Determination of native capsular polysaccharide structures of Steptococcus pneumoniae serotypes 39, 42 and 47F and comparison to genetically or serologically related strains

Petersen, Bent O.; Meier, Sebastian; Paulsen, Berit Smestad; Redondo, Antonio R.; Skovsted, Ian C.

Published in:
Carbohydrate Research

Link to article, DOI:
10.1016/j.carres.2014.06.018

Publication date:
2014

Document Version
Early version, also known as pre-print

Link back to DTU Orbit

Citation (APA):
Petersen, B. O., Meier, S., Paulsen, B. S., Redondo, A. R., & Skovsted, I. C. (2014). Determination of native capsular polysaccharide structures of Steptococcus pneumoniae serotypes 39, 42 and 47F and comparison to genetically or serologically related strains. Carbohydrate Research. DOI: 10.1016/j.carres.2014.06.018

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Title: Determination of native capsular polysaccharide structures of Streptococcus pneumoniae serotypes 39, 42 and 47F and comparison to genetically or serologically related strains

Article Type: Full Length Article

Keywords: NMR; capsular polysaccharide; serotype 39; serotype 42; serotype 47F; streptococcus pneumoniae

Abstract: The diversity of capsular polysaccharides of the bacterial pathogen Streptococcus pneumoniae leads to at least 91 different serotypes. While the genetic loci for capsular biosynthesis have been characterized for all serotypes, the determination of resultant polysaccharide structures remains incomplete. Here, we report the chemical structures of the capsular polysaccharides of serotypes 39, 42 and 47F from the genetic cluster 4, and discuss the structures in the context of structures from serologically and genetically related serotypes. Antigenic determinants can be defined in this manner.

The structure of the serotype 39 capsular polysaccharide is

\[ D \beta-D-Galp-(1,6) \]
\[ \rightarrow 6)\beta-D-Galf-(1\rightarrow3)\beta-D-Galp-(1\rightarrow4)\beta-D-GalpNAc-(1\rightarrow3)\alpha-D-Galp-(1\rightarrow1)-D-ribitol-(5\rightarrowOPO3\rightarrow \]

and has identical composition to the capsular polysaccharide 10A, but two different linkages.

The serotype 42 structure

\[ C \beta-D-Galf-(1,3)^* \]
\*[\(30Ac,35%; 60Ac,65\%\)]

and has identical composition to the capsular polysaccharide 10A, but two different linkages.

The serotype 42 structure
**(30% 6-OAc, 70% 5,6-di-OAc)**

closely resembles the genetically related serotype 35A, which does not contain residue A.
The structure of the serotype 47F capsular polysaccharide

\[
\rightarrow 6)\beta-D-Galf-(3,5-di-OAc)-(1\rightarrow 3)\beta-D-Galp-(1\rightarrow 6)\beta-D-Galf-(2-OAc)-(1\rightarrow 3)\alpha-D-Galp-(1\rightarrow 2)\rightarrow D-\text{ribitol}(5\rightarrow)\rightarrow A \quad B \quad C \quad D \quad E
\]

is somewhat different from a recently determined structure from the same serogroup, while containing a structural motif that is reflected in serotype 35A and 42 capsular polysaccharide structures, thus explaining the cross-reactivity of serotype 47F with the typing serum 35a.
Dear Editor, Dear Prof. Field

Please find enclosed the revised version of the manuscript

“Determination of native capsular polysaccharide structures of *Streptococcus pneumoniae* serotypes 39, 42 and 47F and comparison to genetically or serologically related strains”

by Bent O. Petersen, Sebastian Meier, Berit Smestad Paulsen, Antonio R. Redondo and Ian C. Skovsted. We have revised the text according to the thorough correction of referee 2. Regarding the suggestion by referee 1, we would prefer to stick to the abbreviated text versions of the structures in the abstracts and have chemical structures in the main text and as TOC figure. Detailed point-by-point lists of the changes are detailed in a separate file. In addition, we have added one author affiliation. Thank you and the referees for your exquisitely rapid help in improving the manuscript!

Yours sincerely,

Bent Petersen
Ref 1:

In the interest of space I think we stick to the text versions and have chemical structures in the main text and as Table of contents figure. The submission makes it difficult to assemble the structures correctly when cutting and pasting into submission form.

Ref 2:

We apologize for the inconsistencies with respect to the phosphodiester linkages. The structures depicted in the figures were correct, the text has been corrected accordingly.

We include a sentence in paragraph 3.3 to state that

“Phosphodiester linkages were determined and validated from independent observations, specifically the splitting of $^{13}$C signals through $^{2}J_{13C31P}$ couplings, from $^{1}H^{31}P$ correlations and from chemical shifts, particularly the upfield shift of $^{13}$C signals split by $^{2}J_{13C31P}$ in phosphodiester groups.”

Other things:
- P2 114 reads “can be approximated”
- Table 2 and 3 are fixed,
- P2 l30 mannitol is corrected
- P 4-5, is fused
- P6 should have read and now reads “anomeric signals” instead of “anomeric regions”
- P6 l31,37,44 are corrected
- P7 l5 is corrected
- P8 3.4 and 3.5 are incorporated into the discussion, which now reads

"The analysis of polysaccharide structures remains a low-throughput method that has been outpaced by the speed of genetic and genomic methods. Structural analysis, however, remains essential in demonstrating, how genetic loci for polysaccharide biosynthesis correlate with resultant polysaccharide structures through the action of encoded enzymatic activities. Here, we have determined the structures of capsular polysaccharides for three serotypes of the\textit{S.pneumoniae} genetic cluster 4. The structures determined herein help fill in the remaining blanks of capsular polysaccharides structure in \textit{Streptococcus pneumoniae}.

