Chemical Structure of the Capsular Polysaccharides (CPS) of *Streptococcus pneumoniae* Types 39, 47F and 34 by NMR Spectroscopy and their Relation to CPS10A

SUPPLEMENTAL TEXT and Figures

Details of NMR assignments of native CPS39 and de-O-acetylated CPS39. The detailed strategies of the assignment for the native and de-O-acetylated forms of the polysaccharide were quite similar, the only differences being in the chemical shifts for residue **C**. (Table 1) For the native CPS, the assignment for Residue **A** (β-Gal*f*) started from the ^{13}C and ^{1}H of **A**1 with **A**-H2 located by DQF-COSY and **A**-C2 by HSQC from the latter resonance. Assignment of **A**-H3 was complicated by strong coupling to **A**-H4 but the ^{13}C resonances were well resolved and could be assigned by a combination of HSQC-TOCSY and HMBC since the vicinal coupling in the HMBC of **A**-H1 gave a very strong cross peak to **A**-C4 with a smaller peak for **A**-C3. HSQC-TOCSY gave strong cross peaks from **A**-H1 to **A**-C2 and **A**-C3 but only a very weak correlation to **A**-C4. HSQC-TOCSY from **A**-C4 gave a cross peak at 3.98-3.96 ppm which could be either **A**-H5 or **A**-H6 or both. At this ^{1}H shift there was an HSQC cross peak with a methylene signal at 67.22 and 3.95 ppm that we identified as **A**6. That **A**5 occurred at 70.22 ppm in ^{13}C and 3.98 ppm ^{1}H with the ^{1}H signal overlapping with **A**-H6 could be shown by HSQC-TOCSY with no ^{13}C decoupling during acquisition. In this experiment, direct peaks are split by ^{13}C-^{1}H single-bond coupling (140 Hz) so that TOCSY relay peaks, which are not split, can be used to determine accurate ^{1}H chemical shifts without interference from the direct peak.

The assignment for residue **C** (β-Gal*f*) of the native CPS39 polysaccharide followed arguments similar to those for residue **A** with the difference that for this case all three signals for H2, H3 and H4 were strongly coupled. But the combination of HMBC and HSQC-TOCSY were used to identify **C**-C2, **C**-C3 and **C**-C4 and HSQC can serve to find accurate ^{1}H shifts. HSQC-TOCSY from the resonances of both **C**-C3 and from **C**-C4 showed cross peaks in the region of

4.23 ppm which correlated with methylene protons at 4.220 and 4.240 ppm in edited HSQC with the corresponding ^{13}C chemical shift of 66.79 ppm (**C**6). The TQF-COSY spectrum showed correlation of the methylene proton resonances at 4.22, 4.24 with 4.036 thus identifying **C**-H5 corresponding in HSQC to **C**-C5 at 69.01 ppm. Evidence for the assignment of the position of the O-acetyl group in native CPS39 came from the carbonyl selective HMBC spectrum which showed scalar coupling of the acetate methyl signal at 2.14 ppm with the carbonyl carbon resonance at 174.9 ppm. This peak in turn showed correlation with the proton signals at 4.24 and 4.22 ppm that were assigned as **C**-H6, 6' proving that residue **C** had a 6-O-acetyl substituent.

Residue **E (**α-Gal*p*) was assigned beginning from **E**1 at 4.962 ppm and 99.93 ppm with **E**-H2 assigned at 3.898 ppm by COSY from H1. **E**-C2 was identified by HSQC and by HSQC-TOCSY to be at 68.30 ppm and **E**-H3 had a strong cross peak at 4.005 ppm with **E**-H1 in TOCSY. **E**-C3 was identified at 80.38 ppm by HSQC and HSQC-TOCSY. **E**-H4 showed a weak and narrow cross peak with **E**-H1 at 4.263 ppm in TOCSY and **E**-C4 was located at 70.01ppm by HSQC. HMBC from the resonance of **E**-H1 showed a cross peak with **E**-C3, which was expected for α-Gal*p*, along with peaks at 69.42 ppm and 71.49 ppm. The latter peak, which was assigned as **E**-C5, showed HSQC to 3.997 (**E**-H5) and HSQC-TOCSY to 3.744 ppm which was assigned as the methylene **E**-H6 corresponding to **E**-C6 at 62.01 ppm. The 69.42 ppm peak in HMBC will be shown below to arise from linkage to **F**1.

