Discovery and description of a new serogroup 7 Streptococcus pneumoniae serotype, 7D, and structural analysis of 7C and 7D

Kjeldsen, Christian; Slott, Sofie; Elverdal, Pernille L.; Sheppard, Carmen L.; Kapatai, Georgia; Fry, Norman K.; Skovsted, Ian C.; Duus, Jens Ø.

Published in:
Carbohydrate Research

Link to article, DOI:
10.1016/j.carres.2018.04.011

Publication date:
2018

Document Version
Peer reviewed version

Accepted Manuscript

Discovery and description of a new serogroup 7 *Streptococcus pneumoniae* serotype, 7D, and structural analysis of 7C and 7D

Christian Kjeldsen, Sofie Slott, Pernille L. Elverdal, Carmen L. Sheppard, Georgia Kapatai, Norman K. Fry, Ian C. Skovsted, Jens Ø. Duus

PII: S0008-6215(18)30160-5

DOI: 10.1016/j.carres.2018.04.011

Reference: CAR 7556

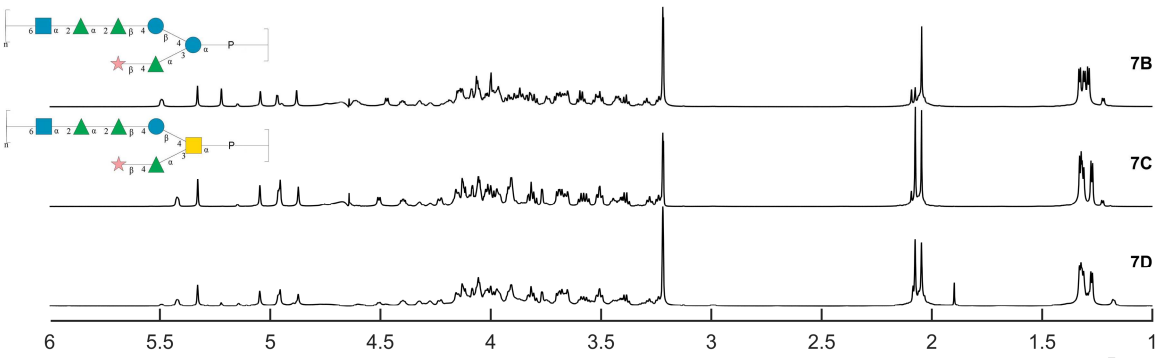
To appear in: *Carbohydrate Research*

Received Date: 12 March 2018

Revised Date: 18 April 2018

Accepted Date: 19 April 2018

Please cite this article as: C. Kjeldsen, S. Slott, P.L. Elverdal, C.L. Sheppard, G. Kapatai, N.K. Fry, I.C. Skovsted, Jens.Ø. Duus, Discovery and description of a new serogroup 7 *Streptococcus pneumoniae* serotype, 7D, and structural analysis of 7C and 7D, *Carbohydrate Research* (2018), doi: 10.1016/j.carres.2018.04.011.



Discovery and description of a new serogroup 7 *Streptococcus pneumoniae* serotype, 7D, and structural analysis of 7C and 7D

Christian Kjeldsen[a], Sofie Slott[a], Pernille L. Elverdal[b], Carmen L. Sheppard[c], Georgia Kapatai[c], Norman K. Fry[c], Ian C. Skovsted[b], Jens Ø. Duus[a,*]

[a]Department of Chemistry, Technical University of Denmark, Kgs. Lyngby, Denmark

[b]SSI Diagnostica A/S, Hilleroed, Denmark

[c]Respiratory and Vaccine Preventable Bacterial Reference Unit, Public Health England – National Infection Service, Colindale, London, UK

*Corresponding author. jduus@kemi.dtu.dk, +45 45252451

Declarations of interest: None.

Abstract

Streptococcus pneumoniae is characterised into 92 serotypes based on antigenic reactions of commercial rabbit sera to the capsular polysaccharides. During development of a bioinformatic serotyping tool (PneumoCaT), an isolate exhibited a novel codon at residue 385 of the glycosyltransferase gene *wcwK* encoding a distinct amino acid, which differentiates genogroup 7. Investigation by repeat serotyping and Quellung reaction revealed a novel pattern of factor sera with the isolate reacting very strongly with 7f, but also with 7e factor sera. The structure of the capsular polysaccharide was determined by NMR spectroscopy to be an approximately 5:1 combination of the structures of 7C and 7B, respectively, and the structure of 7C was also elucidated. All data from whole genome sequencing, NMR spectroscopy, production of antisera and serotyping of the novel 7 strain shows that it is a new serotype, which will be named in the Danish nomenclature as 7D.

Keywords: NMR spectroscopy, structure elucidation, *Streptococcus pneumoniae*, whole genome sequence, hybrid serotype

Author e-mail adresses: Christian Kjeldsen, chkje@kemi.dtu.dk; Sofie Slott, sofie_slott@hotmail.com; Pernille Landsbo Elverdal, PEL@ssidiagnostica.com; Carmen Sheppard, Carmen.Sheppard@phe.gov.uk; Georgia Kapatai, gkapatai@gmail.com; Norman Fry, Norman.Fry@phe.gov.uk; Ian Christian Skovsted, ICS@ssidiagnostica.com; Jens Øllgaard Duus, jduus@kemi.dtu.dk.

1 Introduction

Streptococcus pneumoniae (pneumococcus) is a major cause of morbidity and mortality worldwide. The pneumococcal capsular polysaccharide (CPS) which surrounds the organism is a major virulence factor, enabling the cell to evade phagocytosis [1]. The currently licensed 23–serotype valent plain polysaccharide (PPV23) and 7, 10 and 13-valent protein-conjugate vaccines (PCV7, PCV10 and PCV13) cover the most important circulating serotypes.

To date, 92 serotypes of pneumococcus have been described based on reaction of their capsule with the commercial rabbit antisera [2,3] available from SSI Diagnostica A/S; in this scheme immunogenically cross-reactive serotypes are assigned to serogroups in the Danish serotype classification system. As the capsule is the target of all currently licensed vaccines, accurate surveillance of pneumococcal serotypes is essential for the evaluation of the efficacy of the vaccines and to inform national vaccine policies. Vaccine pressure has resulted in an increase in the prevalence of non-vaccine serotypes and could also result in emergence of novel serotypes as the organism evolves to escape the vaccine.

However, many additional novel serotypes have been proposed based on genotypic analysis [4,5] or reaction with monoclonal antibodies [6]. Some of the described novel variants based on genotypic differences actually produce no difference in the structure of the polysaccharide, for example the putative serotype 6E which in fact produces a polysaccharide identical to serotype 6B [7], and the genetic variant of 23B which, although it has low sequence homology in the capsular operon and is from an entirely different genetic lineage to the originally described 23B, produces an identical polysaccharide and is therefore the same phenotypic serotype [8]. In other cases, the capsular polysaccharide produced appears to be either a hybrid or variable expression of distinct polysaccharide structures. This has been seen in the case of rare serogroup 6 variant isolates [9,10] and 11A variants [11], though in neither of these cases it is possible to determine a difference to standard 6A, 6B or 11A serotypes, using the widely available commercial typing sera.

During development and evaluation of a new bioinformatic serotyping tool (PneumoCaT, Pneumococcal Capsule Typing) [12] by Public Health England (PHE) ,to aid in surveillance of pneumococcal serotypes using whole genome sequence (WGS) data, a novel variant within serogroup 7 was discovered. The source of the pneumococcal isolate demonstrating a novel sequence was from an ear swab from an outpatient attending a London hospital in 2012, submitted for antibiotic susceptibility testing. The WGS data from this strain have been submitted to the European National Archive (designated PHESPD0846 - https://www.ebi.ac.uk/ena/data/view/ERS1194144).

