research Center (University of Georgia). After methanolysis of the polysaccharide in 4 N trichloroacetic acid at 100 °C (4 h), the resulting monosaccharide methylglycosides were derivatized and analyzed on an Alltech AT1 fused silica capillary column. Linkage analysis was done by GC-MS of the partially methylated alditol acetates. Samples were permethylated with CH₃I in anhydrous NaOH in DMSO. The partially methylated polysaccharide was hydrolyzed in 2 M trichloroacetic acid at 121 °C, reduced with NaBD₄, acetylated in pyridine/acetic anhydride, and analyzed on a DB-1 capillary column.

Aqueous HF was used to cleave the phosphate from selected samples prior to the analyses. Polysaccharide was treated in 48% aqueous HF at 4 °C for 2 days, then evaporated in a stream of nitrogen followed by neutralization with NH₄OH. Salts were removed by Biogel P2 gel permeation chromatography, which resolved two peaks of carbohydrate. Both peaks were analyzed along with the intact polysaccharide.

Structural Characterization of Polysaccharides by NMR Spectroscopy—NMR spectra of purified polysaccharides were recorded as in previous studies (10, 19) with a Bruker DRX500 with a cryoprobe and a DRX700 using standard acquisition software. Generally, a 1–5 mg sample of polysaccharide was exchanged twice by lyophilization from 3 ml of 99.96% D₂O and dissolved in 0.6 ml of 99.99% D₂O for a 5-mm sample tube or a lesser volume for a Shigemitsu tube. Chemical shifts were recorded at a probe temperature of 25 °C or 45 °C relative to internal acetone (¹H, 2.225 ppm; ¹³C, 31.07 ppm). All the data were processed using NMRPipe, NMRDraw, NMRView, and Sparky software. Double quantum filtered homonuclear coherence spectroscopy (COSY) and total correlation spectroscopy (TOCSY) along with gradient triple quantum filtered spectra (TQF-COSY) were carried out to assign the scalar coupled protons of each monosaccharide residue. ¹³C chemical shifts were assigned by heteronuclear single quantum coherence spectroscopy (HSQC) and combination HSQC-TOCSY. Inter-residual linkages were determined by nuclear Overhauser spectroscopy (NOESY) with mixing times of 100 ms and 300 ms and by long-range C-H heteronuclear multiple bond correlation spectroscopy (HMBC). All chemical shifts reported were measured from natural abundance ¹³C-¹H HSQC spectra to avoid chemical shift and lineshape distortion by ¹H strong coupling that is common in carbohydrates. A number of these spectra were acquired at relatively high resolution (5 Hz) in the indirect (¹³C) dimension by means of folding and acquisition of 2048 FID, which was performed to increase the information content of

TABLE 1
Streptococci and plasmids used in this study

Strain or plasmid	Description	Source
<i>S. pneumoniae</i> 34355	Wild-type CPS10F	Statens Serum Institute
<i>S. pneumoniae</i> 8334	Wild-type CPS10A	ATCC
<i>S. oralis</i> C104	Wild-type RPS4Gn	(31)
<i>S. oralis</i> YC1	<i>S. oralis</i> C104 with <i>ermAM</i> in place of <i>wefM</i>	(10)
<i>S. oralis</i> YC6	<i>S. oralis</i> C104 with <i>ermAM</i> in place of <i>wefE</i>	(10)
<i>S. oralis</i> YC7	<i>S. oralis</i> C104 with <i>ermAM</i> in place of <i>wefL</i>	This study
<i>S. oralis</i> YC8	<i>S. oralis</i> C104 with <i>ermAM</i> in place of <i>wzy</i>	This study
pJY	Expression vector, Kan ^r	(10)
pJY-11	pJY expressing <i>wcrB</i> of <i>S. pneumoniae</i> 34355	This study
pJY-12	pJY expressing <i>wcrF</i> of <i>S. pneumoniae</i> 34355	This study
pJY-13	pJY expressing <i>wzy</i> of <i>S. pneumoniae</i> 34355	This study
pJY-14	pJY expressing <i>wcrH</i> of <i>S. pneumoniae</i> 34355	This study
pJY-15	pJY expressing <i>wcrD</i> of <i>S. pneumoniae</i> 8334	This study

HSQC-TOCSY spectra. This was possible because the natural line width for these polysaccharides is relatively narrow.

Molecular Methods—Previously described methods (10) were used to prepare *S. oralis* YC7 and YC8 by replacement of *wefL* or *wzy* in *S. oralis* C104 with a nonpolar *erm* cassette and pJY-derived plasmids expressing genes of interest from *S. pneumoniae* (Table 1). The PCR primers (supplemental Table S1) used to prepare these plasmids were designed to amplify not only the *S. pneumoniae* gene of interest but also its upstream Shine-Dalgarno sequence.

Bacterial Adhesion—The bacteria overlay technique, performed with fluorescein isothiocyanate-labeled bacteria (20), was used to compare adhesion of *A. naeslundii* 12104 to different streptococci. Briefly, nitrocellulose membranes were spotted with decreasing numbers of streptococci as described for dot immunoblotting and dried overnight. Prior to use, membranes were blocked 2 h by incubation in 20 mM phosphate-buffered (pH 7.4) saline containing 0.1 mM CaCl₂ and 5% bovine serum albumin. Blocked membranes were overlaid with 40 ml of fluorescein-labeled *A. naeslundii* 12104 (5 × 10⁸/ml) in the same buffer, incubated at room temperature for 2 h with occasional gentle mixing, washed with 20 mM Tris-buffered (pH 7.5) saline containing 0.05% Tween 20 to remove unattached bacteria, and scanned with a Typhoon scanner to detect adherent bacteria.

RESULTS

Structure of *S. pneumoniae* CPS10F—Based on the previously reported structure of CPS10F (Fig. 1), we expected that the presence of phosphate in this polysaccharide would adversely affect the results of carbohydrate composition and linkage analysis. Consequently, we analyzed both intact CPS10F and the HF-treated polysaccharide. Aqueous HF treatment effectively cleaves phosphate; however, it also cleaves furanoside and to some extent other glycosidic linkages, which may account for the recovery of HF-treated CPS10F in two fractions following Biogel P2 column chromatography (Table 2). The only components identified from intact CPS10F or

Molecular Basis of *S. pneumoniae* CPS10F Structure

TABLE 2
Carbohydrate composition and linkage analysis of *S. pneumoniae* CPS10F

Component	Carbohydrate composition (% total carbohydrate)		
	Intact CPS	HF-treated CPS	
		Fraction 1	Fraction 2
Ribitol	10.6	31.8	13.2
Galactose	71.8	29.7	67.7
Galactosamine	16.7	38.5	18.4

PMAA ^a	Linkage Analysis (% of peak area response)		
	Intact CPS	Fraction 1	Fraction 2
2-Ac-1,3,4,5 tetra-Me-ribitol	3.6	21	11
1,2,4,5-tetra-Ac-3-Me-ribitol	4.7	0	0
<i>t</i> -Gal _f	16	0	0
<i>t</i> -Gal _p	6.4	51	67
4-Gal _p	4.4	2.8	2.1
3-Gal _p	41	14	12
<i>t</i> -GalNAc	0	7.9	7.1
4-GalNAc	0	3.2	2.8
4,6-GalNAc	13	0	0

^a PMAA, partially methylated alditol acetates or sugar identities derived from the corresponding PMAA.

either HF-treated fraction were ribitol, galactose, and galactosamine (Table 2).

