

Comparative Structural and Molecular Characterization of *Streptococcus pneumoniae* Capsular Polysaccharide Serogroup 10^{*S}

Received for publication, April 28, 2011, and in revised form, August 18, 2011. Published, JBC Papers in Press, August 22, 2011, DOI 10.1074/jbc.M111.255422

Jinghua Yang[‡], Moon H. Nahm[§], C. Allen Bush[¶], and John O. Cisar^{‡1}

From the [‡]Oral Infection and Immunity Branch, NIDCR, National Institutes of Health, Bethesda, Maryland 20892, the [§]Department of Pathology, University of Alabama at Birmingham, Birmingham, Alabama 35249, and the [¶]Department of Chemistry and Biochemistry, University of Maryland Baltimore County, Baltimore, Maryland 21250

Streptococcus pneumoniae serogroup 10 includes four cross-reactive capsular polysaccharide (CPS) serotypes (10F, 10A, 10B, and 10C). In the present study, the structures of CPS10B and CPS10C were determined by chemical and high resolution NMR methods to define the features of each serotype. Both CPS10C and CPS10F had β 1–6-linked Galf branches formed from the termini of linear repeating units by *wzy*-dependent polymerization through the 4-OH of subterminal GalNAc. The only difference between these polysaccharides was the *wcrC*-dependent α 1–2 or *wcrF*-dependent α 1–4 linkages between Gal and ribitol-5-phosphate. The presence of one linkage or the other also distinguished the repeating units of CPS10B and CPS10A. However, whereas these polysaccharides both had β 1–3-linked Galf branches linked to GalNAc, only CPS10A had additional β 1–6-linked Galp branches. These Galp branches and the reaction of a CPS10A-specific monoclonal antibody were eliminated by deletion of *wcrG* from the *cps10A* locus. In contrast, deletion of this gene from the *cps10B* locus had no effect on the structure of CPS10B, thereby identifying *wcrG* as a pseudogene in this serotype. The β 1–3-linked Galf branches of CPS10A and CPS10B were eliminated by deletion of *wcrD* from each corresponding *cps* locus. Deletion of this gene also eliminated *wcrG*-dependent β 1–6-linked Galp branches from CPS10A, thereby identifying *WcrG* as a branching enzyme that acts on the product of *WcrD*. These findings provide a complete view of the molecular, structural, and antigenic features of CPS serogroup 10, as well as insight into the possible emergence of new serotypes.

The capsular polysaccharides (CPS)² of *Streptococcus pneumoniae* are of interest both as virulence determinants in the

pathogenesis pneumococcal infections and as protective antigens for induction of serotype-specific immunity. Over 90 CPS serotypes are currently recognized (1, 2), many of which comprise serogroups, the evolution of which is generally attributed to immune selection. A comprehensive genetic framework for the characterization of these polysaccharides recently became available following identification of the chromosomal loci (*cps*) for CPS biosynthesis from reference strains of all serotypes (3). General functions were predicted for most of the nearly 2000 genes in these loci including the 342 encoded glycosyltransferases, comprising 92 homology groups (4). In addition, a number of genes for transferases were tentatively assigned to linkages in the corresponding CPS structures, including those of CPS10A and CPS10F (4). These assignments provided starting points for earlier studies of CPS serogroup 10 (5, 6), which in turn set the stage for the present study to complete the structural and molecular characterization of this serogroup.

CPS serotypes 10F and 10A, the two most common members of serogroup 10, were distinguished over 70 years ago by the reactions of cross-absorbed rabbit antisera (7, 8). From subsequent surveys, serotype 10A was identified as the 20th most common *S. pneumoniae* serotype isolated from cases of invasive disease, and on this basis, CPS10A was selected as the representative of serogroup 10 in the 23-valent vaccine (9). More recently, two other members of this serogroup, serotypes 10B and 10C, were revealed by their reactions with available factor antisera and distinguished from these serotypes and from each other by cross-absorption studies (10). At the genetic level, high shared synteny was found between the *cps10A* locus and *cps10B* and the *cps10C* locus and *cps10F* (11). In fact as originally described (3), the loci in each syntenous pair appeared to be genetically identical.

The initial assignments of specific genes to linkages in CPS10A and CPS10F (4) were based on the available structures of these polysaccharides determined from chemical data (1). Although the proposed structure of CPS10A was confirmed by NMR (12), similar studies were not performed with CPS10F. Our interest in the latter polysaccharide arose from the striking similarity of the *cps10F* locus to the locus of a *Streptococcus oralis* coaggregation receptor polysaccharide designated RPS4Gn (5), which, unlike CPS10F, functioned as a cell surface receptor for interbacterial interactions between members of the dental plaque biofilm community. Comparative molecular studies of these closely related but functionally distinct polysac-

* This work was supported, in whole or in part, by National Institutes of Health Grant AI-31473 (to M. H. N.). This work was also supported by the Intramural Research Program of the NIDCR.

^S The on-line version of this article (available at <http://www.jbc.org>) contains supplemental Tables S1 and S2 and Figs. S1–S3.

¹ To whom correspondence should be addressed: Bldg. 30, Rm. 3A-301, 30 Convent Dr., NIDCR, NIH, 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; CSP, competence-stimulating peptide; TOCSY, total correlation spectroscopy; HSQC, heteronuclear single quantum coherence spectroscopy; HMBC, heteronuclear multiple bond correlation spectroscopy; HSQMBC, heteronuclear single quantum multiple bond correlation; HF, hydrofluoric acid.

Characterization of *S. pneumoniae* CPS Serogroup 10

TABLE 1
S. pneumoniae strains and plasmids

Strain or plasmid	Description	Source or reference
Wild type strains		
10061/38	CPS serotype 10A (SSISP 10A/1) ^a	Statens Serum Institute
423/82	CPS serotype 10B (SSISP 10B/2) ^a	Statens Serum Institute
Gro Norge	CPS serotype 10C (SSISP 10C/2) ^a	Statens Serum Institute
34355	CPS serotype 10F (SSISP 10F/2) ^a	Statens Serum Institute
Mutant strains		
JA1	Strain 10061/38 with <i>ermAM</i> in place of <i>wcjG</i>	This study
JA2	Strain 10061/38 with <i>ermAM</i> in place of <i>wcrG</i>	This study
JA3	Strain 10061/38 with <i>ermAM</i> in place of <i>wcrD</i>	This study
JB1	Strain 423/82 with <i>ermAM</i> in place of <i>wcjG</i>	This study
JB2	Strain 423/82 with <i>ermAM</i> in place of <i>wcrG</i>	This study
JB3	Strain 423/82 with <i>ermAM</i> in place of <i>wcrD</i>	This study
Plasmids		
pJY	Expression vector, Kan ^r	Ref. 5
pJY-16	pJY expressing <i>wcrG</i> of strain 10061/38	This study
pJY-17	pJY expressing <i>wcrG</i> of strain 423/82	This study

^a Statens Serum Institute *S. pneumoniae* strain number is given in parentheses.

charides revealed four discrepancies in the apparent roles of similar genes in *S. pneumoniae* and *S. oralis*, all of which were resolved by high resolution NMR of CPS10F (6). One discrepancy involved the linkage between Gal and ribitol-5-phosphate, which had been reported as α 1–2 in both CPS10A and CPS10F. The finding from NMR that this linkage was α 1–4 in CPS10F indicated that the corresponding gene in the *cps10F* was not *wcrC* as in serotype 10A but instead was an allele that is now designated *wcrF* (6). Importantly, distribution of *wcrC* and *wcrF* among the four CPS10 serotypes suggested that the presence of one allele or the other distinguished *cps10A* from closely related *cps10B* and *cps10C* from closely related *cps10F*. This in turn implied that the linkage between Gal and ribitol-5-phosphate distinguished each pair of polysaccharide structures. To test these predictions, we have now determined the structures of CPS10B and CPS10C by high resolution NMR. The results reveal the expected linkages between Gal and ribitol-5-phosphate. However, they also reveal an unexpected difference between CPS10B and CPS10A that we now show depends on the presence of a previously unrecognized pseudogene in the *cps10B* locus. The findings, which complete the structural and molecular characterization of CPS serogroup 10, provide an unparalleled view of the features that define each serotype.

EXPERIMENTAL PROCEDURES

Bacterial Strains and Culture Conditions—Wild type and mutant *S. pneumoniae* strains (Table 1) were cultured in brain-heart infusion broth (OXOID Ltd., UK) or agar supplemented with 5% heat-inactivated horse serum (Sigma) and 0.5 μ g ml⁻¹ erythromycin as needed for the maintenance of antibiotic-resistant strains.

