Available online at www.sciencedirect.com
SCIENCE DIRECT®
Carbohydrate Research 340 (2005) 403–409
Carbohydrate RESEARCH

Full NMR assignment and revised structure for the capsular polysaccharide from *Streptococcus pneumoniae* type 15B

Christopher Jones* and Xavier Lemercinier

Laboratory for Molecular Structure, National Institute for Biological Standards and Control, Blanche Lane, South Mimms, Herts EN6 3QG, UK

Received 22 July 2004; received in revised form 23 September 2004; accepted 2 December 2004
Available online 13 January 2005

Abstract—The capsular polysaccharide from *Streptococcus pneumoniae* Type 15B is a component of the 23-valent polysaccharide vaccine against pneumococcal disease. We report full NMR assignments for the native and de-*O*-acetylated polysaccharide, and confirm that the phosphorylated substituent is glycerol-2-phosphate rather than phosphocholine, located on O-3 of the side chain β-Gal*p* residue. The polysaccharide is *O*-acetylated on the terminal α-Gal residue, distributed between *O*-2, *O*-3, *O*-4 and *O*-6 in a ratio of 6:12:12:55, with approximately 15% of the repeat units not *O*-acetylated.

Keywords: *Streptococcus pneumoniae*; Capsular polysaccharide; Vaccine; Glycerol phosphate

1. Introduction

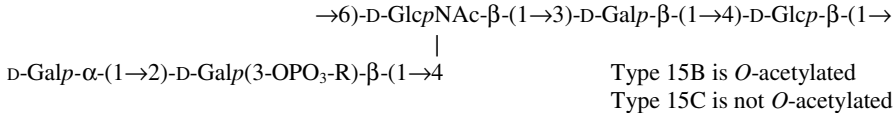
The pathogenic bacterium *Streptococcus pneumoniae* remains a major cause of death, both in the developed and developing worlds, and with the increasing incidence of antibiotic resistance in this organism, there has been a renewed interest in prophylaxis through vaccination. A vaccine, consisting of a mixture of the capsular polysaccharides (CPSs) from 23 of the 90 serotypes of *S. pneumoniae* has been available for more than 20 years,[1] and there has been increased usage of this vaccine recently, particularly for the elderly. At the same time, polysaccharide–protein conjugate vaccines against pneumococcal infection have been developed: such vaccines are also effective against invasive disease in infants and, to a lesser extent, against otitis media.[2–4] Quality control methods to assess the identity and purity of these polysaccharides previously depended on a combination of classical wet chemical methods to establish composition, and immunological tests.[5] These methods are gradually being replaced by the use of NMR spectroscopy,[6–9] which provides more complete information on composition, structure and purity in a single experiment. The use of the NMR methodology is more readily accepted if the structures are defined and full NMR assignments are available.

The CPS from *S. pneumoniae* Type 15B (Pn15B, Danish nomenclature: Type 54 in US nomenclature) is a component of the 23-valent polysaccharide vaccine, and has been found to be a common serotype causing invasive childhood infections in Bangladesh.[10] The structure of the repeat units of the Types 15B and 15C CPSs were investigated by Jansson et al.,[11] using classical degradative procedures, and were reported to be as shown below, structure **1** where R is choline in about 20% of the repeat units. Approximately 70% of the repeat units in Pn15B were *O*-acetylated, although the location was not defined.[11] This structure is identical to the tetrasaccharide repeat unit of the Type 14 CPS,[12] but with an additional α-Gal*p* residue and a

Abbreviations: COSY, correlation spectroscopy; CPS, capsular polysaccharide; HSQC, heteronuclear single quantum correlation spectroscopy; NOE, nuclear Overhauser enhancement; NOESY, NOE spectroscopy; ROESY, rotating frame NOE spectroscopy; TOCSY, total correlation spectroscopy.