Serogroup 7 is an important serogroup as it contains the vaccine type 7F which is included in the currently licensed PPV23, PCV10 and PCV13 vaccines. Serogroup 7 contains four recognised serotypes, 7F, 7A, 7B and 7C, of which 7F is by far the most common worldwide, representing 98.8% of serogroup 7 isolates in UK serotype surveillance for epidemiological years 2006-2014 with serotypes 7A, 7B, and 7C, representing 0.2%, 0.2%, and 0.8% respectively (PHE surveillance data).

The polysaccharide repeat unit structures of serotypes 7F [13], 7A [14] and 7B [15] are known, shown in figure 1. All of the serotypes contain cell wall polysaccharide (CWPS), also known as teichoic acid, as well as their respective CPS, and the structure of the CWPS has been elucidated [16]. The antigenic formula of serotype 7F and 7A is similar and the only structural difference between them is the presence of a terminal side-chain β-D-galactose residue in 7F. Serotype 7B only has one antigenic component in common with serotype 7F and 7A and does not have *O*-acetyl groups [17]. These assignments all relied primarily on NMR spectroscopy, and to our knowledge there was no published structure or structural data for serotype 7C. In the present study, it was realised that in order to describe structure of the newly isolated serotype, it was necessary to first make a full structural identification of 7C.

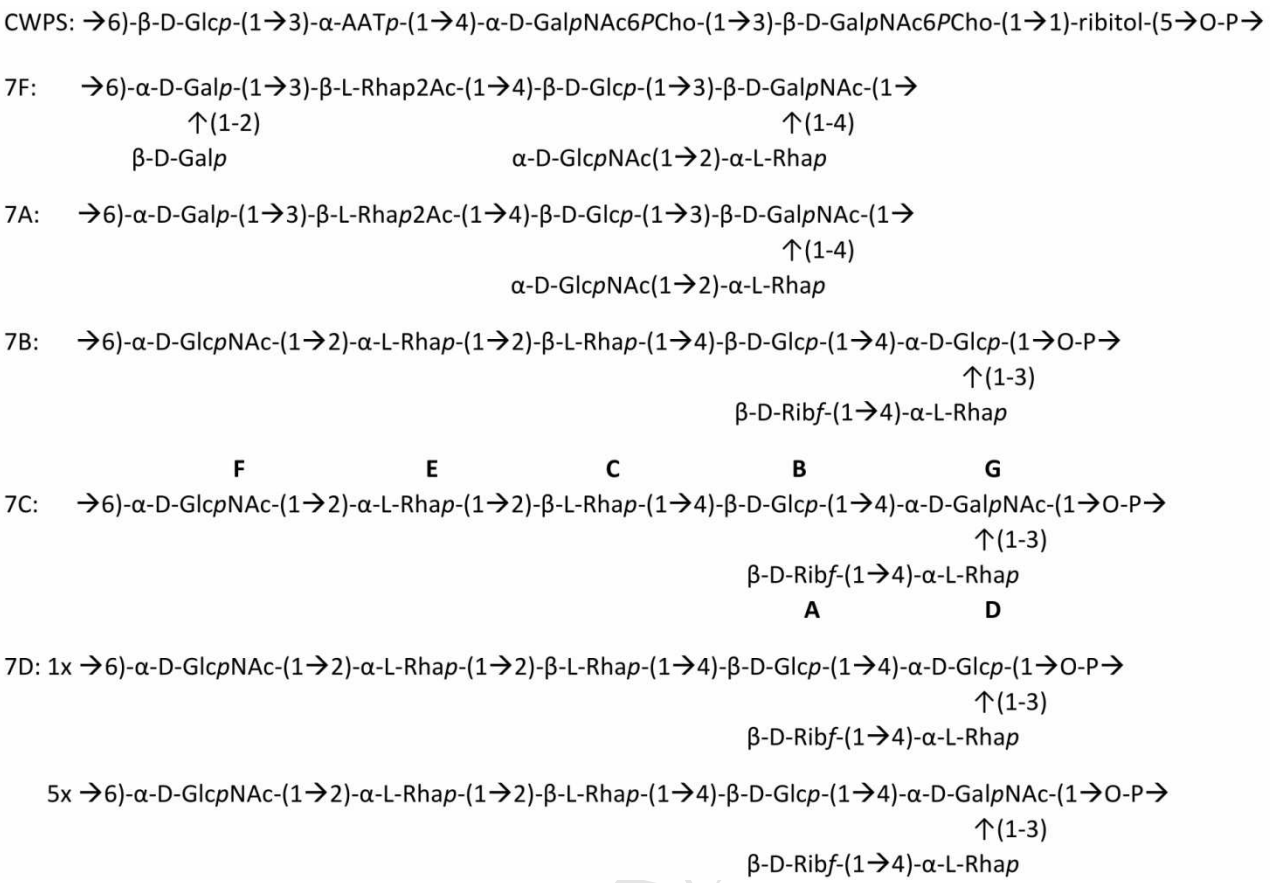


Figure 1: Structures of the serogroup 7 capsular polysaccharides as well as CWPS. The **A**-**G** labels for 7C have been highlighted, and are used for reference in the structure elucidation.

The genetic arrangement of the *cps* operon in serogroup 7 was described by Mavroidi *et al.* [18] in 2007 and the *cps* loci forms two syntenic pairs, 7F/7A and 7B/7C and the two pairs cluster apart from each other on amino acid sequence comparisons [18]. The lack of the side chain in 7A compared to 7F can be explained by a frameshift mutation in *wcwD* which codes for a glycosyltransferase [19].

During development of PneumoCaT, nine single-nucleotide polymorphism (SNP) positions in *wcwK*, a glycosyltransferase, which can be used to differentiate serotype 7B from serotype 7C and also serotype 40 which is closely genetically related to 7B and 7C [12], were described. Nucleotide differences at residues 46 and 385 of *wcwK* are considered as more significant, as both result in distinct amino acids in the three serotypes differentiated in this way.

When the PneumoCaT was evaluated on panels of previously serotyped isolates, one 7C isolate (PHESPD0846) displayed an atypical SNP pattern which did not fit the expected patterns in the Capsular Type Variant (CTV) database used by the tool to define serotypes [12].

This paper describes the serological, genetic and structural properties of this novel variant, which is to be named in the Danish nomenclature as serotype 7D. We also describe the structure of the capsular polysaccharide of serotype 7C which has not been published previously. Furthermore, as the NMR assignment of the different capsular polysaccharides have been obtained using different conditions, and in some cases have incomplete assignments, the full assignment of 7F, 7A (supporting information) and 7B (table 2) was also performed.

2 Methods

2.1 DNA extraction and sequencing

The isolate was grown overnight on horse blood agar (PHE Media Services) with 5% CO_2. DNA was extracted from an entire agar plate of growth using the QIAsymphony SP automated instrument (Qiagen, Hilden, Germany) and QIAsymphony DSP DNA Mini Kit, using the manufacturer's recommended tissue extraction protocol for Gram negative bacteria (including a 1 hour pre-incubation with proteinase K in ATL buffer and RNAse A treatment). DNA concentrations were measured using the Quant-iT dsDNA Broad-Range Assay Kit (Life Technologies, Paisley, UK) and GloMax® 96 Microplate Luminometer (Promega, Southampton, UK). DNA was sent for whole genome sequencing by Illumina sequencing using the PHE Genomic Services and Development Unit (Colindale, UK) 29. Illumina Nextera DNA libraries were constructed and sequenced using the Illumina HiSeq 2500.