Antibodies and Immunochemical Methods—Factor 10b and 10d rabbit antisera were purchased from Statens Serum Institute (Copenhagen, Denmark). Prior to use in dot immunoblotting, these antisera were absorbed with CPS-negative *S. pneumoniae* mutant strains JA1 and JB1 (Table 1) to remove antibodies to non-CPS cellular antigens. Absorptions were performed as previously described (5) by incubating pneumococci harvested from 3 ml of overnight broth cultures of each mutant strain with 5 μ l of factor antiserum in 0.5 ml of PBS containing 4 mg/ml BSA. Dot immunoblotting was performed as previ-

ously described (6) to detect binding of absorbed factor antisera (1/1000 dilution) or previously described (13) mouse mAb Hyp10AM6 (1:5 dilution of hybridoma cell culture supernatant) to decreasing numbers of streptococci spotted on nitrocellulose membranes. Bound antibody was detected with peroxidase-conjugated goat anti-mouse or anti-rabbit IgG (Bio-Rad) and a metal-enhanced DAB substrate kit (Pierce).

Isolation of Polysaccharides—CPS was isolated from culture supernatants of *S. pneumoniae* serotype 10B strain 423/82, serotype 10C strain Gro Norge, and mutant strains JA2, JA3, JB2, and JB3 (Table 1) following a previously described protocol (6) that included DEAE Sephacel (GE Healthcare) anion exchange column chromatography. CPS10A was purchased from Statens Serum Institute (Copenhagen, Denmark).

Chemical Methods for Carbohydrate Composition and Linkage Analysis—Glycosyl composition analysis was done by GC-MS of monosaccharide TMS derivatives at the Complex Carbohydrate Research Center (University of Georgia). After methanolysis of the polysaccharide, the resulting monosaccharide methylglycosides were derivatized and analyzed on a Supelco EC-1 capillary column. Linkage analysis was done by GC-MS of the partially methylated alditol acetates. The samples were dissolved in dimethyl sulfoxide and treated with dry NaOH and methyl iodide. Following workup, the permethylated samples were hydrolyzed in 2 M trifluoroacetic acid at 121 °C in a sealed tube, reduced with NaBD₄, and acetylated using acetic anhydride. They were analyzed on a 30-m Supelco 2330 bonded phase capillary column by electron impact mass spectrometry.

Pretreatment of some of the polysaccharide samples with 48% aqueous hydrofluoric acid (HF) at 4 °C for 2 days was used to cleave phosphodiester linkages that were anticipated for these polymers. This treatment also is known to cleave furanose linkages of sugars.

Structural Characterization of Polysaccharides by NMR Spectroscopy—NMR spectra of purified polysaccharides were recorded as in previous studies (6, 14) with a Bruker DRX 500 with a cryoprobe and a DRX700 using standard acquisition software. Generally, a 1–5-mg sample of polysaccharide was exchanged twice with 3 ml of 99.96% D₂O, lyophilized, and

dissolved in 0.6 ml of 99.99% D₂O for a 5-mm sample tube or a lesser volume for use in a Shigemi tube. Chemical shifts were recorded relative to internal acetone (¹H, 2.225 ppm; ¹³C, 31.07 ppm). All of the data were processed using NMRPipe, NMRDraw, NMRView, and Sparky software. Double quantum filtered homonuclear COSY and total correlation spectroscopy (TOCSY) along with gradient triple quantum filtered COSY spectra 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 NOESY and combined HSQC-NOESY with mixing times of 300 ms. Long range coherence-based linkage determination was done by C-H heteronuclear multiple bond correlation spectroscopy (HMBC) and by its single-quantum analog, phase-sensitive gradient heteronuclear single quantum multiple bond correlation (HSQMBC) (15). All of the chemical shifts were measured from natural abundance ¹³C-¹H HSQC spectra to avoid chemical shift and line shape distortion by the ¹H strong coupling that is common in carbohydrates. A number of single quantum heteronuclear spectra were acquired at relatively high resolution (5 Hz) in the indirect (¹³C) dimension by means of folding and acquisition of 2048 free induction decays, which was performed to increase the information content of the spectra. This was possible because the natural line width for these polysaccharides is relatively narrow.

Competence-stimulating Peptides (CSP)—Amino acid sequences of CSP of *S. pneumoniae* strains 10061/38 (*i.e.* MKNTVKLEQFVALKEKDLQKIKGGEMRISRILDFLFLRKK) and 423/82 (*i.e.* MKNTVKLEQFVALKEKDLQKIKGGEMRLSKFFRDFILQRKK) were predicted from *comC* gene sequences as described previously (5). The sequence of mature CSP of strain 10061/38, which is underlined above, was the same as CSP-2 from *S. pneumoniae* A66 (16) and was generously provided by Don Morrison (University of Illinois at Chicago). The mature CSP of strain 423/82, which is also underlined, was synthesized using automated 9-fluorenylmethoxy carbonyl chemistry and purified by high performance liquid chromatography (CEBR Research Central, National Institutes of Health).

Molecular Methods—Mutant *S. pneumoniae* strains (Table 1) were prepared by transformation of wild type strains 10061/38 or 423/82 with appropriately designed PCR constructs containing the nonpolar *ermAM* cassette flanked by ~1.0-kb gene-targeting sequences for homologous recombination with identical chromosomal sequences located upstream and downstream of the gene of interest. Transforming DNA was prepared by overlap extension PCR performed as previously described (5) with the primers listed in [supplemental Table S1](#). Transformation of strains 10061/38 and 423/82 was performed as previously described (16, 17) with minor modifications. Briefly, overnight brain-heart infusion broth cultures with 5% heat-inactivated horse serum were diluted 1:100 in fresh medium and incubated 1 or 2 h at 37 °C until A_{600 nm} reached values from 0.02 to 0.03. Transformation reactions were set up from such cultures by adding BSA (0.16%), CaCl₂ (0.01%), CSP-2 (400 ng ml⁻¹), or CSP of strain 423/82 (250 ng

ml⁻¹) and transforming DNA (1 μg ml⁻¹) to the final concentrations that are indicated in parentheses. Following incubation for 150 min at 37 °C, reaction mixtures were plated on brain-heart infusion agar supplemented with 5% heat-inactivated horse serum and erythromycin (0.5 μg ml⁻¹) to select for transformants. The location of *ermAM* cassette in each mutant strain was verified by PCR.

The plasmids listed in Table 1 were prepared by PCR amplification and cloning of *wcrG* from *S. pneumoniae* 10061/38 or 423/82 into plasmid pJY (5) using the primers listed in [supplemental Table S2](#). The integrity of each cloned gene was verified by sequencing. Transformation of pneumococci with recombinant plasmids and selection for transformants was performed as described above using brain-heart infusion plates that contained kanamycin (750 μg ml⁻¹).

RESULTS

Chemical Studies of CPS10B and CPS10C—Carbohydrate composition analysis of CPS10B and CPS10C revealed ~9% ribitol, 70% galactose, and 20% galactosamine for each polysaccharide. Linkage analysis identified three times more terminal Gal_f from intact CPS10B than from the HF-treated polysaccharide sample, which suggested the presence of Gal_f branches. 3-Linked Gal_p was detected from intact CPS10B, whereas terminal Gal_p was detected for the HF-treated sample, which suggested a phosphodiester or furanoside linkage to the 3-OH of Gal_p. 3,4-Linked Gal_NAc was detected from the intact polysaccharide, and 4-linked Gal_NAc was detected from the HF-treated sample, which suggested a phosphodiester or furanoside linkage to the 3-OH of Gal_NAc. Linkage results for ribitol from the intact sample were not interpretable because of interference of the phosphate linkage with both methylation and acid hydrolysis. However, results from the HF-treated sample suggested either 2- or 4-linked ribitol, which are indistinguishable because of the symmetry of ribitol. Linkage analysis of CPS10C gave results similar to those described for CPS10B, with the exception that 4,6-substituted Gal_NAc was detected from intact CPS10C versus 4-linked Gal_NAc from the HF-treated sample, which suggested a phosphodiester or furanoside linkage to the 6-OH of Gal_NAc. The results from chemical studies of CPS10B and CPS10C were all compatible with the structures of these polysaccharides determined from NMR spectroscopic analysis.