* Corresponding author. Tel.: +44 1707 641211; fax: +44 1707 646730; e-mail: cjones@nibsc.ac.uk

0008-6215/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.carres.2004.12.009



1

Type 15B is *O*-acetylated
Type 15C is not *O*-acetylated

phosphorylated substituent. However, the earlier studies of Venkateswaran et al. had suggested the presence of 1 M equiv of glycerol per repeat unit, and that the *O*-acetyl group is located on one of the Gal residues, probably at an O-6 position.[13] The presence of a phosphocholine substituent in a pneumococcal CPS is unusual, and Abeygunawarda et al. report a revised structure for the Pn15B CPS,[9] without experimental evidence, containing a glycerol-2-phosphate substituent rather than phosphocholine. The pneumococcal CPSs contain, as a ubiquitous contaminant,[14] pneumococcal C-polysaccharide which contains, usually but not invariably, two phosphocholine substituents per repeat unit,[15] and which may have given rise to a misassignment of the structure.

We report here the full assignment of the ^{1}H, ^{13}C and ^{31}P NMR spectra of the native and de-*O*-acetylated Type 15B CPS, confirm the revised structure of Abeygunawardana et al., and define the location of the *O*-acetyl group.

2. Experimental

The polysaccharide samples were from material prepared for vaccine manufacture. Type 15B CPS (ca. 3 mg in 300 μL) was de-*O*-acetylated by treatment with 200 mM sodium hydroxide for 90 min at room temperature. The sample was desalted by passage through a pad of cation-exchange resin (BioRad AG50Wx8, 100–200 mesh, H^+ form), neutralised with ammonia and lyophilised. CPS or de-*O*-acetylated CPS (ca. 3–5 mg) were dissolved on 0.5 mL of deuterated water (M&G Chemicals, Glossop, UK) and dried in vacuo, then redissolved in 320 μL of the same deuterated water and introduced into a 5 mm Shigemi susceptibility-matched NMR tube (Shigemi, Japan). The sample was depolymerised by extensive sonication.[16] Spectra were recorded as previously described,[17] using Varian Unity 500 and 600 and Varian Inova 500 spectrometers at an indicated probe temperature of 50 °C. Standard Varian pulse sequences were used, except for the HSQC experiment, which used the pulsed field gradient method of Wider and Wüthrich[18] and the HSQC–TOCSY and HSQC–NOESY experiments, which are variants of the sequences of Crouch et al.[19] Chemical shifts are referenced to internal sodium trimethyl-silylpropionate-d_4 (TSP-d_4) at 0 ppm (^{1}H) or −1.8 ppm (^{13}C).[20] Phosphorus chemical shifts are referenced against external 85% phosphoric acid at 0 ppm.

3. Results and discussion

3.1. Characterisation of the de-*O*-acetylated CPS

A sample of the Pn15B CPS was de-*O*-acetylated under conditions which did not result in extensive loss of the phosphorylated substituent, and the NMR spectra assigned using a conventional group of experiments. The one-dimensional ^{1}H spectrum showed, inter alia, one *N*-acetyl resonance and five resonances in the anomeric region. One resonance at lowfield, 5.41 ppm, was assigned to an α-linked residue whilst the other anomeric resonances showed coupling constants consistent with being β-linked (Table 1). The multiplicity of the anomeric resonances showed no evidence of the presence of long-range ^{1}H–^{31}P coupling, which might occur if the residue were 2-phosphorylated.

Assignments for the spin systems were obtained using COSY, TOCSY and HSQC experiments (Fig. 1), with additional data obtained from HSQC–TOCSY, HSQC–NOESY (Fig. 2), long-range ^{1}H–^{13}C correlation experiments and ROESY experiments. These assignments are listed in Table 1. The substitution positions of the sugars were confirmed from the down-field shifts of the ^{13}C NMR signals for carbons at the glycosylated positions, compared to model systems. The low-field glycerol-2-phosphate H-2 was assigned using as starting points proton resonances observed in ^{1}H–^{31}P correlation experiments. The spectra were consistent with a major component (ca. 80%) carrying a glycerol-2-phosphate substituent and a minor component (ca. 20%) where the glycerol residue had been lost. Since the spectra also indicated the presence of free glycerol, the minor component almost certainly arose from cleavage of the phosphodiester linkage during de-*O*-acetylation. These data are not compatible with a structure in which there is a low degree of substitution by phosphocholine residues.