The COSY cross peak from the **D**-H1 (β-GalNAc) resonance at 4.747 ppm was used to locate **D**-H2 at 4.162 ppm with the corresponding **D**-C2 at 52.97 ppm characteristic of the position of the acetamido substituent. **D**-H3 (3.896 ppm) and **D**-C3 (79.04 ppm) were located by TOCSY and HSQC- TOCSY from the **D**-H1 signal. TOCSY from **D**-H1 showed a narrow cross peak, in addition to those with **D**-H2 and **D**-H3, at 4.279 ppm which we assign as **D**-H4 with **D**-

C4 identified by HSQC at 75.78 ppm. HSQC-NOESY from **D**-H1 showed a cross peak, in addition to peaks assigned to **D**-C3, as expected for β-pyranosides, and to **E**-C3 due to glycosidic linkage. An additional peak in this spectrum at 74.37 ppm was assigned to **D**-C5 corresponding in HSQC to **D**-H5 at 3.955 ppm. HSQC-TOCSY correlated the **D**-C5 resonance with a methylene group at 3.992, 4.097 ppm and 70.98 ppm in the ^{13}C dimension that we assigned as **D**6 position. Carbonyl HMBC showed a cross peak between the acetyl methyl ^{1}H resonance at 2.048 ppm with the amide carbonyl ^{13}C (176.0 ppm) that in turn correlated with the resonance assigned to **D**-H2 confirming this as an acetamido sugar.

Beginning from the anomeric ^{1}H signal of **G**1 (β-Gal*p*) at 4.436 ppm, one could assign **G**-H2 (3.511 ppm) by COSY and this resonance was sufficiently well resolved from **G**-H3 and **G**-H4 that the latter resonances at 3.628 and 3.921 ppm can also be readily assigned by COSY. These assignments were confirmed by TOCSY and the corresponding ^{13}C signals clearly identified by HSQC and HSQC-TOCSY. Although the ^{1}H correlation could not be extended to H5 and H6, the **G**-C5 resonance was detected by a cross peak with **G**-H1 in HSQC-NOESY spectra at 75.92 ppm and **G**-H5, **G**-H6 and **G**-C6 were all assigned by HSQC and HSQC-TOCSY.

The chemical shift of **B**-H1 (β-Gal*p*) was close to that of **D**-H1 but at the probe temperature (20^{o}C) used in the experiments we report, the signals are resolved by 0.02 ppm, sufficient to distinguish cross peaks belonging to residues **B** and **D** (Fig. S1). COSY showed **B**-H2 at 3.68 ppm but TOCSY indicated this resonance to be strongly coupled to that of **B**-H3 complicating their assignment. The two ^{13}C signals at this ^{1}H chemical shift could be distinguished by HSQC-NOESY for which a cross peak between **B**-H1 and **B**-C3 (81.15 ppm) was observed as expected for a β-Gal*p* and none was observed for **B**-C2 at 71.28 ppm. TOCSY

from **B-**H1 showed, in addition to the cross peak for overlapping **B-**H2 and **B-**H3, a weaker and narrower cross peak for **B-**H4 at 4.082 ppm which was correlated with **B-**C4 by HSQC at 69.41 ppm. HSQC-NOESY from **B-**H1 identified **B-**C5 at 75.80 ppm that was correlated by HSQC to **B-**H5 at 3.669 ppm. HSQC-TOCSY from the **B-**C5 resonance, which was adequately resolved from that of **G-**C5, showed a cross peak to the frequency of the same methylene signal (3.77, 3.79 ppm) as the resonance assigned to **G**6. The conclusion that the signals of **G**6 and **B**6 coincide in both ^{1}H and ^{13}C was supported by the observation that the intensity of the HSQC signal was approximately twice as strong as that of the nearby cross peak for **E**6.