NMR Spectroscopic Analysis of CPS10A—We anticipated that the structure of CPS10B from strain 423/82 would be closely related to that of CPS10A from *S. pneumoniae* 10061/38 based on similarities between the chromosomal loci of these polysaccharides (11). However, the structure of CPS10A was determined by modern high resolution NMR spectroscopy (12) of the polysaccharide included in the 23-valent vaccine (Pneumovax 23; Merck & Co., Inc.). Thus, to confirm identity, we recorded and compared the ¹H-¹³C HSQC spectrum of strain 10061/38 CPS10A (Statens Serum Institute) with that of the Merck polysaccharide. The anomeric region of strain 10061/38 CPS10A (Fig. 1A) contained six resonances, as expected, that corresponded to the sugar residues previously identified (12) in the CPS10A repeating subunit. The ¹³C and ¹H resonances in the remainder of the spectrum (data not shown) were also in

Characterization of *S. pneumoniae* CPS Serogroup 10

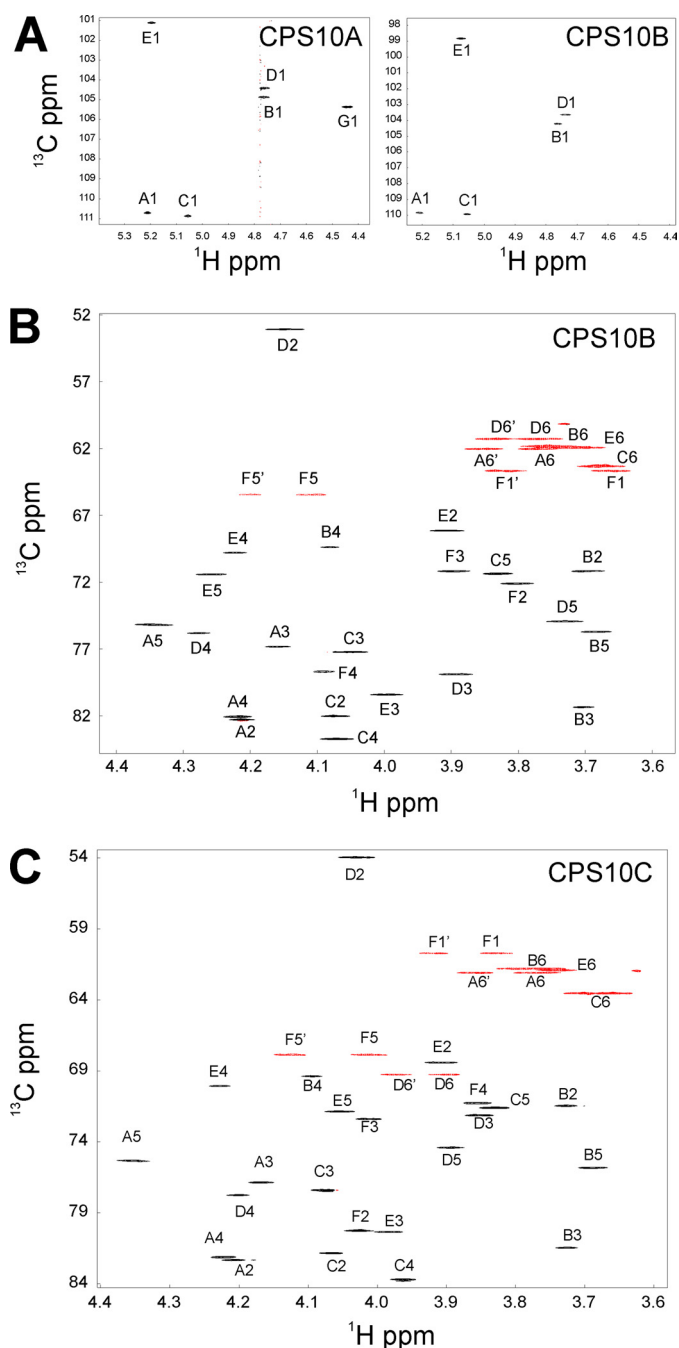


FIGURE 1. Multiplicity edited ^1H - ^{13}C HSQC spectra recorded at 25 °C; anomeric regions of CPS10A and CPS10B (A) and central regions of CPS10B (B) and CPS10C (C) spectra. Negative contours (red) indicate methylene groups.

satisfactory agreement with the data reported by Jones (12), indicating that the structures of CPS10A from the two strains examined are identical. The notation of residue letters used in Figs. 1 and 2 and Table 2 were chosen to parallel that used in previous studies of *S. oralis* receptor polysaccharide and *S. pneumoniae* CPS10F (6).

Structure of CPS10B from *S. pneumoniae* 423/82—Whereas six resonances were noted in the anomeric region of the ^1H - ^{13}C HSQC spectrum of CPS10A, only five were present in this region of the CPS10B spectrum (Fig. 1A). The resonance absent from the CPS10B spectrum was the one labeled G1 in the

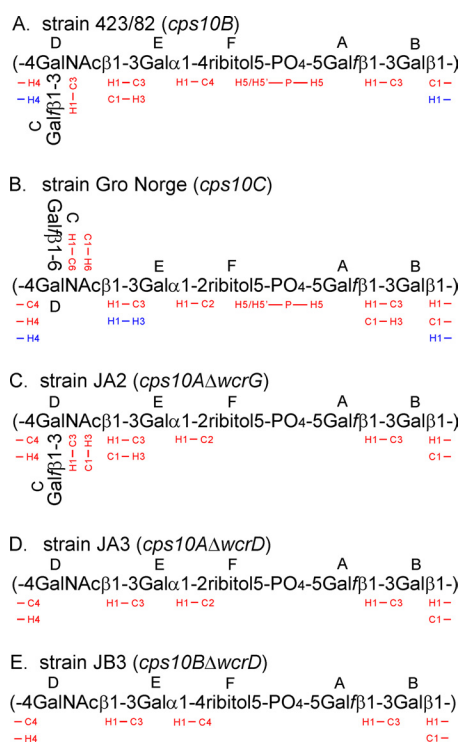


FIGURE 2. HMBC (red) and NOE (blue) inter-residue connectivities of CPS isolated from *S. pneumoniae* wild type and mutant strains.

CPS10A spectrum, which corresponds to the β -Galp side chain of CPS10A (Table 2). We began our analysis of the CPS10B spectrum by assigning the resonances for residue A. The downfield ^{13}C chemical shift of A-C1 (109.86) suggested a furanoside. COSY from A-H1 identified A-H2 at 4.210 ppm, and TOCSY indicated similar chemical shifts for H2, H3, and H4 of residue A. Although similar, these resonances were readily assigned by HSQC-TOCSY and HSQMBC because the ^{13}C chemical shifts were distinct. Following the pattern commonly observed with β -galactofuranosides (14), A-H1 showed a strong HSQC-TOCSY cross-peak to A-C2 and a weaker cross-peak to A-C3, whereas long range ^1H - ^{13}C correlation spectra showed a very strong cross-peak to A-C4, a weaker cross-peak to A-C3, and a very weak cross-peak to A-C2. Further assignments in this spin system were made from HSQC-TOCSY cross-peaks from A-C3 and A-C4 to A-H5 at 4.342 ppm accompanied by a strong cross-peak from A-H5 to A-C6, which was shown to be a methylene group in edited HSQC. This assignment was confirmed by triple quantum filtered COSY of A-H5, A-H6,6'.

Residue B is a β -galactopyraoside with B1 at 4.758 and 104.22 ppm (Table 2). COSY identified B-H2 at 3.70 ppm but TOCSY from B-H1 indicated that B-H2 might overlap with B-H3, a speculation supported by HSQC peaks at 71.22 and 81.34 ppm in this ^1H column. HSQC-TOCSY from B-H1 showed strong cross-peaks at both ^{13}C frequencies. Although B-H2 and B-H3 occurred at the same chemical shift, assignment of the corresponding ^{13}C atoms was possible by HSQC-TOCSY from B-H1 at short mixing time, which showed a stronger cross-peak for B-H2 than from B-H3 or by HSQC-NOESY from B-H1 because a cross-peak was expected only for B-H3 and not for B-H2. TOCSY from B-H1 showed a weak

TABLE 2

Residue by residue comparison of HSQC ¹H and ¹³C chemical shifts for *S. pneumoniae* serogroup 10 wild type and mutant strains

Chemical shifts from NMR spectra were recorded at 60 °C for Merck CPS10A (12), 45 °C for CPS of strains 34355 (6) and JA2, and 25 °C for CPS of strains 423/82, Gro Norge, JA3, and JB3.