2.2 Bioinformatic methods

Raw Illumina reads were pre-processed as described by Kapatai *et al*. [12]. KmerID (https://github.com/phe-bioinformatics/kmerid) was used to confirm species identification as part of the PHE WGS workflow. Multi-locus Sequence Typing on short-read data was performed using Metric Oriented Sequence Typing [20].

PneumoCaT (https://github.com/phe-bioinformatics/PneumoCaT) was used to predict capsular type from WGS data as described by Kapatai *et al.* [12]. In brief, processed reads are initially mapped to capsular locus sequences for all 94 serotypes (92 serotypes plus 2 molecular subtypes) which predicts serotype or genogroup based on mapping coverage (> 90%). If genogroup is predicted the analysis uses the capsular type variant database and the reads are mapped to genogroup relevant genes allowing for variant analysis. Serotype can be predicted if 100% match with an available variant profile is achieved.

2.3 Assembly-based sequence analysis

Genomic reads were assembled using SPAdes (version 2.5.1) *de novo* assembly software [21] with the following parameters 'spades.py --careful -1 strain.1.fastq.gz -2 strain.2.fastq -t 4 -k 33,55,77,85,93'. The resulting contigs.fasta file was annotated using Prokka [22] and the capsular locus was extracted using the Artemis Comparison Tool [23].

2.4 Serotyping

Isolates were initially serotyped at the Pneumococcal Reference Laboratory, PHE - National Infection Service, London, using slide agglutination. Briefly, cultures were grown in 5 ml of MAST Todd Hewitt Broth (PHE media services) for four hours, at 35°C with 5% CO_2, centrifuged at 453 x*g* for 30 min, and the supernatant removed. Cell pellets were re-suspended in a small residual volume of broth and subjected to slide agglutination tests with latex absorbed pool antisera (ImmuLex™ Pneumotest kit) or standard group and factor sera (SSI Diagnostica A/S, Hilleroed, Denmark) [24].

Isolates were serotyped at SSI Diagnostica using the gold-standard method for Pneumococcus serotyping - Quellung reaction and with latex agglutination [2,25]. The cultures were grown in serum broth (SSI Diagnostica A/S, Hilleroed, Denmark) overnight. Two to four µL of culture were added onto a glass slide and mixed with the same amount of either pool, group or factor antisera (SSI Diagnostica A/S, Hilleroed, Denmark). The mixture was then observed for capsule swelling reaction in a phase contrast microscope, and typed using the Key to pneumococcal factor serum from SSI Diagnostica A/S [26].

2.5 Production of antiserum at SSI Diagnostica A/S

A vaccine was developed from the novel strain, and rabbits were immunised. Bleeds from the rabbits were taken throughout the immunisation period, and the titer of the reaction between antiserum and the novel strain was monitored using the Quellung reaction.

The titer of the antiserum was found by making a twofold dilution of the antisera from each bleed and mixing it with either serotype 7F, 7A, 7B, 7C and the novel strain.

2.6 Capsular polysaccharide purification

The polysaccharides were isolated from cultivated pneumococcus bacteria of the relevant serotypes. The 7D pneumococcus was grown in serum bouillon. The cells were harvested by centrifugation and lysed using deoxycholate. The solution was adjusted to pH 5.0 and the

precipitate, deoxycholate and cell debris, was removed by centrifugation. Proteins were precipitated by adding $CaCl_2$ to reach 0.1% concentration followed by ethanol to reach 25% concentration and removed by filtration. Polysaccharides were precipitated by increasing the concentration of ethanol to 80% and recovered by centrifugation. Subsequently, the precipitate was dissolved in 0.05 M acetate buffer, pH 5.0 and DNAse and RNAse were added to the solution. The pH was adjusted and trypsin was added to decompose any remaining protein residues. Finally, the polysaccharides were freeze-dried after diafiltration.

2.7 NMR spectroscopy

Polysaccharide samples (10 mg) were dissolved in 99.9% D_2O (500 µL, Sigma Aldrich, Munich, Germany), and spectra were recorded at 40 °C on a Bruker Avance (798.80 MHz for ^{1}H and 200.88 MHz for ^{13}C) equipped with a 5 mm TCI $^{1}H/(^{13}C, ^{15}N)$ cryoprobe using acetone as reference (2.22 ppm and 30.89 ppm for ^{1}H and ^{13}C, respectively). ^{1}H-NMR spectra were acquired with presaturation, sweep widths of 6393.9 Hz and 32768 points zero filled to 65536 points. ^{13}C-NMR spectra were acquired with power-gated decoupling of ^{1}H, sweep widths of 48076.9 Hz and 32768 points zero filled to 65536 points. The DQF-COSY spectra were acquired with sweep widths of 6393.9 and 6390.4 Hz in F2 and F1, respectively, and 4096x512 points zero filled to 8192x1024 points. The TOCSY spectra were acquired width the same sweep widths and number of points as the DQF-COSY using MLEV17 spinlock sequence with a mixing time of 60 ms and 31887.8 Hz spinlock power. The NOESY spectra were acquired with presaturation and the same sweep widths and number of points as the DQF-COSY spectra with a mixing time of 200 ms. The multiplicity edited ^{1}H-^{13}C HSQC were recorded with 2048x512 points zero filled to 4096x1024 points with sweep widths of 6393.9 and 35354.2 Hz in F2 and F1, respectively. The ^{1}H-^{13}C HSQC-TOCSY spectra were acquired with 1024x512 points zero filled to 4096x1024 points and the same sweep widths as the HSQC spectra using the DIPSI2 spinlock sequence with a mixing time of 60 ms and 31887.8 Hz spinlock power. The long range optimised ^{1}H-^{13}C HMBC spectra, optimised for 8 Hz couplings, were recorded with 4096x512 points zero filled to 8192x1024 points with sweep widths of 6393.9 and 44192.7 Hz in F2 and F1, respectively. The DOSY NMR spectra were recorded with 32768 points with sweep widths of 9615.4 Hz using gradients from 0.96 to 47.19 G/cm in 16 increments.

The ^{31}P NMR was recorded at 40 °C on a Bruker 400 AvanceIII with a 5 mm Smartprobe. The 1D ^{31}P NMR spectra was recorded with sweep widths of 12931.0 Hz and 65536 points zero filled to 131072 points. The ^{1}H-^{31}P HMBC was optimised for 8 Hz coupling constants with 2406.9 and 1944.2 Hz in the F2 and F1, respectively, and 4096x512 points zero filled to 8192x1024.

The CLIP-HQSC [27] spectra was measured at 25 °C on a Bruker AvanceIIIHD with a 5 mm prodigy cryoprobe using 4795.4 and 16668.5 Hz in F2 and F1, respectively, and 2048x512 points zero filled to 4096x1024 points.

With the exception of the CLIP-HSQC, all two-dimensional spectra were recorded using standard Bruker pulse sequences and were acquired using TopSpin 2.1. The CLIP-HSQC was recorded using TopSpin 3.2. All NMR data were processed using TopSpin 3.5.

3 Results

When run through the PneumoCaT pipeline, the isolate demonstrated a novel codon at residue 385 which results in a novel amino acid change (CTT (Leu) compared to ACT (Thr) for serotype 40, TTT (Phe) for 7B and TGT (Cys) for 7C), the rest of the capsular variant pattern was otherwise the same as expected for 7B in the CTV database [12]. When tested by slide agglutination at PHE a non-recognised pattern of reaction was seen with the serogroup 7 factor serum, giving reactions with both factor 7e and 7f.