The ribitol residue, **F**, has two methylene groups and there were two methylene signals not assigned to the hexoses in the edited HSQC spectrum. One of them, at 4.080, 3.989 in ^{1}H and 67.39 in ^{13}C was assigned as **F**5 by correlation in the ^{31}P HSQC spectrum which also correlated ^{31}P with **A-**H6. This is the phosphodiester linkage in the polysaccharide backbone. The other methylene at 3.596, 3.965 ppm in ^{1}H and 69.42 ppm in ^{13}C was therefore assigned as **F**1. The TQF-COSY spectrum showed correlation between these methylene protons and one at 4.066 ppm which we assigned to **F-**H2. In addition, the TQF-COSY showed correlation between the **F-**H5,H5' resonance with 3.928 ppm which was therefore assigned as **F-**H4. The proton signals assigned to both **F-**H2 and H4 correlated in HSQC with the same ^{13}C chemical shift at 71.64 ppm and HSQC-TOCSY at this frequency showed a relay peak to 3.828 ppm implying that this remaining signal should be assigned to **F-**H3 which correlated with a ^{13}C chemical shift of 72.22 ppm. HSQC-TOCSY at this ^{13}C frequency shows that it is correlated with both **F-**H2 and **F-**H4 confirming that it is in the residue **F** spin system.

The above considerations completed the assignment of all the signals in the HSQC spectrum of Fig. 1A so the polysaccharide structure can be determined from long-range ^{1}H-^{13}C

coupling data. Our HMBC data showed correlation between **A**-H1 with **B**-C3 confirming the linkage of β-Gal*f* to position 3 of β-Gal*p*, residue **B**. Correlation of **B**-H1 with **D**-C4 as well as correlation of **B**-C1 with **D**-H4 confirmed the linkage of **B** with **D**. HMBC of **C**-H1 with **D**-C3 showed this unsubstituted residue to be a side chain. Similarly, the HMBC cross peak between **G**-H1 and **D**-C6 located the second side chain on the tri-substituted GalNAc residue, **D**. HMBC between **D**-H1 and **E**-C3 showed the linkage of the β-GalNAc to the 3-position of the α-Gal*p* residue. There was an HMBC cross peak between **E**-H1 and the ^{13}C frequency of 69.42 ppm but the resolution of the HMBC spectrum was not adequate to distinguish between **F**-C1 and **G**-C4. However, the linkage of **E**1 to **F**1 was proven by HMBC between **E**-C1 and **F**-H1, H1'. The assignment of the resonances of the de-O-acetylated CPS39 (Fig. 1B) closely paralleled the arguments given for the native form. The most notable differences arose from the change in the δ ^{1}H of the **C**6 methylene group for which the δ ^{1}H is 0.5 ppm upfield from the position of **C**-H6 in the O-acetylated native CPS (Table 1).

Details of NMR assignment of native CPS47F. The most unusual feature of the NMR spectrum of native CPS47F was the large chemical shift dispersion caused by the presence of three O-acetyl groups in each repeating unit, which caused the chemical shifts of several of the sugar ring protons to fall well within the downfield region usually occupied by anomeric proton signals (Fig. 3). The assignments for residue **E**, α-Gal, began from **E**1 at 5.247 ppm ^{1}H and 100.03 ppm ^{13}C with **E**-H2 identified by the COSY cross peak with H1 at 3.965 ppm. HSQC-TOCSY cross peaks of H1 appeared at 68.2, 69.97 and 77.92 ppm; the first of those was identified as **E**-C2 by its HSQC cross peak with **E**-H2. An HMBC cross peak between **E**-H1 and 77.92 ppm provided assignment of this peak to **E**-C3 leaving the HMBC cross peak at 72.02 ppm, which was assigned to **E**-C5 along with the peak at 80.7 ppm. This latter cross peak was

shown to represent the linkage of residue **E** to **F**-C2. The third HSQC-TOCSY cross peak at 69.97 ppm was then assigned to **E**-C4. An HSQC-TOCSY cross peak between **E**-C5 and methylene signals at 3.744 ppm ^{1}H and 61.80 ppm ^{13}C was assigned as **E**6.

