JB Journals.ASM.org

Structures of Capsular Polysaccharide Serotypes 35F and 35C of *Streptococcus pneumoniae* Determined by Nuclear Magnetic Resonance and Their Relation to Other Cross-Reactive Serotypes

C. Allen Bush,[a] John O. Cisar,[b] Jinghua Yang[c]

Department of Chemistry and Biochemistry, University of Maryland Baltimore County, Baltimore, Maryland, USA[a]; Microbial Receptors Section, Laboratory of Cell and Developmental Biology, National Institutes of Health, Bethesda, Maryland, USA[b]; State Key Laboratory of Mycology, Institute of Microbiology, Chinese Academy of Sciences, Beijing, People's Republic of China[c]

ABSTRACT

The structures of *Streptococcus pneumoniae* capsular polysaccharides (CPSs) are essential for defining the antigenic as well as genetic relationships between CPS serotypes. The four serotypes that comprise CPS serogroup 35 (i.e., types 35F, 35A, 35B, and 35C) are known to cross-react with genetically related type 20, 29, 34, 42, or 47F. While the structures of CPS serotype 35A (CPS35A) and CPS35B are known, those of CPS35F and CPS35C are not. In the present study, the serotypes of CPS35F and CPS35C were characterized by high-resolution heteronuclear magnetic resonance (NMR) spectroscopy and glycosyl composition analyses to reveal the following repeat unit structures:

CPS35F $[\text{-6Gal}f\text{-(2-OAc)-}\beta\text{1-3Gal}\alpha\text{1-2ribitol-5-PO}_4^-\text{-3Gal}f\beta\text{1-3Gal}\beta\text{1-}]_n$

CPS35C $[\text{-6Gal}f\beta\text{1-1mannitol-6-PO}_4^-\text{-3Gal}\beta\text{1-3Gal}f\beta\text{1-3Glc}\beta\text{1-}]_n$ with $\text{Glc}\alpha\text{1}$ attached to position 2

where OAc indicates O-acetylated. Importantly, CPS35F, the immunizing serotype for the production of group 35 serum, more closely resembles CPS34 and CPS47F than other members of serogroup 35. Moreover, CPS35C is distinct from either CPS35F or CPS35B but closely related to CPS35A and identical to de-O-acetylated CPS42. The findings provide a comprehensive view of the structural and genetic relations that exist between the members of CPS serogroup 35 and other cross-reactive serotypes.

IMPORTANCE

Cross-reactions of diagnostic rabbit antisera with *Streptococcus pneumoniae* capsular polysaccharide serotypes are generally limited to members of the same serogroup. Exceptions do, however, occur, most notably among a group of nonvaccine serotypes that includes the members of serogroup 35 (i.e., types 35F, 35A, 35B, and 35C) and other genetically related types. The presently determined structures of *S. pneumoniae* serotypes 35F and 35C complete the structural characterization of serogroup 35 and thereby provide the first comprehensive description of how different members of this serogroup are related to each other and to types 29, 34, 42, and 47F. The structural and genetic features of these serotypes suggest the existence of three distinct capsular polysaccharide subgroups that presumably emerged by immune selection in the human host.

The virulence of *Streptococcus pneumoniae* depends in part on the production of capsular polysaccharides (CPSs) that protect invading bacteria from phagocytosis and subsequent killing by host neutrophils (1). When administered in vaccines, these polysaccharides elicit the formation of opsonic antibodies that confer serotype-specific protective immunity. Currently available CPS-based vaccines include the 23-valent polysaccharide vaccine and the 13-valent conjugate vaccine. The widespread use of these vaccines, while highly effective in preventing invasive pneumococcal disease, may also favor the emergence of nonvaccine, replacement serotypes (2) or possible new serotypes by recombination of existing strains (3). Surveillance of both vaccine and nonvaccine pneumococcal serotypes is thus essential for monitoring and maintaining vaccine efficacy.

Twenty-four of the 92 presently recognized CPS serotypes (4–6) are distinct in the Danish system of nomenclature, while the remaining 68 are distributed among 21 serogroups that contain from two to five members each. For the most part, cross-reactions between serotypes are limited to members of the same CPS serogroup. Exceptions do, however, occur, most notably among a group of nonvaccine serotypes (7) that includes the members of CPS serogroup 35 (i.e., 35F, 35A, 35B, and 35C) and other individual types (i.e., 20, 29, 34, 42, and 47F). Examples of cross-reactions between these types include group 35 serum with types 42 and 47F (8), factor 34b serum with type 35F, factor 20b serum

Received 17 March 2015 **Accepted** 30 May 2015

Accepted manuscript posted online 8 June 2015

Citation Bush CA, Cisar JO, Yang J. 2015. Structures of capsular polysaccharide serotypes 35F and 35C of *Streptococcus pneumoniae* determined by nuclear magnetic resonance and their relation to other cross-reactive serotypes. J Bacteriol 197:2762–2769. doi:10.1128/JB.00207-15.

Editor: O. Schneewind

Address correspondence to Jinghua Yang, yangjh@im.ac.cn.

doi:10.1128/JB.00207-15

with type 35A, factor 29b serum with type 35B, and factor 20b and 42a sera with type 35C (9). Structures are available for many of these types (6), including CPS serotype 35A (CPS35A) (10) and CPS35B (11) but not for CPS35F and CPS35C. The absence of structures for the last two types limits description of the group, which was initially defined by cross-reactions of antiserum against *S. pneumoniae* serotype 35F with subsequently identified types (8, 12).

The 88 *S. pneumoniae* CPS serotypes that are synthesized by the Wzy-dependent pathway have been compared genetically by cluster analysis of structure-determining genes in the chromosomal loci (*cps*) that direct the synthesis of these polysaccharides (9). This analysis revealed eight genetic clusters, one of which (i.e., cluster 4) contained the 15 members of serogroups 10, 33, 35, and 47 and individual types 13, 20, 29, 34, 36, 39, 42, and 43. Interestingly, the members of CPS serogroup 35 did not cluster together but instead were more closely related to serotypes outside this group (i.e., type 35F was more closely related to types 47F and 34, types 35A and 35C were more closely related to type 42, and type 35B was more closely related to type 29). To better assess the genetic as well as antigenic relations between these types, we determined the structures of CPS35F and CPS35C to complete the characterization of serogroup 35.

MATERIALS AND METHODS

Glycosyl composition analysis. All procedures associated with glycosyl composition analysis were performed at the University of Georgia Complex Carbohydrate Research Center. Initially, samples (400 μg) of CPS35F, CPS35C, and CPS42 (Statens Serum Institute, Copenhagen, Denmark) were analyzed before and after pretreatment with 48% aqueous hydrofluoric acid (HF) at 4°C for 48 h to cleave phosphodiester bonds. Methyl glycosides were prepared from dried samples by methanolysis in 1 M HCl in methanol at 80°C (17 h), followed by re-*N*-acetylation with pyridine and acetic anhydride in methanol (for detection of amino sugars). Samples were then per-O-trimethylsilylated (TMS) by treatment with Tri-Sil (Pierce) at 80°C (0.5 h). Gas chromatography (GC)-mass spectrometry analysis of TMS methyl glycosides was performed on an Agilent 7890A GC interfaced to a 5975C MSD system, using an Agilent DB-1 fused-silica capillary column (30 m by 0.25 mm [inside diameter]). Peaks were identified by comparison with monosaccharide standards, which included ribitol and mannitol.