The novel 7 isolate was sent to SSI Diagnostica A/S, Denmark for confirmation of serotype. The isolate was confirmed to have an unusual pattern of factor sera reactions by Quellung test. Subsequently, it was grown in serum broth and then tested with ImmuLex Pneumotest (SSI Diagnostica A/S, Denmark), which is a latex agglutination test. The novel 7 isolate agglutinated with Pool C and P, which shows that the isolate belonged to serogroup 7. The result was then confirmed with Group 7 antiserum, and further serotyped using factor 7b, 7c, 7e and 7f antisera. The novel 7 strain showed capsule swelling with factor 7e and 7f, the capsule swelling was very strong with factor 7f. The Key to pneumococcal factor serum from SSI Diagnostica A/S [26], adapted in table 1, shows that serotype 7B only reacts with factor 7e and serotype 7C only reacts with factor 7f, so when the novel 7 isolate shows a reaction with both factor 7e and 7f, this indicates that it is a new serotype within serogroup 7.

Table 1: Serogroup 7 agglutination results.

Type	Reaction in factor serum			
	7b	7c	7e	7f
7F	+++			
7A	(+)	+++		
7B			+++	
7C				+++
7D			+	+++

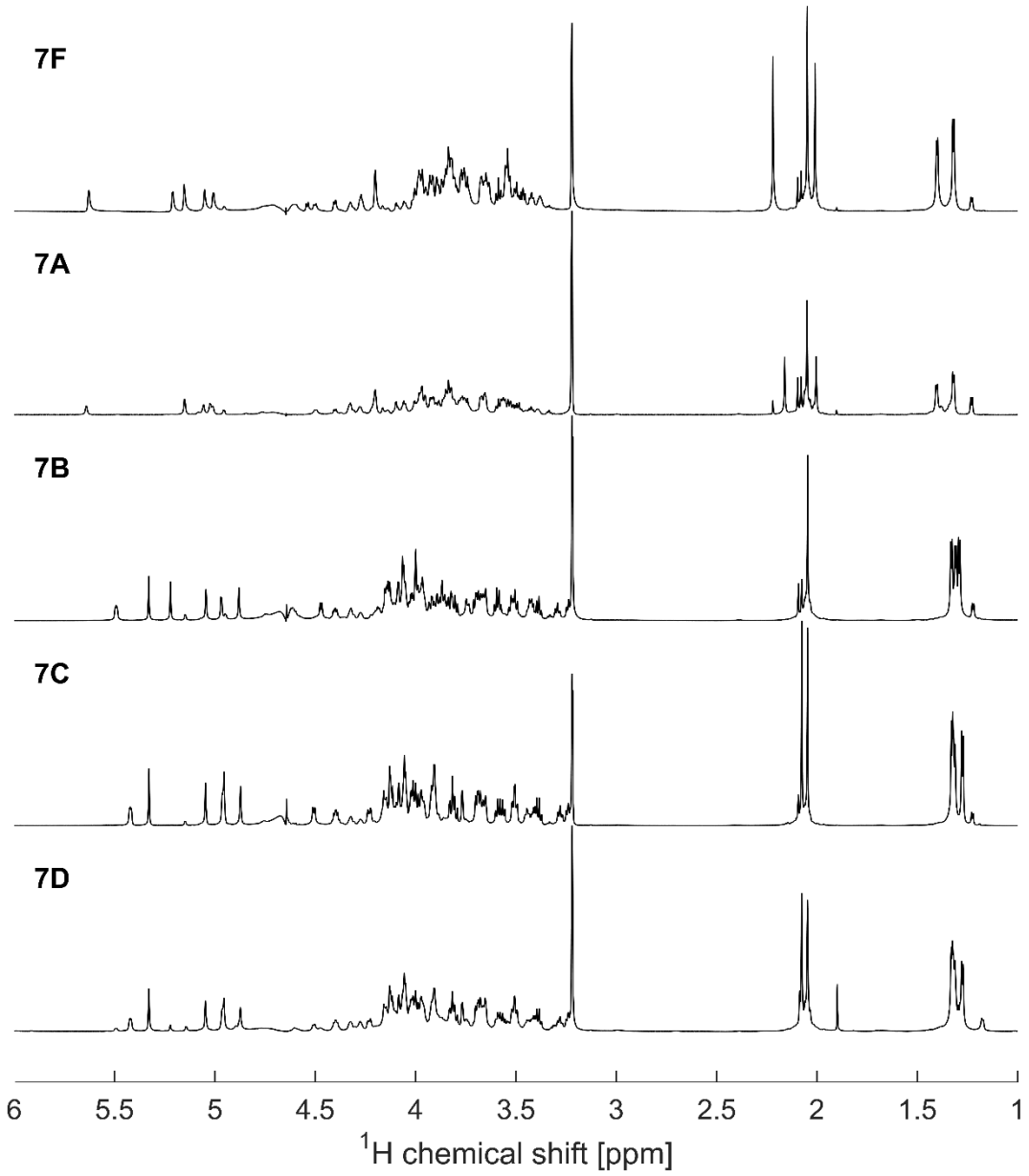
Rabbits were immunised with a vaccine made from the novel 7 strain, and when examining the antiserum from the rabbits it showed reactions with both serotype 7B, 7C and the novel strain. The antiserum showed a stronger reaction with the novel strain than with the serotype 7B and 7C strains. This also indicated that the strain was a novel strain, which shares antigenic epitopes with both serotypes 7B and 7C.

Further genomic analysis at PHE showed that the novel 7 isolate was multi-locus sequence type (ST) 1533 (allelic profile 8,10,2,1,1,26,1), which was the same as four other serotype 7B isolates in the PneumoCaT evaluation analysis, and has 3 of 7 alleles in common with the most common serotype 7F sequence type in PHE data, ST191 (allelic profile 8,9,2,1,6,1,17). Comparison of the 7B capsular locus (CR931641.1) and the extracted capsular locus for the novel 7 isolate revealed that the two sequences differ by a single non-synonymous SNP at position 385 of *wcwK.*

In order to elucidate the structural background for the serological results, the strains were grown and polysaccharides were isolated. The yield of the 7D CPS purification was 83.6 mg from 5 L of cell broth, and using NanoDrop the protein impurities was determined to be 1.4%. The remaining serogroup 7 strains are commercially available (SSI Diagnostica A/S).

The NMR investigation started by inspection of standard 1D ^{1}H spectra, and all samples (7B, 7C and the new serotype) resulted in good NMR spectra as seen in figure 2, indicating that the polysaccharides in the concentrations used demonstrate good solubility. Initial NMR data of the CPS of the novel 7 strain revealed that it had many similarities to the spectra of 7B and 7C. As the structure for 7C was not previously described, it was necessary to fully assign this structure before focusing on the novel 7D strain. The spectra for 7C contained signals corresponding to seven different anomeric signals, see figure 3, one of which had a phosphodiester linkage, and these were labelled **A** to **G** in order of descending ^{13}C chemical shift. Furthermore, signals corresponding to five methyl groups were observed, three of which had almost identical chemical shifts corresponding to the methyl of 6-deoxy sugars. The remaining two were slightly more downfield

239 with chemical shifts corresponding to the methyl of acetyl groups. The assignment of the different
240 carbohydrate units were performed using DQF-COSY, NOESY, TOCSY, HSQC, HSQC-TOCSY
241 and HMBC.