Linkage (methylation) analysis of intact CPS10F (Table 2) revealed two acetylated forms of ribitol (*i.e.* 2-Ac-1,3,4,5 tetra methyl- and 1,2,4,5-tetra-Ac-3-methyl ribitol); however, only the 2-acetylated derivative was identified from the HF-treated polysaccharide, thereby suggesting that ribitol was linked through either the 2 or 4 position. Terminal Gal_f was only identified from intact CPS, presumably because unsubstituted furanoside (Fig. 1) was cleaved by HF. A noteworthy finding involved the identification of 4,6-substituted GalNAc from the intact polysaccharide rather than 3,6-substituted Gal_p as was expected from the currently available structure of CPS10F (Fig. 1). The failure to detect 4,6-substituted GalNAc from HF-treated CPS10F suggested cleavage of a furanoside (or phosphate) from GalNAc. HF treatment of the polysaccharide also reduced the amount of 3-substituted Gal_p while increasing the amount of terminal galactopyranoside (*t*-Gal_p), thereby indicating substitution of Gal_p at the 3-position by a furanoside (or phosphate). The findings from chemical analysis were augmented by results from NMR spectroscopy.

The HSQC spectrum of the low field anomeric region (not shown) revealed five distinct signals indicating five sugar residues in the CPS10F repeating subunit. These were labeled **A**, **B**, **C**, **D**, and **E** for the purpose of assignment in Fig. 2 and Table 3. The letters, although arbitrary, were chosen to correspond to the notation used in previous studies of RPS4Gn from *S. oralis* C104 (21) and RPS5Gn from *S. oralis* SK144 (10). Spectra recorded at 45 °C showed better resolution than those at room temperature because of greater polysaccharide mobility at the higher temperature.

For residue **A**, signals for H1 through H3 were readily assigned by COSY (supplemental Fig. S1) and TOCSY (supplemental Fig. S2) and the corresponding ¹³C were assigned by HSQC. The low field positions of the ¹³C shifts and the small values of J_{HH} identified this residue as a furanoside. A strong HMBC cross peak (supplemental Fig. S3) was noted between H1 and C4 characteristic of furanosides and the

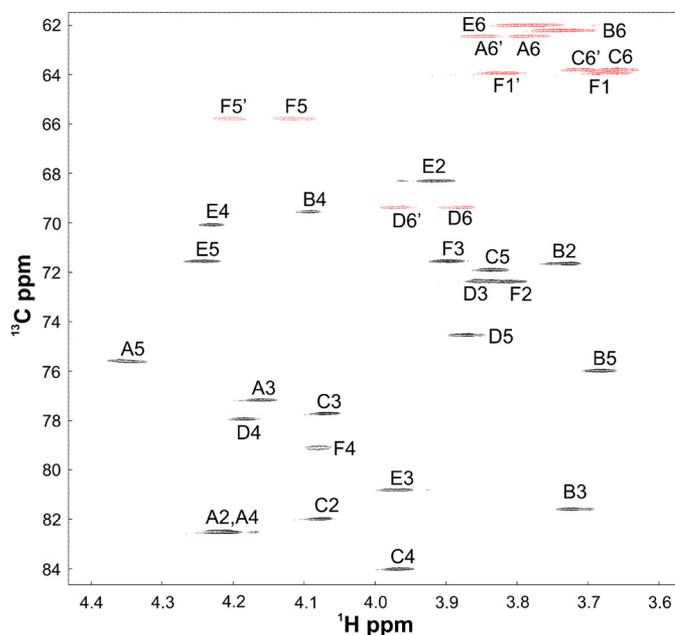


FIGURE 2. Central region of the multiplicity edited ¹H-¹³C HSQC spectrum of *S. pneumoniae* 34355 CPS10F at 45 °C. Negative contours (red) indicate methylene groups.

assignments confirmed by two-dimensional HSQC-TOCSY. Both HMBC and HSQC-TOCSY cross peaks (not shown) were observed between C4 and a pair of protons at 3.86 and 3.78 ppm that were identified as methylene protons by an edited HSQC spectrum. These signals were identified as H6' and H6 of residue **A** (*i.e.* A-H6' and A-H6, respectively) by TQF-COSY (not shown), which also identified A-H5 (4.355 ppm). The latter peak exhibited TOCSY signals to H6 and H6' as well as to H4 and H3. Although the **A**-2 and **A**-4 resonances overlapped in the spectrum shown in Fig. 2, these resonances were resolved in spectra recorded at 25 °C. Residue **A** of CPS10F was identified β-Gal_f by comparison of chemical shifts to those of methylgalactofuranosides (22).

For residue **B**, H2 was identified by COSY spectra and TOCSY indicated that H2 and H3 were strongly coupled. HSQC-TOCSY at short mixing times (20 ms) identified **B**-C2 and **B**-C3 at 71.68 and 81.59 ppm, respectively. A narrow TOCSY cross peak between H1 and H4 (supplemental Fig. S2) identified the equatorial H4 at 4.094 ppm. An HMBC cross peak (supplemental Fig. S3) between the signal assigned by HSQC as **B**-C4 (69.56 ppm) placed **B**-H5 at 3.689 ppm, an assignment confirmed by NOE between H1 and H5. **B**-C5 (75.99) assigned by HSQC exhibited HSQC-TOCSY to methylene peaks at 3.77 and 3.80, which were assigned as H6 and H6'. The H-H coupling constants and chemical shifts identified this residue as β-Gal_p.

For residue **C**, the downfield shift of the anomeric signal at 5.057, 109.14 ppm suggested a furanoside. The C-H2 signal was identified by COSY at 4.077 ppm and C-C2 by HSQC at 82.00 ppm. The strong HMBC cross peak between H1 and C4 (84.04 ppm) was also characteristic of a furanoside. C-H2 and C-H3 were strongly coupled; however, HMBC cross peaks (supplemental Fig. S3, upper panel) from H1 to C-C2 (82.00 ppm) and C-C3 at 77.72 ppm identified these resonances.

TABLE 3

Residue by residue comparison of HSQC ¹H and ¹³C chemical shifts for CPS10F of *S. pneumoniae* 34355, RPS4Gn of *S. oralis* C104, and polysaccharides isolated from plasmid-bearing mutant strains of *S. oralis* C104

Chemical shifts from NMR spectra recorded at 45 °C for *S. pneumoniae* CPS10F and 25 °C for *S. oralis* polysaccharides.