Strain (cps)	Residue	H1/C1	H2/C2	H3/C3	H4/C4	H5/C5	H6/C6
	(A)						
Merck	-5Galββ1-	5.225	4.215	4.189	4.210	4.378	3.784, 3.862
(cps10A)		110.17	82.73	77.30	82.67	75.77	62.82
423/82		5.210	4.210	4.158	4.216	4.342	3.768, 3.844
(cps10B)		109.86	82.29	76.84	82.03	75.18	62.08
Gro Norge		5.214	4.205	4.170	4.222	4.353	3.772, 3.858
(cps10C)		109.86	82.37	76.89	82.13	75.35	62.09
34355		5.222	4.210	4.166	4.214	4.355	3.792, 3.857
(cps10F)		109.97	82.54	77.18	82.52	75.61	62.46
JA2		5.218	4.213	4.181	4.213	4.359	3.772, 3.855
(cps10AΔwcrG)		109.81	82.28	76.80	82.24	75.32	62.24
JA3		5.218	4.205	4.158	4.222	4.343	3.781, 3.855
(cps10AΔwcrD)		109.88	82.35	76.96	82.17	75.31	62.04
JB3		5.213	4.205	4.165	4.219	4.348	3.773, 3.858
(cps10BΔwcrD)		109.84	82.33	76.87	82.17	75.33	62.06
	(B)						
Merck	-3Galpβ1-	4.760	3.692	3.692	4.085	3.656	3.762
(cps10A)		104.41	71.73	81.8	69.84	76.06	62.24
423/82		4.758	3.700	3.702	4.077	3.684	3.740, 3.764
(cps10B)		104.22	71.22	81.34	69.42	75.74	61.82
Gro Norge		4.646	3.726	3.722	4.091	3.687	3.758, 3.796
(cps10C)		104.85	71.47	81.48	69.39	75.84	61.80
34355		4.637	3.743	3.728	4.094	3.689	3.775, 3.806
(cps10F)		105.05	71.68	81.59	69.56	75.99	62.00
JA2		4.764	3.705	3.702	4.079	3.684	3.762
(cps10AΔwcrG)		104.27	71.21	81.37	69.38	75.72	61.85
JA3		4.656	3.740	3.740	4.089	3.713	3.766, 3.789
(cps10AΔwcrD)		105.03	71.39	81.35	69.36	75.87	61.81
JB3		4.656	3.741	3.739	4.090	3.712	3.763, 3.786
(cps10BΔwcrD)		105.10	71.37	81.44	69.36	75.87	61.84
	(C)						
Merck	Galββ1-	5.073	4.078	4.055	4.068	3.841	3.684
(cps10A)		110.52	82.53	77.92	84.68	72.02	63.93
423/82		5.056	4.072	4.052	4.071	3.830	3.678
(cps10B)		109.96	82.01	77.21	83.68	71.36	63.32
Gro Norge		5.066	4.064	4.077	3.963	3.830	3.662, 3.698
(cps10C)		108.94	81.87	77.42	83.73	71.62	63.55
34355		5.057	4.077	4.072	3.971	3.841	3.673, 3.720
(cps10F)		109.14	82.00	77.72	84.03	71.92	63.81
JA2		5.070	4.078	4.057	4.072	3.838	3.687
(cps10AΔwcrG)		110.06	82.07	77.37	84.02	71.49	63.42
	(D)						
Merck	-6,-3,-4GalNAcβ1-	4.788	4.145	3.928	4.269	3.938	3.979, 4.099
(cps10A)		103.76	53.54	79.29	76.51	74.84	71.22
423/82	-3,-4GalNAcβ1-	4.737	4.144	3.891	4.274	3.726	3.772, 3.833
(cps10B)		103.64	53.10	78.91	75.84	74.95	61.32
Gro Norge	-6,-4GalNAcβ1-	4.687	4.029	3.852	4.199	3.891	3.896, 3.973
(cps10C)		103.95	53.98	72.15	77.77	74.42	69.27
34355	-6,-4GalNAcβ1-	4.679	4.027	3.845	4.186	3.877	3.888, 3.970
(cps10F)		103.99	54.20	72.38	77.95	74.53	69.39
JA2	-3,-4GalNAcβ1-	4.774	4.140	3.914	4.272	3.744	3.776, 3.84
(cps10AΔwcrG)		103.54	53.13	78.95	75.85	74.99	61.38
JA3	-4GalNAcβ1-	4.679	4.024	3.855	4.212	3.712	3.802, 3.857
(cps10AΔwcrD)		104.02	54.04	72.15	77.48	74.95	61.36
JB3	-4GalNAcβ1-	4.695	4.025	3.856	4.212	3.712	3.798, 3.856
(cps10BΔwcrD)		104.02	54.07	72.17	77.52	74.98	61.38
	(E)						
Merck	-3Galpα1-	5.203	3.938	4.006	4.272	4.057	3.762
(cps10A)		100.36	68.88	80.70	70.29	72.25	62.24
423/82		5.073	3.906	3.994	4.222	4.243	3.711, 3.757
(cps10B)		98.87	68.11	80.33	69.82	71.40	61.96
Gro Norge		5.199	3.915	3.981	4.227	4.055	3.738
(cps10C)		100.07	68.44	80.37	70.06	71.86	61.92
34355		5.080	3.926	3.970	4.235	4.246	3.740
(cps10F)		99.18	68.31	80.84	70.07	71.58	62.21
JA2		5.201	3.925	3.988	4.217	4.057	3.736
(cps10AΔwcrG)		99.89	68.42	80.22	69.92	71.83	61.81
JA3		5.070	3.895	3.984	4.221	4.262	3.729
(cps10AΔwcrD)		98.81	68.19	80.36	69.83	71.38	61.95
JB3		5.194	3.902	3.986	4.212	4.056	3.731, 3.756
(cps10BΔwcrD)		100.02	68.52	80.13	69.98	71.85	61.83
	(F)						
Merck	-2Ribitol5-	3.856, 3.931	4.017	3.879	3.998	4.022, 4.145	
(cps10A)		61.36	80.81	71.91	73.18	68.33	
423/82	-4Ribitol5-	3.659, 3.819	3.805	3.900	4.084	4.113, 4.199	
(cps10B)		63.73	72.09	71.17	78.78	65.52	
Gro Norge	-2Ribitol5-	3.828, 3.919	4.025	4.013	3.855	4.018, 4.123	
(cps10C)		60.73	80.28	72.41	71.29	67.88	
34355	-4Ribitol5-	3.677, 3.826	3.814	3.902	4.086	4.120, 4.203	
(cps10F)		63.95	72.39	71.56	79.14	65.79	
JA2	-2Ribitol5-	3.834, 3.921	4.019	3.996	3.866	4.01, 4.132	
(cps10AΔwcrG)		60.80	80.30	72.55	71.38	67.87	
JA3	-2Ribitol5-	3.66, 3.822	3.791	3.891	4.097	4.11, 4.205	
(cps10AΔwcrD)		63.66	72.08	71.22	78.70	65.45	
JB3	-4Ribitol5-	3.833, 3.917	4.022	4.011	3.860	4.014, 4.125	
(cps10BΔwcrD)		60.71	80.29	72.37	71.29	67.85	
	(G)						
Merck	Galβ1-	4.435	3.532	3.627	3.943	3.676	3.762
(cps10A)		104.84	72.09	74.23	69.94	76.30	62.24

Downloaded from www.jbc.org at university of maryland baltimore county, on October 7, 2011

Characterization of *S. pneumoniae* CPS Serogroup 10

cross-peak with a narrow line shape at 4.077 ppm, which was assigned as **B-H4**, an equatorial proton. HSQC-TOCSY from **B-H1** showed a weak peak at 69.42 ppm that was assigned as **B-C4**. NOE cross-peaks were observed from **B-H4** to **B-H3** and to a peak at 3.68 ppm that was assigned to **B-H5**. HSQC provided assignment of **B-C5** at 75.74 ppm, from which HSQC-TOCSY cross-peaks were observed to methylene protons at 3.74 and 3.76 ppm belonging to **B-H6,6'**.