242

243 Figure 2: ^{1}H NMR spectra of serogroup 7 serotypes. From top to bottom: 7F, 7A, 7B, 7C, 7D.

244 One of these anomeric signals, **A**, was identified as originating from a furanoside, while the
245 remaining six were identified as pyranosides. The anomeric configuration of the six pyranosides
246 were determined using the $^1J_{H,C}$ coupling constants [28], shown in supplementary table S1, and

four of them, **D**, **E**, **F** and **G**, were identified as having α-configuration, $^{1}J_{H1,C1}$ of ~170 Hz, while the remaining two, **B** and **C**, were identified as having β-configuration, $^{1}J_{H1,C1}$ of ~160 Hz.

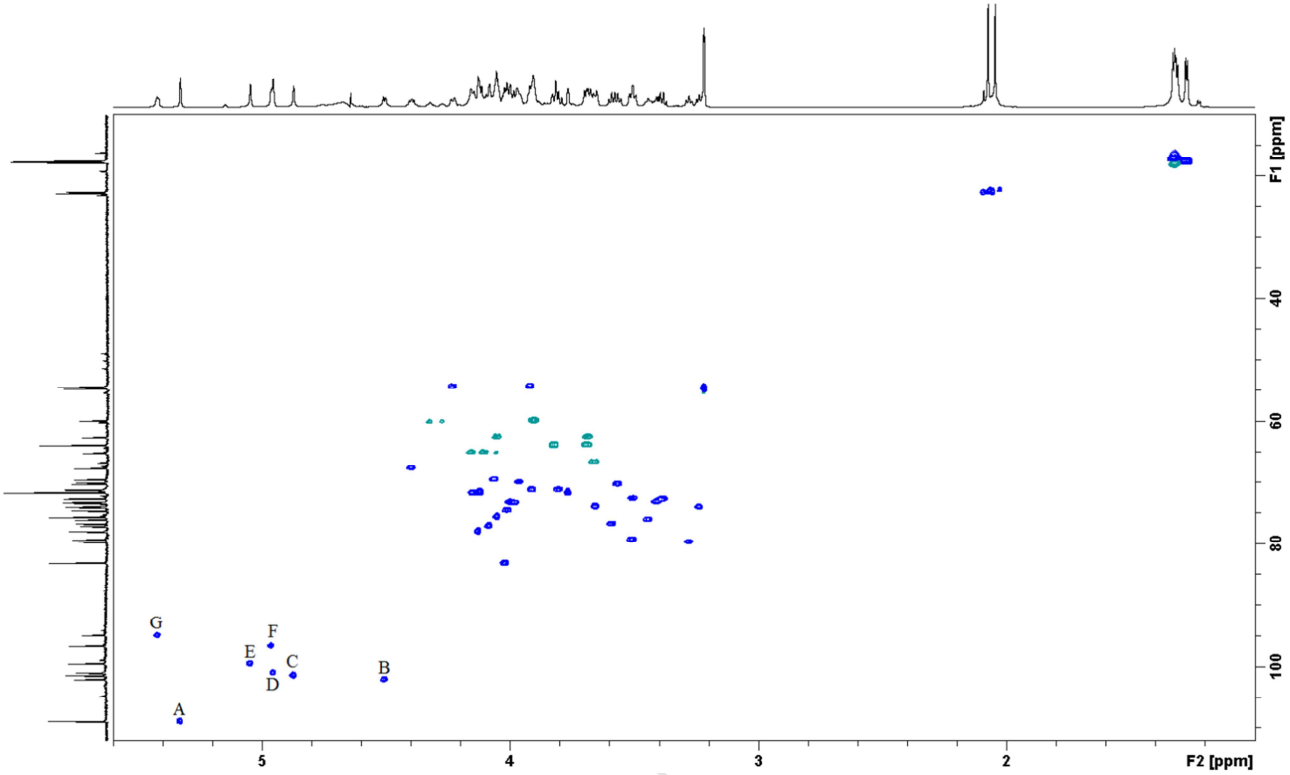


Figure 3: HSQC spectrum of the 7C sample. The ^{1}H and ^{13}C projections are from separate experiments. The labels correspond to the anomeric positions of the seven monosaccharides.

The three pyranosides **C**, **D** and **E** had very similar couplings constants (table S1 in supplementary information), including the $^{3}J_{H1,H2}$ coupling constant, even though only one was of β-configuration, and all three were determined to be monosubstituted rhamnoses. The presence of three rhamnoses would also explain the three upfield methyl groups described earlier. The β-rhamnose **C** was determined to be 2-substituted, as evident by its downfield ^{13}C chemical shift, as was the α-rhamnose **E** while the other α-rhamnose, **D**, was 4-substituted. The rhamnoside configurations were determined by the $^{3}J_{H,H}$ coupling constants, which corresponded to 6-deoxy mannose configurations with the 2-position proton being equatorial, and the 3, 4 and 5-position protons being axial.

The pyranoside **B** was determined to be a β-glucose, as it only has large $^{3}J_{H,H}$ coupling constants, and was found to be 4-substituted, based on the downfield chemical shift of the respective position. The pyranoside **F** was also of glucose configuration, but contained a 2-*N*-acetyl, and had α-configuration. **F** was found to be 6-substituted with a phosphodiester. The last pyranoside, **G**, was determined to be an *N*-acetyl α-galactosamine and was 3,4-substituted as well as being

anomerically bound to the aforementioned phosphodiester. This galactoside configuration was determined mainly from the $^{3}J_{H,H}$ coupling constants, as seen in supplementary table S1, as determination of configuration could not rely on chemical shifts as they were quite heavily shifted due to the disubstitution compared to unsubstituted α-galactosamine [29]. Finally, the furanoside **A** was determined to be a terminal β-ribose as the chemical shifts matched those of methylated β-ribofuranose [29], as well as those of the corresponding β-ribofuranose in 7B [15].

Finally connection of the different monosaccharide units was determined using HMBC and NOESY correlations (table S2 in supplementary data). Starting from the 6-substituted α-Glc*p*NAc **F**, the anomeric proton had HMBC and NOESY correlations to the 2-position of the α-rhamnose **E**, as well as NOESY correlations to the anomeric proton of **E**, further confirming the 2-substitution. From the anomeric position of **E** there were HMBC and NOESY correlations to the 2-position of the β-rhamnose **C**, the anomeric of which had HMBC and NOESY correlations with the 4-position of the β-glucose **B**. This in turn had correlations to the 4-position of the disubstituted α-Gal*p*NAc, which were the last part of the backbone of the repeating unit as the anomeric position of **G** were connected to the same phosphodiester as the 6-position of **F**, which was confirmed by ^{1}H, ^{31}P HMBC (figure S17 in supplementary data). The terminal β-ribose **A** had HMBC and NOESY correlations to the 4-position of the α-rhamnose **D**, which in turn had correlations with the 3-position of **G**. This means the CPS structure of 7C was almost identical to that of 7B, with the only difference being a α-Gal*p*NAc (**G**) in 7C instead of an α-glucose in 7B, and all of the signals from **A**, **C**, **E**, and **F** were practically identical to their counterparts in 7B. The signals from **B** and **D**, which were the two carbohydrates directly connected to the α-Gal*p*NAc **G**, are shifted to some extent relative to the corresponding residues in 7B. The full assignment of the 7C CPS is given in table 2. During the elucidation of 7B, Jansson *et al*. [15] determined monosaccharide composition, assigned fragments obtained by partial acid hydrolysis and determined absolute configuration, and due to the high similarity to the CPS in 7B and the good agreement with assigned chemical shifts, the sugars in 7C were assumed to have the same absolute configuration.