Strain	Residue	H1/C1	H2/C2	H3/C3	H4/C4	H5/C5	H6/C6
34355	β -Galp (A)	5.222	4.210	4.166	4.214	4.355	3.792,3.857
		109.97	82.54	77.18	82.52	75.61	62.46
C104		5.218	4.208	4.101	4.108	3.990	3.97
		110.05	82.26	77.46	83.36	70.10	67.18
YC7(p)Y-11)		5.210	4.211	4.176	4.229	<u>4.355^a</u>	<u>3.778,3.866</u>
YC1(p)Y-12)		109.85	82.35	76.81	82.09	<u>75.25</u>	<u>62.01</u>
YC8(p)Y-13)		5.220	4.210	4.100	4.107	3.983	3.962,3.985
YC6(p)Y-15)		110.09	82.25	77.51	83.52	70.29	67.26
YC8(p)Y-13)		5.229	4.213	4.104	4.104	3.990	3.954,3.986
YC6(p)Y-15)		110.05	82.25	77.52	83.37	70.15	67.21
YC6(p)Y-15)		5.215	4.207	4.103	4.104	3.988	3.955,3.981
YC6(p)Y-15)		110.05	82.27	77.46	83.36	70.07	67.14
34355	β -Galp (B)	4.637	3.743	3.728	4.094	3.689	3.775,3.806
C104		105.05	71.68	81.59	69.56	75.99	62.00
C104		4.507	3.675	3.742	4.107	3.730	3.78
YC7(p)Y-11)		103.91	70.84	81.09	69.38	75.96	61.85
YC7(p)Y-11)		4.512	3.681	3.741	4.117	3.733	3.771
YC1(p)Y-12)		103.90	70.81	81.40	69.33	75.92	61.85
YC1(p)Y-12)		4.508	3.677	3.745	4.108	3.733	3.786
YC8(p)Y-13)		103.94	70.83	81.08	69.41	75.95	61.89
YC8(p)Y-13)		<u>4.653</u>	<u>3.731</u>	<u>3.732</u>	<u>4.091</u>	<u>3.694</u>	<u>3.786</u>
YC6(p)Y-15)		<u>104.80</u>	<u>71.52</u>	81.11	69.37	75.89	61.77
YC6(p)Y-15)		4.496	3.666	3.737	4.104	3.72	3.775
YC6(p)Y-15)		103.85	70.84	81.08	69.35	75.90	61.80
34355	β -Galp (C)	5.057	4.077	4.072	3.971	3.841	3.673,3.720
C104		109.14	82.00	77.72	84.03	71.92	63.81
C104		5.071	4.068	4.100	4.006	4.027	3.767,4.077
YC7(p)Y-11)		108.62	81.81	77.51	83.96	70.38	71.91
YC7(p)Y-11)		5.073	4.073	4.102	4.004	4.027	3.765,4.074
YC1(p)Y-12)		108.64	81.83	77.49	83.94	70.34	71.93
YC1(p)Y-12)		5.071	4.091	4.097	4.011	4.026	3.761,4.085
YC8(p)Y-13)		108.68	81.77	77.57	83.96	70.47	71.98
YC8(p)Y-13)		5.067	4.064	4.075	3.963	<u>3.832</u>	<u>3.669,3.702</u>
YC6(p)Y-15)		108.88	81.83	77.31	83.64	<u>71.56</u>	<u>63.50</u>
YC6(p)Y-15)		5.093	4.082	4.050	4.052	4.010	3.754,4.059
YC6(p)Y-15)		109.94	82.23	77.68	83.78	70.38	71.84
34355	β -GalNAc (D)	4.679	4.027	3.845	4.186	3.877	3.888,3.970
C104		103.99	54.20	72.38	77.95	74.53	69.39
C104		4.635	3.944	3.752	3.945	3.829	3.775,3.911
YC7(p)Y-11)		103.93	53.36	71.56	68.59	74.54	68.08
YC7(p)Y-11)		4.635	3.944	3.753	3.947	3.840	3.775,3.913
YC1(p)Y-12)		103.98	53.34	71.47	68.66	74.50	68.02
YC1(p)Y-12)		4.629	3.953	3.753	3.944	3.821	3.769,3.924
YC8(p)Y-13)		103.94	53.39	71.56	68.68	74.54	68.19
YC8(p)Y-13)		4.677	4.034	3.856	<u>4.204</u>	3.904	<u>3.905,3.973</u>
YC6(p)Y-15)		103.93	53.92	72.13	<u>77.68</u>	74.36	<u>69.24</u>
YC6(p)Y-15)		4.705	<u>4.047</u>	<u>3.803</u>	4.069	<u>3.706</u>	<u>3.773</u>
YC6(p)Y-15)		103.68	<u>52.42</u>	<u>79.03</u>	68.65	<u>75.52</u>	<u>61.73</u>
34355	α -Galp (E)	5.080	3.926	3.970	4.235	4.246	3.740
C104		99.18	68.31	80.84	70.07	71.58	62.21
C104		4.959	3.890	3.963	4.204	3.980	3.74
YC7(p)Y-11)		99.92	68.26	80.22	70.08	71.43	62.06
YC7(p)Y-11)		4.962	3.890	3.967	4.206	3.981	3.743
YC1(p)Y-12)		99.90	68.26	80.20	70.09	71.44	62.06
YC1(p)Y-12)		<u>5.074</u>	3.889	3.979	4.239	4.266	3.740
YC8(p)Y-13)		<u>98.71</u>	68.04	80.25	69.95	71.35	62.02
YC8(p)Y-13)		4.963	3.903	3.976	4.213	3.975	3.746
YC6(p)Y-15)		99.92	68.25	80.43	70.09	71.45	62.05
YC6(p)Y-15)		4.961	3.895	3.986	4.216	3.977	3.733
YC6(p)Y-15)		99.92	68.32	80.11	69.97	71.43	61.95
34355	Ribitol (F)	3.677,3.826	3.814	3.902	4.086	4.120,4.203	
C104		63.95	72.39	71.56	79.14	65.79	
C104		3.599,3.965	4.054	3.818	3.933	3.990,4.085	
YC7(p)Y-11)		69.48	71.63	72.29	71.66	67.40	
YC7(p)Y-11)		3.598,3.962	4.061	3.818	3.933	4.020,4.126	
YC1(p)Y-12)		69.41	71.62	72.3	71.67	67.53	
YC1(p)Y-12)		<u>3.661,3.824</u>	3.796	3.900	<u>4.105</u>	4.088,4.157	
YC8(p)Y-13)		<u>63.73</u>	72.13	71.13	<u>78.61</u>	65.26	
YC8(p)Y-13)		3.597,3.970	4.057	3.824	3.938	4.002,4.079	
YC6(p)Y-15)		69.46	71.62	72.26	71.64	67.37	
YC6(p)Y-15)		3.595,3.965	4.056	3.817	3.926	3.982,4.071	
YC6(p)Y-15)		69.46	71.61	72.30	71.61	67.41	

^a Chemical shifts that are underlined differ significantly from those of wild-type RPS4Gn (strain C104).

Molecular Basis of *S. pneumoniae* CPS10F Structure

HSQC-TOCSY from C-H4 (not shown) confirmed the assignments of C-C2 and C-C3. HSQC-TOCSY from C-C3 and C-C4 identified C-H5 as well as C-H6 and C-H6' that were recognized as methylene protons in edited HSQC. Like residue A, residue C was assigned as β -GalF with the anomeric configuration based on chemical shifts (22).

For residue D, with signal D-1 at 4.679, 103.99 ppm, H2 was identified by COSY. HSQC located D-C2 at 54.20 ppm, a chemical shift characteristic of an amino sugar, which suggested that residue D was the GalNAc identified by chemical analysis (Table 2). H3 of this residue was located by homonuclear TOCSY as well as by a HSQC-TOCSY cross peak with D-C2 (data not shown). A sharp cross peak at 4.18 ppm in the same row of this spectrum, as well as in supplemental Fig. S2, was assigned as the equatorial proton, D-H4. An HMBC cross peak observed between the signal assigned by HSQC to D-C4 (77.95 ppm) and a signal at 3.887 ppm suggested assignment of the latter peak to D-C5, which was supported by NOE (Fig. S4) from D-H4. HSQC-TOCSY from D-C5 identified methylene peaks in edited HSQC for D-H6 and D-H6'. Any possible confusion resulting from the overlapping chemical shifts of D-H5 and D-H6 (3.888 ppm) was resolved in HSQC-TOCSY spectra (data not shown) run without ^{13}C decoupling during acquisition. In such spectra, direct peaks were split by $^1\text{J}_{\text{CH}}$ allowing the correct ^1H chemical shift of the relay peaks to be accurately determined. This residue was identified as β -GalNAc on the basis of the anomeric chemical shifts and the large coupling of H1 and H2.