The assignment for residue **A**, β-Gal*f*, began with **A**1 at 5.301 ppm ^{1}H and 110.04 ppm ^{13}C. A strong HSQC-TOCSY cross peak between **A**H1 and 79.61 ppm identified this as **A**-C2, which was correlated by HSQC to **A**-H2 at 4.359 ppm. That peak also showed a COSY cross peak with **A**-H1. The **A**-H2 resonance also showed a COSY cross peak at 4.880 ppm identified as **A**-H3, a position of O-acetylation. There was a strong HMBC cross peak between **A**-H1 and 81.15 ppm, which was assigned to **A**-C4 and a weaker cross peak at 79.75 ppm that was correlated in HSQC with **A**-H3. An HSQC cross peak between the **A**-C4 resonance and 4.498 ppm identified this as **A**-H4. An HSQC-TOCSY cross peak between **A**-C4 and 5.385 ppm identified this as the position of **A**-H5, a second position of O-acetylation of this residue. There was also a cross peak of the **A**-C4 resonance with a methylene proton resonance at 4.051 ppm which was identified as **A**-H6 correlated by HSQC with **A**-C6 at 64.40 ppm.

A minor difficulty in the assignments for residue **B (**β-Gal) arose from overlap of the anomeric proton signal at 4.506 with that of **A**H4. There were three COSY cross peaks at that chemical shift but since two of them were already assigned to **A**-H3 (4.880 ppm) and **A**-H5 (5.385 ppm), the remaining one at 3.666 ppm should be assigned as **B**-H2 which was correlated in HSQC to **B**-C2 at 70.86 ppm. There were HSQC-TOCSY cross peaks between the **B**-H1 resonance position at peaks assigned to **A**-C3 and to **A**-C5 in addition to one at 80.90 which we assigned to **B**-C3. HSQC-TOCSY connected this latter peak to the narrow resonance of **B**-H4. An HSQC-NOESY cross peak was observed between the **B**-H1 resonance and **B**-C5 at 76.08 ppm, an assignment confirmed by HSQC-NOESY between **B**-H4 and B-C5. An HSQC-TOCSY

cross peak was observed between this ^{13}C chemical shift and a methylene signal at 3.779 ppm. The **B**6 peak was slightly resolved from that of **E**6.

C1, whose anomeric signal is at 5.386 ppm ^{1}H and 107.89 ppm ^{13}C, has its ^{1}H signal overlapping with that of a ring proton, **A**-H5. HSQC-TOCSY cross peaks occur at 5.386 ppm with several ^{13}C resonances already assigned to the residue **A** spin system but the peak at 84.78 ppm was assigned to **C**-C2. HMBC cross peaks were observed between 5.386 ppm and 84.78 ppm but this was not **C**-C2, which was expected to give only a weak peak but it rather arose from **C**-C4 at the same chemical shift, which was expected to show strong three-bond correlation with **C**-H1. The HMBC cross peak at 76.56 ppm was identified as **C**-C3 in addition to a peak at 77.92 ppm assigned to the linkage to **E**-C3. **C**H5 was assigned at 4.049 ppm by a cross peak in HSQC-TOCSY with the **C**-C3 signal. **C**-C5 was located at 70.64 ppm by HSQC from **C**-H5. An HSQC-TOCSY cross peak was observed between **C**-C5 and methylene signals at 3.780 and 4.069 ppm assigned to **C**-H6 that were correlated with **C**-C6 at 71.78 ppm. The downfield chemical shift of **C**-C6 resulted from linkage at this position from residue **B**.

Given the NMR assignment of the four sugar residues of native CPS47F, there remained the assignment of the ribitol (resiude **F)** which began with **F**-H5 assigned by ^{31}P HSQC to the methylene group with ^{1}H signals at 4.057 and 3.943 ppm along with the corresponding ^{13}C at 67.83 ppm. There is an HSQC-TOCSY cross peak between this latter frequency and 3.845 ppm along with an indication of a second cross peak at 4.02 ppm which overlaps with the auto peak of **F**H5' at 4.057 ppm. In an HSQC-TOCSY spectrum without ^{13}C decoupling during acquisition, the direct peak of **F**H5' is split by 1-bond coupling revealing a strong cross peak at 4.023 ppm. Therefore, we assign this cross peak and its HSQC partner at 72.39 ppm as **F**4. The weaker cross peak at 3.845 ppm is due to **F**H3, which correlates in HSQC with **F**C3 at 71.21 ppm. The

single remaining unassigned methylene resonance at 3.921 and 3.845 ppm in ^{1}H and 60.83 ppm in ^{13}C was assigned to **F**1 and an HSQC-TOCSY cross peak identified **F**-H2 at 4.052 ppm with **F**-C2 at 80.62 ppm.