NMR spectroscopy. Nuclear magnetic resonance (NMR) studies were performed as previously described (13, 14) with native CPS35F, CPS35C, and CPS42 and with de-O-acetylated (de-OAc) samples of CPS35F and CPS42, prepared by overnight incubation of native CPS in a mild base (0.1 M NH_4OH) at 4°C. Prior to NMR, the polysaccharides (3 to 10 mg) were lyophilized twice from 99.8% D_2O and taken up in 99.996% D_2O. Spectra were recorded at 25°C in a Bruker DRX 500 spectrometer with a cryoprobe. A DRX700 spectrometer was used for ^{1}H-detected ^{31}P spectra. All proton and carbon chemical shifts were referenced relative to internal acetone using δ ^{1}H of 2.225 ppm and δ ^{13}C of 31.07 ppm.

Multiplicity-edited heteronuclear single quantum coherence (HSQC) was used to distinguish methylene from methine groups, which was useful for identification of the C-6 groups of hexoses as well as the C-1 and C-5 or C-6 groups of alditols in polysaccharides. The common homonuclear two-dimensional NMR methods of double quantum filtered coherence spectroscopy (DQF-COSY), total coherence spectroscopy (TOCSY), and nuclear Overhauser effect spectroscopy (NOESY) were augmented by triple quantum filtered coherence spectroscopy (TQF-COSY), which was useful for correlating methylene protons with adjacent positions. The utility of the hybrid method, HSQC-TOCSY, in the crowded carbohydrate spectra was enhanced by the use of a high digital resolution in the indirect dimension (^{13}C). With an increased number of data points (n = 2,048) and by folding of these single quantum spectra, the useful resolution was improved to approximately 3 Hz, approaching the natural ^{13}C line width. A second hybrid pulse sequence, HSQC-NOESY, also recorded at a high digital resolution in ^{13}C, was especially valuable for correlating C-5 of β-Gal*p* with H-1 and H-4. Heteronuclear multiple bond correlation (HMBC) spectra and a single quantum analog (HSQMBC) were used for identifying linkage positions and assignments of residues. A variation of the HMBC sequence in which the ^{13}C carrier was placed in the region of the carbonyl ^{13}C (175 ppm) was used for the reliable correlation of acetyl and amide methyl groups with proton resonances in the sugar ring. All NMR data were processed by the NMRPipe and NMRDraw software packages (NMRScience) with analysis by the NMRView program (One Moon Scientific).

RESULTS

Structure of CPS35F. Glycosyl composition analysis of native CPS35F indicated 98.6 mol% galactose and 0.8 mol% ribitol with negligible amounts of mannose, glucose, and GlcNAc. HF-treated CPS35F revealed 77.4 mol% Gal and 21.5 mol% ribitol. The increase in the amount of ribitol can be attributed to cleavage of ribitol-phosphate by HF. These findings are consistent with the results from NMR described below, which indicate four residues of Gal and one ribitol-phosphate in the CPS35F repeating unit. All Gal residues were assumed to be the normal D-isomers, like those in closely related CPS34 (9, 15). The ^{1}H-^{13}C HSQC spectrum of CPS35F contained a peak at 2.137 ppm in ^{1}H and 21.27 ppm in ^{13}C, signifying an *O*-acetyl substituent that is connected to residue C at the 2 position (see below). The anomeric region of the spectrum showed several peaks of differing intensities (results not shown), suggesting that O-acetylation of CPS35F might be heterogeneous. Therefore, to simplify the NMR spectrum, a sample of native polysaccharide was treated with mild base, as described in Materials and Methods, to produce the de-O-acetylated form. As expected, the ^{1}H spectrum of this sample lacked the peak at 2.137 ppm attributed to O-acetylation.

The anomeric region of the HSQC spectrum of de-O-acetylated CPS35F (Fig. 1A) showed four peaks that were assigned the letters A, B, C, and E, chosen to correspond to the four Gal residues of genetically related CPS47F (16). Like CPS47F, the anomeric resonance of residues A and C were assigned to β-Gal*f* on the basis of their downfield ^{13}C chemical shifts (Table 1). H-2 of each residue was assigned by DQF-COSY cross peaks with H-1, and the corresponding C-2 was assigned by HSQC. Long-range ^{1}H-^{13}C coupling correlation by HMBC can often be used in the assignment of C-3 and C-4 of furanosides, but for this system, with its limited chemical shift dispersion, the single quantum analog, HSQMBC, with its pure antiphase line shape gave more satisfactory spectra with an improved resolution in the ^{13}C dimension (Fig. 2). Strong cross peaks of H-1 with C-4 along with lesser peaks for C-3 identified these assignments for both Gal*f* residues (i.e., residues A and C). The remaining resonances for these residues were obtained by HSQC-TOCSY.

Residue E was identified to be α-Gal*p* by the narrow DQF-COSY cross peak of H-1 with H-2. The HSQMBC spectrum showed strong cross peaks of H-1 with C-3 and C-5 expected for the α-anomer. E6-H was identified by HSQC-TOCSY cross peaks with C5-C. Residue B, with B1 being at 4.503 ppm in ^{1}H and at 103.92 ppm in ^{13}C, was identified to be β-Gal*p* by the large coupling constant between H-1 and H-2 in DQF-COSY. The remaining resonances of residue B were assigned by HSQC-TOCSY and HSQC-NOESY.

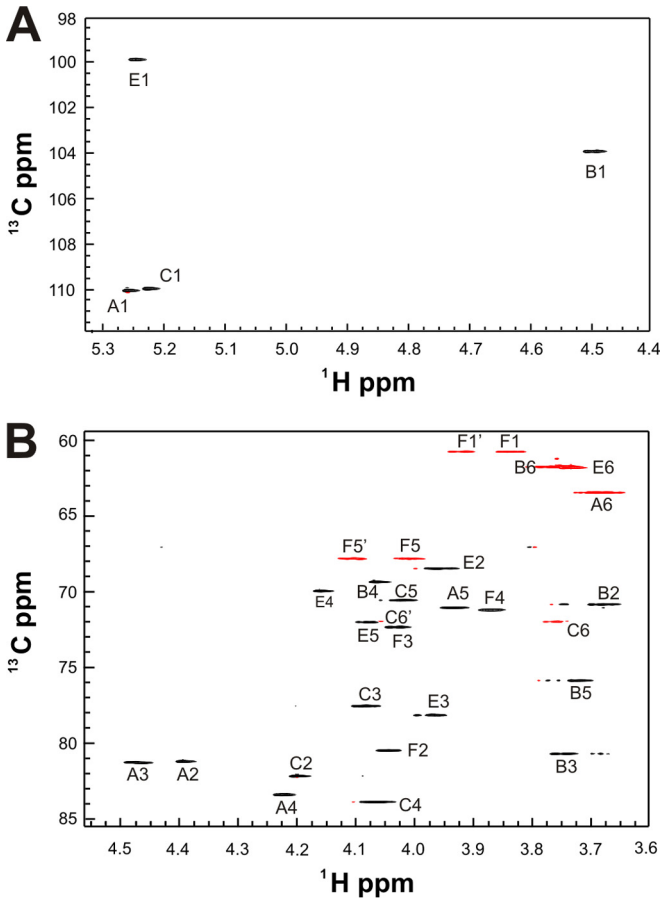

FIG 1 Anomeric region (A) and central region (B) of multiplicity-edited HSQC spectrum of de-O-acetylated CPS35F. Peaks in red represent negative intensity, signifying methylene group cross peaks.