The chemical shifts of residue **C**, with **C1** at 5.056 and 109.96 (Table 2), indicated a second furanoside residue. COSY identified **C-H2** at 4.072, but as was the case with residue **A**, TOCSY from **C-H1** indicated overlap with **C-H3**. Long range H-C correlation by HSQMBC showed a very strong cross-peak from **C-H1** to 83.68 ppm, which is expected for **C-C4** in a β -galactofuranoside. This peak was correlated by HSQC to a 4.071-ppm resonance that was assigned as **C-H4**. HSQC-TOCSY from **C-H1** showed a strong cross-peak with **C-C2** at 82.01 ppm and a weaker peak at 77.21 ppm that was assigned as **C-C3**. Further assignments of the spin system of residue **C** were provided by HSQC-TOCSY from **C-C4** to the overlapping signals of **C-H2** and **C-H3** as well as peaks at 3.68 and 3.83. The peak at 3.68 ppm corresponded in multiplicity-edited HSQC to a methylene group with the ^{13}C at 63.32 ppm. That ^{13}C resonance was assigned as **C-C6** because it showed HSQC-TOCSY to **C-H4**, as well as to the 3.83 ppm resonance that was assigned as **C-H5**.

For residue **D**, COSY showed correlation of **D-H1** to **D-H2** at 4.144 ppm. This resonance was correlated by HSQC to **D-C2** at 53.10 ppm, a chemical shift indicative of an amino sugar, presumably the GalNAc residue identified by chemical analysis. HSQC-TOCSY from **D-H1** gave cross-peaks to **D-C2**, as well as to **D-C3** at 78.91 and a weaker peak to **D-C4** at 75.84 ppm. TOCSY from **D-H1** gave a cross-peak to **D-H2**, another strong cross-peak to **D-H3** at 3.89 ppm, and a narrow peak at 4.274 that was characteristic of the equatorial H4 expected for β -GalNAc. HSQC-NOESY was observed from **D-H1** to **D-C3** at 78.91 ppm and to **D-C5** at 74.95 ppm, consistent with the β -anomeric configuration. HSQC-TOCSY from **D-C5** to methylene protons at 3.77 and 3.83 ppm was used to assign **D-H6,6'** and the corresponding **D-C6** at 61.32 ppm.

Residue **E** with **E1** at 5.073 and 98.87 ppm (Table 2) was identified as an α -sugar by $^1J_{\text{CH}} = 173.5$ Hz and by the small homonuclear coupling to **E-H2** in COSY spectra. **E-C2** and **E-C3** were assigned by a combination of HSQC and HSQC-TOCSY. TOCSY from **E-H1** also gave a narrow peak expected for the H4 of an α -galactopyranoside at 4.222 ppm. HMBC from **E-H1** gave cross-peaks to **E-C3** and **E-C5** as expected for this α -sugar, which provided the assignment of **E5** at 4.243 and 71.40 ppm. HSQC-TOCSY from this isolated ^1H peak gave a cross-peak to 61.96 ppm, a methylene group assigned as **E6**.

Assignment of the resonances expected for the five sugar residues in the repeat unit left the remaining signals to be assigned to the ribitol (residue **F**). This assignment was initiated with the methylene resonances of **F-H5,5'**, which were assigned as 4.113 and 4.199 ppm by correlation with ^{31}P HSQC as described below. HSQC-TOCSY from **F-C5** at 65.52 ppm identified **F-H4** at 4.084 ppm; HSQC placed **F-C4** at 78.78 ppm. HSQC-TOCSY at short mixing time (10 ms) from **F-C4** gave cross-peaks to **F-H5** and **F-H5'** as well as to **F-H3** at 3.90 ppm,

which was associated with a ^{13}C resonance at 71.17 ppm. HSQC-TOCSY at longer mixing times also identified a resonance at 3.805 ppm as part of this spin system, and this was assigned as **F-H2**. HSQC-TOCSY from the associated ^{13}C resonance at 72.09 ppm identified the methylene resonance of **F-H1,1'** at 3.659 and 3.819 ppm, which was correlated with the **F-C1** at 63.73 ppm. This assignment accounts for all the NMR signals seen in the HSQC spectrum in Fig. 1B.

Given the complete ^1H and ^{13}C spectral assignments for the CPS10B and identification of the sugar residues, the complete chemical structure was readily determined from the connectivities indicated in NOE and three-bond C-H correlation spectra. Both HMBC and HSQMBC spectra showed cross-peaks between **A-H1** and **B-C3**, proving the Gal β 1-3Gal ρ linkage. HMBC and HSQMBC cross-peaks between **B-C1** and **D-H4** as well as NOE cross-peaks between **B-H1** and **D-H4** were used to prove the Gal ρ 1-4GalNAc linkage. The HMBC cross-peak between **C-H1** and **D-C3** was obscure because of overlap with the cross-peak between **E-H1** and **F-C4**. However, the higher resolution in ^{13}C chemical shift provided by the HSQMBC experiment clearly distinguished these two important cross-peaks, establishing the Gal β 1-3GalNAc linkage. HMBC between **D-H1** and **E-C3** established the GalNAc β 1-3Gal linkage. The correlation of **E-H1** to **F-C4** was resolved by the HSQMBC spectrum, proving the Gal α 1-4ribitol linkage. The phosphodiester linkage between **F5** and **A5** was evident from the ^{31}P HSQC spectrum, which showed cross-peaks between the phosphate and **A-H5** (4.342 ppm) and **F-H5,5'** at 4.113 and 4.199 ppm, completing the structure (Fig. 2A).

Structure of CPS10C from S. pneumoniae Gro Norge—As expected from the carbohydrate composition analysis, the HSQC spectrum of CPS10C contained a peak in the methyl group region (results not shown) at 2.039 ppm (^1H) and 23.14 ppm (^{13}C) that was characteristic of the amide group in GalNAc. Also observed was a substoichiometric peak ($\sim 30\%$) at 2.140 ppm (^1H) and 21.27 ppm (^{13}C), which indicated partial *O*-acetylation of CPS10C. Although not previously noted (6), a similar peak was present in the HSQC spectrum of CPS10F but not in presently recorded spectra of CPS10A or CPS10B. The exact position(s) of *O*-acetyl groups in CPS10C (and CPS10F) could not be determined because of the weak signal from the carbonyl of this partial substituent in HMBC.

The appearance of the anomeric region of the HSQC spectrum of the CPS10C was similar to those for CPS10B (Fig. 1A) and CPS10F (6) showing five sugar residues in the repeating unit. The only difference involved the chemical shifts assigned to α -Gal (**E1**), which were near 5.2 and 100.3 ppm for CPS10C as had been noted for CPS10A versus ~ 5.07 and 99 ppm for CPS10B and CPS10F (Table 2). These similarities suggested that the linkage to ribitol in CPS10C might be α 1-2 as in CPS10A rather than the α 1-4 as in CPS10B and CPS10F. To evaluate this speculation, we carried out a full assignment of the NMR spectra and linkage assignment based on C-H coupling and NOE data to provide a rigorous proof of structure. The evidence for the structure of CPS10C is quite analogous to that provided for CPS10B. Thus, what follows is only a brief summary of the data used for the signal assignments (Table 2), along with solutions to specific problems that are most obvious from

examination of the data. For the assignment of signals in residue A (Gal β), the problem resulting from the similar chemical shifts of H2, H3, and H4 was solved using the well resolved chemical shifts of the attached ^{13}C atoms, which were assigned by long range C-H correlation and HSQC-TOCSY. The methylene group of A6 was correlated to A-C3 and A-C4 by HSQC-TOCSY. Another problem was that B-H2 and B-H3 of CPS10C displayed the same chemical shift. As described above for CPS10B, the corresponding ^{13}C atoms of CPS10C were assigned by HSQC-TOCSY at short mixing time. TOCSY from H1 identified the sharp resonance of equatorial B-H4 and B-H5, which was identified in NOESY from the resolved peak of B-H4. HSQC-TOCSY from B-C5 identified the methylene group of B6. The resonances of side chain residue C (β -Gal β) were assigned by essentially the same strategy as that used for residue A (β -Gal β). Residue D was identified as β -GalNAc by the distinctive chemical shift of D-C2. The assignment of the remaining resonances for D followed the same approach as that used for residue B, which also has the β -galacto configuration. For residue E (α -Gal), the 2 and 3 positions were assigned by a combination of TOCSY and HSQC-TOCSY, whereas E-H4 was identified by the narrow cross-peak with E-H1. HMBC from E-H1 gave cross-peaks to E-C3 and E-C5 as expected for this sugar, but with the complication that E-C3 (80.37 ppm) was very close to another signal at 80.27 ppm that was assigned as F-C2. A correct assignment was possible using HSQC-TOCSY and phase-sensitive gradient HSQMBC folded to give ^{13}C resolution of 5 Hz. HSQC-TOCSY from E-C5 showed a cross-peak with methylene protons at 3.73 ppm that were correlated with E-C6 at 61.92 ppm.