The 1D ^{31}P spectrum showed two resonances, a sharp high-field resonance at −0.4 ppm, assigned as the phosphodiester group of the glycerophosphate substituent, and a broader higher field resonance at 0.2 ppm, which probably arises from a phosphomonoester. The relative intensities of these resonances were approximately 3:1, and consistent with partial loss of glycerol occurring during the treatment, the observation of unlinked glycerol in the ^{1}H–^{13}C HSQC spectrum, and of minor heterogeneity in the spin system of the terminal α-Gal residue. The ^{1}H–^{31}P correlation spectrum contained one major cross-peak at $\delta_C = -0.4$ and $\delta_H = 4.26$ ppm consistent with the super-position of correlations between

Table 1. Proton and ^{13}C assignments for de-*O*-acetylated Pn15B, and selected coupling constants

Residue		H-1 C-1 $^3J_{H1,H2}$	H-2 C-2	H-3 C-3	H-4 C-4	H-5 C-5	H-6 C-6	H-6	NAc
→4,6GlcNAcβ1→	D	4.73 103.92 7.6 Hz	3.81 56.42	3.79 74.21	4.03 76.33	3.76 73.14	4.26 68.63	4.08	2.05
→3Galpβ1→	A	4.46 104.08 9.8 Hz	3.61 71.12	3.74 83.46	4.19 69.31	3.72 75.99	3.78 62.04	3.78	
→4Glcβ1→	E	4.58 103.54 7.3 Hz	3.39 73.76	3.70 75.45	3.66 79.88	3.63 75.84	4.00 61.32	3.82	
→2,3Galp(3-P)β1→	B	4.76 102.68 7.6 Hz	3.91 72.84	4.26 77.08	4.23 69.07	3.75 75.72	3.78 62.04	3.78	
Terminal Galpα1→	C	5.41 98.33 3.8 Hz	3.79 69.44	4.02 70.34	4.04 70.40	4.38 71.74	3.74 62.33	3.74	
					δ_P				
Gro-2-*P*	F	3.76/3.76 62.72	4.29 78.25	3.76/3.76 62.72	0.40				

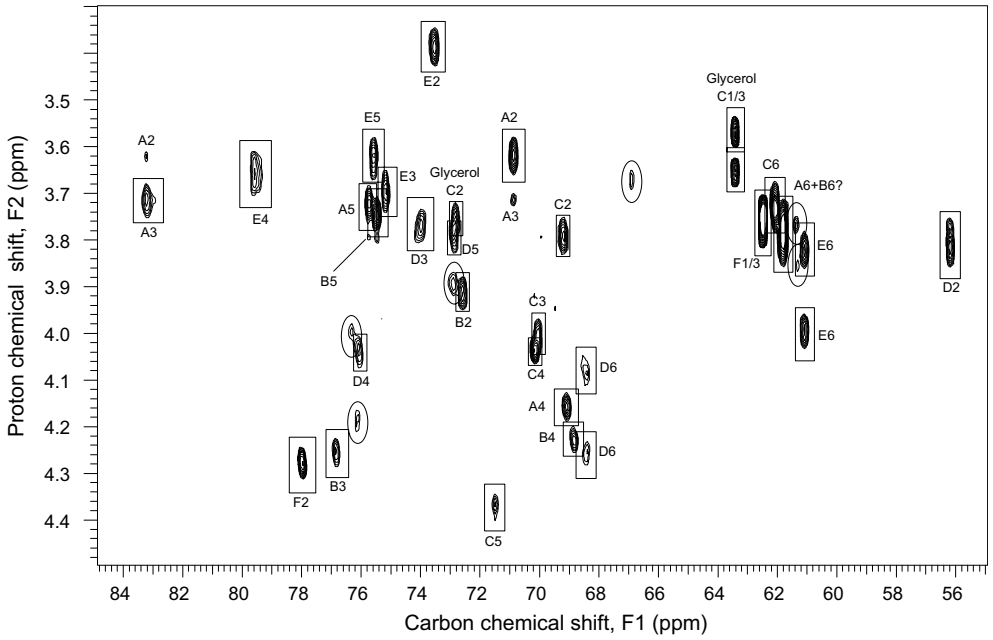


Figure 1. Partial 500 MHz HSQC spectrum of the ring carbon region of the spectrum of de-*O*-acetylated Pn15B, obtained at 50 °C. Individual cross-peaks are labelled. Cross-peaks enclosed in ovals are thought to arise from the side chain β-Gal residue in repeat units lacking the glycerol phosphate substituent. The residues are denominated as in Table 1.