Serotype 39 supposedly has no known closely related serotype, but shows cross-reaction with typing serum 10d that reacts with the serotype 10A of known capsular polysaccharide structure.\textsuperscript{21} Serotype 10A and 39 capsular polysaccharides contain only one different glycosidic linkage and one different linkage in the
phosphodiester bridge. In addition, the β-D-Galf branch attached to O3 of β-D-GalNac (Figure 1) is found to be mono-acetylated in our serotype 39 capsular polysaccharide preparation. Notwithstanding, the structures of serotype 10A and 39 capsular polysaccharides turn out to be identical in composition, which was validated by NMR spectroscopy of the serotype 10A capsular polysaccharide. Surprisingly, a previous cluster analysis of streptococcal cps biosynthetic loci did not suggest the close compositional resemblance, which is reflected in the cross-reaction of serotype 39 with typing serum 10d.

The cps biosynthetic locus of serotype 42 is closely related to the cps biosynthetic loci of serotypes 35A and 35C. For serotype 35A, a structure of the capsular polysaccharide has previously been described. The structures of serotypes 35A and 42 are compared in Figure 4B. The close genetic resemblance of the cps biosynthetic loci of serotypes 35A and 42 is reflected by the close resemblance of capsular polysaccharide structures, which are identical apart from the presence of a branching (1→2)-α-D-galactopyranose attached a β-D-galactopyranose in the phosphodiester linkage in serotype 42. The presence of a mannitol previously only reported for serotype 35A structure is validated in the analysis of serotype 42 capsular polysaccharide. Nevertheless, the 35a typing serum reportedly does not react with serotype 42, but shows some reaction with serotype 47F. The structural similarity of the serotypes 35A/42 and 47F is limited to a -β-D-Galp-(1→3)-β-D-GlcP-(1→6)-β-D-Galf motif in serotypes 35A/42 that is also found in serotype 47F and is encircled in Figure 4, explaining the reaction of pneumococcal typing serum 35a with serotype 47F.

The serotype 47F capsular polysaccharide structure is additionally noteworthy in its comparison to a recently determined structure from the same serogroup, the serotype 47A structure (Figure 4). While the serotype 47A structure contains a doubly branched heptasaccharide repeating unit, the serotype 47F contains a linear pentasaccharide. The common antigenic determinant shared by the serotype 47A and 47F capsular polysaccharides is thus part of the conserved -α-D-Galp-(1→2)-D-ribitol-(5→OPO₃→6)-β-D-Galf-(3,5-di-OAc)-(1→3)-β-D-Galp-(1→ motif. The structure of serotypes 47F sheds light on the antigenic determinants in strains with genetically distant cps biosynthetic loci 47F and 47A. Genetically, the serotype 47F
The recently renewed interest in uncharacterized capsular polysaccharide structures predictably will close the remaining gaps in the understanding of capsular polysaccharide biodiversity and its relation to the \textit{cps} biosynthetic loci.\textsuperscript{14-16} Such an improved understanding is warranted in the light of rapid pneumococcal evolution as a consequence clinical interventions and pneumococcal genomic plasticity. This plasticity includes capsule-switching events that encompass the formation of vaccine-escape phenotypes.\textsuperscript{9} The prediction of phenotypic consequences of genomic changes in the \textit{cps} locus should benefit from studies of capsular polysaccharide biodiversity and antigenic determinants as provided herein.”

- P8 l 53-57 Comparison of 35A and 42 corrected and moved to the discussion, as above.
- 35A structure is referenced.
- P9 l 5 now reads “The structural similarity of the serotypes 35A/42 and 47F is limited to a -\(\beta\)-D-Galf\((1\rightarrow3)\)-\(\beta\)-D-Glcp\((1\rightarrow6)\)-\(\beta\)-D-Galf motif in serotypes 35A/42 that is also found in serotype 47F and is encircled in Figure 4, explaining the reaction of pneumococcal typing serum 35a with serotype 47F.\textsuperscript{13}”
- Figures and Tables have been corrected as requested.
- Thank you for the exquisitely thorough correction!
*Graphical Abstract (for review)
Click here to download Graphical Abstract (for review): TOC.eps
• Structures were obtained for serotype 39, 42 and 47F capsular polysaccharides.
• Serotype 39 polysaccharide differs from 10A in two linkages.
• Serotype 42 has close structural resemblance to 35A, but contains a branch.
• Serotype 47F contains a structural motif resembling serotype 35A and 42.
• The resemblance explains the reaction of typing serum 35a with serotype 47F.
Determination of native capsular polysaccharide structures of *Streptococcus pneumoniae* serotypes 39, 42 and 47F and comparison to genetically or serologically related strains

Bent O. Petersen, a,b* Sebastian Meier, a,c Berit Smestad Paulsen, d Antonio R. Redondo e and Ian C. Skovsted e

aCarlsberg Laboratory, 10 Gamle Carlsberg vej, DK-1799 Copenhagen V
bNovo Nordisk A/S, Novo Nordisk Park, DK-2760 Måløv
cDepartment of Chemistry, Technical University of Denmark, Kemitorvet, Building 201, DK-2800 Kgs. Lyngby
dFarmasøytisk institutt, Sem Selands vei 3, N-0316 Oslo, Norway
eStatens Serum Institut, 5 Artillerivej DK-2300 Copenhagen S

*Corresponding author: Bent O. Petersen
E-mail: bp@crc.dk

Keywords: NMR; capsular polysaccharide; serotype 39; serotype 42; serotype 47F; streptoccocus pneumoniae
Abstract
The diversity of capsular polysaccharides of the bacterial pathogen Streptococcus pneumoniae leads to at least 91 different serotypes. While the genetic loci for capsular biosynthesis have been characterized for all serotypes, the determination of resultant polysaccharide structures remains incomplete. Here, we report the chemical structures of the capsular polysaccharides of serotypes 39, 42 and 47 from the genetic cluster 4, and discuss the structures in the context of structures from serologically and genetically related serotypes.