For the assignment of residue F (i.e., ribitol), which lacks an anomeric resonance, we began with the assignment of the methylene F5 with ^{1}H at 4.010 and 4.100 ppm and ^{13}C at 67.82 ppm. Those ^{1}H resonances corresponded to peaks in a ^{1}H-^{31}P HSQC spectrum due to the phosphodiester bond. The HSQC-TOCSY cross peak between F5-C and F4-H provided assignment of the latter resonance at 3.869 ppm, and F4-C was located at 71.25 ppm by HSQC. HSQC-TOCSY provided a further correlation to F3-H at 4.040 ppm, but HSQC identified two candidate peaks for F3-C at 80.51 and 72.39 ppm, making it difficult to distinguish which one was F3-C. At this point all the resonances in the spectrum except for F1, F2, and F3 had been assigned. The sole remaining unassigned methylene peak in the edited HSQC spectrum was assigned to F1, while F2 and F3 were distinguished and assigned by HSQC-TOCSY.

Once all the resonances in the ^{1}H-^{13}C HSQC spectrum were assigned (Fig. 1; Table 1), the interresidue linkages were apparent from the long-range ^{1}H-^{13}C correlation detected in the well-resolved HSQMBC spectrum. Figure 2 shows not only the intraresidue cross peaks of A1-H, C1-H, and E1-H used in the assignment of those residues, as discussed above, but also cross peaks of A1-H with B3-C, C1-H with E3-C, and E1-H with F2-C. The corresponding NMR correlations are indicated in the structure of CPS35F in Fig. 3A. Although they are not included in Fig. 2, there are well-resolved cross peaks of B1-H with C6-C and of B1-C with C6-H and C6′-H, as well as HSQC-NOESY cross peaks between B1-C and C6,6′-H, that confirm the linkage between those two residues, as shown in Fig. 3A.

Given the complete NMR assignment of the de-O-acetylated form and the sugar backbone structure of CPS35F (Fig. 1; Table 1), assignment of the native O-acetylated form, including the position and extent of acetylation, was straightforward. Native CPS35F is a mixture of the de-O-acetylated form with a form that contains a single *O*-acetyl substituent at the 2 position of β-Gal*f* (i.e., residue C), which is linked to α-Gal*p*. Therefore, the anomeric region of the HSQC spectrum of the native form contains resonances that match those of all the residues of the de-O-acetylated form. All the correlations used to assign the resonances of de-O-acetylated CPS35F were observed for the corresponding resonances of the native form. In addition to these, two other resonances were present in the anomeric region of the spectrum of native CPS35F that were assigned to the acetylated form of CPS35F; one at 5.247, 99.93 ppm was assigned to residue E1, and the other at 5.385, 107.90 ppm was assigned to residue C1 (Table 1). The C1 peaks of acetylated and unacetylated CPS35F were approximately the same height in both the 1-dimensional ^{1}H and HSQC spectra of native CPS35F, suggesting 50% O-acetylation of the sample.

O-acetylation of residue C at the 2 position was evident from the large downfield ^{1}H shift in the cross peak between C1-H and C2-H at 5.055 ppm in the DQF-COSY spectrum of native CPS35F. There were also significant shifts of resonances at C3, C4, and C5 (Table 1). Further proof of the position of O-acetylation was provided by cross peaks between the carbonyl ^{13}C at 174.05 ppm and both the acetyl methyl group protons at 2.137 ppm and residue C2-H at 5.055 ppm (Table 1).

Structure of CPS35C. Glycosyl composition analysis of native CPS35C showed 34.8 mol% Gal, 47.7 mol% Glc, and 15.3 mol% mannitol, along with small amounts of Man and GlcNAc. After treatment with cold HF, the composition analysis was 40.8 mol% Gal, 37.9 mol% Glc, and 15.1 mol% mannitol, along with a small amount of GlcNAc. These data are consistent with the results from NMR, which showed that the CPS35C repeat contains six residues: three Gal residues, two Glc residues, and one mannitol residue. Although the absolute configuration of these sugars was not determined, we expect that they are the normal D-isomers, like those in closely related CPS35A (9, 10).

The 1-dimensional ^{1}H NMR spectra of CPS35C showed several small peaks in the 2.1-ppm region rather than the strong sharp peaks characteristic of *O*-acetyl methyl groups seen in spectra of native CPS35F and CPS42, as discussed below. The anomeric region of the ^{1}H-^{13}C HSQC spectrum showed five prominent peaks, which were interpreted to indicate five sugar residues in the CPS35C repeating unit (Fig. 4A). Several smaller peaks with intensities that were 10 to 15% of those of the five major peaks were also present; these smaller peaks were attributed to the presence of minor contaminating polysaccharides. ^{1}H-^{31}P HSQC spectra contained major peaks that were attributed to phosphodiester bonds in CPS35C along with minor unidentified contaminating polysaccharides.

The five anomeric sugar residues in the CPS35C repeat were assigned the letter designations A to E, respectively. The anomeric ^{1}H and ^{13}C signals of residues A and C at 4.354, 109.01 ppm and at 5.056, 108.78 ppm, respectively, were identified to be β-Gal*f* resi-

TABLE 1 Residue-by-residue comparison of HSQC ^{1}H and ^{13}C chemical shifts for *S. pneumoniae* CPS34, CPS35F, CPS35F-de-OAc, CPS47F, CPS47F-de-OAc, CPS35C, CPS42, and CPS42-de-OAc[a]