phosphate and the glycerol H-2, the β-Gal H-3 and a long range correlation to the β-Gal H-4. Less intense long-range correlations to the β-Gal H-2 (3.912 ppm) and the glycerol H-1s and H-3s (3.77 ppm) were observed. A minor resonance from a cyclophosphate group at $\delta_P = 15$ ppm was again consistent with phosphodiester hydrolysis occurring during de-*O*-acetylation.

The sequence and linkage between the sugar residues was determined principally from an HSQC–NOESY experiment, obtained at 50 °C and with a 50 ms mixing time, combined with data from the comparison of the down-field shifts in the ^{13}C spectrum with model systems,[21] which indicated sites of glycosylation, and NOE data from a ROESY experiment. These data are

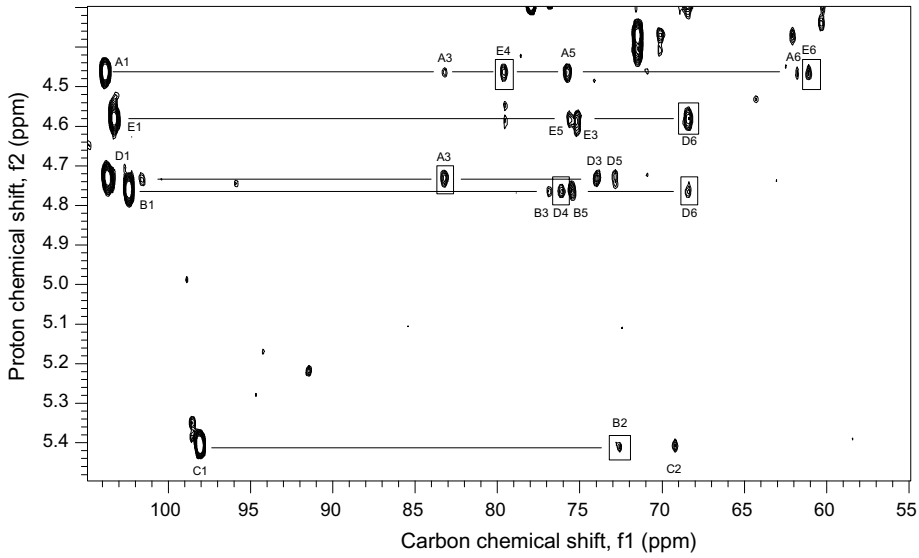


Figure 2. Partial 500 MHz HSQC–NOESY spectrum of the de-*O*-acetylated Pn15B CPS, showing correlations from the anomeric protons. The cross-peaks are labelled with residue and number, and inter-residue correlations which define linkage and sequence are highlighted by boxes. The residues are denominated as in Table 1.

summarised in Table 2. This HSQC–NOESY spectrum showed clear inter-residue correlations, which confirmed the saccharide structure **1**, with R = glycerol-2-phosphate, as reported by Abeygunawardana et al.[9] The glycerol-2-phosphate substituent appears to be present in essentially stoichiometric amounts.

Eight samples of Pn15B examined by NMR over a period of seven years contained a resonance at δ_H = 3.232 ppm, which can be assigned as arising from phosphocholine. Integration of this resonance, compared to the adjacent β-Glc H-2 resonance gave an approximate molar ratio of phosphocholine to the CPS repeat unit of 4.6% (range 2.8–11%). The NMR experiment was not optimised for quantitation: however, this low phosphocholine content is consistent with contamination of the CPS with C-polysaccharide, rather than involvement of the phosphocholine in the Pn15B repeat unit. Similarly, the 202 MHz ^{31}P spectrum of the native CPS allowed the degree of contamination of the CPS with pneumococcal C-polysaccharide to be estimated: integration of the resonances suggested the presence of ca. 2.5% mol/mol of the repeat unit, consistent with the estimate from the integration of the choline *N*-methyl resonance. Taken together, these data argue against the presence of phosphocholine substituents in the CPS.