Antigenic determinants can be approximated in this manner.

The structure of the serotype 39 capsular polysaccharide is

\[ \text{D\(\beta\)-D-Galp-(1,6)} \]

\[ \text{\rightarrow 6)-\beta\)-D-Galp(1→3)-\beta\)-D-Galp(1→4)-\beta\)-D-GalpNAc(1→3)-\alpha\)-D-Galp(1→1)-D-ribitol(5→OPO_3→} \]

\[ \text{A B C D E F G} \]

\[ *(3\text{OAc,35%}; 6\text{OAc,65%}) \]

and has identical composition to the capsular polysaccharide 10A, but two different linkages.

The structure of the serotype 42 structure

\[ \text{B} \]

\[ \text{\rightarrow 3)-\beta\)-D-Galp(1→3)-\beta\)-D-Galp(1→3)**-\beta\)-D-Glcp(1→6)-\beta\)-D-Galp(1→1)-D-mannitol(6→OPO_3→} \]

\[ \text{\rightarrow 3)-\beta\)-D-Galp(1→3)-\beta\)-D-Galp(1→3)**-\beta\)-D-Glcp(1→6)-\beta\)-D-Galp(1→1)-D-mannitol(6→OPO_3→} \]

\[ \text{A}\alpha\)-D-Glcp(1,2) \]

\[ **(30\% 6-OAc, 70\% 5,6-di-OAc) \]

closely resembles the genetically related serotype 35A, which does not contain residue A.

The structure of the serotype 47F capsular polysaccharide

\[ \text{\rightarrow 6)-\beta\)-D-Galp(3,5-di-OAc)-(1→3)-\beta\)-D-Galp(1→6)-\beta\)-D-Galp(1→3)-\alpha\)-D-Galp(1→2)-D-ribitol(5→OPO_3→} \]

\[ \text{A B C D E} \]

is somewhat different from a recently determined structure from the same serogroup, while containing a structural motif that is reflected in serotype 35A and 42 capsular polysaccharide structures, thus explaining the cross-reactivity of serotype 47F with the typing serum 35a.
1. Introduction

*Streptococcus pneumoniae*, a Gram-positive bacterial pathogen, is a major cause of pneumonia, septicaemia, meningitis and otitis media in humans. As such, *Streptococcus pneumoniae* (also known as pneumococcus) causes a significant number of deaths both in children and among elderly people each year.\(^1\,\,^2\) *Streptococcus pneumoniae* is an encapsulated bacterium, and the genetic locus encoding the capsule has provided the proof that DNA carries the hereditary information.\(^3\) The polysaccharide capsule forms a hydrated outermost layer of *S. pneumoniae* with a thickness on the order of 200-400 nm. The surface exposed capsule polysaccharide is a target for the host immunological response.\(^4\,\,^5\,\,^6\) In order to evade the host immune system, pneumococcus has evolved at least 91 defined serotypes with immunochemically distinct capsular polysaccharides.\(^7\) These polysaccharides are virulence factors, form the basis of the observed patterns of immunological reactivity, are involved in bacterial adherence and biofilm formation, form the basis of vaccines due to their antigenicity, and are hallmarks of the evolution of bacterial polysaccharide diversity.\(^4\,\,^8\)

Reportedly, pneumococcal evolution encompassing capsule switching events can occur on surprisingly short timescales, especially in response to clinical interventions.\(^9\,\,^10\)

*S. pneumoniae* was first isolated more than 130 years ago and capsular polysaccharides were first isolated almost a century ago and extensively studied since.\(^3\,\,^11\,\,^12\) Nevertheless, the different chemical structures of capsular polysaccharides (CPS) are not yet established for all serotypes.\(^8\) While the capsular polysaccharide structures thus remain incompletely characterized, the sequences of the capsular biosynthetic genes of all pneumococcus serotypes have been determined, yielding insights into the genetic relatedness of the *S. pneumoniae* capsular biosynthetic loci.\(^4\,\,^13\) Unsurprisingly, the study of the repeating unit structure of the *S. pneumoniae* capsular polysaccharide has regained recent interest\(^14\,\,^15\,\,^16\) following the genetic description of the *cps* biosynthetic loci for all serotypes.\(^4\,\,^13\) The description of the *S. pneumoniae* capsular polysaccharides has a role to play in establishing relations between the *cps* genetic locus and the capsular polysaccharide structure,\(^7\) and in providing insights into the epitopes of capsular polysaccharides detected by the immune system.\(^14\) Here, we chose to conduct the complete assignment of \(^1\)H and \(^13\)C resonances in NMR spectra of the native polysaccharides of serotype 39, 42 and 47F. These serotypes of unknown capsular polysaccharide structure belong the same major genetic cluster, previously designated as cluster 4. Cluster 4 is remarkable insofar, as it encodes capsular polysaccharide structures that belong to different serogroups, despite the genetic similarity of the *cps* locus.\(^13\)
Hence, cluster 4 is of central interest in terms of capsular polysaccharide evolution and immunochemistry. For the three serotypes 39, 42 and 47F of cluster 4, the chemical structures of the repeating units in the capsular polysaccharides are reported herein (Figure 1). The structures are compared to known structures of polysaccharides in the same serogroup, or to polysaccharides of genetically closely related serotypes from the same cluster. Such comparisons help define the shared antigenic determinants within serogroups, even in the absence of apparent genetic similarity, and the different antigenic determinants in genetically related serogroups.