Table 2: Table containing the assignments of 7C, 7B and 7D as it contains both repeating units in a 5:1 ratio.

7C table	**1**	**2**	**3**	**4**	**5**	**6**	**NAc-Me**	**C=O**
β-Rib*f*-(1- **A**	5.330 108.84	4.054 75.63	4.124 71.64	4.024 83.17	3.691/3.824 63.77	N/A	N/A	N/A
-4)-α-Rha*p*-(1- **D**	4.956 100.90	3.771 71.39	3.916 71.24	3.510 79.31	4.401 67.71	1.276 17.60	N/A	N/A
O-P-6)-α-Glc*p*NAc-(1- **F**	4.964 96.63	3.921 54.39	3.808 71.18	3.571 70.25	4.154 71.72	4.113/4.154 65.17	2.049 22.60	N/A 174.99
-2)-α-Rha*p*-(1- **E**	5.049 99.48	4.089 77.10	3.967 70.01	3.508 72.62	4.066 69.56	1.324 17.54	N/A	N/A
-2)-β-Rha*p*-(1- **C**	4.874 101.33	4.131 77.95	3.659 73.93	3.39 72.69	3.414 73.19	1.317 17.77	N/A	N/A
-4)-β-Glc*p*-(1- **B**	4.509 102.01	3.240 74.02	3.594 76.71	3.283 79.69	3.447 76.06	3.688/4.054 62.63	N/A	N/A
-3,4)-α-Gal*p*NAc-(1-O-P- **G**	5.422 94.94	4.234 54.45	4.014 74.56	3.976 73.37	4.002 73.25	3.909 59.94	2.076 22.78	N/A 174.87
7B table	**1**	**2**	**3**	**4**	**5**	**6**	**NAc-Me**	**C=O**
β-Rib*f*-(1-	5.330 108.85	4.062 75.61	4.146 71.61	4.022 83.10	3.699/3.837 63.89	N/A	N/A	N/A
-4)-α-Rha*p*-(1-	5.222 100.89	3.999 71.17	3.989 71.28	3.524 79.64	4.401 67.50	1.290 17.61	N/A	N/A
O-P-6)-α-Glc*p*NAc-(1-	4.968 96.56	3.933 54.35	3.804 71.11	3.596 70.13	4.145 71.67	4.119/4.188 64.99	2.048 22.56	N/A 174.96
-2)-α-Rha*p*-(1-	5.046 99.42	4.087 76.98	3.967 69.96	3.506 72.53	4.062 69.53	1.331 17.47	N/A	N/A
-2)-β-Rha*p*-(1-	4.881 101.24	4.130 77.87	3.657 73.90	3.385 72.64	3.421 73.14	1.330 17.54	N/A	N/A
-4)-β-Glc*p*-(1-	4.473 102.00	3.237 74.05	3.586 76.59	3.292 79.54	3.433 76.01	3.687/4.066 62.61	N/A	N/A
-3,4)-α-Glc*p*-(1-O-P-	5.493 95.95	3.746 73.26	4.003 75.29	3.869 72.81	3.967 72.86	3.893 59.90	N/A	N/A

(N/A for non applicable)

After having elucidated the structure of 7C it was evident from the 7D spectra, as seen in the 1D spectra in figure 2 and overlay of the anomeric region of the 2D HSQC spectra, figure 4, that 7D was a combination of the repeating unit structures of 7B and 7C. The relative ratio was approximate 1:5 between 7B and 7C, as measured by integrals from ^{1}H NMR, ^{31}P NMR (figure S8 in supplementary data) and HSQC. In an attempt to deduce if the repeating units corresponding to 7B and 7C are in the same polymer, a DOSY NMR experiment, which separate the NMR signals based on diffusion rate, was performed. The DOSY data showed some very small differences for signals unique to 7B and 7C, but the overall pattern were unclear, so no conclusion could be drawn. However, the experiment clearly showed that the CWPS polysaccharide is significantly smaller in size having approximately one order of magnitude faster diffusion. It cannot be determined based on the current data if the polysaccharide for 7D is one long chain with both structural elements present, or if it is a mixture of the two structures in separate chains. The peaks corresponding to the CWPS [16], which is present in all the samples, was in higher concentration in the 7D sample, and the sample also had a higher salt concentration compared to the 7B and 7C samples. As CWPS is highly charged, containing both phosphocholine groups and backbone phosphodiesters, some of the signals have shifted markedly compared to the CWPS present in the 7B and 7C samples. This illustrates the importance of considering differences in sample preparation methodologies and reagent concentrations when comparing chemical shifts between different studies.

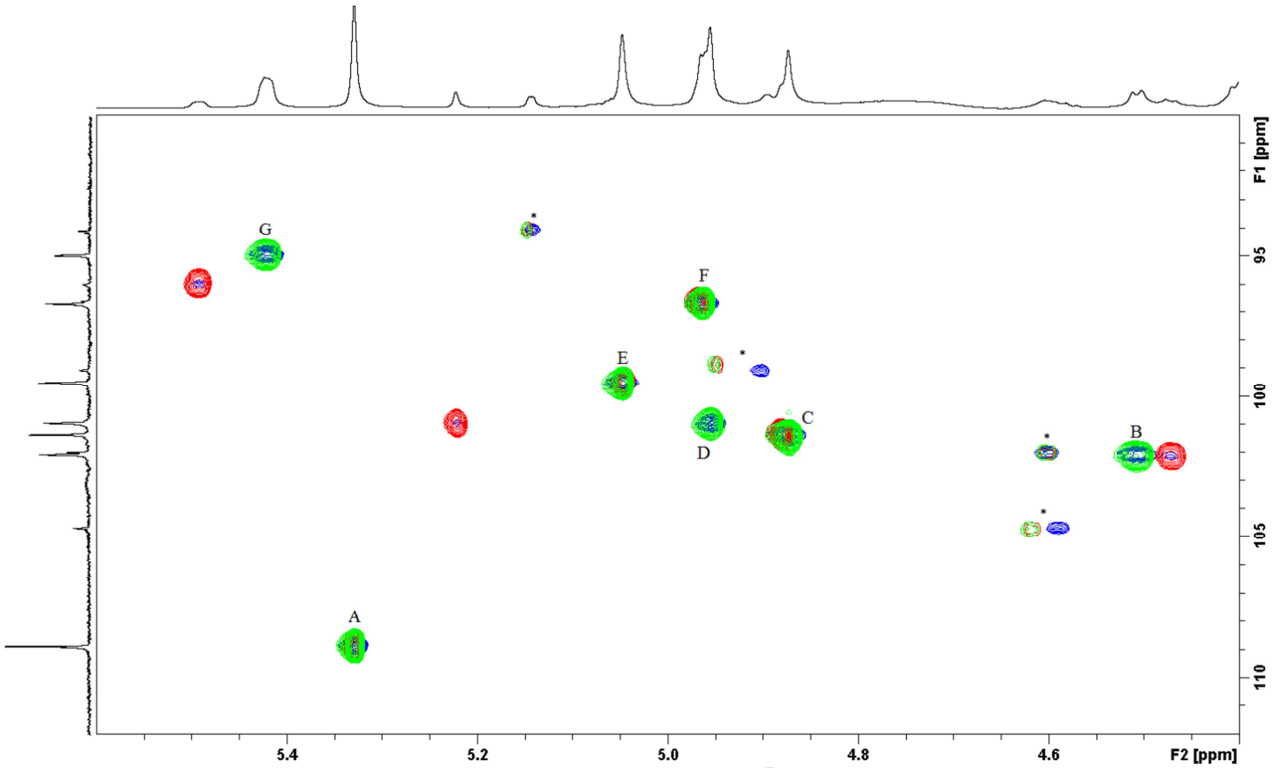


Figure 4: HSQC of 7D (blue) with 7B (red) and 7C (green) superimposed. The projections are from separate 7D ^{1}H and ^{13}C NMR, note that the signals from 7C are more intense in 7D compared to signals originating from 7B. The signals marked with * are from CWPS and the A-G labels correspond to the 7C labels in figure 3 and table 2.