3.2. Location of the *O*-acetyl group in Pn15B

Base catalysed de-*O*-acetylation of the sample and comparison of the integrals of appropriate resonances[8] indicated that approximately 85% of the repeat units were substituted. Comparing 12 samples of material intended for vaccine manufacture showed a range between 80%

Table 2. Inter-residues NMR correlations observed for the de-*O*-acetylated Pn15B CPS defining the sequence of and linkage between the residues

Residue	Residue	Significant downshift shifts of ^{13}C resonances[a]	NOE correlations in ROESY spectrum	^{1}H–^{13}C correlations in HSQC–NOESY spectrum
A	→3)-Gal*p*-β-(1→H-1	C-3: 9.68 ppm	**E** H-4 and **E** H-5	**E** C-4 and **E** C-6
A	→3)-Gal*p*-β-(1→C-1			**E** H-4
B	→2,3)-Gal*p*-β-(1→H-1	C-3: 3.30 ppm	**D** H-4	**D** C-4
B	→2,3)-Gal*p*-β-(1→C-1			**D** H-4
C	Galp-α-(1→H-1	None	**B** H-1, **B** H-2 and **B** H-3	**B** C-2
C	Galp-α-(1→C-1			**B** H-2
D	→4,6)-GlcNAc-β-(1→H-1	C-4: 5.27 ppm and C-6: 6.78 ppm	**A** H-2 and **A** H-3	**A** C-3
D	→4,6)-GlcNAc-β-(1→C-1			**A** H-3, **D** H-3 and **D** H-5
E	→4)-Glc-β-(1→H-1	C-4: 9.17 ppm	**D** H-6	**D** C-6 and **D** C-4
E	→4)-Glc-β-(1→C-1			**D** H-6 and **D** H-6′

[a] Compared to the chemical shift of the resonance in the monosaccharide, using values reported by Jansson et al.[21]

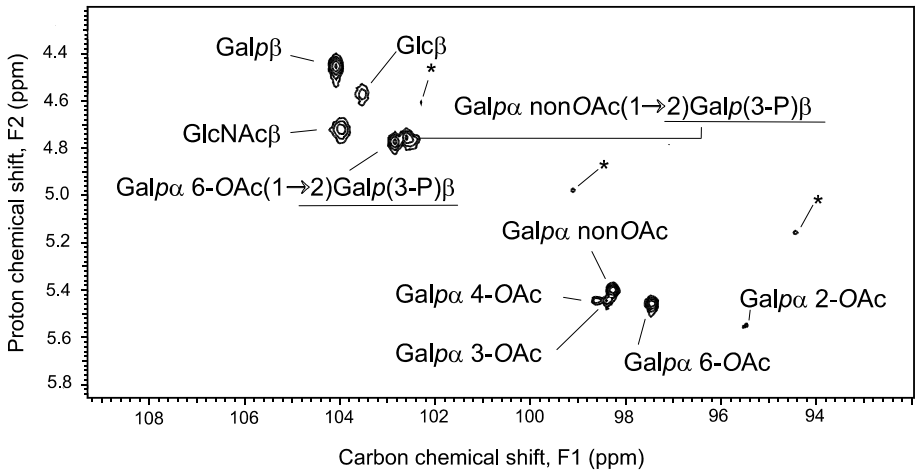


Figure 3. Partial 500 MHz HSQC spectrum of the native Pn15B showing the cross-peaks from the anomeric region of the spectrum. Key resonances are labelled, whilst cross-peaks highlighted with a star arise from pneumococcal C-polysaccharide.