			Chemical shift (ppm)[b]					
Polysaccharide	Residue	Structure	1-H 1-C	2-H 2-C	3-H 3-C	4-H 4-C	5-H 5-C	6-H 6-C
CPS34	A	3-β-Gal*f*	5.302 109.15	4.354 81.02	4.473 81.58	4.301 83.94	3.947 71.29	3.699 63.80
CPS35F		3-β-Gal*f*	5.261 110.03	4.396 81.24	4.4 71 81.32	4.222 83.45	3.932 70.10	3.690 63.50
CPS47F		6-β-Gal*f*	5.301 110.04	4.359 79.61	**4.88 79.75**	4.498 81.14	**5.385 71.84**	4.051 64.40
CPS47F-de-OAc		6-β-Gal*f*	5.217 110.08	4.210 82.26	**4.096 77.47**	4.102 83.40	**3.983 70.16**	3.950, 3.981 67.19
CPS35C		3-β-Gal*f*	5.354 109.01	4.464 79.95	4.310 85.08	4.316 83.48	3.955 71.49	3.676, 3.714 63.78
CPS42[c]		3-β-Gal*f*	5.363 109.00	4.471 80.00	4.312 85.08	4.355 83.23	4.179 68.79	**4.190, 4.244 66.69**
CPS42[c]		3-β-Gal*f*	5.362 108.81	4.459 79.65	4.100 85.54	4.542 82.24	**5.541 70.89**	**4.324, 4.360 64.33**
CPS42-de-OAc		3-β-Gal*f*	5.354 109.01	4.464 79.95	4.310 85.08	4.316 83.48	3.955 71.49	3.676, 3.714 63.78
CPS34	B	3-α-Glc*p*	5.093 98.56	3.724 72.03	3.859 79.60	3.484 68.64	3.807 73.19	3.790, 3.88 61.48
CPS35F		3-β-Gal*p*	4.503 103.92	3.683 70.87	3.739 80.72	4.065 69.38	3.734 75.87	3.732, 3.784 61.69
CPS47F		3-β-Gal*p*	4.506 103.90	3.666 70.78	3.761 80.90	4.083 69.25	3.718 76.08	3.777 61.85
CPS47F-de-OAc		3-β-Gal*p*	4.504 103.87	3.668 70.86	3.737 81.10	4.101 69.37	3.724 75.95	3.769, 3.788 61.85
CPS35C		3-β-Gal*p*	4.786 103.22	3.829 72.97	4.256 76.85	4.223 68.43	3.747 75.45	3.773 61.69
CPS42		3-β-Gal*p*	4.693 103.56	3.823 73.10	4.252 76.82	4.230 68.23	3.716 75.29	3.765, 61.44
CPS42-de-OAc		3-β-Gal*p*	4.786 103.22	3.829 72.97	4.256 76.85	4.223 68.43	3.747 75.45	3.773 61.69
CPS34	C	2-β-Gal*f*	5.362 108.34	4.271 87.35	4.270 76.06	4.103 82.55	4.082 68.61	**4.247 66.46**
CPS35F		6-β-Gal*f*	5.385 107.90	**5.055 84.77**	4.246 76.49	4.136 84.79	4.065 70.51	3.758, 71.99
CPS35F-de-OAc		6-β-Gal*f*	5.228 109.95	**4.202 82.22**	4.082 77.58	4.064 83.91	4.027 70.59	3.759, 4.064 71.99
CPS47F		6-β-Gal*f*	5.386 107.89	**5.061 84.78**	4.293 76.56	4.151 84.78	4.049 70.64	3.780, 4.069 71.78
CPS47F-de-OAc		6-β-Gal*f*	5.231 109.94	**4.207 82.24**	4.077 77.65	4.069 83.90	4.022 70.71	3.771, 4.074 71.84
CPS35C		6-β-Gal*f*	5.056 108.78	4.135 81.69	4.090 77.59	4.018 83.94	4.027 70.59	3.756, 4.059 72.23
CPS42		6-β-Gal*f*	5.057 108.79	4.136 81.73	4.090 77.59	4.020 83.94	4.026 70.60	3.760, 40065 72.28
CPS42-de-OAc		6-β-Gal*f*	5.056 108.78	4.135 81.69	4.090 77.59	4.018 83.94	4.027 70.59	3.756, 4.059 72.23
CPS35C	D	3-β-Glc*p*	4.542 103.36	3.443 74.49	3.675 82.00	3.418 68.93	3.483 76.73	3.728, 3.920 61.59
CPS42		3-β-Glc*p*	4.541 103.36	3.443 74.57	3.665 81.92	3.406 68.80	3.488 76.72	3.719, 3.916 61.55
CPS42-de-OAc		3-β-Glc*p*	4.542 103.36	3.443 74.49	3.675 82.00	3.418 68.93	3.483 76.73	3.728, 3.920 61.59
CPS34	E	3-α-Gal*p*	5.238 100.31	3.957 68.52	4.008 77.92	4.143 70.28	4.091 72.00	3.747 61.80
CPS35F		3-α-Gal*p*	5.247 99.93	3.97 68.50	4.001 77.85	4.149 70.00	4.079 72.00	3.739 61.85
CPS35F-de-OAc		3-α-Gal*p*	5.247 99.88	3.971 68.50	3.970 78.16	4.159 70.00	4.080 72.05	3.739 61.85
CPS47F		3-α-Gal*p*	5.247 100.03	3.965 68.49	4.016 77.92	4.164 69.97	4.083 72.03	3.744 61.80
CPS47F-de-OAc		3-α-Gal*p*	5.250 99.95	3.962 68.50	3.977 78.17	4.168 69.98	4.082 72.02	3.733, 3.754 61.84
CPS35C		α-Glc*p*	5.564 97.72	3.515 72.67	3.841 73.71	3.340 70.94	4.096 72.55	3.741, 3.910 61.82
CPS42		α-Glc*p*	5.544 97.83	3.506 72.66	3.832 73.71	3.335 70.88	4.087 72.60	3.727, 3.910 61.78
CPS42-de-OAc		α-Glc*p*	5.564 97.72	3.515 72.67	3.841 73.71	3.340 70.94	4.096 72.55	3.741, 3.910 61.82
CPS34	F	2-Ribitol-5	3.847, 3.922 60.77	4.047 80.80	3.869 71.24	4.044 72.35	4.008, 4.096 67.79	
CPS35F		1-Ribitol-5	3.841, 3.920 60.78	4.042 80.50	4.040 72.36	3.869 71.25	4.010, 4.100 67.82	
CPS47F		2-Ribitol-5	3.845, 3.926 60.83	4.052 80.62	4.023 72.39	3.845 71.21	3.943, 4.057 67.83	
CPS47F-de-OAc		2-Ribitol-5	3.840, 3.928 60.80	4.052 80.57	3.859 71.23	4.031 72.38	3.977, 4.084 67.75	
CPS35C		1-Man-ol-6	3.688, 4.038 70.57	3.872 69.78	3.878 70.30	3.877 70.46	3.878 69.21	4.095, 4.163 67.99
CPS42		1-Man-ol-6	3.664, 4.033 70.55	3.870 69.73	3.869 70.27	3.878 70.38	3.877 69.11	4.100, 4.160 67.96
CPS42-de-OAc		1-Man-ol-6	3.688, 4.038 70.57	3.872 69.78	3.878 70.30	3.877 70.46	3.878 69.21	4.095, 4.163 67.99

[a] Chemical shifts of CPS34 (16), CPS35F, CPS47F (16), CPS35C, and CPS42 were recorded at 25°C. de-OAc, de-O-acetylated sample.
[b] Values in boldface highlight positions of O-acetylation. Pairs of values separated by a comma are H, H′ values.
[c] Different degrees of O-acetylation are present on residue A of native CPS42.

dues on the basis of the downfield ^{13}C shifts, a proposal that was confirmed in the detailed assignment of the ^{1}H-^{13}C spectrum. DQF-COSY cross peaks from the anomeric ^{1}H signals identified A2-H and C2-H, with A2-C and C2-C being obtained from the HSQC spectrum. HSQC-TOCSY cross peaks were observed between H-1 and C-2 and C-3 along with weaker cross peaks with C-4 for both residues A and C. The HMBC spectrum showed stronger cross peaks between H-1 and C-4 and weaker cross peaks between H-1 and C-3. HSQC-TOCSY cross peaks of C-3 and C-4 with H-5 and H-6 of both residues A and C completed the spectrum assignments of the two Gal*f* residues.