For residue E, with the E-1 signal at 5.080, 99.18 ppm, cross peaks from H1 to H2 in COSY (supplemental Fig. S1) and between H1 and H3 in TOCSY (supplemental Fig. S2) identified the proton assignments and HSQC was used for assignment of E-C2 and E-C3. HMBC cross peaks (supplemental Fig. S3, lower panel) observed from E-C2 to both E-H3 and E-H4 located these proton resonances. This residue was identified as α -GalP by the small $J_{\text{H1-H2}}$, the large NOE between H1 and H2 (supplemental Fig. S4) and the narrow peak assigned to H4, which resulted from the small scalar coupling of this equatorial proton to H3 and H5. HMBC cross peaks were observed between E-H1 and 80.84 ppm (E-C3) and to peaks at 71.57 and 79.14 ppm, one of which was expected to be E-C5 (19). An HMBC cross peak observed between E-C4 (70.07 ppm) and 4.246 ppm was assigned as E-H5 to assist in the identification of E-C5. A signal at 62.21 ppm attributed to a methylene group by edited HSQC showed an HSQC-TOCSY cross peak with E-H5, identifying the former signal as E-C6 corresponding to E-H6 at 3.740 ppm.

The interpretation of the NMR data presented above, combined with the carbohydrate composition and linkage analyses (Table 2) identified five sugar residues in the repeating subunit of CPS10F. However, a number of resonances in the HSQC spectrum of this polysaccharide (Fig. 2) remained unassigned due to the presence of ribitol phosphate (Table 1), which lacks an anomeric signal. Ribitol (residue F) was identified in multiplicity-edited HSQC spectra of its methylene groups in the 1 and 5 positions. The spectrum shown in Fig. 2 contained two such negative signals (^{13}C shifts of 63.95 and 65.79 ppm) that were not assigned to sugars. In ^{31}P HSQC spectra (not shown),

cross peaks were observed at 4.211 and 4.090 ppm, which corresponded to the negative peak at a ^{13}C shift of 65.79 ppm in Fig. 2. Accurate ^1H chemical shifts for the signal assigned to C5 of ribitol were determined from the ^{13}C HSQC spectrum. HSQC-TOCSY peaks between the F-C5 signal and 4.086 ppm identified F-H4 and HSQC provided a ^{13}C shift of 79.14 for F-C4. An HSQC-TOCSY cross peak was observed between this latter signal and 3.90 ppm, a resonance which was assigned as F-H3; F-C3 was located at 71.56 ppm by HSQC. Although this ^{13}C chemical shift was identical to that of E-C5, the HSQC-TOCSY cross peak with E-H6 was easily distinguished from that of F-H2 at 3.814 ppm as well as that of F-H4. HSQC spectra located F-C2 at 72.39, a chemical shift close to that of D-C3. Nevertheless, HSQC-TOCSY cross peaks for F-H1 (3.677 ppm) and F-H1' (3.826 ppm) in this row were readily distinguished from those arising from residue D. Our assignment of the C-H pairs summarized in Table 3 accounts for all the signals observed in the HSQC spectrum shown in Fig. 2.

Given the complete NMR assignment of *S. pneumoniae* CPS10F (Table 3), determination of the linkages between the residues by HMBC and NOE data were straight forward. In our new and revised structure of CPS10F (Fig. 3A), the scalar coupling between ^{13}C of one residue with ^1H of the adjacent residue, indicated in red, provided unambiguous proof of linkage positions because 3-bond scalar coupling follows the chemical bonds. Also indicated in blue in Fig. 3, are ^1H - ^1H proximities derived from NOE data (Fig. S4); all of these support the proposed structure. The position of the phosphodiester linkage joining the 5-position of ribitol (F) to C5 of the β -GalF (A) was revealed by ^{31}P - ^1H HSQC spectra (not shown), which showed strong correlation of the ^{31}P signal with A-H5 and with F-H5 and H5'.

Molecular Basis of Polysaccharide Structure and Reactivity—The three linkages that distinguish the structures of CPS10F from RPS4Gn (Fig. 3, A and B) have dramatic effects on the reactions of these polysaccharides both as antigens and as receptors for adhesion of type 2 fimbriated *A. naeslundii* (Fig. 4). To define these differences at the molecular level, we characterized selected genes from the *cps10F* or *cps10A* loci for their ability to alter the structure and reactivity of RPS4Gn produced by *S. oralis* C104.

The phosphodiester linkages between ribitol-5-phosphate and GalF in CPS10F and RPS4Gn were previously suggested to depend on *wcrB* in *S. pneumoniae* 34355 (11) and *wefL* in *S. oralis* C104 (10). To test these hypotheses, we replaced *wefL* in *S. oralis* C104 with a nonpolar *erm* cassette, to obtain *S. oralis* YC7. Surprisingly, the loss of *wefL* reduced but did not abolish anti-RPS immunoreactivity. The end point of strain YC7 was ~ 10 -fold lower than wild-type strain C104 in dot immunoblotting performed with anti-2Gn/4Gn reactive IgG (Fig. 4), which binds the GalF β 1–6GalNAc region of RPS4Gn (6), and 100-fold lower with anti-RPS4Gn reactive IgG, which binds the Gal α 1–1ribitol region of RPS4Gn (10). Considered together, these findings suggested reduced cell surface production of an antigenically altered polysaccharide. Although we did not isolate the YC7 polysaccharide for structural characterization, we suspected that it was a variant of RPS4Gn devoid of ribitol-5-phosphate. Formation of such a polysaccharide in the absence of