All of the remaining peaks were assigned to residue F (ribitol) beginning from the methylene resonance of F-H5,5', which was assigned as 4.018 and 4.123 by correlation with ^{31}P HSQC as described below. HSQC-TOCSY from F-C5 identified F-H4 at 3.855 ppm, and HSQC placed F-C4 at 71.29 ppm. Assignment of F3 and F2 was complicated by overlap of the ^1H resonances of F-H2, F-H3, and F-H5, all of which were in the 4.01–4.02-ppm region. We tentatively assign F-H2 as a cross-peak at 4.02 ppm in HSQC-TOCSY, connecting this peak with the ^{13}C signal of the one remaining unassigned methylene group at 60.73 ppm (F-C1). This assignment of F-H2 was supported by triple quantum filtered COSY from F-H1,1' at 3.83 and 3.92 ppm. There were unassigned ^{13}C signals at 80.28 and 72.41 ppm in the HSQC spectrum near 4.02 ppm that were assigned to F-C2 and F-C3, respectively, by specialized HSQC-TOCSY spectra. In HSQC-TOCSY with no decoupling during acquisition, we observed on the row at 80.28 ppm (F-C2) cross-peaks to F-H3 at 4.012 ppm and to F-H1 at 3.828 ppm, whereas on the row at 72.41 ppm (F-C3), we observed a cross-peak to F-H4 at 3.855 ppm and to F-H5' at 4.123 ppm. In a separate experiment done with short mixing time (10 ms), we observed on the F-C2 row the cross-peak to F-H1 and on the F-C3 row the cross-peak to F-H4.

The linkages for the CPS10C were all assigned by long range C-H correlation (HMBC and HSQMBC) augmented by HSQC-NOESY (Fig. 2B). The only difficulty was presented by the assignment of the HMBC cross-peak between D-H1 and E-C3, which was ambiguous because of the small difference in the

chemical shifts of E-C3 and F-C2. These frequencies were, however, resolved in the single-quantum spectra at 80.37 and 80.27 ppm, respectively. Thus, the linkage of α -Gal (residue E) was unambiguous in the HSQMBC spectrum, which showed resolved cross-peaks between E-H1 and both E-C3 and F-C2. The former intra-ring cross-peak is expected for this α -sugar, and the latter (*i.e.* E-H1 to F-C2) showed the Gal α 1-2ribitol linkage. The NOE cross-peak between D-H1 and E-H3 supported the assignment of the GalNAc β 1-3Gal linkage. The phosphodiester linkage between F5 and A5 was evident from the ^{31}P HSQC spectrum, which showed cross-peaks between the phosphate and A-H5 (4.3523 ppm) and F-H5,5' (4.018 and 4.123 ppm), completing the structure of CPS10C (Fig. 2B).

Molecular Basis of Branching in CPS10A and CPS10B—Synthesis of the β 1-6-linked Gal β branches in CPS10A was previously suggested to depend on *wcrG* (4), which occurs in both the *cps10A* and *cps10B* loci (3). To examine the role of this gene, we replaced it in strains 10061/38 (serotype 10A) and 423/82 (serotype 10B) with a nonpolar *ermAM* cassette and isolated CPS from the resulting mutant strains (*i.e.* JA2 and JB2, respectively) for structural characterization. The anomeric region of the high resolution NMR spectra of CPS isolated from mutant strain JA2 (*cps10A* Δ *wcrG*) were very similar to that recorded for CPS10A (Fig. 1A) except that the resonance at 4.435, 104.84 assigned to the β -Gal β side chain (residue G) was absent. The NMR spectrum of strain JA2 CPS (supplemental Fig. S1) was completely assigned (Table 2) by the methods described for CPS10B and CPS10C. The only difficulties arose from the near coincidence of the peaks for A2 and A4 that were barely resolved in the HSQC spectrum with the ^{13}C chemical shifts differing by only 0.05 ppm, which is near the limit of the resolution of our spectrum (\sim 6 Hz). The HSQC peak had double intensity, and HSQC-TOCSY without ^{13}C decoupling revealed two peaks. HSQC-TOCSY and HSQMBC from H1 provided the assignments given in Table 2. The resonances of B6 and E6 also differed by only 6 Hz in the ^{13}C dimension, but HSQC-TOCSY cross-peaks to B-H5 and E-H5 showed the correct assignment. The linkage positions were determined as shown in Fig. 2C by HSQMBC and by HSQC-NOESY at high resolution, which was important because the chemical shifts of E-C3 (80.22 ppm) and F-C2 (80.30 ppm), although close, were resolved by these techniques. Thus, based on results from NMR, deletion of *wcrG* from the *cps10A* locus eliminated β 1-6-linked Gal β branches from CPS10A. In contrast, the NMR spectrum of CPS isolated mutant strain JB2 (*cps10B* Δ *wcrG*), which is not shown, was indistinguishable from that of serotype 10B parental strain 423/82 (Fig. 1B). Thus, deletion of *wcrG* had no effect on the structure of CPS10B.

Pneumococci are typically serotyped by results from Neufeld capsular polysaccharide swelling tests performed with commercially available factor antisera. These antisera are prepared by immunization of rabbits with whole bacteria of one serotype and absorption of the resulting antiserum with whole cells of a related serotype(s) (10). In the present study, the reactions of factor 10d and factor 10b antisera with wild type strains of each CPS10 serotype and mutant constructs of serotype 10A strain 10061/38 and serotype 10B strain 423/82 were compared by dot immunoblotting (see Fig. 4). However, prior to these compari-

Characterization of *S. pneumoniae* CPS Serogroup 10

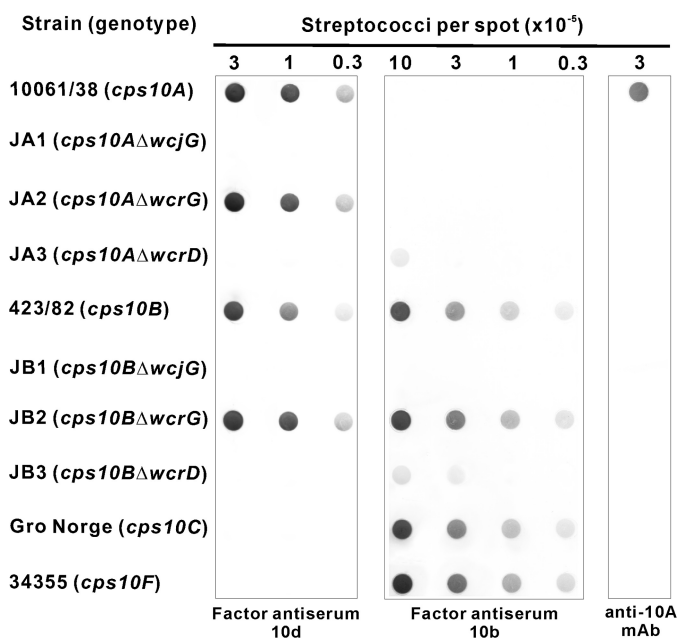


FIGURE 3. Dot immunoblotting of *S. pneumoniae* wild type and mutant strains with different factor antisera or anti-10A mAb (i.e. Hyp10AM6). Membranes were spotted with decreasing numbers of streptococci, incubated with different primary antibodies, washed, and developed with peroxidase-conjugated goat anti-rabbit or anti-mouse IgG and substrate.

sons, it was necessary to remove antibodies against non-CPS cellular antigens. This was done by absorbing each factor antiserum with CPS-negative mutant strains JA1 and JB1, which were prepared from strains 10061/38 and 423/82 by *ermAM* replacement of *wcjG*, the gene for the initial transferase in CPS biosynthesis. Following these absorptions, the reactions of each factor antiserum in dot immunoblotting were CPS-specific and paralleled results from Neufeld capsular polysaccharide swelling tests (10). Thus, factor antiserum 10d, prepared from antiserum against serotype 10A by absorption with serotypes 10C and 10F, reacted with strains 10061/38 (CPS10A) and 423/82 (CPS10B) but not with strains Gro Norge (CPS10C) and 34355 (CPS10F). Similarly, factor antiserum 10b, prepared from antiserum against serotype 10F by absorption with serotype 10A, failed to react with strain 10061/38 (CPS10A) but reacted with wild type strains bearing CPS10B, CPS10C, or CPS10F. Importantly, the reactions of factor 10d or factor 10b antiserum with *wcrG* mutant strains JA2 and JB2 were indistinguishable from those observed with the corresponding CPS10A- and CPS10B-producing wild type strains (Fig. 3). In contrast, CPS10A-specific mAb (Hyp10AM6) reacted with CPS10A-producing strain 10061/38 but not with any other strain included in the present study (Fig. 3).