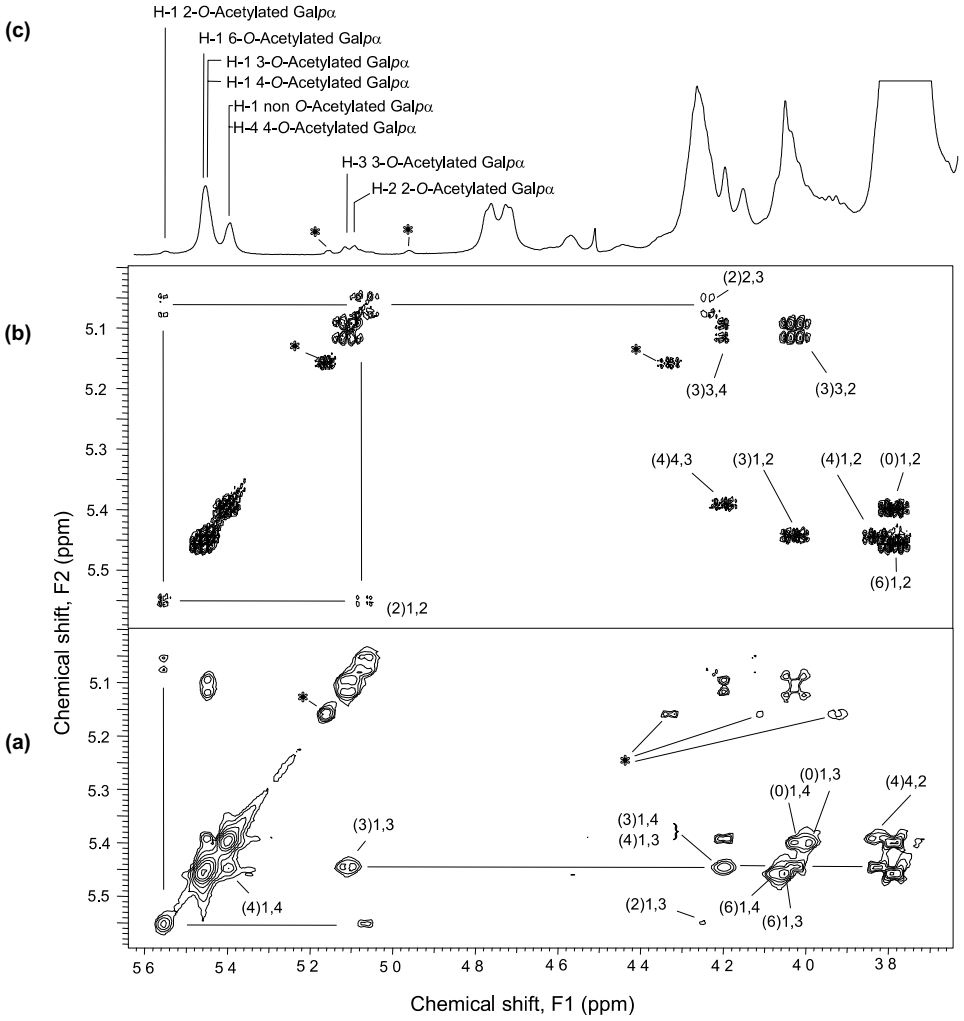


Figure 4. Partial 500 MHz of the native Pn15B CPS. (a) TOCSY spectrum with 80 ms mixing time and (b) COSY spectrum showing correlations involving the anomeric protons of the variously *O*-acetylated α-Gal*p* residue, labelled with the position of *O*-acetylation in parentheses and the assignments. (0) Indicates the non-*O*-acetylated residue. Cross-peaks are labelled with their assignments. Only cross-peaks not present in the COSY spectrum are labelled in the TOCSY spectrum. The one-dimensional proton spectrum is shown in (c). Peaks arising from pneumococcal C-polysaccharide are labelled with an asterisk.

Table 3. Partial NMR assignments for the terminal α-Gal*p* and subterminal β-Gal*p* residue with different *O*-acetylation patterns

Residue		H-1 C-1	H-2 C-2	H-3 C-3	H-4 C-4	H-5 C-5	H-6 C-6	H-6	OAc	^{31}P
Gal*p*(2OAc)α1→	6%	5.55 95.48	5.07	4.23	4.12				2.24 21.82	
Gal*p*(3OAc)α1→	12%	5.45 98.40	4.03	5.10	4.20	4.44	3.76	3.71	2.18 21.49	
Gal*p*(4OAc)α1→	12%	5.45 98.60	3.83	4.19	5.40 72.96				2.18 21.49	
Gal*p*(6OAc)α1→	55%	5.47 97.45	3.78	4.05	4.07	4.58 69.47	4.28 65.28	4.28	2.14 21.42	0.183
Gal*p*(3OAc)α1→2**Gal*p***β1→		4.76	3.93	4.30						
Gal*p*(6OAc)α1→2**Gal*p***β1→		4.78 102.83	3.95	4.25						