Molecular Basis of *S. pneumoniae* CPS10F Structure

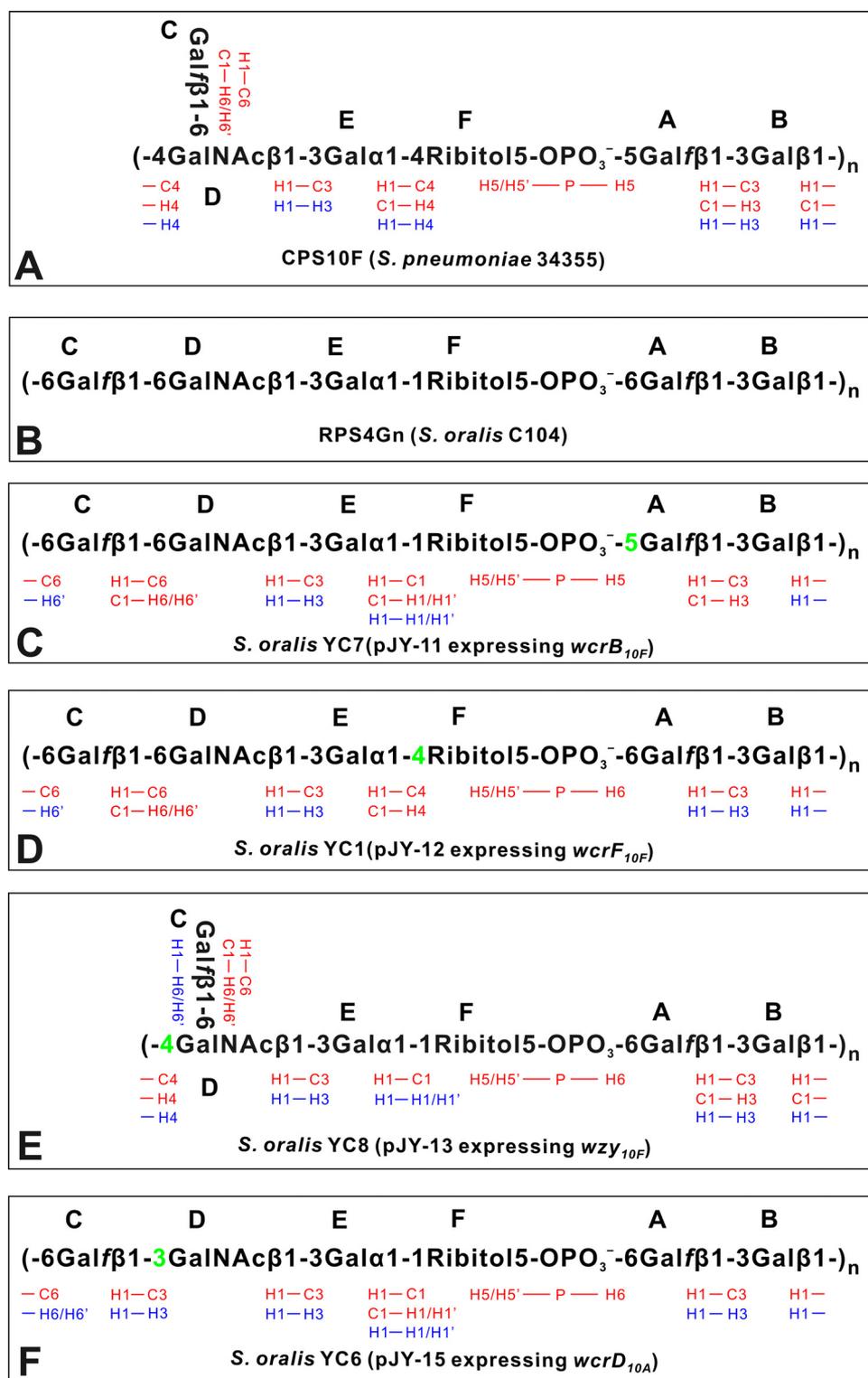


FIGURE 3. HMBC (red) and NOE (blue) inter-residue connectivities of CPS10F (A) and polysaccharides engineered from RPS4Gn (B) by plasmid-based complementation of different *S. oralis* mutant strains with genes from *S. pneumoniae* (C–F). The linkage position that distinguishes each engineered polysaccharide from RPS4Gn is indicated in green.

WefL can be explained by the WefM-mediated transfer α -Gal to the 6-OH of Galf, which would resemble the 1-OH of ribitol and thus, may act as an acceptor (Fig. 5B). In any case, immunoreactivity of *S. oralis* YC7 was fully restored by transformation of this strain with *wcrB*-expressing pJY-11 (Fig. 4).

The ^{13}C and ^1H chemical shifts recorded for the anionic cell wall polysaccharide isolated from *S. oralis* YC7(pJY-11) (supplemental Fig. S5) were fully assigned (Table 3) by the same coherence transfer methods (COSY, TOCSY, HSQC-TOCSY, HMBC, HSQC, and TQF-COSY) used above for CPS10F. This assignment was similar to that of parent strain *S. oralis* C104 RPS, except for the phosphodiester linkage positions A5 and A6 (Table 3). The positions of glycosidic linkages were identified by 3-bond ^1H - ^{13}C scalar coupling in HMBC spectra and confirmed by proton proximity in NOE spectra (Fig. 3C). The phosphodiester linkage was determined by ^1H - ^{31}P HSQC spectra. The results showed that *S. oralis* YC7(pJY11) RPS (Fig. 3C) was identical to wild-type RPS4Gn (Fig. 3B) except for the phosphodiester linkage to the 5-OH of Galf (residue A). *S. oralis* C104 and *S. oralis* YC7(pJY11) were indistinguishable in dot immunoblotting and bacteria overlay experiments (Fig. 4), thereby indicating that the different linkages from ribitol-5-phosphate to Galf did not measurably affect cell surface antigenicity or adhesion of *A. naeslundii*.

The linkages between Gal and ribitol in CPS10F and RPS4Gn were previously associated with the gene that is currently designated *wcrF* (Fig. 5) in *S. pneumoniae* 34355 (11) and with *wefM* in *S. oralis* C104 (10). To compare these genes, we replaced *wefM* of *S. oralis* C104 with a nonpolar *erm* cassette and transformed the resulting RPS⁻ strain (*S. oralis* YC1) with pJY-12 expressing *wcrF* from *S. pneumoniae* 34355. The ^{13}C and ^1H chemical shifts of the anionic polysaccharide produced by *S. oralis* YC7(pJY-12) (supplemental Fig. S6) were fully assigned as described above and were similar to those of *S. oralis* C104 RPS except for residues E1, F1, and F4 (Table 3). The linkages between residues were determined by HMBC spectra and by ^1H - ^{31}P HSQC spectra (Fig. 4D). The results showed that *S. oralis* YC1(pJY-12) RPS was identical to RPS4Gn except for the α 1–4 linkage between Gal to ribitol. As shown by results from dot immunoblotting of

Molecular Basis of *S. pneumoniae* CPS10F Structure

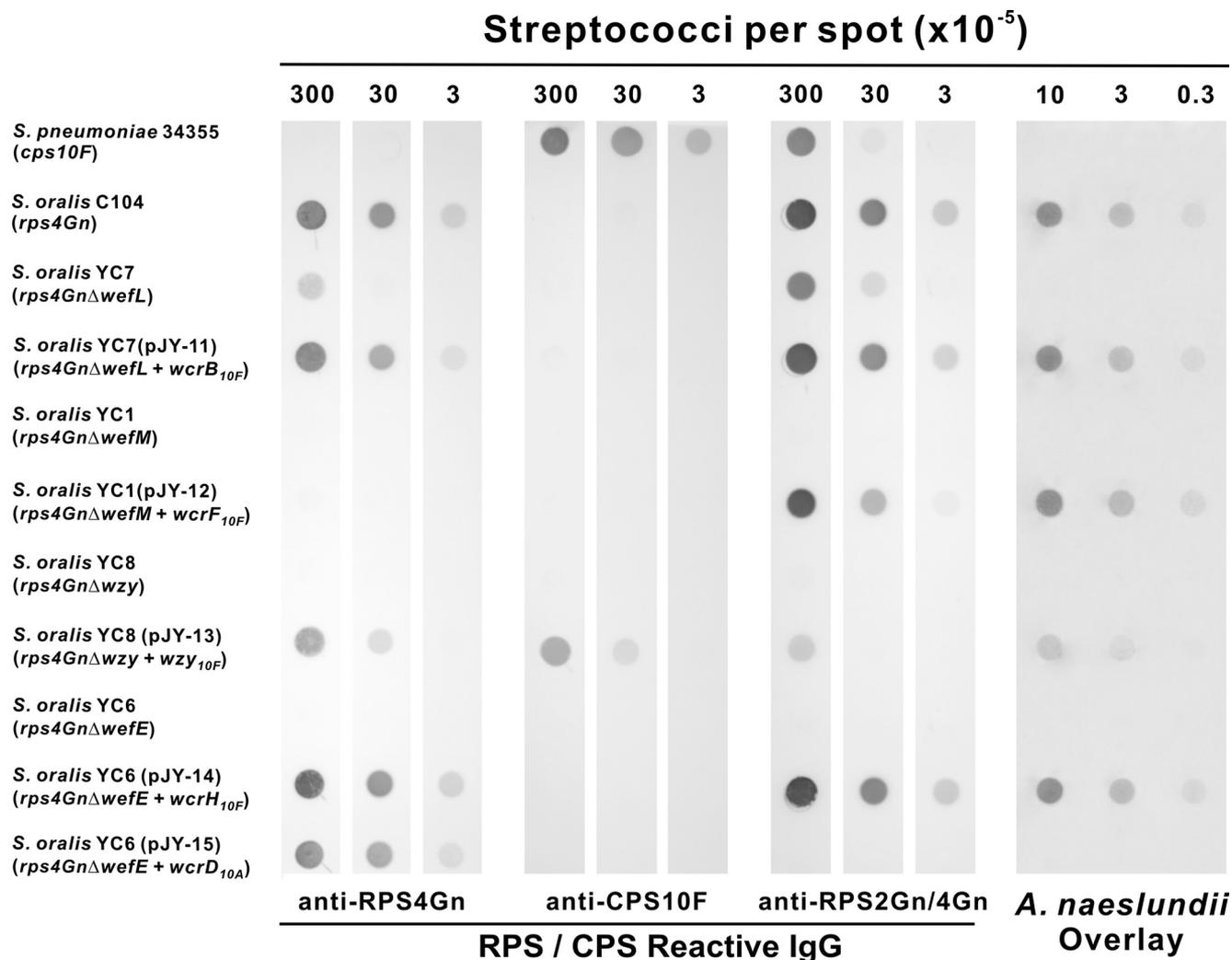


FIGURE 4. Characterization of streptococcal surface polysaccharides as antigens by dot immunoblotting and as receptors for adhesion of *A. naeslundii* by bacteria overlay. For dot immunoblotting, membranes were incubated with RPS or CPS reactive IgG, washed, and developed with peroxidase-conjugated goat anti-rabbit IgG and substrate. For bacteria overlay, membranes were incubated with fluorescein-labeled *A. naeslundii*, washed, and scanned with a fluorescence imager to detect bound bacteria.