Although the positions of the three O-acetyl groups of CPS47F were suggested by the large downfield shifts of the ^{1}H resonances at the substitution positions, information that was more reliable was provided by scalar coupling correlation in carbonyl selective HMBC data. Importantly, this showed cross peaks between 177.11 ppm in ^{13}C with 5.061 ppm (**C**-H2) and with the acetyl methyl protons at 2.136 ppm. Likewise, we observed cross peaks at 177.24 ppm with **A**-H5 at 5.385 ppm and with the acetyl methyl methyl protons at 2.173 ppm as well as at 177.40 ppm with both **A**-H3 (4.880 ppm) and the acetyl methyl protons at 2.143 ppm.

De-O-acetylated CPS47F. The similarity of the anomeric chemical shifts of α-Gal*p* (residue **E)** and the two Gal*f* residues (**A** and **C)** required careful analysis of HSQC and HSQC-TOCSY spectra recorded at resolution of ~3Hz in the ^{13}C dimension in order to obtain reliable resonance assignments. We began with the assignment of residue **E** (α-Gal*p*), for which **E**1 is at 5.250 ppm ^{1}H and 99.95 ppm in ^{13}C to which the anomeric ^{1}H resonances of **A** and **C** were very close but distinguishable (Table 1 and Fig. 3B) HSQC-TOCSY cross peaks of **E**-H1 were observed at 68.50, 78.17 and at 68.98 ppm. The first cross peak was assigned as **E**-C2 by HSQC with 3.962 ppm which showed a COSY cross peak with **E**-H1. The second peak was identified as **E**-C3 by an HMBC cross peak with **E**-H1. The remaining relatively weak peak has a narrow ^{1}H line width, which identified it as **E**-H4. The HMBC spectrum from **E**-H1 had, in addition to the resonance of **E**-C3, a peak at 72.05 ppm identified as **E**-C5 which was correlated by HSQC-TOCSY with a methylene resonance assigned as **E**6 at 3.733, 3.754 ppm in ^{1}H and 61.84 in ^{13}C.

The assignment of the resonances of residue **B** (β-Gal*p*) began with **B**1 at 4.504 ppm ^{1}H

and 103.87 ppm in ^{13}C. A COSY cross peak with **B-**H1 identified **B-**H2 at 3.668 ppm. HSQC-TOCSY peaks were observed between **B-**H1 and 70.86 ppm, 81.10 ppm and 69.39 ppm. That the first peak was **B-**C2 was shown by an HSQC cross peak to **B-**H2. The peak at 81.10 ppm was identified as **B-**C3 by an HSQC-NOESY cross peak with **B-**H1. The weaker HSQC-TOCSY cross peak at 69.37 ppm was assigned as **B-**C4 by HSQC to a narrow ^{1}H resonance at 4.101 ppm that was expected as **B-**H4 for β-Gal. Although no HSQC-NOESY peak was observed between **B-**H1 and **B-**C5, **B**H5 was assigned at 3.724 ppm by an HSQC-NOESY cross peak with **B-**H4. After assignment of **B-**C5 by HSQC, an HSQC-TOCSY cross peak between the latter and the methylene group at 3.769 and 3.788 ppm in ^{1}H and 61.85 ppm in ^{13}C identified **B**6, which overlapped with **E**6.

Knowledge of the assignment of residues **B** and **E** facilitated assignment of residues **A** and **C** since their linkages were previously determined in the native CPS47F. In the COSY spectrum, both **A-**H1 and **C-**H1 showed cross peaks to the 4.21 ppm so assignment of **A**2 and **C**2 cannot depend on this homonuclear spectrum. But HSQC-TOCSY cross peaks of **A-**H1 were distinct from those of **C-**H1. **A-**H1 had cross peaks at 82.16 ppm (strong), 77.47 ppm and 83.40 ppm; the assignments were deduced from cross peaks in the HMBC spectra that showed a strong cross peak to 83.40 ppm that was identified as **A-**C4. A weaker HMBC cross peak of **A-**H1 at 77.47 identified this as **A-**C3 following a coupling pattern typical of β-galactofuranosides. The peak at 82.26 ppm assigned as **A-**C2 was correlated by HSQC to a ^{1}H resonance 4.210 ppm.