2. Experimental

2.1 Sample preparation and materials

The capsular polysaccharides of serotypes 39, 42 and 47F from *Streptococcus pneumoniae* were purified from the corresponding Statens Serum Institute (SSI) collection strains. Purifications were performed using centrifugation of autolysates and choloform-butanol extractions, digests with DNase, RNAse and trypsin, followed by ethanol precipitation. Purity and immunological identity were assessed by immunological assays using typing sera and by measuring protein and DNA contaminations (that were found to be below 5% w/w in each case). The purified capsular polysaccharides were freeze-dried and subsequently dissolved in D$_2$O (99.9%; Cambridge Isotope Laboratories, Andover, MA, USA) to concentrations of approximately 1.5% w/v (10 mg in 600 μl). These samples were transferred to 5 mm NMR sample tubes.

2.2 NMR spectroscopy, processing and analysis

Repeating unit structures of the capsular polysaccharides of serotypes 39, 42 and 47F were determined using NMR spectroscopy. Spectra were recorded on an 800 MHz Bruker Avance II spectrometer equipped with a TCI cryoprobe (Bruker, Karlsruhe, Germany). All spectra were recorded at 310 K and referenced relative to acetone (δ$^1$H=2.225 ppm and δ$^{13}$C=30.89 ppm). The two-dimensional NMR spectra included DQF-COSY (4096×512 complex data points with acquisition times of 574 ms and 72 ms, respectively), NOESY with 800 ms mixing time (2048×512 complex data points sampling 307 ms and 77 ms acquisition),
TOCSY with a 10 kHz spin lock field that was applied for 60 ms (4096×512 complex data points with 574 ms and 72 ms acquisition time), a multiplicity edited $^1$H-$^{13}$C (1024×512 complex points sampling 153 ms and 18 ms acquisition times), $^1$H-$^{13}$C HMBC (2048×512 complex data points sampling 286 ms and 13 ms acquisition times) and a $^1$H-$^{13}$C HSQC TOCSY (1024×512 complex points sampling 143 ms and 21 ms acquisition times, respectively). 1D $^1$H and $^{13}$C spectra were recorded by sampling 16384 complex points during acquisition times of 1.57 s and 341 ms, respectively. In addition, $^1$H-$^{31}$P HSQC spectra were recorded at 310 K on a Bruker DRX400 spectrometer equipped with a 5 mm broadband observe (BBO S1) probe. All NMR spectra were processed with Topspin 2.1 (Bruker) with extensive zero filling in all dimensions. The one-dimensional $^1$H and $^{13}$C NMR spectra were processed with an exponential window functions with a line broadening of 0.3 Hz for $^1$H spectra and 1 Hz for $^{13}$C spectra. All 2D spectra were processed with shifted sine bell window functions in both dimensions. All spectral assignments were conducted in Topspin 2.1.

2.3 Monosaccharide analysis

In order to validate spin system identifications with NMR spectroscopy, the monosaccharide compositions of the capsular polysaccharide of serotypes 39, 42 and 47F were independently determined with monosaccharide composition analysis using gas chromatography (GC). GC was performed for the trimethylsilyl-derivatives of the methyl glycosides, as described,\textsuperscript{16} using mannitol as an internal standard.\textsuperscript{17, 18}
3. Results

3.1 Assignment of spin systems

Samples of purified capsular polysaccharides of serotypes 39, 42 and 47F were prepared at concentrations of approximately 1.5 % (w/v) in D$_2$O. Full assignments of the $^1$H and $^{13}$C chemical shifts and structural analyses were conducted on these samples. One-dimensional $^1$H NMR spectra indicated adequate purity and molecular tumbling (sufficiently long $^1$H T$_2$ times) of the preparations for detailed structural studies (Figure 2). In addition, anomeric signals were well dispersed at high NMR magnetic field (18.7 T, 800 MHz) due to the chemical diversity of the heteropolymeric pneumococcal capsular polysaccharides. Acetylation occurs in all three capsular polysaccharides.

The individual spin systems of the constituting monosaccharide units were identified using the anomeric region of homonuclear 2D DQF-COSY and TOCSY experiments. An overlay of these spectral regions is shown in Figure 3 for the three serotypes 39, 42 and 47F, designating the spin systems of constituting sugar units with letters. Assignments of spin systems showed that the repeating unit of the serotype 39 capsular polysaccharide is a doubly branched heptasaccharide containing β-galactofuranose in addition to α- as well as β-galactopyranose units, as well as β-N-acetylglucosamine and ribitol. The repeating unit is polymerized through phosphate diester linkages. The repeating unit of the serotype 42 capsular polysaccharide is a singly branched hexasaccharide containing β-galactofuranose in addition to β-galactopyranose, α- and β-glucopyranose and mannitol. Also the serotype 42 capsular polysaccharide is polymerized through a phosphodiester linkage. Lastly, the repeating unit of the serotype 47F capsular polysaccharide is a linear pentasaccharide containing β-galactofuranose, β-galactopyranose and α-galactopyranose in addition to ribitol. Acetylation sites were determined from HMBC and NOE signals as well as characteristic chemical shift perturbations to the $^1$H and $^{13}$C signals at the site of attachment. Some heterogeneity is found due to the presence of varying degrees of O-acetylation in serotypes 39 and 42 and chemical shift assignments were conducted for repeating units with different acetylation patterns in these polysaccharides. All spin systems were identified with sugar and sugar alcohol units by their chemical shift patterns and patterns of vicinal scalar couplings between protons ($3J_{HH}$) as described.$^{16}$
The acyclic pentose alcohol ribitol is found in the serotype 39 and 47F capsular polysaccharides of this study, while mannitol is found in serotype 42. Ribitol was previously identified in the repeating units of serogroups 6, 10, 11, 13, 24, 29, 33, 34, 35 and 47, where it was assigned the D-configuration. Arabinitol was suggested as a unit of serotype 17 and mannitol as a component of serotype 35A,¹⁹ which has a cps locus that is genetically related to the serotype 42 studied herein. Other serogroups reportedly contain phosphodiester linkages to glycerol and choline.⁸ The complete chemical shift assignments of the capsular polysaccharide spin systems of serotypes 39, 42 and 47F are tabulated in Tables 1-3.