S. oralis C104 and *S. oralis* YC1(pJY-12) (Fig. 4), the presence of Gal α 1-4ribitol in the polysaccharide of the later strain abolished the reaction of anti-4Gn RPS specific IgG without increasing anti-CPS10F immunoreactivity, which was negative for both strains. Results from parallel bacteria overlay experiments performed with these strains showed that both supported adhesion of *A. naeslundii*.

The structure of CPS10F (Fig. 3A) suggested that the Galf branch in this polysaccharide was formed by the polymerase dependent-linkage of adjacent repeating units through subterminal GalNAc. To test this proposal, we replaced *wzy* in *S. oralis* C104 with a nonpolar *erm* cassette and transformed the resulting RPS⁻ strain (*S. oralis* YC8) with pJY-13 expressing *wzy* from the *cps10F* locus. The yield of anionic polysaccharide from *S. oralis* YC8(pJY-13) was ~10-fold less than expected (*i.e.* 3 mg from an 18-liter culture), which increased the proportion of contaminating cell wall material, as indicated by the presence of a few unassigned signals in the HSQC spectrum of this sample (supplemental Fig. S7). However, these signals did not prevent complete assignment of the ¹H and ¹³C signals associated

with the specific polysaccharide produced by *S. oralis* YC8(pJY-13). The signals assigned to residues A, E, and F of this polysaccharide were similar to those of *S. oralis* C104 RPS4Gn while those assigned as B1 and B2, C5 and C6, and D4 and D6 were similar to the corresponding signals of *S. pneumoniae* CPS10F (Table 3). The inter-residue connectivities of the *S. oralis* YC8(pJY-13) polysaccharide established the structure shown in Fig. 3E. Results from dot immunoblotting (Fig. 4) of this construct showed reduced binding of both anti-RPS antibodies but increased binding of anti-CPS10F reactive IgG, thereby identifying the *S. pneumoniae* polymerase as an important molecular determinant of CPS10F immunoreactivity. Bacteria overlay experiments (Fig. 4) showed weak but significant adhesion of *A. naeslundii* to *S. oralis* YC8(pJY-13), which was confirmed by standard coaggregation assays performed with these strains. In control experiments, *A. naeslundii* WVU45M, which lacks type 2 fimbriae (23), failed to coaggregate with either *S. oralis* YC8(pJY-13) or any other strain listed in Fig. 4.

We also tested *wcrH* from the *cps10F* cluster (Fig. 5) for complementation of the *wefE* deletion in RPS⁻ *S. oralis* YC6.

Molecular Basis of *S. pneumoniae* CPS10F Structure

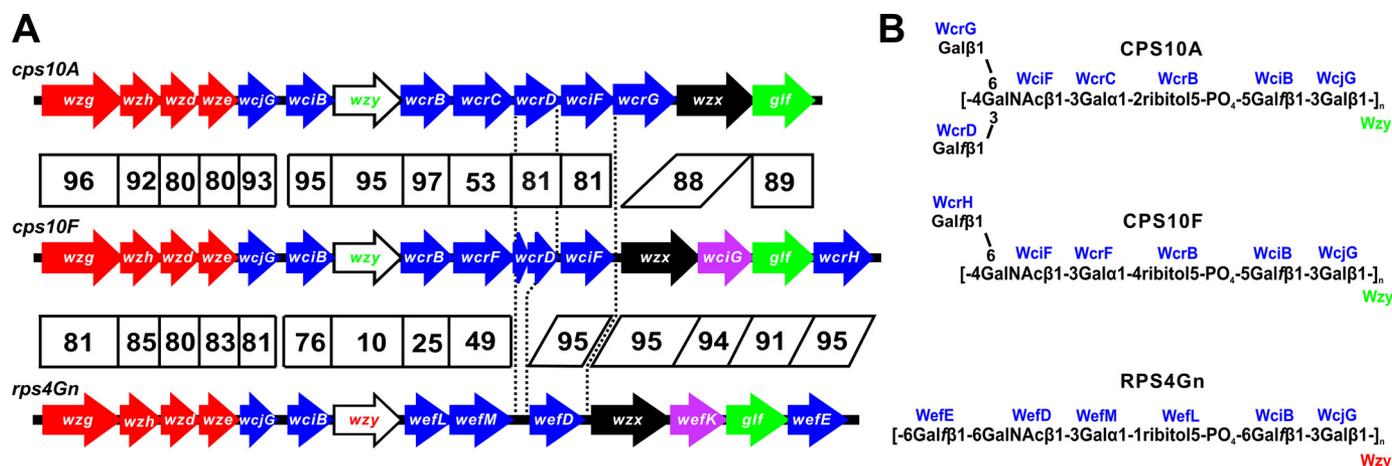


FIGURE 5. **Molecular and structural comparisons of *S. pneumoniae* CPS10A and CPS10F and *S. oralis* RPS4Gn.** *A*, chromosomal loci of these polysaccharides (GenBank™ CR931649, CR931652, and EF587719, respectively) showing percent amino acid sequence identities between encoded regulatory proteins (red), glycosyl or ribitol-phosphate transferases (blue), polymerases (white), flippases (black), acetyltransferases (purple), and enzymes for nucleotide sugar biosynthesis (green). Dotted lines connect sequences at the 5'- or 3'-ends of *wcrD* or *wcrG* in the *cps10A* locus with homologous sequences in the *cps10F* and *rps4Gn* loci that occur within the *wcrD* pseudogene of the *cps10F* locus or the intergenic regions that flank *wciF/wefD*. *B*, corresponding polysaccharide structures showing the biosynthetic role of each encoded transferase (blue) and polymerase (green or red).

NMR spectra recorded for the polysaccharide isolated from *S. oralis* YC6(pJY-14) were indistinguishable from those of wild-type RPS4Gn (results not shown). Likewise, the reactions of this construct and *S. oralis* C104 were identical in dot immunoblotting and bacteria overlay experiments (Fig. 4). Thus, *wcrH* and *wefE* appeared to represent the same gene in different species. RPS⁻ *S. oralis* YC6 was then transformed with pJY-15 expressing *wcrD*, the gene associated (11) with the β1–3-linked Galf branch in CPS10A (Fig. 5). The chemical shifts recorded for the polysaccharide isolated from *S. oralis* YC6(pJY-15) (supplemental Fig. S8) were similar to those of *S. oralis* C104 RPS4Gn for residues A, B, E, and F but different for D2, D3, D5, and D6 (Table 3). Determination of the linkage positions by ¹H-¹³C HMBC and by ¹H-³¹P HSQC indicated that the *S. oralis* YC6(pJY-15) polysaccharide was identical to RPS4Gn except for the β1–3 linkage between residues C and D (Fig. 3F). In dot immunoblotting (Fig. 4), *S. oralis* YC6(pJY-15) was labeled with anti-RPS4Gn reactive IgG but not with anti-RPS2Gn/4Gn reactive IgG. Importantly, this construct failed to support adhesion of *A. naeslundii* (Fig. 4). In comparable studies, we transformed *S. oralis* YC6 with a plasmid expressing *wcrG* for the β1–6-linked Galp branch in CPS10A (Fig. 5B); however, the resulting construct did not produce an immunoreactive cell surface product (results not shown).