The last two rows of this table refer to the subterminal β-Gal*p* residue, in bold.

and 90%. The spectra of the native *O*-acetylated polysaccharide were compared with those of the de-*O*-acetylated sample to determine the site(s) of *O*-acetylation. These spectra are considerably more complex. The 202 MHz ^{31}P spectrum contained four resonances from the CPS (and three from pneumococcal C-polysaccharide[15]) at 0.183 ppm (54% relative abundance), 0.031 ppm (12%), −0.017 ppm (21%) and −0.091 ppm (13%). All these chemical shifts are appropriate for phosphodiester linkages, and suggest at least four different environments for the phosphate group. The HSQC spectrum contained four major and a minor cross-peak in the α-anomeric region (Fig. 3), three *O*-acetyl methyl resonances at 2.24, 2.19 and 2.15 ppm, and an additional C-6 methylene group. This suggests that the *O*-acetylation is on the terminal α-Gal residue. The spectrum of the native CPS was partially assigned using conventional homonuclear correlation techniques (Fig. 4), and the assignments relevant to the *O*-acetylated terminal α-Gal*p* residue are listed in Table 3. These indicate that repeat units exist in which of each the four hydroxyl groups is substituted. The amounts of each form were estimated by integration of cross-peaks in the TOCSY spectrum, which is reasonably accurate if the coupling constants and relaxation rates are the same in each species. Assignments for the various *O*-acetyl methyl resonances arise from NOEs observed in the NOESY spectrum.

Glycerol-2-phosphate substituents are known in other pneumococcal CPSs. Of the other three pneumococcal group 15 CPS, the Type 15C CPS differs from Pn15B in that it lacks the *O*-acetyl group,[11] and type variation can occur spontaneously.[12,22] The Types 15F and 15A CPS have linear pentasaccharide repeat units where the phosphorylated substituent on the Gal *O*-3 is reported to be either choline (Type 15F)[23] or glycerol-2-phosphate (Type 15A).[24] A glycerol-2-phosphate substituent is also present in the pneumococcal Type 23F CPS.[25]

→3)-α-D-Galp-(1→2)-D-β-Gal*p*(3-OPO$_3$-R)-(1→4)-β-D-GlcNAc-(1→3)-β-D-Gal*p*-(1→4)-β-D-Glcp-(1→

Pn15F	R = H or choline (at 20% substitution)
Pn15A	R = glycerol-2- (at 70% substitution): 2 *O*-acetyl groups per repeat unit.

Although the pneumococcal Types 27, 32F and 32A contain phosphocholine substituents,[26,27] the degree of substitution is high and, in our view, the structure of the pneumococcal Type 15F CPS should be re-investigated to confirm the identity of the phosphorylated substituent.

Acknowledgements

We thank Merck and Co., Inc. for the gift of the polysaccharides, and the MRC Biomedical NMR Centre, Mill Hill, for access to the 600 MHz spectrometer.

References

1. Robbins, J. B.; Austrian, R.; Lee, C.-J.; Rastogi, S. C.; Schiffman, G.; Henrichsen, J.; Makela, P. H.; Broome, C. V.; Facklam, R. R.; Tiesjema, R. H.; Parke, J. C. *J. Infect. Dis.* **1983**, *148*, 1136–1159.
2. Black, S.; Shinefield, H.; Fireman, B.; Lewis, E.; Ray, P.; Hansen, J. R.; Elvin, L.; Ensor, K. M.; Hackell, J.; Siber, G.; Malinoski, F.; Madore, D.; Chang, I. H.; Kohberger, R.; Watson, W.; Austrian, R.; Edwards, K. *Ped. Infect. Dis. J.* **2000**, *19*, 187–195.
3. Eskola, J.; Kilpi, T.; Palmu, A.; Jokinen, J.; Haapakoski, J.; Herva, E.; Takala, A.; Käyhty, H.; Karma, P.; Kohberger, R.; Siber, G.; Mäkelä, H. *New Engl. J. Med.* **2001**, *344*, 403–409.
4. Kilpi, T.; Ahman, H.; Jokinen, J.; Lankinen, K. S.; Palmu, A.; Savolainen, H.; Grönholm, M.; Leinonen, M.; Hovi, T.; Eskola, J.; Käyhty, H.; Bohidar, N.; Sadoff, J. C.; Mäkelä, P. H. *Clin. Infect. Dis.* **2003**, *37*, 1155–1164.
5. European Pharmacopoeia, Directorate for the Quality of Medicines of the Council of Europe (EDQM): Strasbourg 4th ed., 2001; pp 2205–2206.