### 3.2 Monosaccharide composition

In addition to the identification of NMR spin systems by COSY and TOCSY experiments, monosaccharide compositions were independently probed by gas-chromatographic analysis of methyl-glycosides as their trimethylsilyl derivatives using mannitol as an internal standard.¹⁶-¹⁸ In order to detect possible mannitol constituents, the analyses were repeated in the absence of mannitol, yielding a clear mannitol signal for serotype 42. Quantitation of mannitol in the serotype 42 sample was then conducted by comparing analyses in the presence and in the absence of mannitol as an internal standard. The indirect quantitation yielded molar fractions of 36.0% glucose, 32.4% galactose, 3.5% ribitol and 27.7% mannitol for serotype 42, thus underlining the presence of mannitol instead of ribitol. Residual ribitol presumably arises due to polysaccharide impurities from the cell wall polysaccharide rather than from the capsular polysaccharide.²⁰

In the absence of detected mannitol, direct quantification for the serotype 39 capsular polysaccharide sample yielded a composition of 14.1% ribitol, 59.7% galactose, 23.4% N-acetylglactosamine and traces of N-acetylglucosamine impurities. Monosaccharide analysis for the serotype 47F capsular polysaccharide sample yielded a composition of 24.3% ribitol and 70.9% galactose in addition to traces of N-acetyl galactosamine and of N-acetylglucosamine. An overestimation of glucose, N-acetylglactosamine and ribitol is likely due to the presence of contaminating impurities of cell wall polysaccharide containing these residues.²⁰ The lower than expected amount of galactose in gas chromatographic analysis is consistent with reduced furanose recovery during methanolysis in our setup.¹⁶ Absolute configurations of the monosaccharide units were not determined, but can in all likelihood be identified with the naturally occurring D-enantiomers of glucose, glucosamine and galactose.
3.3 Sequential assignment of spin systems: the repeating unit structures

The sequential arrangement and branching of monosaccharide units in the repeating unit of the capsular polysaccharides was determined by $^1$H-$^1$H-13C HMBC and $^1$H-$^1$H NOESY experiments. All linkages were unambiguously established both by interresidue nOes and by heteronuclear multibond correlations between $^1$H and natural abundance $^{13}$C across glycosidic bonds. These interresidue connectivities can be found tabulated in the Tables 1, 2 and 3 alongside all $^1$H and $^{13}$C chemical shifts. Phosphodiester linkages were determined and validated from independent observations, specifically the splitting of $^{13}$C signals through $^2J_{13C^31P}$ couplings, from $^1$H-$^{31}$P correlations and from chemical shifts, particularly the upfield shift of $^{13}$C signals split by $^2J_{13C^31P}$ in phosphodiester groups. Resultant capsular polysaccharide structures are depicted in Figure 1, displaying the repeating units of serotypes 39, 42 and 47F.

4. Discussion

The analysis of polysaccharide structures remains a low-throughput method that has been outpaced by the speed of genetic and genomic methods. Structural analysis, however, remains essential in demonstrating, how genetic loci for polysaccharide biosynthesis correlate with resultant polysaccharide structures through the action of encoded enzymatic activities. Here, we have determined the structures of capsular polysaccharides for three serotypes of the S. pneumoniae genetic cluster 4. The structures determined herein help fill in the remaining blanks of capsular polysaccharides structure in Streptococcus pneumoniae.

Serotype 39 supposedly has no known closely related serotype, but shows cross-reaction with typing serum 10d that reacts with the serotype 10A of known capsular polysaccharide structure.21 Serotype 10A and 39 capsular polysaccharides contain only one different glycosidic linkage and one different linkage in the phosphodiester bridge. In addition, the $\beta$-D-Galf branch attached to O3 of $\beta$-GalNac (Figure 1) is found to be mono-acetylated in our serotype 39 capsular polysaccharide preparation. Notwithstanding, the structures of serotype 10A and 39 capsular polysaccharides turn out to be identical in composition, which was validated by NMR spectroscopy of the serotype 10A capsular polysaccharide. Surprisingly, a previous cluster analysis of streptococcal cps biosynthetic loci
did not suggest the close compositional resemblance, which is reflected in the cross-reaction of serotype 39 with typing serum 10d.