4 Discussion

With modern developments in both serological and genomic techniques the ability to detect novel variants of previously described serotypes has improved and many potential novel pneumococcal serotypes have been reported. However the question remains, when is a novel variant a real “new serotype”? In the case of pneumococcus, the phenotypical serotype is the most important determinant for surveillance purposes as vaccine-driven immunity against the pneumococcus is based on serotype-specific immune responses to polysaccharide and conjugated-polysaccharide vaccines. Therefore we would argue that the determination of a “serotype” should be aligned with the immunological properties of the organism in order to accurately reflect the efficacy of a vaccination program.

As observed in the agglutination test, the novel serotype described here was determined to be part of serogroup 7 and acted as a hybrid between serotype 7C and 7B, although more like 7C than 7B. After full structure elucidation of the new CPS, as well as the previously isolated 7C, it was evident that this serological observation was is in good agreement with the structure elucidation where the 7D CPS was identified as an approximately 5:1 combination of the 7C and 7B CPS repeating units, figure 4, respectively. The full assignment for 7D, as well as 7C and 7B, can be found in table 2.

However, whether the 7C and 7B CPS repeating units are both present as separate CPS, or if it is mixed into a hybrid CPS is still unknown. The mixed structure would also serve to explain why 7D reacts with antiserum of rabbits immunised with 7B or 7C, as both repeating units are present, and explain its stronger reaction with factor 7f than 7e in the agglutination test, as it has higher amounts of the repeating unit from 7C than from 7B.

As the structure of the CPS of 7C was previously unknown, it was fully assigned, see table 2, and it was determined that the only difference between 7B and 7C was the presence of a branched α-Gal*p*NAc in 7C instead of the branched α-Glc*p* in 7B, see figure 1. As result of this, the NMR spectra of 7B and 7C are quite similar, with the largest differences, aside from the changed monosaccharide, being from the positions directly attached to this monosaccharide. The only part of the serogroup 7 CPS that is maintained in all five serotypes is the presence of the α-Glc*p*NAc-1,2-α-Rha*p*, although in 7F and 7A these are on a branch off of the backbone and in 7B and 7C, and thus also 7D, it is part of the backbone. Another part of the CPS that is only partly maintained across the serogroup is the β-L-Rha*p*-1,4-β-D-Glc*p*, which is present in 7B, 7C and 7D, whereas in 7F and 7A the β-L-Rha*p* is 2-*O*-acetylated. Additionally, all the samples contained the same CPWS, but varied slightly in chemical shift, likely due to slightly different CWPS concentrations.

The new serotype, 7D in the Danish serotype system, was both serologically, genetically and structurally different from other serotypes, although the CPS was a hybrid of 7C/7B. Its hybrid nature was also evident from the rabbit sera results, in which 7D reacted like both 7C and 7B, but mostly like 7C. Compared to the reported hybrids in serogroup 6 [9,10,30], in which 6F, 6G and 6H are hybrids of 6A/6C, 6B/6D and 6A/6B, respectively, though they do not give distinct reactions using typing sera. In the case of 6F, which genetically similar to 6A, the enzyme that differs between production of 6A or 6C had a single point mutation resulting in the ability to produce both. This is similar to the case for 7D, which is most similar to 7B, but the enzyme responsible for the α-Glc*p* unique to 7B has been mutated to be capable of transferring both α-Glc*p* and α-Gal*p*NAc, making the CPS a mixture of 7C and 7B. While there is no direct evidence of the two repeating units being covalently bound, the overall results indicates that they are, as the expressed glycosyltransferase unique to 7D can transfer both the α-Glc*p* and the α-Gal*p*NAc, and antiserum from 7D immunised rabbits also reacted with 7B and 7C.

In the future, in order to avoid confusion on the naming of pneumococcal “serotypes”, we would propose an ideal solution would be to have an official overarching body with rules for accepting new serotype designations. For example, that the isolate should exhibit both a genetic difference and a serological difference as demonstrated by rabbit sera (enabling possible commercial supply

of typing sera to identify it), plus elucidation of a distinct structural difference to other existing serotypes. We would further propose that other variants noted by genetic analysis or only demonstrated by use of specific monoclonal antibodies, should be named in a different way to avoid confusion. To realise this goal would require international agreement on typing methods and nomenclature for both serological and genotypic capsular types. With serotypes designated only if there are widely available serologically based methods to determine them and enable global comparison between laboratories.

5 Acknowledgements

C.K. acknowledges the Novo Nordisk foundation for funding: Biotechnology-based synthesis and production research, program grant no 5371. The 800 MHz NMR spectra were recorded on the 800 MHz spectrometer of the Danish National Instrument Center for NMR Spectroscopy of Biological Macromolecules at the Carlsberg Laboratory and the Technical University of Denmark. We thank Katrine Helander Pedersen (SSI Diagnostica, Denmark) for doing the serotyping and immunisation of rabbits during this study.

6 References

[1] E. AlonsoDeVelasco, A.F. Verheul, J. Verhoef, H. Snippe, Streptococcus pneumoniae: virulence factors, pathogenesis, and vaccines., Microbiol. Rev. 59 (1995) 591–603. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=239389&tool=pmcentrez&rendertype=abstract.

[2] J. Henrichsen, Six newly recognized types of Streptoccus pneumoniae, J. Clin. Microbiol. 33 (1995) 2759–2762.

[3] M.C. McEllistrem, M.H. Nahm, Novel pneumococcal serotypes 6C and 6D: Anomaly or harbinger, Clin. Infect. Dis. 55 (2012) 1379–1386. doi:10.1093/cid/cis691.

[4] A.L. van Tonder, J.E. Bray, S.J. Quirk, G. Haraldsson, K.A. Jolley, M.C. Maiden, S. Hoffmann, S.D. Bentley, H. Ásgeir, H. Erlendsdóttir, K.G. Kristinsson, A.B. Brueggemann, Putatively novel serotypes and the potential for reduced vaccine effectiveness: capsular locus diversity revealed among 5,405 pneumococcal genomes, Microb. Genomics. 2 (2016) 90. doi:10.1099/mgen.0.000090.

[5] K.A. Geno, J.S. Saad, H. Nahm, Discovery of Novel Pneumococcal Serotype 35D, a Natural WciG-Deficient Variant of Serotype 35B, J. Clin. Microbiol. 55 (2017) 1416–1425. doi:10.1128/JCM.00054-17.

[6] J. Yu, M.D.G.S. Carvalho, B. Beall, M.H. Nahm, A rapid pneumococcal serotyping system based on monoclonal antibodies and PCR, J. Med. Microbiol. 57 (2008) 171–178. doi:10.1099/jmm.0.47549-0.

[7] R.L. Burton, K.A. Geno, J.S. Saad, M.H. Nahm, Pneumococcus with the “6E” cps Locus Produces Serotype 6B Capsular Polysaccharide, J. Clin. Microbiol. 54 (2016) 967–971. doi:10.1128/JCM.03194-15.