The assignment of residue E as an α-pyranoside with anomeric resonances at 5.564 and 97.72 ppm in ^{1}H and ^{13}C, respectively, began with identification of E2-H at 3.515 ppm by a narrow DQF-COSY cross peak with E1-H. The HSQC-TOCSY cross peak at 72.67 ppm was confirmed to be E2-C by an HSQC cross peak with E2-H. Additional HSQC-TOCSY cross peaks with E1-H included a strong peak at 73.71 ppm (E3-C) and a weak peak at 70.94 ppm

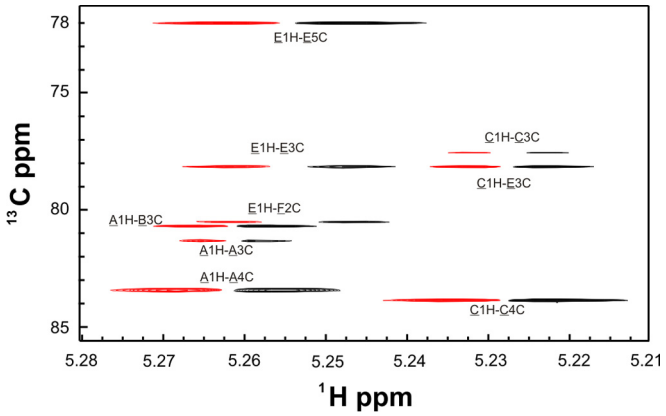

FIG 2 Partial HSQMBC spectrum of de-O-acetylated CPS35F showing selected intraresidue and interresidue C-H correlations as antiphase cross peaks.

(E4-C). The assignment of E3-C was confirmed by the HSQMBC cross peak expected for an α-pyranoside with E1-H, and the cross peak at 72.55 ppm was assigned to E5-C. That residue E is Glc*p* and not Gal*p* was shown by a strong DQF-COSY cross peak between equatorial E4-H and E5-H. In addition, there was a strong HSQC-TOCSY between E4-C, E5-H, and E6,6′-H.

For residue B, which is β-Gal*p*, the anomeric resonances were at 4.776 and 103.22 ppm in ^{1}H and ^{13}C, respectively; B2-H was located at 3.829 ppm by DQF-COSY, and B2-C was located at 72.97 ppm by HSQC. HSQC-TOCSY cross peaks of B1-H were seen at 72.97 ppm (B2-C) and at 78.86 ppm (B3-C), with a third weaker peak being seen at 68.43 ppm (B4-C). An HSQC-NOESY cross peak between B1-H and 78.86 ppm supported assignment of the peak as B3-C, and a second cross peak between B1-H and 75.45 ppm supported assignment of the peak as B5-C. HSQC-TOCSY cross peaks of B5-C point to B6-H at 3.773 ppm and the corresponding B6-C at 61.69 ppm, where the overlap with E6 was sorted out with HSQC-TOCSY.

For residue D, which is β-Glc*p*, D2-H was located by DQF-COSY and D2-C was located by HSQC. HSQC-TOCSY cross peaks with D1-H were observed at 74.49 ppm (D2-C), and at 82.00 ppm, a signal was identified to be D3-C by the occurrence of an HSQC-NOESY cross peak with D1-H at that position. A third HSQC-TOCSY cross peak with D1-H was seen at 68.93 ppm (D4-C), and a fourth at 76.73 ppm was assigned to D5-C, since an HSQC-NOESY cross peak was also seen there. The β-Glc with all equatorial protons also resulted in a very weak HSQC-TOCSY

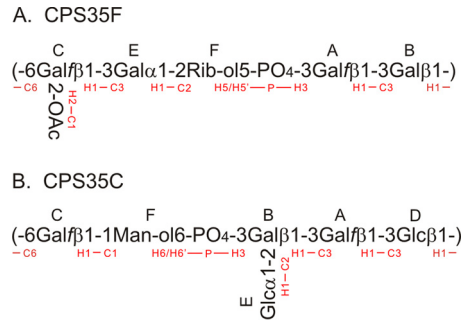

FIG 3 Structures and HMBC interresidue connectivities of CPS35F (A) and CPS35C (B).

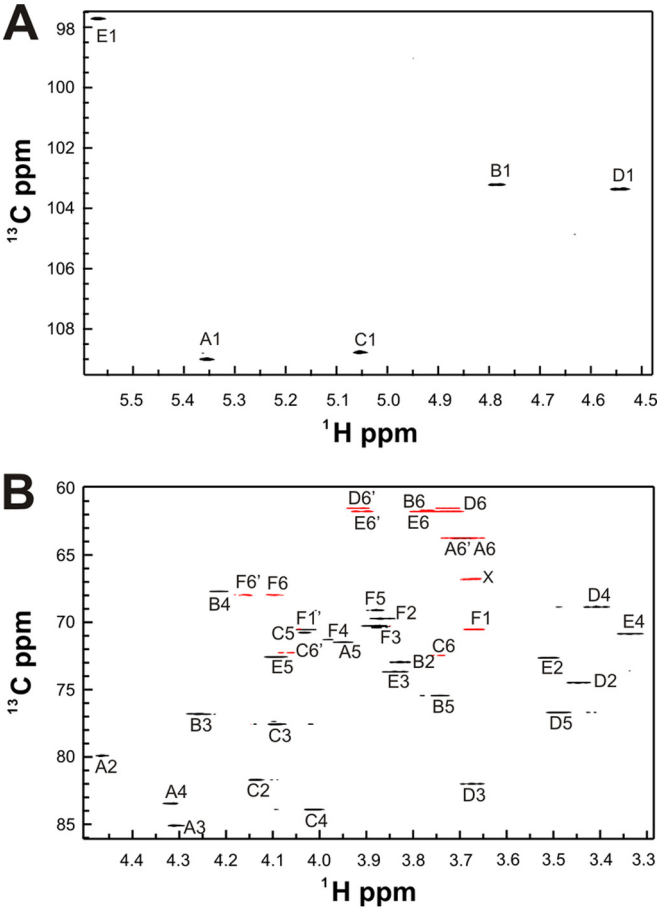

FIG 4 Anomeric region (A) and central region (B) of multiplicity-edited HSQC spectrum of CPS35C. Peaks in red represent negative intensity, signifying methylene group cross peaks.

cross peak of D1-H with D6-C, completing the assignment of residue D.