The **C-**H1 resonance showed HSQC-TOCSY cross peaks at 82.24 ppm (strong), 77.65 ppm and 83.90 ppm. The HMBC spectrum showed the strongest cross peak between **C-**H1 and 83.90 ppm indicating that this was **C-**C4. A weaker HMBC cross peak at 77.65 ppm identified this as **C-**C3. Continuing with the assignment of residue **C**, the HSQC-TOCSY spectrum

showed cross peaks between the resonances of **C**-C3 and **C**-C4 with that of **C**-H2 and also 4.02 ppm as well as to methylene protons at 3.771 and 4.074 ppm. The latter pair were assigned as **C**-H6 and 6' and the former to **C**-H5. The observation of TQF-COSY cross peaks among these three confirmed the assignment and HSQC provided the assignments of **C**-C5 and **C**-C6. The remainder of the assignment of residue **A** followed the same coupling pattern as the case of residue **C**. HSQC-TOCSY cross peaks were observed between the resonances of **A**-C3 and **A**-C4 with 4.20 ppm (**A**H2) and 3.98 ppm. This latter cross peak resulted from overlapping signals of **A**-H5 and **A**-H6 as shown in the HSQC spectra by two unassigned signals in the ^{13}C dimension at this ^{1}H chemical shift, 70.16 ppm and 67.19 ppm, the latter a methylene group. The assignment of these two resonances respectively to **A**-C5 and **A**-C6 was confirmed by these two rows in HSQC-TOCSY without ^{13}C decoupling during acquisition. In each case, the direct peaks were split by one-bond coupling revealing the correct chemical shift of the relayed peak.

The remaining resonances should be assigned to the ribitol, residue **F**. Since we knew the linkage position of the α-Gal (**E**) to the ribitol from the structure of the native CPS47F, we could begin the assignment with **F**-C2 which was identified by an HMBC cross peak with **E**-H1 observed at 80.57 ppm. An HSQC-TOCSY cross peak between this resonance and the methylene proton signals at 3.840 and 3.928 ppm identified **F**1, an assignment that was confirmed by TQF-COSY between these two and **F**-H2 at 4.052 ppm. Although we expected to see an HSQC-TOCSY cross peak between **F**-C2 and **F**-H3, this latter resonance overlapped the resonance of **F**-H1. But we completed the assignment with correlations from the resonance of **F**5, the single remaining unassigned methylene resonance that was located at 3.977, 4.084 ppm in ^{1}H and 67.75 ppm in ^{13}C. An HSQC-TOCSY cross peak was observed between this latter ^{13}C resonance and that of **F**-H4 at 4.031 ppm accompanied by a weaker cross peak at 3.859 ppm.

HSQC cross peaks correlated **F**-H4 with **F**C4 at 72.38 ppm and an HSQC-TOCSY cross peak with this latter resonance located **F**-H3 at 3.859 ppm.

The linkages and the structure of CPS47F. The linkage between residues **A** and **B** was indicated by an HMBC cross peak between **A**-H1 and **B**-C3 in both the native and the de-O-acetylated CPS47F. The linkage between residues **B** and **C** was indicated by an HMBC cross peak between **B**-H1 and **C**-C6 in the de-O-acetylated CPS. An HMBC cross peak was observed between **C**-H1 and **E**-C3 in both the native and de-O-acetylated CPS47F. There was an HMBC cross peak between the resonance of **E**-H1 and **F**-C2 in both the native and de-O-acetylated CPS. The phosphodiester linkage between **F**5 and **A**6 was evident from ^{31}P HSQC spectra of CPS47F which showed a strong cross peak to **F**-H5 and a lesser cross peak with **F**-H4. The ^{31}P HSQC with **A**-H6 overlaps that with **F**-H5' but there was a weaker cross peak with the well resolved **A**-H5.