The *cps* biosynthetic locus of serotype 42 is closely related to the *cps* biosynthetic loci of serotypes 35A and 35C. For serotype 35A, a structure of the capsular polysaccharide has previously been described. The structures of serotypes 35A and 42 are compared in Figure 4B. The close genetic resemblance of the *cps* biosynthetic loci of serotypes 35A and 42 is reflected by the close resemblance of capsular polysaccharide structures, which are identical apart from the presence of a branching (1→2)-α-D-galactopyranose attached a β-D-galactopyranose in the phosphodiester linkage in serotype 42. The presence of a mannitol previously only reported for serotype 35A structure is validated in the analysis of serotype 42 capsular polysaccharide. Nevertheless, the 35a typing serum reportedly does not react with serotype 42, but shows some reaction with serotype 47F. The structural similarity of the serotypes 35A/42 and 47F is limited to a -β-D-Galf-(1→3)-β-D-Glcp-(1→6)-β-D-Galf motif in serotypes 35A/42 that is also found in serotype 47F and is encircled in Figure 4, explaining the reaction of pneumococcal typing serum 35a with serotype 47F.

The serotype 47F capsular polysaccharide structure is additionally noteworthy in its comparison to a recently determined structure from the same serogroup, the serotype 47A structure (Figure 4). While the serotype 47A structure contains a doubly branched heptasaccharide repeating unit, the serotype 47F contains a linear pentasaccharide. The common antigenic determinant shared by the serotype 47A and 47F capsular polysaccharides is thus part of the conserved -α-D-Galp-(1→2)-D-ribitol-(5→OPO₃→6)-β-D-Galf-(3,5-di-OAc)-(1→3)-β-D-Galp-(1→ motif. The structure of serotypes 47F sheds light on the antigenic determinants in strains with genetically distant *cps* biosynthetic loci 47F and 47A. Genetically, the serotype 47F *cps* locus is closely related to 35F, for which no capsular polysaccharide structure has been described, so far. The recently renewed interest in uncharacterized capsular polysaccharide structures predictably will close the remaining gaps in the understanding of capsular polysaccharide biodiversity and its relation to the *cps* biosynthetic loci. Such an improved understanding is warranted in the light of rapid pneumococcal evolution as a consequence clinical interventions and pneumococcal genomic plasticity. This plasticity includes capsule-switching events that encompass the formation of vaccine-escape phenotypes. The prediction of phenotypic consequences of genomic changes in the *cps* locus should benefit
from studies of capsular polysaccharide biodiversity and antigenic determinants as provided herein.

**Acknowledgement**

The 800 MHz NMR spectra were recorded at the Danish National Instrument Center for NMR spectroscopy of Biological Macromolecules at the Carlsberg Laboratory. The authors gratefully acknowledge Hoai Aas and Yuanfeng Zou for performing the monosaccharide analysis.
**Figure Legends**

**Figure 1** Chemical structures of the repeating units of Streptococcus pneumoniae serotype 39, 42 and 47F capsular polysaccharides. Units are annotated by letters as in Tables 1-3.

**Figure 2.** 1D $^1$H NMR spectra of native *S. pneumoniae* serotype 39, 42 and 47F capsular polysaccharides. Resolved signals, primarily anomic protons and methyl groups of $O$-acetyl and $N$-acetyl groups are labeled.

**Figure 3.** Overlay of the anomic region of 2D DQF-COSY and TOCSY spectra of the *S. pneumoniae* serotype 39 (A), 42 (B) and 47F (C) capsular polysaccharides. Spin systems are named as in Figure 1, numbers indicate the atom numbers within the residues designated by letters.

**Figure 4** (A) Comparison of *S. pneumoniae* serotype 47F capsular polysaccharide to the recently determined serotype 47A capsular polysaccharide. Despite of their similar antigenic properties, serotypes 47F and 47A are both genetically dissimilar and disparate in their capsular polysaccharide structures. (B) Comparison of *S. pneumoniae* serotype 42 capsular polysaccharide to capsular polysaccharides from the genetically closely related serotype 35A structure lacking only a $\alpha$-Glc(1→2) branch. Encircled is a structural motif that is consistent with the cross-reaction of 35a typing serum with the serotype 47F.
References