The identification of mAb Hyp10AM6 as a probe for Galp branches in CPS10A opened a convenient approach for comparing the status of *wcrG* from *S. pneumoniae* serotypes 10A and 10B. Thus, binding of this mAb to mutant strain JA2 (*cps10AΔwcrG*) was fully restored by pJY-16, which expressed *wcrG* from serotype 10A strain 10061/38 but was not restored by pJY-17, which expressed the same gene from serotype 10B strain 423/82 (results not shown). Importantly, comparable binding of this mAb to serotype 10B strain 423/82 was also noted following transformation of this strain with pJY-16.

To assess the role of *wcrD*, we replaced this gene in strains 10061/38 (serotype 10A) and 423/82 (serotype 10B) with a non-polar *ermAM* cassette. When compared by dot immunoblotting (Fig. 3), the resulting mutants, strains JA3 and JB3, respectively, were unreactive with factor 10d antiserum but weakly reactive with factor 10b antiserum. The yields of CPS isolated from culture supernatants of these mutant strains were much lower (i.e. 10–100-fold) than those from wild type or other mutant strains examined in the present study. However, the amounts (~3 mg) of material isolated from strains JA3 and JB3 were sufficient for NMR studies. The anomeric regions (data not shown) of ^1H - ^{13}C HSQC NMR spectra of these polysaccharides indicated four sugar residues in the repeating subunit of CPS isolated from either JA3 or strain JB3. The spectra of these polysaccharides (supplemental Figs. S2 and S3) were completely assigned (Table 2) using the same NMR methods described above for CPS10B and CPS10C and inter-residue linkages determined by C-H long range coupling correlation (Fig. 2, D and E). The structures of the two polysaccharides were identical except for the linkages between residues E and F (i.e. α -Gal and ribitol), which were α 1–2 for the CPS of strain JA3 and α 1–4 for the CPS of strain JB3. Thus, deletion of *wcrD* eliminated β 1–3-linked Galp branches from CPS10A and CPS10B as well as *wcrG*-dependent Galp branches from CPS10A.

DISCUSSION

The structures of CPS10B and CPS10C determined from reference strains of *S. pneumoniae* in conjunction with results from previous studies (4–6) provide a comprehensive view of the molecular, structural, and antigenic features of CPS serogroup 10 (Fig. 4), as well as insight into the possible emergence of new CPS10 serotypes. Each *cps10* locus contains 11 common genes. These include four regulatory and processing genes (i.e. *wzg*, *wzh*, *wzd*, and *wze*), which occur at the 5'-end of each locus (not shown in Fig. 4A); *wcjG*, *wciB*, *wcrB*, *wciF*, and *wzy* for the common structural features of these polysaccharides; *wzx* for a flippase; and *glf* for galactofuranose mutase, the enzyme that converts UDP-Galp to UDP-Galf, an essential CPS precursor. The remaining six genes, which are distributed differently among serotypes, include *wcrC* and *wcrF* for the different linkages between Gal and ribitol-5-phosphate; *wcrD*, *wcrG*, and *wcrH* for different branches; and *wciG* for a putative *O*-acetyltransferase that presumably accounts for partial *O*-acetylation of CPS10C and CPS10F at positions that remain to be determined. The findings associate each of the four presently known CPS10 serotypes with a different combination of *wcrC*- or *wcrF*-dependent α 1–2 or α 1–4 linkages between Gal to ribitol-5-phosphate and *wcrD*- or *wcrH*-dependent β 1–3- or β 1–6-linked Galp branches from GalNAc. In addition, they show that *wcrG*-dependent β 1–6-linked Galp branches are present in CPS10A but not in CPS10B. Importantly, these branches, although specifically detected by binding of CPS10A-specific mAb Hyp10AM6, were not detected by factor 10b and 10d antisera (Fig. 3), thereby raising the possibility that isolates identified as *S. pneumoniae* serotypes 10A or 10B may include additional, unrecognized serotypes.

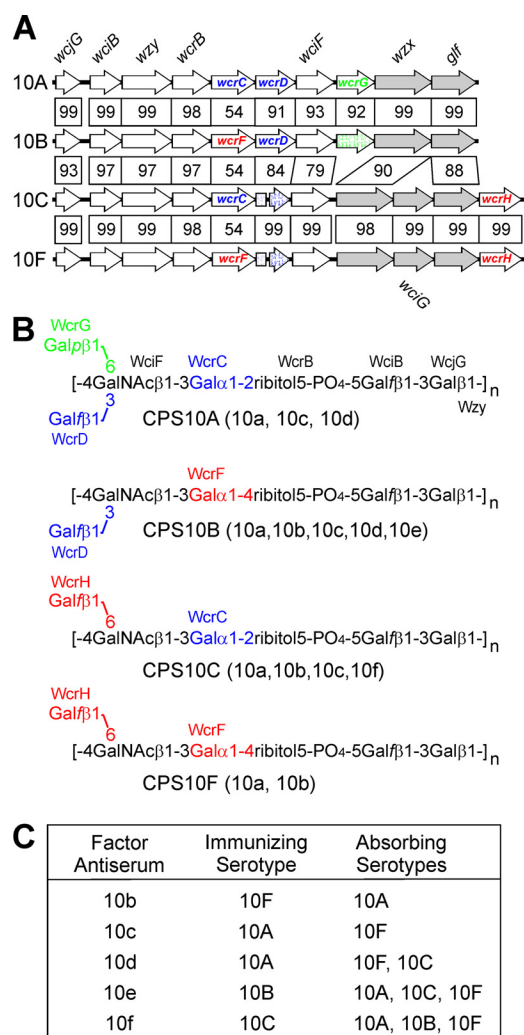


FIGURE 4. Molecular basis of *S. pneumoniae* serogroup 10 CPS structure and serotype specificity. A, structure-determining regions of *cps10* loci (GenBank™ CR931649, CR931650, CR931651, and CR931652, respectively) showing percentage identities between predicted glycosyl or ribitol-phosphate transferases and polymerases (unfilled), pseudogenes (blue or green speckles), and other genes (shaded gray) for a flippase (*wzx*), galactofuranose mutase (*glf*), and putative *O*-acetyltransferase (*wciG*). All of the genes are identified in the *cps10A* locus. Genes associated with structural differences between serotypes are labeled in green, blue, or red. B, assignment of encoded transferases and polymerase to all features of CPS10A and to the distinguishing features of other serotypes. CPS10C and CPS10F are partially *O*-acetylated, but the position(s) of *O*-acetylation are not known. The antigenic formula of each CPS10 serotype, as determined from reactions of factor antisera (10), is indicated in parentheses. Factor 10a is common to all four serotypes. C, specificities of factor antisera defined by immunizing and absorbing CPS10 serotypes.

The specificities of the factor 10b and 10d antisera used to identify CPS serotypes 10A and 10B (Fig. 4C) can now be inferred from the structures of different polysaccharides and from the history of CPS serogroup 10. The structural difference between CPS10F and CPS10A, the first two members of CPS serogroup 10 (7, 8), involves the branches in these polysaccharides as well as the linkages between Gal and ribitol-5-phosphate (Fig. 4B). Consequently, factor 10b and 10c antisera, produced by immunization with one serotype (*i.e.* 10F or 10A) and absorption with the other, are both expected to cross-react with CPS10B and CPS10C because of either the common branches or linkages between Gal and ribitol-5-phosphate in these poly-

saccharides. Following the identification of serotypes 10B and 10C (10), production of factor 10c antiserum was discontinued in favor of factor 10d antiserum, which was prepared by absorbing antiserum against serotype 10A with both serotype 10F and 10C. The inclusion of serotype 10C as an absorbing antigen is expected to remove antibodies directed against Gal α 1-2ribitol-containing epitopes of CPS10A and leave antibodies against the different branches of this polysaccharide. Thus, unlike mAb Hyp10AM6, the reaction of factor 10d antiserum is not abolished by the loss of *wcrG*-dependent β 1-6-linked Gal p branches from mutant strain JA2 (and CPS10B) but is abolished by the loss of both Gal p and Gal f branches from the CPS of *wcrD* mutant strain JA3 (Fig. 3). In addition to the positive reaction of bacteria with factor 10d antiserum, the identification of serotype 10A pneumococci depends on a negative reaction with factor 10b antiserum. The latter antiserum, which is prepared by absorbing antiserum against serotype 10F with serotype 10A, does not react with CPS10A but does react with the Gal α 1-4ribitol region of CPS10B. Not surprisingly, the structures of different polysaccharides suggest alternative strategies for the production of serotyping reagents. For example, it may well be possible to produce a factor antiserum with a serotype specificity that resembles that of mAb Hyp10AM6 by absorbing antiserum against serotype 10A with serotypes 10B and 10C. The present findings also reveal a likely explanation for the long recognized cross-reaction of factor 10d antiserum and serotype 39 *S. pneumoniae* (18), which was also found to bind mAb Hyp10AM6 (13). These reactions correlate with the presence of *wcrD*, *wciF*, and *wcrG* in the *cps39* locus, as well as *wzy* that belongs to the same homology group as *wzy* of serogroup 10, which suggests that CPS39 (like CPS10A) has β 1-6 Gal p and β 1-3 Gal f branches linked to GalNAc. Studies are underway to determine the structure of CPS39, which should in turn offer insight into the structures of other genetically related CPS serotypes (11).