Details of the NMR spectroscopy and structure of CPS34. The ^{1}H-^{13}C HSQC spectrum of CPS34 showed 4 resonances that were assigned residue letters **A, B, C** and **E** (Fig. S2) in accordance with those of other similar polysaccharides discussed above. There was also one peak characteristic of an O-acetyl methyl group at 2.145 ppm (^{1}H) and 21.16 ppm (^{13}C) whose intensity was comparable to three times that of the anomeric peaks indicating that the acetylation is very nearly stoichiometric. For residue **E,** identified as α-Gal, **E**1 to **E**4 were readily assigned by a combination of COSY, TOCSY, HSQC and HSQC-TOCSY. The HMBC cross peak expected for α-Gal was seen between **E**-H1 and **E**-C3 along with a peak assigned as **E**-C5 which in turn showed an HSQC-TOCSY cross peak to the methylene protons of **E**-H6. The resonances of residue **B**, identified as α-Glc, were assigned by methods similar to those for residue **E**.

257 Residues **A** and **C** were both identified as β-Gal*f* by characteristic chemical shifts and
258 coupling constants of atoms 1 to 4. **A**-H2 was identified by a COSY cross peak with **A**-H1
259 which showed an HSQC cross peak with **A**-C2. HMBC showed a characteristically strong cross
260 peak of **A**-H1 with **A**-C4 and a less intense cross peak with **A**-C3. Given HSQC-TOCSY data
261 confirming these assignments, we used HSQC-TOCSY from **A**-C3 and **A**-C4 to identify **A**-H5
262 with **A**-C5 identified by HSQC. An HSQC-TOCSY cross peak between the **A**-C5 resonance and
263 methylene protons of **A**-H6 completed the assignment of this residue.
264 The resonances of residue **C** were assigned by similar methods but with the complication of
265 exact overlap of the resonances **C**-H2 and **C**-H3. But the resonances of **C**-C2 and **C**-C3 were
266 well resolved and the accurate proton chemical shifts were evident in HSQC-TOCSY without
267 decoupling during acquisition. HSQC-TOCSY cross peaks from both **C**-C3 and **C**-C4 to 4.08
268 ppm identified **C**-H5 at that frequency and HSQC-TOCSY from **C**-C5 identified methylene
269 protons at 4.25 ppm as **C**H6. The downfield location of **C**-H6 results from O-acetylation at that
270 position as shown by HMBC cross peaks of the carbonyl carbon of the acetyl group with both its
271 methyl group and **C**-H6.
272 Since we anticipated finding a phosphodiester bond in the backbone of this polysaccharide,
273 ^{1}H-^{31}P HSQC spectra were recorded to determine its location. The spectra showed cross peaks at
274 two different ^{31}P chemical shifts. The more intense peaks were assigned to the type-specific
275 capsular polysaccharide, the main constituent of these polysaccharide preparations, along with
276 peaks of lesser intensity assignable to low levels of contaminating polysaccharide. Our ^{1}H-^{31}P
277 HSQC spectra showed a ^{1}H resonance at 4.47 ppm that we identified with **A**-H3, the linkage
278 position of the phosphodiester along with resonances of a methylene group at 4.096 and 4.008
279 ppm which we identified as H5 of the ribitol residue. The methylene protons identified as **F**-

H5', and **F**-H5 correlate with the ^{13}C signal at 67.79 which showed an HSQC-TOCSY cross peak with 4.04 ppm. There were two unassigned ^{13}C resonances at this ^{1}H chemical shift, one at 80.80 ppm and the other at 72.35 ppm, either of which could be **F**-C4. HSQC-TOCSY from the former showed cross peaks at 3.84 and 3.92 ppm which did not match **F**-H5,5' but cross peaks from the latter appeared at 4.09, 4.01 and at 3.869 ppm. Therefore, if we assign the first two as the **F**5 protons, then the third could be assigned as **F**-H3 with corresponding ^{13}C signal at 71.24 ppm. The HSQC-TOCSY cross peaks from the 80.80 ppm ^{13}C resonance matched well with the proton resonances of the last remaining methylene group which were therefore assigned as **F**1 and implying that the C-H pair at 4.047 and 80.80 ppm was **F**2 completing the assignment of the ^{1}H and ^{13}C resonances of residue **F**. Given the complete assignment of the ^{13}C and ^{1}H resonances of CPS34 (See Table 1 and Fig. S2) the structure of the CPS could be deduced from the residue linkages shown by NMR.