[8] G. Kapatai, C.L. Sheppard, L.J. Troxler, D.J. Litt, J. Furrer, M. Hilty, N.K. Fry, Pneumococcal 23B molecular subtype identified using whole genome sequencing, Genome Biol. Evol. 9 (2017) 2145–2158. doi:10.1093/gbe/evx092.

[9] M.B. Oliver, M.P.G. van der Linden, S.A. Küntzel, J.S. Saad, M.H. Nahm, Discovery of streptococcus pneumoniae serotype 6 variants with glycosyltransferases synthesizing two differing repeating units, J. Biol. Chem. 288 (2013) 25976–25985. doi:10.1074/jbc.M113.480152.

[10] C.L. Sheppard, B. Pichon, R.C. George, L.M.C. Hall, Streptococcus pneumoniae isolates expressing a capsule with epitopes of both serotypes 6A and 6B, Clin. Vaccine Immunol. 17 (2010) 1820–1822. doi:10.1128/CVI.00335-10.

[11] J.J. Calix, M.H. Nahm, A New Pneumococcal Serotype, 11E, Has a Variably Inactivated *wcjE* Gene, J. Infect. Dis. 202 (2010) 29–38. doi:10.1086/653123.

[12] G. Kapatai, C.L. Sheppard, A. Al-Shahib, D.J. Litt, A.P. Underwood, T.G. Harrison, N.K. Fry, Whole genome sequencing of Streptococcus pneumoniae: development, evaluation and verification of targets for serogroup and serotype prediction using an automated pipeline., PeerJ. 4 (2016) e2477. doi:10.7717/peerj.2477.

[13] M. Moreau, J.C. Richards, M.B. Perry, Application of high-resolution NMR spectroscopy to the elucidation of the structure of the specific capsular polysaccharide of Streptococcus pneumoniae type 7F, Carbohydr. Res. 182 (1988) 79–99. doi:10.1016/0008-6215(88)84093-X.

[14] I. Backman-Marklund, P.-E. Jansson, B. Lindberg, J. Henrichsen, Structural studies of the capsular polysaccharide from Streptococcus pneumoniae type 7A, Carbohydr. Res. 198 (1990) 67–77. doi:10.1016/0008-6215(90)84277-2.

[15] P.E. Jansson, J. Lindberg, K.M. Swarna Wimalasiri, J. Henrichsen, The structure of the capsular polysaccharide from Streptococcus pneumoniae type 7B, Carbohydr. Res. 217 (1991) 171–180. doi:10.1016/0008-6215(91)84127-Z.

[16] I.C. Skovsted, M.B. Kerrn, J. Sonne-Hansen, L.E. Sauer, A.K. Nielsen, H.B. Konradsen, B.O. Petersen, N.T. Nyberg, J. Duus, Purification and structure characterization of the active component in the pneumococcal 22F polysaccharide capsule used for adsorption in pneumococcal enzyme-linked immunosorbent assays, Vaccine. 25 (2007) 6490–6500. doi:10.1016/j.vaccine.2007.06.034.

[17] J.P. Kamerling, Pneumococcal polysaccharides: A chemical view., in: A. Tomasz (Ed.), Streptococcus Pneumoniae Mol. Biol. Mech. Dis., Mary Ann Liebert, New York, NY, 2000: pp. 81–114.

[18] A. Mavroidi, D.M. Aanensen, D. Godoy, I.C. Skovsted, M.S. Kaltoft, P.R. Reeves, S.D. Bentley, B.G. Spratt, Genetic relatedness of the Streptococcus pneumoniae capsular biosynthetic loci, J. Bacteriol. 189 (2007) 7841–7855. doi:10.1128/JB.00836-07.

[19] D.M. Aanensen, A. Mavroidi, S.D. Bentley, P.R. Reeves, B.G. Spratt, Predicted functions and linkage specificities of the products of the Streptococcus pneumoniae capsular biosynthetic loci, J. Bacteriol. 189 (2007) 7856–7876. doi:10.1128/JB.00837-07.

[20] R. Tewolde, T. Dallman, U. Schaefer, C.L. Sheppard, P. Ashton, B. Pichon, M. Ellington, C. Swift, J. Green, A. Underwood, MOST: a modified MLST typing tool based on short read sequencing, PeerJ. 4 (2016) e2308. doi:10.7717/peerj.2308.

[21] A. Bankevich, S. Nurk, D. Antipov, A.A. Gurevich, M. Dvorkin, A.S. Kulikov, V.M. Lesin, S.I. Nikolenko, S. Pham, A.D. Prjibelski, A. V Pyshkin, A. V Sirotkin, N. Vyahhi, G. Tesler, M.A. Alekseyev, P.A. Pevzner, SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing., J. Comput. Biol. 19 (2012) 455–77. doi:10.1089/cmb.2012.0021.

[22] T. Seemann, Prokka: Rapid prokaryotic genome annotation, Bioinformatics. 30 (2014) 2068–2069. doi:10.1093/bioinformatics/btu153.

[23] T.J. Carver, K.M. Rutherford, M. Berriman, M.A. Rajandream, B.G. Barrell, J. Parkhill, ACT: The Artemis comparison tool, Bioinformatics. 21 (2005) 3422–3423. doi:10.1093/bioinformatics/bti553.

[24] H.-C. Slotved, C.L. Sheppard, T. Dalby, A. van der Ende, N.K. Fry, E. Morfeldt, O. Nyholm, A. Rokney, M. Ron, L. Siira, K.J. Scott, A. Smith, L. Thom, M. Toropainen, D.F. Vestrheim, External Quality Assurance for Laboratory Identification and Capsular Typing of Streptococcus pneumoniae, Sci. Rep. 7 (2017) 13280. doi:10.1038/s41598-017-13605-8.

[25] H.-C. Slotved, M. Kaltoft, I.C. Skovsted, M.B. Kerrn, F. Espersen, Simple, Rapid Latex Agglutination Test for Serotyping of Pneumococci (Pneumotest-Latex), J. Clin. Microbiol. 42 (2004) 2518–2522. doi:10.1128/JCM.42.6.2518.

[26] Key to pneumoccoccal factor serum, (2013). http://www.ssidiagnostica.com/upload/files/english/brochures/pneumococcus/BrochurePneumococcal factor antisera key 18058.pdf (accessed March 11, 2018).

[27] A. Enthart, J.C. Freudenberger, J. Furrer, H. Kessler, B. Luy, The CLIP/CLAP-HSQC: Pure absorptive spectra for the measurement of one-bond couplings, J. Magn. Reson. 192 (2008) 314–322. doi:10.1016/j.jmr.2008.03.009.

[28] K. Bock, C. Pedersen, A study of 13CH coupling constants in hexopyranoses, J. Chem. Soc., Perkin Trans. 2. (1974) 293–297. doi:10.1039/p29740000293.

[29] P.K. Agrawal, NMR Spectroscopy in the structural elucidation of oligosaccharides and glycosides, Phytochemistry. 31 (1992) 3307–3330. doi:10.1016/0031-9422(92)83678-R.

[30] I.H. Park, K.A. Geno, J. Yu, M.B. Oliver, K.H. Kim, M.H. Nahm, Genetic, biochemical, and serological characterization of a new pneumococcal serotype, 6H, and generation of a pneumococcal strain producing three different capsular repeat units, Clin. Vaccine Immunol. 22 (2015) 313–318. doi:10.1128/CVI.00647-14.

- Discovery of the 7D serotype and description of its hybrid capsular polysaccharide
- The new serotype was identified as a hybrid between 7B and 7C
- Structural elucidation of the capsular polysaccharides by NMR spectroscopy
- From genetics to serotyping and structure, the results support the new serotype