The mannitol (i.e., residue F) identified by carbohydrate composition analysis lacks a characteristic anomeric resonance and thus was located by its phosphodiester linkage using ^{1}H-^{31}P HSQC, which showed the correlation of ^{31}P to resonances at 4.164, 4.088, and 4.256 ppm. The first two corresponded to the ^{1}H frequencies of a methylene resonance in multiplicity-edited ^{1}H-^{13}C HSQC at 67.99 ppm in ^{13}C, which was assigned to F6. The ^{31}P cross peak at 4.256 ppm results from the linkage to B3, for which B3-H is at that chemical shift. An HSQC-TOCSY cross peak between the F6-C resonance and 3.878 ppm was used to identify F5-H. There were four unassigned methine peaks in HSQC very near that ^{1}H chemical shift that must represent F5 through F2, since all other peaks in the spectrum had been assigned. The sole remaining unassigned methylene resonance at 70.57 ppm in ^{13}C and at 3.668 and 4.038 ppm in ^{1}H was assigned to F1. The peak at 69.21 ppm showed HSQC-TOCSY cross peaks with F6,6′-H and, thus, was assigned to F5-C. The ^{13}C chemical shifts of F2, F3, and F4 were assigned by HSQC-TOCSY without ^{13}C decoupling during ^{1}H acquisition. In such spectra, rows at the ^{13}C chemical shifts of F2-C, F3-C, and F4-C exhibited direct peaks that were split by the 1-bond C-H coupling of approximately 140 Hz, thereby re-

vealing accurate ^{1}H chemical shifts of relayed cross peaks to provide the assignments presented in Table 1.

With the complete assignment of the ^{1}H-^{13}C NMR spectrum (Fig. 4B; Table 1), the structure of the CPS35C polysaccharide was deduced from the long-range C-H correlation determined by the HMBC and HSQMBC spectra. Linkages from residues B to A and from residues A to D were shown by cross peaks in the HMBC spectra between B1-H and A3-C and between A1-H and D3-C, respectively. Linkages from residues D and C and from residues C and F were shown by both the HMBC and HSQMBC spectra (data not shown). The phosphodiester linkage between residues F and B was defined by the cross peak between B3-H at 4.256 ppm in the ^{1}H-^{31}P HSQC spectra. Finally, the linkage between side chain residue E and residue B was clear from the HSQMBC spectra, whose 6-Hz resolution is sufficient to resolve cross peaks of E1-H with E5-C (72.55 ppm), E3-C (73.71 ppm), and B2-C (72.97 ppm). The structure of CPS35C and the interresidue correlations in the NMR spectra are given in Fig. 3B.

Structure of CPS42. Petersen et al. (17) published the structure of CPS42 prior to completion of the present study. Since the published structure and our independently determined structure are essentially identical, differing only in the extent of O-acetylation, we have truncated our description of this polysaccharide.

We found that treatment of CPS42 with a mild base to remove *O*-acetyl groups gave a product whose HSQC spectrum was identical to that of CPS35C, showing that the carbohydrate backbones of these two polysaccharides are identical. To determine the positions of O-acetylation of the backbone of the native CPS42, we used the carbonyl HMBC spectra to show that there are two forms of the CPS: one form with O-acetylation at both A5 and A6 and a second form with acetylation only at A6, consistent with published results (17). Our data indicated, however, that there is a third form that has no O-acetylation and that is thus identical to CPS35C.

The relative amounts of the three forms of CPS42 were estimated from the HSQC peaks of the A5 and A6 positions, which were well resolved for all three forms (Table 1). The relative heights of these peaks suggested 54% di-OAc form, 30% mono-OAc form, and 16% nonacetylated form. The two acetyl methyl peaks of the diacetylated form were similar in size and were about twice the size of the acetyl methyl peak of the mono-O-acetylated form, and all of these were 6-OAc. It is significant that we did not see a measurable amount of a 5-OAc form, implying that O-acetylation at that position requires the presence of a 6-OAc group. This result differs slightly from that of Petersen et al. (17), who found the major 5,6-OAc form and the 6-OAc form but not the nonacetylated form.

DISCUSSION

The structures of *S. pneumoniae* CPS35F and CPS35C determined in the present study and those of CPS35A and CPS35B characterized previously (10, 11) provide the first comprehensive description of serogroup 35 (Fig. 5). Surprisingly, the only features shared by the four members of this group are the β1-3-linked Gal*f* units at the second and fifth positions of each repeating unit. The intervening structures distinguish CPS35F (i.e., -3Galα1-2ribitol-5-PO_4^--) either from CPS35A or CPS35C (i.e., -1mannitol-6-PO_4^--3Galβ-) or from CPS35B (-1ribitol-5-PO_4^--4GalNAcβ-). Other differences between these types include the initial sugar, which is Gal*p* in type 35F and Glc*p* in the other types; the unique Glc branch of type 35C; and the O-acetylation of conserved Gal*f*. We expected that CPS35C would contain a Glc branch (like CPS42) and be fully O-acetylated (like CPS35A), based on comparisons of *cps* loci and antigenic formulas (Fig. 5). While the Glc branch was revealed by NMR of commercially available CPS35C, there was no evidence of O-acetylation (i.e., less than 10%). The possibility that samples of CPS35C prepared under different conditions are O-acetylated cannot, however, be excluded, especially since the genes (i.e., *wciG* and *wcjE*) for this modification are virtually identical in types 35A, 35C, and 42. Clearly, further studies are needed to explain why these types have different patterns of O-acetylation and also whether this contributes to differences in antigenicity.

Cross-reactions of rabbit antiserum against CPS serotype 35F with other members of serogroup 35 and closely related type 47F are represented by factor 35a in the antigenic formulas of these types (Fig. 5B). Interestingly, these cross-reactions do not appear to correlate with major features of these polysaccharides, which is perhaps not surprising, in view of the structural differences noted between CPS35F and other members of serogroup 35. Instead, the cross-reactions of factor 35a serum may depend on structural similarities that are more subtle, such as -6Gal*f*-(2-OAc)β- in types 35F, 47F, 35A, and 35B and perhaps other features in CPS35C that remain to be identified. In contrast, cross-reactions of other diagnostic sera, including factor serum 35b (i.e., anti-35F serum absorbed with type 35A and 34), factor serum 35c (i.e., anti-35A serum absorbed with types 35F), factor serum 42a, and, as previously suggested (11), factor serum 29b, can be correlated with common features in the repeating units of different polysaccharides. The use of these cross-reactions (rather than those of anti-type 35F) to define CPS serogroups is clearly possible and may well have occurred had all presently recognized types been available to Kauffmann and coworkers (12) at the time that these investigators initiated serological studies of pneumococci that led to the Danish system of nomenclature (8). While a more direct correlation of CPS35 serotypes with structural (and genetic) types would be aesthetically pleasing, this would probably not simplify identification of individual types, which would still have to be tested for their reactions with different group- and factor-specific sera prepared in rabbits. The recognition of three structural (and genetic) subgroups among these nonvaccine serotypes is, however, of interest from an evolutionary standpoint, as these subgroups display the genetic, structural, and antigenic properties expected of authentic serogroups that emerged by immune selection within the human host. With this in mind, it would be of interest to compare the reactions of convalescent-phase human sera and diagnostic rabbit sera with members of these groups.