DISCUSSION

The structure of CPS10F (Fig. 3A), like those of previously characterized CPS10A (24) and RPS4Gn (21) (Fig. 5B), has now been established by chemical and high resolution NMR methods. In addition, the molecular difference between these polysaccharides was defined by carbohydrate engineering of *S. oralis* RPS4Gn (Fig. 3) with genes from the *cps10F* and *cps10A* loci (Fig. 5A). The results of these studies provide new insight into the possible evolution of RPS from CPS and resolve inconsistencies that arose in earlier molecular studies (11) from errors in the available structure of CPS10F (Fig. 1). Based on the present findings, it is clear that the gene *wcrB*, which occurs in both the *cps10A* and *cps10F* loci, is associated with the same

structural feature in the corresponding polysaccharides, namely, the linkage of ribitol-5-phosphate to the 5-OH of Galf, rather than 6-OH as in RPS4Gn (Fig. 5). It is also clear from the corrected structure of CPS10F (Fig. 5B), that the genes previously designated *wcrC* in both the *cps10A* and *cps10F* loci (11, 13), are in fact, distinct. This was firmly established by the ability of each gene to change the α1–1 linkage from Gal to ribitol in RPS4Gn. Thus, expression of *wcrC* from the *cps10A* locus in *wefM* deficient *S. oralis* C104 (*i.e.* strain YC1) changed this linkage to α1–2 in our previous study (10), and expression of the corresponding gene from the *cps10F* locus changed the same linkage to α1–4 in the present study (Fig. 3D). In view of these findings, the gene previously designated *wcrC* in the *cps10F* locus has now been given the designation *wcrF* in the Bacterial Polysaccharide Gene Database (25), as indicated in Fig. 5. Finally, the present findings associate *wcrH* of *S. pneumoniae* and *wefE* of *S. oralis* with the β1–6 transfer of Galf to GalNAcβ and attribute formation of the Galf branch in CPS10F to the subsequent Wzy-dependent linkage of adjacent repeating units, which is comparable in CPS10F and CPS10A. The only structure-determining genes that are not considered in Fig. 5 are *wciG* and closely related *wefK*, which encode putative *O*-acetyltransferases. In previous studies, we associated *wefK* of *S. oralis* 10557 with partial *O*-acetylation of RPS3G (26, 27). However, the involvement of these genes in biosynthesis of either CPS10F or RPS4Gn remains to be established as NMR spectra of these polysaccharides did not reveal any evidence of *O*-acetylation.

Adhesion of *A. naeslundii* to RPS4Gn-bearing *S. oralis* but not to CPS10F-bearing *S. pneumoniae* (Fig. 4) provided experimental evidence for the critical role of β1–6-linked Galf in RPS function. However, the same *wzy*-dependent linkage that abolished adhesion of *A. naeslundii* to *S. pneumoniae* only reduced adhesion to the surface polysaccharide of *S. oralis* YC8(pJY-13) (Fig. 4). Adhesion of *A. naeslundii* to *S. oralis* YC8(pJY-13) can be explained by the action of the *S. pneumoniae* polymerase, which is not expected to affect the recognition domain at the

Molecular Basis of *S. pneumoniae* CPS10F Structure

non-reducing end of each polysaccharide chain. Whereas functional receptors are also expected at the ends of CPS10F chains, the effective cell surface density of these may be reduced on *S. pneumoniae* by the presence of long CPS10F chains and increased on *S. oralis* YC8(pJY-13) by the presence of relatively short chains, a possibility consistent with the low yield of polysaccharide from this construct. In any case, *A. naeslundii* did not attach to *S. oralis* YC6(pJY-15) (Fig. 3F), which has a surface polysaccharide identical to wild-type RPS4Gn except for the *wcrD*-dependent β 1–3 linkage between Galf and GalNAc β 1–3Gal α . Thus, β 1–6-linked Galf in RPS4Gn allowed recognition of adjacent GalNAc β 1–3Gal whereas β 1–3-linked Galf in the polysaccharide of *S. oralis* YC6(pJY-15) blocked recognition. Whether exposure of the host-like feature in wild-type RPS depends on a simple steric effect of the β 1–6 linkage or alternatively, on the flexibility of this linkage (8) and associated conformational effects remains to be determined.

The difference between CPS10F and RPS4Gn as receptors for interbacterial adhesion points to *wzy* replacement as an important step in the proposed evolution of CPS to RPS. Additional steps are suggested from the presence of ancestral *cps10A*-like sequences in both the *cps10F* and *rps4Gn* loci (Fig. 5A). Thus, the intergenic regions between *wciF/wefD* and *wzx* in these loci harbor similar 80 base pair sequences that resemble the 5'-end of *wcrG* in the *cps10A* locus. The loss of *wcrG* from a CPS10A-like serotype would eliminate the β 1–6-linked Galp branch, thereby allowing acquisition of *wcrH/wefE* for a β 1–6-linked Galf branch (Fig. 5B). Evidence for the loss of *wcrD* for a β 1–3-linked Galf branch is also clear from the *wcrD* pseudogene in the *cps10F* locus and from sequences that closely resemble the 5'- and 3'-ends of *wcrD* in the intergenic region between *wefM* and *wefD* in the *rps4Gn* locus. The apparent loss of *wcrG* or *wcrD* and acquisition of *wcrH/wefE* may be explained by selective pressure from the host immune response for the emergence of new CPS serotypes (28, 29). In contrast, conversion of CPS to RPS, via *wzy* replacement in the case of RPS4Gn, may depend on the advantage gained from RPS-mediated interactions with other commensal species, leading to the establishment of mutualism in biofilm communities (30). The time frame for the possible evolution of CPS to RPS, although largely unknown, could extend back to a common ancestor of man and the great apes, based on the host range of modern day *S. pneumoniae* (2). The greater homology seen between the *cps10F* and *rps4Gn* loci (Fig. 5) than between *cps21* and either *rps1Gn* or *rps2G* loci (9) is consistent with the different distributions of the two RPS groups on modern day commensal species. Thus, ribitol phosphate-containing types of RPS, such as RPS4Gn, have only been identified from strains of *S. oralis*, a close relative of *S. pneumoniae* whereas Glc and L-Rha-containing types of RPS (*i.e.* serotypes 1, 2, or 3) occur on strains of *S. oralis* and also on more distantly related species, including *S. gordonii* and *S. sanguinis* (6). Further insights into the evolution of these polysaccharides seem likely based on comparative genomic studies of different RPS-bearing commensal species.