<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6a</th>
<th>6b</th>
<th>N/O-Ac</th>
<th>N/O-Ac</th>
<th>NOE/HMBC</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>5,222</td>
<td>4,208</td>
<td>4,094</td>
<td>4,092</td>
<td>3,977</td>
<td>3,981</td>
<td>3,936</td>
<td></td>
<td></td>
<td>3,676</td>
</tr>
<tr>
<td></td>
<td>109.4</td>
<td>81.5</td>
<td>76.7</td>
<td>82.7</td>
<td>69.4</td>
<td>66.5</td>
<td></td>
<td></td>
<td></td>
<td>80.3</td>
</tr>
<tr>
<td>B</td>
<td>4,751</td>
<td>3,685</td>
<td>3,676</td>
<td>4,070</td>
<td>3,643</td>
<td>3,78</td>
<td>3,76</td>
<td></td>
<td></td>
<td>4,257</td>
</tr>
<tr>
<td></td>
<td>102.9</td>
<td>70.5</td>
<td>80.3</td>
<td>68.6</td>
<td>75.1</td>
<td>61.0</td>
<td></td>
<td></td>
<td></td>
<td>75.1</td>
</tr>
<tr>
<td>C</td>
<td>5,063</td>
<td>4,063</td>
<td>4,055</td>
<td>4,040</td>
<td>4,023</td>
<td>4,248</td>
<td>4,207</td>
<td>2,142</td>
<td></td>
<td>3,899</td>
</tr>
<tr>
<td></td>
<td>109.3</td>
<td>81.2</td>
<td>76.4</td>
<td>83.4</td>
<td>68.3</td>
<td>65.9</td>
<td>20.4</td>
<td>174.1</td>
<td>78.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>65%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>4,425</td>
<td>3,509</td>
<td>3,620</td>
<td>3,919</td>
<td>3,669</td>
<td>3,78</td>
<td>3,76</td>
<td></td>
<td></td>
<td>4,082</td>
</tr>
<tr>
<td></td>
<td>103.8</td>
<td>70.8</td>
<td>73.0</td>
<td>71.0</td>
<td>75.1</td>
<td>61.0</td>
<td></td>
<td></td>
<td></td>
<td>3,973</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>70.2</td>
</tr>
<tr>
<td></td>
<td>3,4,6)-β-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GalNACP</td>
<td>4,746</td>
<td>4,149</td>
<td>3,899</td>
<td>4,257</td>
<td>3,934</td>
<td>4,082</td>
<td>22.3</td>
<td>175.2</td>
<td>3,987</td>
</tr>
<tr>
<td></td>
<td></td>
<td>103.1</td>
<td>52.2</td>
<td>78.3</td>
<td>75.1</td>
<td>73.6</td>
<td>70.2</td>
<td></td>
<td></td>
<td>79.7</td>
</tr>
<tr>
<td>E</td>
<td>4,961</td>
<td>3,902</td>
<td>3,987</td>
<td>4,253</td>
<td>3,986</td>
<td>3,74</td>
<td>3,73</td>
<td></td>
<td></td>
<td>3,606</td>
</tr>
<tr>
<td></td>
<td>99.2</td>
<td>67.5</td>
<td>79.7</td>
<td>69.2</td>
<td>70.7</td>
<td>61.2</td>
<td></td>
<td></td>
<td></td>
<td>3,958</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>68.8</td>
</tr>
<tr>
<td>F</td>
<td>3,958</td>
<td>4,051</td>
<td>3,808</td>
<td>3,924</td>
<td>3,984</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3,606</td>
<td></td>
<td></td>
<td>4,077</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>68.8</td>
<td>71.0</td>
<td>71.6</td>
<td>68.7</td>
<td>66.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>5,205</td>
<td>4,205</td>
<td>4,094</td>
<td>4,092</td>
<td>3,977</td>
<td>3,981</td>
<td>3,936</td>
<td></td>
<td></td>
<td>3,676</td>
</tr>
<tr>
<td></td>
<td>109.4</td>
<td>81.5</td>
<td>76.7</td>
<td>82.7</td>
<td>69.4</td>
<td>66.5</td>
<td></td>
<td></td>
<td></td>
<td>80.3</td>
</tr>
<tr>
<td>B</td>
<td>4,751</td>
<td>3,685</td>
<td>3,676</td>
<td>4,070</td>
<td>3,643</td>
<td>3,78</td>
<td>3,76</td>
<td></td>
<td></td>
<td>4,257</td>
</tr>
<tr>
<td></td>
<td>102.9</td>
<td>70.5</td>
<td>80.3</td>
<td>68.6</td>
<td>75.1</td>
<td>61.0</td>
<td></td>
<td></td>
<td></td>
<td>75.1</td>
</tr>
<tr>
<td>C</td>
<td>5,063</td>
<td>4,063</td>
<td>4,055</td>
<td>4,040</td>
<td>4,023</td>
<td>4,248</td>
<td>4,207</td>
<td>2,142</td>
<td></td>
<td>3,899</td>
</tr>
<tr>
<td></td>
<td>109.3</td>
<td>81.2</td>
<td>76.4</td>
<td>83.4</td>
<td>68.3</td>
<td>65.9</td>
<td>20.4</td>
<td>174.1</td>
<td>78.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>65%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>4,425</td>
<td>3,509</td>
<td>3,620</td>
<td>3,919</td>
<td>3,669</td>
<td>3,78</td>
<td>3,76</td>
<td></td>
<td></td>
<td>4,082</td>
</tr>
<tr>
<td></td>
<td>103.8</td>
<td>70.8</td>
<td>73.0</td>
<td>71.0</td>
<td>75.1</td>
<td>61.0</td>
<td></td>
<td></td>
<td></td>
<td>3,973</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>70.2</td>
</tr>
<tr>
<td></td>
<td>3,4,6)-β-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GalNACP</td>
<td>4,745</td>
<td>4,159</td>
<td>3,907</td>
<td>4,324</td>
<td>3,934</td>
<td>4,082</td>
<td>3,973</td>
<td>2,039</td>
<td>3,987</td>
</tr>
<tr>
<td></td>
<td></td>
<td>102.8</td>
<td>52.2</td>
<td>78.5</td>
<td>74.1</td>
<td>73.6</td>
<td>70.2</td>
<td></td>
<td></td>
<td>79.7</td>
</tr>
<tr>
<td>F</td>
<td>4,961</td>
<td>3,902</td>
<td>3,987</td>
<td>4,253</td>
<td>3,986</td>
<td>3,74</td>
<td>3,73</td>
<td></td>
<td></td>
<td>3,606</td>
</tr>
<tr>
<td></td>
<td>99.2</td>
<td>67.5</td>
<td>79.7</td>
<td>69.2</td>
<td>70.7</td>
<td>61.2</td>
<td></td>
<td></td>
<td></td>
<td>3,958</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>68.8</td>
</tr>
<tr>
<td>G</td>
<td>3,958</td>
<td>4,051</td>
<td>3,808</td>
<td>3,924</td>
<td>3,984</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3,606</td>
<td></td>
<td></td>
<td>4,077</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>68.8</td>
<td>71.0</td>
<td>71.6</td>
<td>68.7</td>
<td>66.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*The two sets of assignments reflect two molecular species with different acetylation patterns of residue C.