All structure-determining genes in the *cps* loci of serogroup 35 and related type 29 can be unambiguously assigned to specific residues and linkages of the corresponding polysaccharide structures (Fig. 5B) in a manner consistent with that used in previous studies of these types (18) and types 34, 42, and 47F (16). CPS structures are now available for all five serotypes with genes designated *wcrJ* (Table 2). In types 35A, 35C, and 42, these genes are associated with the transfer of mannitol-6-phosphate to Gal*p*, while in types 29 and 35B, they are associated with the transfer of ribitol-5-phosphate to GalNAc. The possibility that different transferases catalyze these reactions is suggested by segregation of the corresponding WcrJ sequences into distinct homology subgroups with sequences that are only about 40% identical.

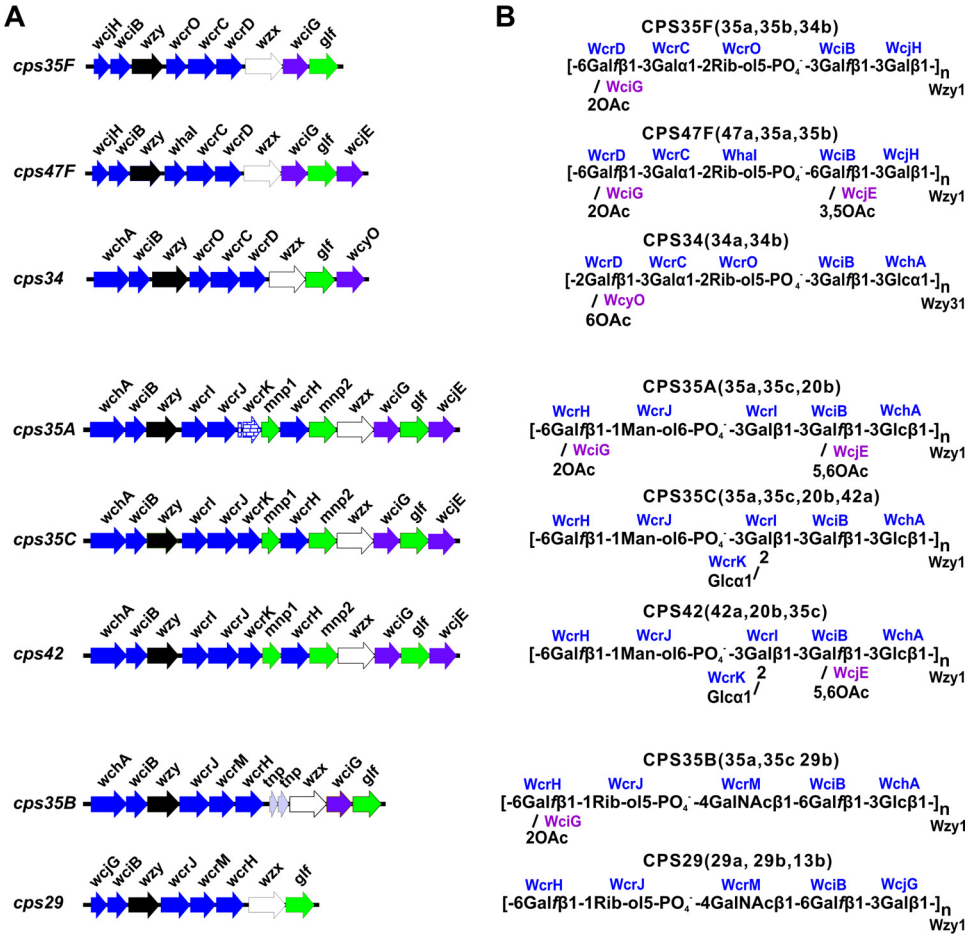

FIG 5 Association of structure-determining genes in the *cps* loci of *S. pneumoniae* serogroup 35 and related serotypes (GenBank accession numbers CR931707, CR931721, CR931703, CR931704, CR931706, CR931715, CR931705, and CR931694) (A) with the corresponding CPS structures determined in the present or previous investigations (10, 11, 16, 17, 19) (B). The antigenic formulas of the CPS serotypes are indicated in parentheses. Colors identify genes for glycosyl or polyalcohol-phosphate transferases (blue), polymerases (black), flippases (white), *O*-acetyltransferases (purple), and precursor biosynthesis (green); the *cps35A* locus contains a *wcrK* pseudogene (blue brick-like pattern). Genes designated *wcrH* or *wcrJ* are associated with features in CPS35A, CPS35C, and CPS42 different from those in CPS35B and CPS29 (Table 2).

Interestingly, the five CPS serotypes with genes designated *wcrJ* also have genes designated *wcrH* (Table 2) that are associated with the transfer of Gal*f* to mannitol-6-phosphate in types 35A, 35C, and 42 but to ribitol-5-phosphate in types 29 and 35B. It is unclear, however, whether the corresponding WcrH homologues, which segregate into two distinct homology subgroups, are members of the same or different complementation groups. In other words, WcrH homologues of different serotypes may act on the structurally similar 1-OH groups of mannitol-6-phosphate and ribitol-5-phosphate, as previously suggested (18), or alternatively, they may act specifically on one polyalcohol-phosphate or the other. Genes designated *wcrH* also occur in the *cps* loci of serotypes 10F and 10C, where they have been shown (14) to direct the transfer of Gal*f* to the C-6-OH of GalNAc (Table 2). The possibility that these genes, which encode a third WcrH subgroup, are also associated with the transfer of Gal*f* to the C-1-OH of mannitol-6-phosphate or ribitol-5-phosphate cannot be excluded at this time. Finally, it is noteworthy that the sequences of the WcrH homologues of CPS serotypes 36 and 43 are distinct from each other as well as from those of other CPS serotypes. Additional WcrH acceptors may thus be suggested from the structures of CPS36 and CPS43, the only serotypes in genetic cluster 4 (9) that remain to be structurally characterized.

TABLE 2 Specificity and serotype distribution of selected *S. pneumoniae* polyalcohol phosphate or glycosyl transferases inferred from known CPS structures

Gene[a]	Structure			*S. pneumoniae* serotypes possessing a gene whose CPS structure is:	
	Donor	Linkage	Acceptor	Known	Unknown
wcrJ	Man-ol	6-P-3	Gal*p*	35A, 35C, 42	
	Rib-ol	5-P-4	GalNAc	29, 35B	
wcrH	Gal*f*	β1-1	Man-ol-6-P	35A, 35C, 42	36, 43
	Gal*f*	β1-1	Rib-ol-5-P	29, 35B	
	Gal*f*	β1-6	GalNAc	10F, 10C	

[a] Genes designated *wcrJ* and *wcrH* encode members of protein homology groups 129 and 51, respectively (20).