The α 1–4 linkage between Gal and ribitol of CPS10F could only be established by high resolution NMR of this polysaccharide as the identification of 2-Ac-1,3,4,5 tetra-Me-ribitol by

methylation analysis of HF-treated CPS10F (Table 2) could indicate either an α 1–2 or α 1–4 linkage due to the symmetry of ribitol. Based on the membership of WcrC and WefM in glycosyltransferase family 4 of the Carbohydrate Active Enzymes database (CAZy), we previously suggested (10) that the divergent N-terminal regions of these proteins formed different acceptor binding sites for ribitol-5-phosphate and the conserved C-terminal regions similar donor binding sites for UDP-Gal. Accordingly, sequence identities of \sim 50% seen between these proteins and presently designated WcrF (Fig. 5A) depend primarily on similarities in the C-terminal regions of these proteins. The genes for these allelic transferases are important molecular determinants of RPS serotype specificity. Thus, the structure and corresponding antigenic difference between RPS4Gn and RPS5Gn was previously shown to depend on the presence of *wefM* in the *rps4Gn* locus versus *wcrC* in the *rps5Gn* locus (10). In addition, the homologous reaction of anti-RPS4Gn specific IgG with *S. oralis* C104 was abolished in the present study by changing the *wefM*-dependent, α 1–1 linkage in RPS4Gn to a *wcrF*-dependent, α 1–4 linkage in the polysaccharide produced by *S. oralis* YC1(pJY-12) (Fig. 4). A puzzling feature of CPS serogroup 10 involved the apparent genetic identity of *cps10A* to *cps10B* and *cps10C* to *cps10F* (13). It is now clear that these closely related loci are distinguished by the presence of *wcrC* or *wcrF*. Thus, *wcrC* occurs in the *cps10A* and *cps10C* loci while *wcrF* occurs in the *cps10B* and *cps10F* loci. Based on the occurrence of these genes, we hypothesize that CPS10B is similar to CPS10A (Fig. 5B) except for the presence of an α 1–4 linkage between Gal and ribitol-5-phosphate and likewise, that CPS10C is similar to CPS10F (Fig. 5B) except for the presence of an α 1–2 linkage between Gal and ribitol-5-phosphate. Comparative structural and molecular studies of these polysaccharides are currently underway to test these predictions.

Acknowledgments—We thank Russell Carlson of the Complex Carbohydrate Research Center, University of Georgia, for composition and methylation analysis, Daron Freedberg of Center for Biologics Evaluation and Research, FDA, Bethesda, MD, for facilities and assistance with the ^{31}P NMR experiments, and the Human Microbiome Project Reference Genome Initiative for information regarding *S. oralis* (U54 HF003273).

REFERENCES

- Bishop, C. J., Aanensen, D. M., Jordan, G. E., Kilian, M., Hanage, W. P., and Spratt, B. G. (2009) *Bmc Biology* 7, 3
- Kilian, M., Poulsen, K., Blomqvist, T., Håvarstein, L. S., Bek-Thomsen, M., Tettelin, H., and Sørensen, U. B. (2008) *PLoS ONE* 3, e2683
- Nyvad, B., and Kilian, M. (1987) *Scand. J. Dent. Res.* 95, 369–380
- Palmer, R. J., Jr., Gordon, S. M., Cisar, J. O., and Kolenbrander, P. E. (2003) *J. Bacteriol.* 185, 3400–3409
- Cisar, J. O., Sandberg, A. L., Abeygunawardana, C., Reddy, G. P., and Bush, C. A. (1995) *Glycobiology* 5, 655–662
- Cisar, J. O., Sandberg, A. L., Reddy, G. P., Abeygunawardana, C., and Bush, C. A. (1997) *Infect. Immun.* 65, 5035–5041
- McIntire, F. C., Crosby, L. K., Vatter, A. E., Cisar, J. O., McNeil, M. R., Bush, C. A., Tjoa, S. S., and Fennessey, P. V. (1988) *J. Bacteriol.* 170, 2229–2235
- Xu, Q., and Bush, C. A. (1996) *Biochemistry* 35, 14521–14529
- Mavroidi, A., Aanensen, D. M., Godoy, D., Skovsted, I. C., Kalltoft, M. S.,

Molecular Basis of *S. pneumoniae* CPS10F Structure

- Reeves, P. R., Bentley, S. D., and Spratt, B. G. (2007) *J. Bacteriol.* **189**, 7841–7855
10. Yang, J., Ritchey, M., Yoshida, Y., Bush, C. A., and Cisar, J. O. (2009) *J. Bacteriol.* **191**, 1891–1900
11. Aanensen, D. M., Mavroidi, A., Bentley, S. D., Reeves, P. R., and Spratt, B. G. (2007) *J. Bacteriol.* **189**, 7856–7876
12. Kamerling, J. P. (2000) in *Streptococcus pneumoniae: Molecular Biology & Mechanisms of Disease* (Tomasz, A., ed), Mary Ann Liebert, Inc., Larchmont, NY
13. Bentley, S. D., Aanensen, D. M., Mavroidi, A., Saunders, D., Rabinowitz, E., Collins, M., Donohoe, K., Harris, D., Murphy, L., Quail, M. A., Samuel, G., Skovsted, I. C., Klotz, M. S., Barrell, B., Reeves, P. R., Parkhill, J., and Spratt, B. G. (2006) *PLoS Genetics* **2**, e31
14. Xu, D. Q., Thompson, J., and Cisar, J. O. (2003) *J. Bacteriol.* **185**, 5419–5430
15. Kabat, E. A. (1961) in *Experimental Immunochemistry* (Kabat, E. A., and Mayer, M. M. ed), Charles C. Thomas, Springfield, IL
16. Xu, Q., Abeygunawardana, C., Ng, A. S., Sturgess, A. W., Harmon, B. J., and Hennessey, J. P., Jr. (2005) *Anal. Biochem.* **336**, 262–272
17. Dubois, M., Gilles, K. A., Hamilton, J. K., Rebers, P. A., and Smith, F. (1956) *Anal. Chem.* **28**, 350–356
18. Karlsson, C., Jansson, P. E., and Skov Sørensen, U. B. (1999) *Eur. J. Biochem.* **265**, 1091–1097
19. Abeygunawardana, C., and Bush, C. A. (1993) *Adv. Biophys. Chem.* **3**, 199–249
20. Walz, A., Odenbreit, S., Mahdavi, J., Borén, T., and Ruhl, S. (2005) *Glycobiology* **15**, 700–708
21. Abeygunawardana, C., Bush, C. A., and Cisar, J. O. (1991) *Biochemistry* **30**, 8568–8577
22. Beier, R. C., Mundy, B. P., and Strobel, G. A. (1980) *Can. J. Chem.* **58**, 2800–2804
23. Cisar, J. O., David, V. A., Curl, S. H., and Vatter, A. E. (1984) *Infect. Immun.* **46**, 453–458
24. Jones, C. (1995) *Carbohydr. Res.* **269**, 175–181
25. Reeves, P. R., Hobbs, M., Valvano, M. A., Skurnik, M., Whitfield, C., Coplin, D., Kido, N., Klena, J., Maskell, D., Raetz, C. R., and Rick, P. D. (1996) *Trends Microbiol.* **4**, 495–503
26. Abeygunawardana, C., Bush, C. A., and Cisar, J. O. (1991) *Biochemistry* **30**, 6528–6540
27. Yoshida, Y., Ganguly, S., Bush, C. A., and Cisar, J. O. (2005) *Mol. Microbiol.* **58**, 244–256
28. Coffey, T. J., Daniels, M., Enright, M. C., and Spratt, B. G. (1999) *Microbiology* **145**, 2023–2031
29. Coffey, T. J., Enright, M. C., Daniels, M., Morona, J. K., Morona, R., Hryniewicz, W., Paton, J. C., and Spratt, B. G. (1998) *Mol. Microbiol.* **27**, 73–83
30. Palmer, R. J., Jr., Kazmerzak, K., Hansen, M. C., and Kolenbrander, P. E. (2001) *Infect. Immun.* **69**, 5794–5804
31. Cisar, J. O., Kolenbrander, P. E., and McIntire, F. C. (1979) *Infect. Immun.* **24**, 742–752