We previously showed that the Gal f branches of CPS10F are formed from the termini of linear oligosaccharide repeating units by *wzy*-dependent polymerization through subterminal GalNAc (6). We have now found that the β 1-6-linked Gal p branches in CPS10A are formed in the oligosaccharide repeating unit of this polysaccharide by WcrG, a predicted member of the Core-2/I branching enzyme family. The identification of WcrG as a branching Gal p transferase was clear from the finding that deletion of *wcrD* from the *cps10A* locus eliminated not only the β 1-3 Gal f branches of CPS10A but also the *wcrG*-dependent β 1-6 Gal p branches, thereby indicating that WcrG transfers Gal p to subterminal GalNAc in the product of WcrD (*i.e.* Gal β 1-3GalNAc β 1-). Whereas deletion of *wcrG* from the *cps10A* locus eliminated Gal p branches from the CPS produced by mutant strain JA2, deletion of this gene from the *cps10B* locus had no effect on the structure of CPS10B produced by mutant strain JB2, thereby suggesting that *wcrG* occurs as a pseudogene in the *cps10B* locus. The inactivity of this pseudogene was confirmed by plasmid-based genetic complementation studies using CPS10A-specific mAb Hyp10AM6 to detect *wcrG*-dependent synthesis of Gal p branches. Thus, mAb Hyp10AM6 bound serotype 10B strain 423/82 following transformation of this strain with a plasmid harboring *wcrG* from

Characterization of *S. pneumoniae* CPS Serogroup 10

serotype 10A strain 10061/38 but not following transformation of mutant strain JA2 (*cps10AΔwcrG*) with a comparable plasmid harboring *wcrG* of serotype 10B strain 423/82. The mutation associated with the loss of *wcrG* function in strain 423/82 does not involve a frameshift like that described for conversion of CPS serotype 15B to 15C (19) but instead appears to involve a change in amino acid coding sequence. Studies are underway to identify the corresponding mutation in strain 423/82 and determine whether the same mutation occurs in other serotype 10B strains.

Although the genetic difference between CPS serotypes 10C and 10F depends on the presence of *wcrC* or *wcrF*, the difference between CPS serotypes 10A and 10B involves these genes as well as the presence or absence of functional *wcrG*. Consequently, in addition to CPS10A and CPS10B, two additional CPS10 serotypes appear to be possible. One of these serotypes would be like that of *wcrG* mutant strain JA2 (Fig. 2C), which was identified as serotype 10A by the positive and negative reactions of factor 10b and 10d antisera, respectively. However, unlike serotype 10A, strain JA2 was unreactive with mAb Hyp10AM6 (Fig. 3). Previous studies (13) of this mAb showed that it reacted with each of 12 clinical isolates identified as serotype 10A. Thus, there is no presently available evidence that isolates identified as CPS serotype 10A by their reactions with factor antisera include some that lack Gal β branches. The other novel CPS10 serotype could be derived by replacement of *wcrC* in the *cps10A* locus with allelic *wcrF* and would thus produce a CPS that is like CPS10A except for the presence of α 1–4 linkages between Gal and ribitol-5-phosphate. This serotype was derived in the present study by expression of functional *wcrG* from pJY-16 in serotype 10B strain 423/82. The resulting construct reacted like CPS10B with factor 10b and 10d antisera but, unlike CPS10B, bound mAb Hyp10AM6. This mAb, although unreactive with serotype 10B reference strain 423/82 (Fig. 3), has not yet been screened for possible reactions with other serotype 10B isolates. Regardless of the outcome of such studies, which are in progress, the present findings indicate that conversion of serotype 10A *S. pneumoniae* to a novel CPS10 serotype is possible, either from the loss of *wcrG* function or from replacement of *wcrC* with allelic *wcrF* via homologous recombination between flanking genes. Whether immunization with CPS10A, a component of the 23-valent vaccine (9), increases selective pressure for either of these events would depend on the specificity of the

human immune response for the corresponding features of this polysaccharide.

Acknowledgments—We thank the Complex Carbohydrate Research Center, University of Georgia, for glycosyl composition and linkage analyses, Daron Freedberg (Center for Biologics Evaluation and Research, FDA, Bethesda, MD) for facilities and assistance with the 31P NMR experiments, and Lotte Munch Lambertsen and Ian Chr. Skovsted (Statens Serum Institute) for generous help.

REFERENCES

1. Kamerling, J. P. (2000) in *Streptococcus pneumoniae: Molecular Biology & Mechanisms of Disease* (Tomasz, A., ed.) pp. 81–114, Mary Ann Liebert, Inc., Larchmont, NY
2. Park, I. H., Pritchard, D. G., Cartee, R., Brandao, A., Brandileone, M. C., and Nahm, M. H. (2007) *J. Clin. Microbiol.* **45**, 1225–1233
3. 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., Kalltoft, M. S., Barrell, B., Reeves, P. R., Parkhill, J., and Spratt, B. G. (2006) *PLoS Genet.* **2**, e31
4. Aanensen, D. M., Mavroidi, A., Bentley, S. D., Reeves, P. R., and Spratt, B. G. (2007) *J. Bacteriol.* **189**, 7856–7876
5. Yang, J., Ritchey, M., Yoshida, Y., Bush, C. A., and Cisar, J. O. (2009) *J. Bacteriol.* **191**, 1891–1900
6. Yang, J., Shelat, N. Y., Bush, C. A., and Cisar, J. O. (2010) *J. Biol. Chem.* **285**, 24217–24227
7. Kauffmann, F., Mørch, E., and Schmith, K. (1940) *J. Immunol.* **39**, 397–426
8. Vammen, B. (1939) *J. Immunol.* **37**, 359–365
9. Robbins, J. B., Austrian, R., Lee, C. J., Rastogi, S. C., Schiffman, G., Henrichsen, J., Mäkelä, P. H., Broome, C. V., Facklam, R. R., and Tiesjema, R. H. (1983) *J. Infect. Dis.* **148**, 1136–1159
10. Henrichsen, J. (1995) *J. Clin. Microbiol.* **33**, 2759–2762
11. Mavroidi, A., Aanensen, D. M., Godoy, D., Skovsted, I. C., Kalltoft, M. S., Reeves, P. R., Bentley, S. D., and Spratt, B. G. (2007) *J. Bacteriol.* **189**, 7841–7855
12. Jones, C. (1995) *Carbohydr. Res.* **269**, 175–181
13. Yu, J., Carvalho Mda, G., Beall, B., and Nahm, M. H. (2008) *J. Med. Microbiol.* **57**, 171–178
14. Abeygunawardana, C., and Bush, C. A. (1993) *Adv. Biophys. Chem.* **3**, 199–249
15. Williamson, R. T., Marquez, B. L., Gerwick, W. H., and Kover, K. E. (2000) *Magn. Reson. Chem.* **38**, 265–273
16. Pozzi, G., Masala, L., Iannelli, F., Manganello, R., Havarstein, L. S., Piccoli, L., Simon, D., and Morrison, D. A. (1996) *J. Bacteriol.* **178**, 6087–6090
17. Whatmore, A. M., Barcus, V. A., and Dowson, C. G. (1999) *J. Bacteriol.* **181**, 3144–3154
18. Mørch, E. (1942) *J. Immunol.* **43**, 177–202
19. van Selm, S., van Cann, L. M., Kolkman, M. A., van der Zeijst, B. A., and van Putten, J. P. (2003) *Infect. Immun.* **71**, 6192–6198