ACKNOWLEDGMENTS

We thank Parastoo Azadi for helpful discussions regarding glycosyl composition analysis, Bruce Johnson for assistance with the NMRView program and assignment of polysaccharide NMR spectra, and Ian Skovsted for helpful discussions of CPS serotyping.

This work was supported by NIH grant P01-HL-107153 (to C.A.B.), the Intramural Research Program of the NIH, NIDCR (to J.O.C.), and grant 31470003 from the National Natural Science Foundation of China (to J.Y.).

REFERENCES

1. **Austrian R.** 2000. The enduring pneumococcus: unfinished business and opportunities for the future, p 3–7. *In* Tomasz A (ed), *Streptococcus pneumoniae*: molecular biology & mechanisms of disease. Mary Ann Liebert, Inc., Larchmont, NY.
2. **Croucher NJ, Harris SR, Fraser C, Quail MA, Burton J, van der Linden M, McGee L, von Gottberg A, Song JH, Ko KS, Pichon B, Baker S, Parry CM, Lambertsen LM, Shahinas D, Pillai DR, Mitchell TJ, Dougan G, Tomasz A, Klugman KP, Parkhill J, Hanage WP, Bentley SD.** 2011. Rapid pneumococcal evolution in response to clinical interventions. Science **331:**430–434. http://dx.doi.org/10.1126/science.1198545.
3. **Salter SJ, Hinds J, Gould KA, Lambertsen L, Hanage WP, Antonio M, Turner P, Hermans PW, Bootsma HJ, O'Brien KL, Bentley SD.** 2012. Variation at the capsule locus, *cps*, of mistyped and non-typable *Streptococcus pneumoniae* isolates. Microbiology **158:**1560–1569. http://dx.doi.org/10.1099/mic.0.056580-0.
4. **Calix JJ, Porambo RJ, Brady AM, Larson TR, Yother J, Abeygunwardana C, Nahm MH.** 2012. Biochemical, genetic, and serological characterization of two capsule subtypes among *Streptococcus pneumoniae* serotype 20 strains: discovery of a new pneumococcal serotype. J Biol Chem **287:**27885–27894. http://dx.doi.org/10.1074/jbc.M112.380451.
5. **Park IH, Pritchard DG, Cartee R, Brandao A, Brandileone MC, Nahm MH.** 2007. Discovery of a new capsular serotype (6C) within serogroup 6 of *Streptococcus pneumoniae*. J Clin Microbiol **45:**1225–1233. http://dx.doi.org/10.1128/JCM.02199-06.
6. **Kamerling JP.** 2000. Pneumococcal polysaccharides: a chemical view, p 81–114. *In* Tomasz A (ed), *Streptococcus pneumoniae*: molecular biology & mechanisms of disease. Mary Ann Liebert, Inc., Larchmont, NY.
7. **Skov Sorensen UB.** 1993. Typing of pneumococci by using 12 pooled antisera. J Clin Microbiol **31:**2097–2100.
8. **Lund E, Henrichsen J.** 1978. Laboratory diagnosis, serology and epidemiology of *Streptococcus pneumoniae*. Methods Microbiol **12:**241–262.
9. **Mavroidi A, Aanensen DM, Godoy D, Skovsted IC, Kaltoft MS, Reeves PR, Bentley SD, Spratt BG.** 2007. Genetic relatedness of the *Streptococcus pneumoniae* capsular biosynthetic loci. J Bacteriol **189:**7841–7855. http://dx.doi.org/10.1128/JB.00836-07.
10. **Beynon LM, Richards JC, Perry MB.** 1997. Identification of the common antigenic determinant shared by *Streptococcus pneumoniae* serotypes 35A and 20 capsular polysaccharides—structural analysis of the *Streptococcus pneumoniae* serotype 35A capsular polysaccharide. Eur J Biochem **250:**163–167. http://dx.doi.org/10.1111/j.1432-1033.1997.00163.x.
11. **Beynon LM, Richards JC, Perry MB, Kniskern PJ.** 1995. Characterization of the capsular antigen of *Streptococcus pneumoniae* serotype 35B. Can J Chem **73:**41–48. http://dx.doi.org/10.1139/v95-006.
12. **Kauffmann F, Morch E, Schmith K.** 1940. On the serology of the pneumococcus-group. J Immunol **39:**397–426.
13. **Abeygunawardana C, Bush CA.** 1993. Determination of the chemical structure of complex polysaccharides by heteronuclear NMR spectroscopy. Adv Biophys Chem **3:**199–249.
14. **Yang J, Shelat NY, Bush CA, Cisar JO.** 2010. Structure and molecular characterization of *Streptococcus pneumoniae* capsular polysaccharide 10F by carbohydrate engineering in *Streptococcus oralis*. J Biol Chem **285:**24217–24227. http://dx.doi.org/10.1074/jbc.M110.123562.
15. **Roberts WK, Buchanan JG, Baddiley J.** 1963. The specific substance from *Pneumococcus* type 34 (41). The structure of a phosphorus-free repeating unit. Biochem J **88:**1–7.
16. **Bush CA, Yang J, Yu B, Cisar JO.** 2014. Chemical structures of *Streptococcus pneumoniae* capsular polysaccharide type 39 (CPS39), CPS47F, and CPS34 characterized by nuclear magnetic resonance spectroscopy and their relation to CPS10A. J Bacteriol **196:**3271–3278. http://dx.doi.org/10.1128/JB.01731-14.
17. **Petersen BO, Meier S, Paulsen BS, Redondo AR, Skovsted IC.** 2014. Determination of native capsular polysaccharide structures of *Streptococcus pneumoniae* serotypes 39, 42, and 47F and comparison to genetically or serologically related strains. Carbohydr Res **395:**38–46. http://dx.doi.org/10.1016/j.carres.2014.06.018.
18. **Aanensen DM, Mavroidi A, Bentley SD, Reeves PR, Spratt BG.** 2007. Predicted functions and linkage specificities of the products of the *Streptococcus pneumoniae* capsular biosynthetic loci. J Bacteriol **189:**7856–7876. http://dx.doi.org/10.1128/JB.00837-07.
19. **Kenne L, Lindberg B.** 1988. The structure of the *Streptococcus pneumoniae* type 29 polysaccharide: a re-examination. Carbohydr Res **184:**288–291. http://dx.doi.org/10.1016/0008-6215(88)80033-8.
20. **Bentley SD, Aanensen DM, Mavroidi A, Saunders D, Rabbinowitsch E, Collins M, Donohoe K, Harris D, Murphy L, Quail MA, Samuel G, Skovsted IC, Kaltoft MS, Barrell B, Reeves PR, Parkhill J, Spratt BG.** 2006. Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. PLoS Genet **2:**e31. http://dx.doi.org/10.1371/journal.pgen.0020031.