6. Lemercinier, X.; Jones, C. *Biologicals* **2000**, *28*, 75–83.
7. Lemercinier, X.; Martinez-Cabrera, I.; Jones, C. *Biologicals* **2000**, *28*, 17–24.
8. Jones, C.; Lemercinier, X. *J. Pharm. Biomed. Anal.* **2002**, *30*, 1233–1247.
9. Abeygunawardana, C.; Williams, T. C.; Sumner, J. S.; Hennessey, J. P. *Anal. Biochem.* **2000**, *279*, 226–240.
10. Saha, S. K.; Rikitomi, N.; Biswas, D.; Watanabe, K.; Ruhulamin, M.; Ahmed, K.; Hanif, M.; Matsumoto, K.; Sack, R. B.; Nagatake, T. *J. Clin. Microbiol.* **1997**, *35*, 785–787.
11. Jansson, P.-E.; Lindberg, B.; Lindquist, U.; Ljungberg, J. *Carbohydr. Res.* **1987**, *162*, 111–116.
12. Lindberg, B.; Lönngren, J.; Powell, D. A. *Carbohydr. Res.* **1977**, *58*, 177–186.
13. Venkateswaram, P. S.; Stanton, N.; Austrian, R. *J. Infect. Dis.* **1983**, *147*, 1041–1054.
14. Talaga, P.; Bellamy, L.; Moreau, M. *Vaccine* **2001**, *19*, 2987–2994.
15. Karlsson, C.; Jansson, P. E.; Skov Sørensen, U. B. *Eur. J. Biochem.* **1999**, *265*, 1091–1097.
16. Szu, S. C.; Zon, G.; Schneerson, R.; Robbins, J. B. *Carbohydr. Res.* **1986**, *152*, 7–20.
17. Lemercinier, X.; Whitley, C.; Jones, C. *Carbohydr. Res.* **2000**, *325*, 192–201.
18. Wider, G.; Wüthrich, R. *J. Magn. Reson. Ser. B* **1993**, *102*, 239–241.
19. Crouch, R. C.; McFadyen, R. B.; Daluge, S. M.; Martin, G. E. *Magn. Reson. Chem.* **1990**, *28*, 792–796.
20. Wishart, D. S.; Bigam, C. G.; Yao, J.; Abildgaard, F.; Dyson, H. J.; Oldfield, E.; Markley, J. L.; Sykes, B. D. *J. Biomol. NMR* **1995**, *6*, 135–140.
21. Jansson, P.-E.; Kenne, L.; Widmalm, G. *Carbohydr. Res.* **1989**, *188*, 169–191.
22. van Selm, S.; van Cann, L. M.; Kolkman, M. A. B.; van der Zeijst, B. A. M.; van Putten, J. P. M. *Infect. Immun.* **2003**, *71*, 6192–6198.
23. Perry, M. B.; Bundle, D. R.; Daoust, V.; Carlo, D. J. *Mol. Immunol.* **1982**, *19*, 235–242.
24. Caroff, M.; Perry, M. B. *Can. J. Biochem. Cell Biol.* **1984**, *62*, 151–161.
25. Richards, J. C.; Perry, M. B. *Biochem. Cell Biol.* **1988**, *66*, 758–771.
26. Bennett, L. G.; Bishop, C. T. *Can. J. Chem.* **1977**, *55*, 8–16.
27. Karlsson, C.; Jansson, P. E.; Sørensen, U. B. *Eur. J. Biochem.* **1998**, *255*, 296–302.