293

SUPPLEMENTAL FIGURES

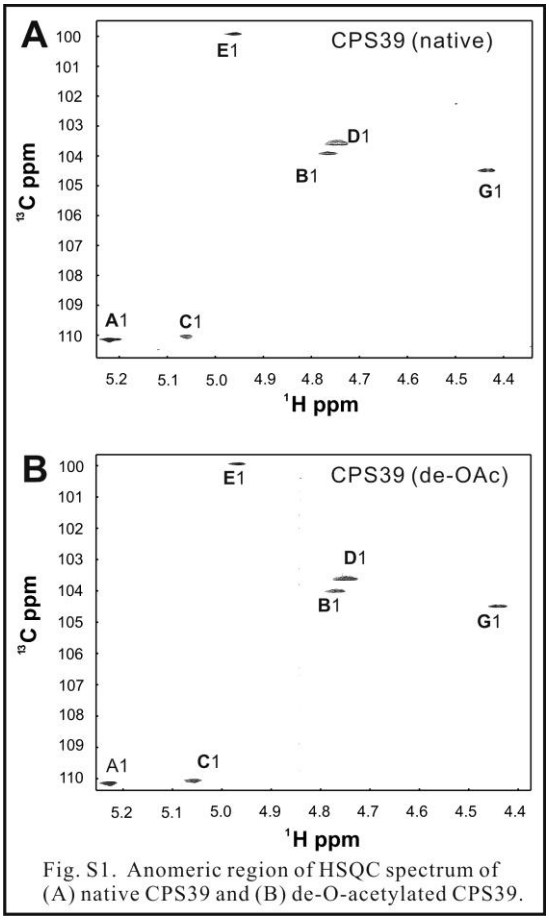

Fig. S1. Anomeric region of HSQC spectrum of (A) native CPS39 and (B) de-O-acetylated CPS39.

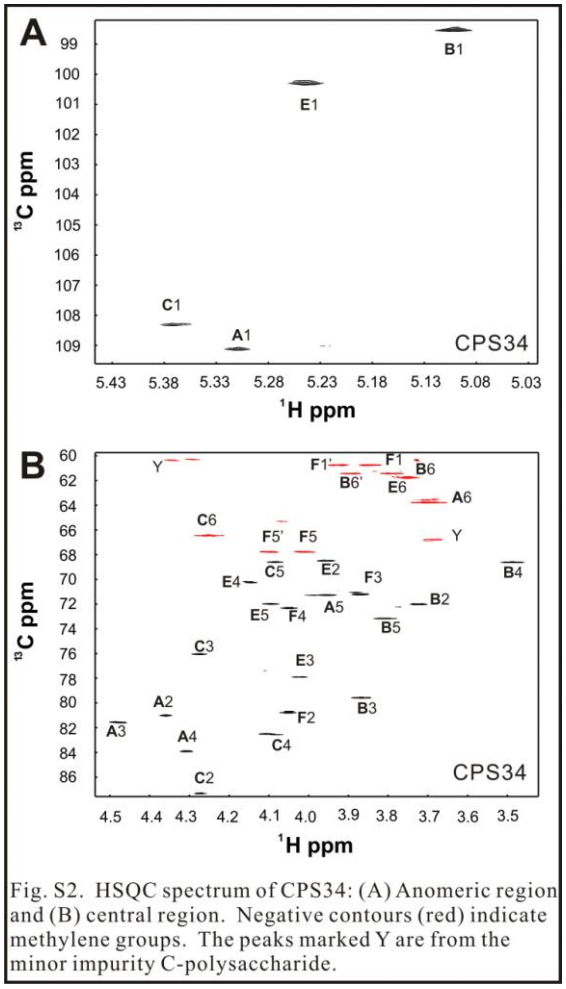

Fig. S2. HSQC spectrum of CPS34: (A) Anomeric region and (B) central region. Negative contours (red) indicate methylene groups. The peaks marked Y are from the minor impurity C-polysaccharide.

294