Table 1. Chemical shift assignments of the serotype 39 capsular polysaccharide.
<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>5'/6</th>
<th>6</th>
<th>OAc</th>
<th>OAc</th>
<th>NOE/HMBC</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>P-3β-Glc(1→2)</td>
<td>5,526</td>
<td>3,501</td>
<td>3,830</td>
<td>3,335</td>
<td>4,095</td>
<td>3,906</td>
<td>3,716</td>
<td>3,819</td>
<td>73,0</td>
</tr>
<tr>
<td></td>
<td>2,3)β-Galp(1→3)</td>
<td>4,693</td>
<td>3,819</td>
<td>4,239</td>
<td>4,228</td>
<td>3,713</td>
<td>3,73</td>
<td>3,73</td>
<td>4,105</td>
<td>85,4</td>
</tr>
<tr>
<td>B</td>
<td>3)6-OAc-β-Gal(1→3)</td>
<td>5,361</td>
<td>4,451</td>
<td>4,105</td>
<td>4,529</td>
<td>5,519</td>
<td>4,368</td>
<td>4,317</td>
<td>2,155</td>
<td>3,667</td>
</tr>
<tr>
<td></td>
<td>70%</td>
<td>108,9</td>
<td>79,7</td>
<td>85,4</td>
<td>82,1</td>
<td>70,9</td>
<td>64,3</td>
<td>21,1</td>
<td>21,0</td>
<td>82,1</td>
</tr>
<tr>
<td>C</td>
<td>3)β-Glc(1→6)</td>
<td>4,539</td>
<td>3,443</td>
<td>3,667</td>
<td>3,410</td>
<td>3,490</td>
<td>3,918</td>
<td>3,724</td>
<td>4,061</td>
<td>72,2</td>
</tr>
<tr>
<td></td>
<td>103,4</td>
<td>74,5</td>
<td>82,1</td>
<td>68,8</td>
<td>76,7</td>
<td>61,6</td>
<td>174,3</td>
<td>174,6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>6)β-Gal(1→1)</td>
<td>5,054</td>
<td>4,131</td>
<td>4,084</td>
<td>4,013</td>
<td>4,022</td>
<td>4,061</td>
<td>3,763</td>
<td>4,023</td>
<td>3,669</td>
</tr>
<tr>
<td></td>
<td>108,8</td>
<td>81,7</td>
<td>77,5</td>
<td>83,9</td>
<td>70,5</td>
<td>72,2</td>
<td>70,5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>1)mannitol(6→P</td>
<td>4,023</td>
<td>3,882</td>
<td>3,855</td>
<td>3,872</td>
<td>3,871</td>
<td>4,162</td>
<td>4,085</td>
<td>82,1</td>
<td>3,669</td>
</tr>
<tr>
<td></td>
<td>3,669</td>
<td>70,5</td>
<td>70,4</td>
<td>69,8</td>
<td>69,2</td>
<td>70,6</td>
<td>68,0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Chemical shift assignments of the serotype 42 capsular polysaccharide.  

* The two sets of assignments reflect two molecular species with different acetylation patterns of residue C.
Table 3. Chemical shift assignments of the serotype 47F capsular polysaccharide.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>6</th>
<th>O-Ac</th>
<th>O-Ac</th>
<th>NOE/HMBC</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P-6)3,5diOAc-β-Gal(1→3)</td>
<td>5,299</td>
<td>4,349</td>
<td>4,885</td>
<td>4,484</td>
<td>5,363</td>
<td>4,05</td>
<td>4,05</td>
<td>2.143</td>
<td>2.017</td>
<td>3,741</td>
</tr>
<tr>
<td></td>
<td>109,8</td>
<td>79,4</td>
<td>79,4</td>
<td>80,8</td>
<td>71,6</td>
<td>64,1</td>
<td></td>
<td>20,9</td>
<td>20,9</td>
<td>80,7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>174,0</td>
<td>173,8</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3)β-Galp(1→6)</td>
<td>4,497</td>
<td>3,665</td>
<td>3,741</td>
<td>4,074</td>
<td>3,708</td>
<td>3,73</td>
<td>3,73</td>
<td>4,054</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>103,6</td>
<td>70,5</td>
<td>80,7</td>
<td>69</td>
<td>75,8</td>
<td>61,5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>71,5</td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6)2OAc-β-Gal(1→3)</td>
<td>5,379</td>
<td>5,061</td>
<td>4,228</td>
<td>4,143</td>
<td>4,038</td>
<td>4,054</td>
<td>3,779</td>
<td>2.134</td>
<td></td>
<td>4,007</td>
</tr>
<tr>
<td></td>
<td>107,6</td>
<td>84,5</td>
<td>76,2</td>
<td>84,4</td>
<td>70,3</td>
<td>71,5</td>
<td></td>
<td>20,9</td>
<td></td>
<td>77,7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>173,7</td>
</tr>
<tr>
<td>D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3)α-Galp(1→2)</td>
<td>5,237</td>
<td>3,966</td>
<td>4,007</td>
<td>4,161</td>
<td>4,068</td>
<td>3,769</td>
<td>3,73</td>
<td>4,035</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>99,7</td>
<td>68,2</td>
<td>77,7</td>
<td>69,6</td>
<td>71,7</td>
<td>61,5</td>
<td></td>
<td></td>
<td></td>
<td>80,3</td>
</tr>
<tr>
<td>E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2)ribof(5→P</td>
<td>3,916</td>
<td>4,035</td>
<td>4,002</td>
<td>3,84</td>
<td>4,054</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3,838</td>
<td>3,939</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>60,6</td>
<td>80,3</td>
<td>72,2</td>
<td>71,0</td>
<td>67,5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>